
Regional modulation of enhancer accessibility
during the establishment of spatial coordinates
in the *Drosophila* embryo

Marta Teresa Božek

München 2018

Disseration zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig–Maximilians–Universität München

Regional modulation of enhancer accessibility
during the establishment of spatial coordinates
in the *Drosophila* embryo

Marta Teresa Bożek

aus

Głuszyca, Polen

2018

Erklärung:

Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Klaus Förstemann betreut.

Eidesstattliche Versicherung:

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 20. September 2018

Marta Teresa Božek.....

Dissertation eingereicht am 20. September 2018

1. Gutachter: Prof. Dr. Klaus Förstemann

2. Gutachter: Prof. Dr. Nicolas Gompel

Mündliche Prüfung am 22. Oktober 2018

Summary

The earliest stages of *Drosophila* embryogenesis are dedicated to the establishment of body axes and specification of spatial coordinates. This process is regulated primarily at the level of transcription initiation by complex gene regulatory networks that operate at the onset of zygotic genome activation. Precise and robust establishment of expression patterns is mediated by axis patterning enhancers. Each enhancer is targeted by multiple activators and repressors, whose varying concentrations along the body axes provide specific positional information that is converted in each nucleus into a distinct transcriptional output.

While the regulatory logics and organizational principles of axis patterning enhancers have been studied mainly in the context of the composition and arrangement of transcription factor binding sites, an increasing number of studies suggests the importance of their chromatin structure. However, the interplay between regulatory activity of enhancers and accessibility as determined by the local chromatin organization is not well understood. It is not clear whether axis patterning enhancers are characterized by a uniform and highly accessible chromatin organization throughout the entire embryo or whether their accessibility varies regionally in accordance with their differential activity.

In order to address this question, I profiled chromatin accessibility by performing ATAC-seq in narrow, genetically tagged domains along the antero-posterior axis of the cellular blastoderm. I demonstrate that chromatin organization of the *Drosophila* embryo shows regional variation immediately after zygotic genome activation. While the position of open chromatin regions remains highly conserved along the AP axis, as much as one quarter of the accessible genome displays significant quantitative modulation of its ATAC-seq signal. Since the most variable regions correspond to the annotated as well as putative axis patterning enhancers, I conclude that differential accessibility is a signature of patterning *cis*-regulatory elements in the *Drosophila* blastoderm.

Regional accessibility changes strongly correlate with the regulatory activity of axis patterning enhancers. When receiving a net activating input and promoting transcription, the enhancers display elevated accessibility. On the other hand, reduced accessibility coincides with a net repressive input, yet it never decreases to background levels of the inaccessible genome. Importantly, axis patterning enhancers are characterized by similar dynamics regardless of their position along the AP axis and, thus, the composition of input transcription factors.

I conclude that chromatin context plays an integral role in the spatial regulation of axis patterning enhancers. I propose a model in which accessibility of the enhancers is

initially established by ubiquitous mechanisms prior to zygotic genome activation, followed by subsequent regional modulation during the operation of gene regulatory networks. I discuss potential mechanisms by which accessibility of enhancers may be directly modified by their interactions with activator and repressor transcription factors. I additionally propose that differential accessibility is a marker of differential regulatory activity and can potentially serve as a metric for *de novo* identification of enhancers patterning complex tissues.

Acknowledgements

I would like to thank Prof. Ulrike Gaul for her support and guidance throughout my PhD project, for motivating and encouraging me to be creative and independent, and offering an environment where I could grow both on a personal and professional level. I am also immensely grateful to Prof. Klaus Förstemann and Prof. Nicolas Gompel for having so much faith in me and my research during the past year. I would like to thank Klaus for taking up the role of my thesis advisor and Nicolas for his help with defining my scientific ideas and putting them in writing. Thank you both for your time and support.

I am very grateful to all the other members of my thesis advisory committee, Prof. Peter Becker, Prof. Erwin Frey and Dr. Ralph Neumüller, for their valuable feedback. I am also indebted to the Graduate School of Quantitative Biosciences Munich for creating a very supportive and stimulating environment. I am especially grateful to Michael and Mara for helping me settle down in Munich.

I would like to thank all the present and former members of the Gaul group. It was a pleasure to work with such supportive colleagues and friends. I would like to especially thank Andrea and Alessio for being my lab companions during the past five years, and Roberto for his invaluable help with data analysis.

Finally, I would like to thank my dear family for their love and support. Thank you, Lorenzo, for being by my side through the good and tough times. You are my inspiration and motivation.

Contents

Summary	vii
Acknowledgements	ix
1 Introduction	1
1.1 Establishment of the <i>Drosophila melanogaster</i> body plan	1
1.1.1 Earliest stages of <i>Drosophila</i> embryogenesis	1
1.1.2 Gene regulatory networks patterning the embryo	2
1.1.3 Organizational features of the AP network	4
1.2 Operational principles of axis patterning enhancers	5
1.2.1 Definition of axis patterning enhancers	5
1.2.2 Discovery of axis patterning enhancers	6
1.2.3 Organization of TF binding sites as a determinant of enhancer activity	6
1.2.4 Chromatin context of axis patterning enhancers	7
1.3 Research questions	9
1.3.1 Alternative models of chromatin organization at axis patterning enhancers	9
1.3.2 Current state of knowledge	10
1.3.3 Outline of the study	10
1.3.4 Summary of the results	11
2 Experimental strategy	13
2.1 Experimental methods	13
2.1.1 INTACT: genetic tagging and affinity purification of specific cell-types	13
2.1.2 ATAC-seq: chromatin accessibility assay	15
2.2 Generation of transgenic strains with tagged AP domains	16
2.2.1 Selection of nuclear tags	16
2.2.2 Selection of axis patterning enhancers	16
2.2.3 Assembly of expression constructs	18
2.3 Validation of transgenic strains	19
2.3.1 Selection of integration sites, basal promoters and nuclear tags	19
2.3.2 Characterization of tagged domains	23
2.4 Generation of ATAC-seq libraries from tagged domains	26

2.4.1	Timing of embryonic collections	26
2.4.2	Experimental workflow	26
2.4.3	Summary of ATAC-seq samples	27
3	Processing and quality control of ATAC-seq libraries	29
3.1	Interpretation of ATAC-seq signal	29
3.1.1	Definitions of ATAC-seq signal	29
3.1.2	Distinction between nucleosome-bound and nucleosome-free regions	30
3.1.3	Size distribution of ATAC-seq libraries	31
3.1.4	Genomic distribution of size-selected ATAC-seq fragments	33
3.2	Identification of accessible genomic regions	36
3.2.1	Definition of ATAC-seq peaks	36
3.2.2	Definition of high-confidence ATAC-seq peaks	36
3.3	Reproducibility of ATAC-seq experiments	36
3.3.1	Comparison between biological replicates	36
3.3.2	Comparison with published chromatin accessibility profiles	37
4	Regional differences in chromatin accessibility along the AP axis	41
4.1	Global variation of accessibility profiles	41
4.1.1	Examination of individual genomic loci	41
4.1.2	Genome-wide assessment of the accessibility variation	41
4.2	Scale of accessibility variation along the AP axis	46
4.2.1	Identification of differential ATAC-seq peaks	46
4.2.2	Evaluation of differential regions	48
4.3	Characterization of differential ATAC-seq peaks	50
4.3.1	Genomic localization of differential ATAC-seq peaks	50
4.3.2	Occupancy by different classes of proteins	51
4.3.3	Overlap of ATAC-seq peaks with annotated <i>cis</i> -regulatory elements	54
4.3.4	Differential accessibility is a signature of axis patterning enhancers .	57
4.3.5	Differential accessibility as a metric for <i>de novo</i> identification of en- hancers	59
5	Accessibility modulation of individual axis patterning enhancers	61
5.1	Elevated accessibility of enhancers coincides with domains of their transcrip- tional activity	61
5.1.1	Modulation of local accessibility at individual enhancers	61
5.1.2	Relative changes of accessibility signal at active and inactive enhancers	61
5.2	Accessibility of axis patterning enhancers is quantitatively correlated with their transcriptional output	65
5.2.1	Deconvolution of ATAC-seq signal from individual tagged domains .	65
5.2.2	Linear correlation between ATAC-seq signal and proportional repre- sentation of active enhancers	66

5.2.3	Active and inactive axis patterning enhancers display distinct accessibility profiles	68
5.3	Inactive enhancers exhibit residual accessibility	69
5.4	Local accessibility modulation within individual enhancers	70
6	Discussion	71
6.1	Summary of key findings	71
6.2	Advantages and limitations of the experimental strategy	72
6.3	Comparison with other studies examining regional modulation of chromatin accessibility in <i>Drosophila</i> embryos	73
6.3.1	Single-cell ATAC-seq by Cusanovich et al. (2018)	73
6.3.2	ATAC-seq on cryo-sliced anterior and posterior domains by Haines and Eisen (2018)	74
6.3.3	Summary	74
6.4	Differential accessibility as a signature of axis patterning enhancers	75
6.4.1	Relationship between activity and accessibility	75
6.4.2	Interpretation of differential accessibility in the context of local chromatin organization	76
6.5	Differential accessibility as a metric for <i>de novo</i> discovery of enhancers	77
6.6	Independent mechanisms for establishment and modulation of enhancer's accessibility	78
6.7	Mechanisms of accessibility modulation at axis patterning enhancers	80
6.8	Outlook	81
6.8.1	Role of local accessibility modulations at axis patterning enhancers	81
6.8.2	Implications for the future studies on axis patterning enhancers	82
7	Materials and Methods	83
7.1	Generation of transgenic strains	83
7.1.1	Elements of the expression cassette	83
7.1.2	Entry vectors	83
7.1.3	Assembly of expression constructs	84
7.1.4	Site-specific genomic integration	84
7.2	Characterization of tagged domains	84
7.2.1	Immunostaining of transgenic embryos	84
7.2.2	Definition of positions of tagged domains along the AP axis	85
7.3	Preparation of ATAC-seq libraries	85
7.3.1	Collection and staging of embryos	85
7.3.2	Affinity-purification of tagged nuclei	85
7.3.3	ATAC-seq on purified nuclei	86
7.3.4	ATAC-seq on genomic DNA	86
7.4	Data analysis	87
7.4.1	ATAC-seq data processing	87
7.4.2	Peak calling	87

7.4.3	Normalization of ATAC-seq signal	87
7.4.4	Correlation of ATAC-seq profiles with published DNase-seq data . .	88
7.4.5	Principal component analysis	88
7.4.6	Identification of differential ATAC-seq peaks	88
7.4.7	Genomic annotations of ATAC-seq peaks	89
7.4.8	Overlap of ATAC-seq peak with published ChIP data	89
7.4.9	Overlap of ATAC-seq peaks with annotated <i>cis</i> -regulatory elements	89
7.4.10	Overlap of tagged domains with activity patterns of AP enhancers .	90
7.4.11	Identification of background regions	90
8	Appendix	91
8.1	Supplementary Material A: DNA sequences	91
8.2	Supplementary Material B: Enhancer drivers	98
8.3	Supplementary Material C: Evaluation of D6 Replicate 1	100
8.4	Supplementary Material D: AP enhancers	103

List of Figures

1.1	Summary of the earliest stages of <i>Drosophila</i> embryogenesis	2
1.2	Outline of the AP gene regulatory network	3
1.3	Principles of gene regulation in the AP network	12
2.1	Summary of the experimental protocol	14
2.2	Design of the expression cassette	17
2.3	Reported activity patterns of the seven enhancers of gap and pair-rule genes used as drivers of the nuclear tag	18
2.4	Summary of a cloning strategy used for the assembly of expression constructs	20
2.5	Localization of the transgenic tag to the nuclear envelope	21
2.6	Comparison of expression levels driven by different basal promoters	22
2.7	RanGAP-3xFLAG is characterized by wider expression domains than UNC84-3xFLAG	23
2.8	List of transgenic strains used in the ATAC-seq experiments	24
2.9	Overview of tagged domains (D1-D7)	25
2.10	Composition of embryonic collections	27
2.11	Detailed outline of the experimental protocol	28
3.1	Two definitions of ATAC-seq signal	30
3.2	Size distribution of ATAC-seq fragments and their position in reference to DNA-bound proteins	31
3.3	Representative size distribution of ATAC-seq libraries from the <i>Drosophila</i> blastoderm	32
3.4	Distribution of different classes of size-selected ATAC-seq fragments over the locus of <i>giant</i>	34
3.5	Correlation of coverage tracks of different classes of size-selected ATAC-seq fragments over the locus of <i>giant</i>	35
3.6	Size distribution of high-confidence ATAC-seq peaks	37
3.7	Correlation between replicate ATAC-seq experiments on tagged domains	38
3.8	Correlation between replicate ATAC-seq experiments on whole-embryos	39
3.9	High similarity between whole-embryo controls and published DNase-seq accessibility profiles from stage-5 embryos	40

4.1	Comparison of ATAC-seq accessibility profiles from different tagged domains at the locus of <i>giant</i>	42
4.2	Principal component analysis of tagged domains and whole-embryo controls	43
4.3	Pair-wise representations of PCA components	44
4.4	Conserved position of highly accessible regions along the AP axis	45
4.5	One quarter of the accessible genome displays significant regional variation in the <i>Drosophila</i> blastoderm	47
4.6	Pair-wise comparisons between individual tagged domains provide increased sensitivity of detection	48
4.7	Number of differential ATAC-seq peaks is proportional to the distance between compared tagged domains	49
4.8	Distribution of ATAC-seq signal fold-changes among the entire set of differential peaks	51
4.9	Genomic localization of ATAC-seq peaks	52
4.10	Overlap of ATAC-seq peaks with chromatin immunoprecipitation (ChIP) data for different classes of proteins	53
4.11	Temporal activity of Vienna Tiles overlapping ATAC-seq peaks	55
4.12	Spatial activity of Vienna Tiles overlapping ATAC-seq peaks	56
4.13	Overlap of differential ATAC-seq peaks with annotated <i>Drosophila cis</i> -regulatory elements	57
4.14	Comparison of annotated and unannotated differential peaks	60
5.1	Modulation of local accessibility at individual AP enhancers	62
5.2	Relative changes of accessibility signal at active and inactive enhancers . .	63
5.3	Comparison of fold-change accessibility changes at three classes of enhancers	64
5.4	Comparison of ATAC-seq signal over <i>run_(+17)</i> enhancer between all embryonic domains	66
5.5	Relationship between ATAC-seq signal measured in tagged domains and the proportion of nuclei with active enhancers	67
5.6	Distribution of correlation coefficients	68
5.7	Comparison of ATAC-seq signal between background genomic regions and axis patterning enhancers	70
6.1	Proposed model of chromatin organization at axis patterning enhancers during the earliest stages of <i>Drosophila</i> embryogenesis	79
8.1	Correlation matrix of ATAC-seq samples	101
8.2	DESeq2 analysis with replicates of D6	102

Chapter 1

Introduction

1.1 Establishment of the *Drosophila melanogaster* body plan

The process of fate specification and cell differentiation during metazoan development relies crucially on spatial and temporal regulation of gene expression [1, 2]. Particularly in the context of early embryogenesis, most of the control takes place at the level of transcription initiation, giving rise to gene regulatory networks (GRNs). GRNs represent complex genetic interactions, manifested at the level of *cis*-regulatory elements (CREs) that receive regulatory inputs from trans-acting transcription factors (TFs) and interpret them into a transcriptional output of their target gene. As a result, GRNs govern the establishment of cell-type specific expression patterns, ultimately defining a blueprint for the embryonic body plan.

The earliest stages of *Drosophila* embryogenesis serve as a canonical model system for studying regulatory logics of developmental GRNs as well as organizational principles of individual *cis*-regulatory elements. At the onset of zygotic genome activation, the networks pattern the embryo along the body axes, by establishing spatial coordinates and specifying future cell identities.

1.1.1 Earliest stages of *Drosophila* embryogenesis

After fertilization, the *Drosophila* embryo undergoes rapid and synchronous rounds of mitotic cleavages, which are not accompanied by cellular divisions (Figure 1.1). During the tenth cleavage cycle, the nuclei migrate to the surface of the embryo, forming a syncytial blastoderm. The subsequent divisions become increasingly longer until cycle 14, which lasts for more than 60 min. During this time, the peripheral nuclei elongate and become separated by invaginating cellular membranes, giving rise to a cellular blastoderm (cycle 14A, stage 5). Subsequently, the embryo undergoes gastrulation to segregate the three presumptive germ layers: mesoderm, endoderm and ectoderm.

stage 1	cycle 1-2	25 min	cleavage stage
stage 2	cycle 3-7	40 min	
stage 3	cycle 8-9	15 min	
stage 4	cycle 10-13	50 min	syncytial blastoderm
stage 5	cycle 14A	40 min	cellular blastoderm
stage 6-7	cycle 14B	20 min	gastrulation

Figure 1.1: Summary of the earliest stages of *Drosophila* embryogenesis. The table represents key events during embryonic development (grey) together with the corresponding stages of embryogenesis and mitotic cleavages, as well as the duration of each stage (after [3, 4, 5]).

The initial stages of embryogenesis are regulated by maternally-deposited transcripts and proteins. Maternal factors mediate mitotic divisions [6] and the initial chromatin organization [7, 8], as well as provide the symmetry-breaking upstream input for the patterning gene regulatory networks [9, 10, 11, 12]. While sustained zygotic transcription starts around mitotic cycle 7, widespread zygotic genome activation (ZGA) occurs only at the beginning of the cleavage cycle 14 [13, 14], which coincides with lengthening of the interphase and cellularization of the blastoderm (mid-blastula transition).

ZGA is immediately followed by the full activation and maturation of expression patterns that are regulated by the patterning GRNs. More than 1400 genes are characterized by spatially-restricted expression at the end of stage 5 [15], conferring unique cell identities as early as the mid-blastula transition [16].

1.1.2 Gene regulatory networks patterning the embryo

Establishment of the *Drosophila* body plan is governed by two regulatory networks that pattern the embryo along the antero-posterior (AP) and dorso-ventral (DV) axes [17]. The DV network specifies three basic tissues: mesoderm, neurogenic ectoderm and dorsal ectoderm [18], while the AP network defines the position and identity of future body

segments [4]. Both networks rely on the positional information that is initially provided by the maternally-deposited regulators and is subsequently processed into increasingly complex and refined patterns of the downstream zygotic regulators. While often discussed separately, the two networks are also characterized by considerable cross-talk [19, 20, 21, 22].

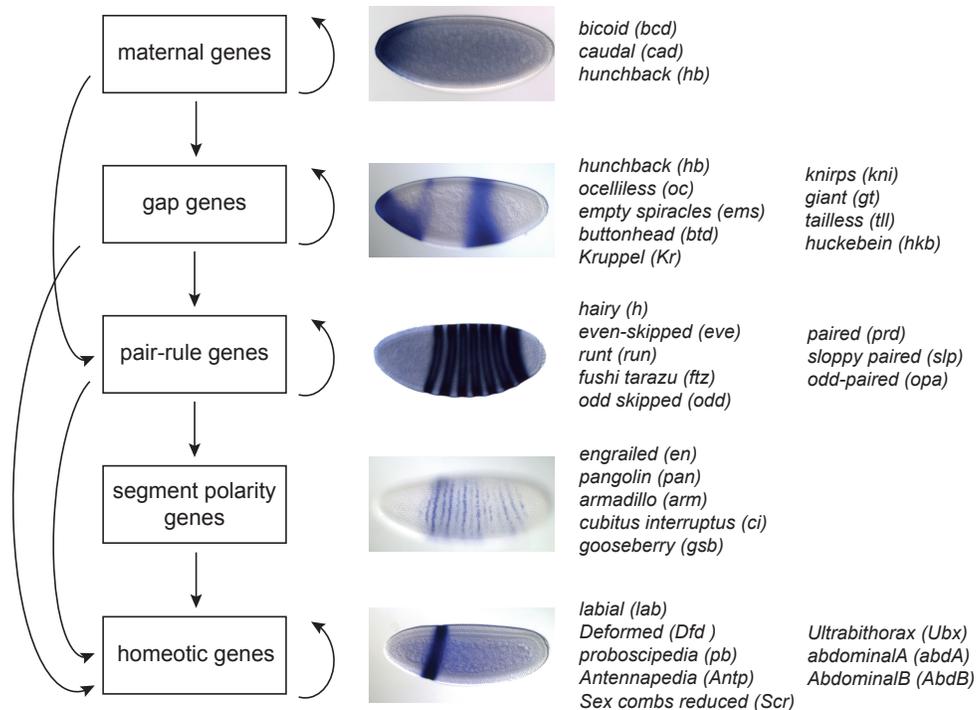


Figure 1.2: Outline of the AP gene regulatory network. The diagram represents genetic interactions between different gene classes, with embryo images showing representative expression patterns. On the right, a list of genes from each tier of the network. Only those genes that encode TFs and are characterized by spatially restricted expression domains are included (after [4]). RNA *in situ* hybridization images represent: *bicoid*, *knirps*, *fushi tarazu*, *engrailed* and *Deformed* (from Berkeley *Drosophila* Genome Project; [23]).

The AP network, also referred to as the segmentation network, is a focus of this study [3, 4]. It has a hierarchical structure, with the top tier of genes consisting of maternally deposited regulators that form broad concentration gradients in the syncytial embryo (Figure 1.2). They control transcription of zygotic gap genes in broad overlapping domains that are around 10-20 nuclei wide. In a combinatorial fashion, maternal and gap factors regulate the downstream pair-rule genes, which are expressed in a periodic pattern of 7-8 narrow stripes, each encompassing around 4 nuclei. Pair-rule genes subsequently control segment-polarity genes, whose expression in 14 stripes becomes established just before the onset of gastrulation. They divide the AP axis into fine segments that correspond to the

position of parasegmental boundaries later in development. At the same time, gap and pair-rule genes regulate homeotic Hox genes.

Maternal to segment polarity genes specify a coordinate system of the embryo, and define the number and polarity of future body segments. On the other hand, Hox genes are required for allocation of distinct morphological identities to each body segment [24]. Maternal, gap and pair-rule genes primarily encode transcription factors and their expression peaks in the cellular blastoderm. Unlike the upstream regulators, only a subset of segment polarity genes represents TFs. Since they are operating after formation of cellular membranes, this class of genes also encodes proteins involved in cell-to-cell signaling. Additionally, expression of segment polarity genes and Hox genes continues until later stages of embryogenesis, which allows for stable maintenance of the developmental programs.

1.1.3 Organizational features of the AP network

The AP network converts positional information provided by a small set of upstream regulators into complex and refined expression patterns of the downstream genes. A key transition in the network is a switch from broad gradients of maternal and gap genes to periodic patterns of pair-rule genes, which lays the basis for the definition of future parasegments. Several properties allow the network to regulate these intricate expression patterns in a precise and robust manner.

First of all, gap and pair-rule genes are controlled by multiple *cis*-regulatory elements placed in a relatively close proximity both upstream and downstream of their target gene (Figure 1.3). Each enhancer defines an independent expression domain. In spite of close localization in the locus, the CREs act autonomously, with no cross-talk between their regulatory inputs [25, 26].

Second, each enhancer is regulated in a combinatorial manner by multiple activating and repressing TFs (Figure 1.3). Their varying concentrations along the AP axis provide specific positional information to the enhancer, which is subsequently integrated and interpreted by the element into a spatially restricted expression pattern. Transcription of the target gene is promoted only in the specific spatial domain of the embryo in which the enhancer receives a net activating input [27].

Third, interpretation of the activating and repressing regulatory input crucially depends on the architecture of enhancers as defined by the number, affinity and organization of TF binding sites (TFBSs). Enhancers respond to input TF in a concentration-dependent manner [28, 29], which is determined by affinity of the respective binding sites [30]. High-affinity sites increase sensitivity of the *cis*-regulatory element to lower concentrations of the factor at the tails of its gradient. Clustering of homotypic binding sites [31, 21] mediates cooperative DNA binding that allows for a sharp on/off response of the enhancer to small concentration changes of a TF along its gradient [32, 33]. The enhancers are also characterized by a considerable overlap between binding sites of activators and repressors, which mediates a direct competition between the opposing regulatory inputs [27, 34]. In addition, repressors encoded by gap genes have been demonstrated to suppress activator sites only within a distance of 50-150 bp (hence referred to a short-range repressors), which

additionally ensures the autonomy of neighbouring enhancers [35, 25].

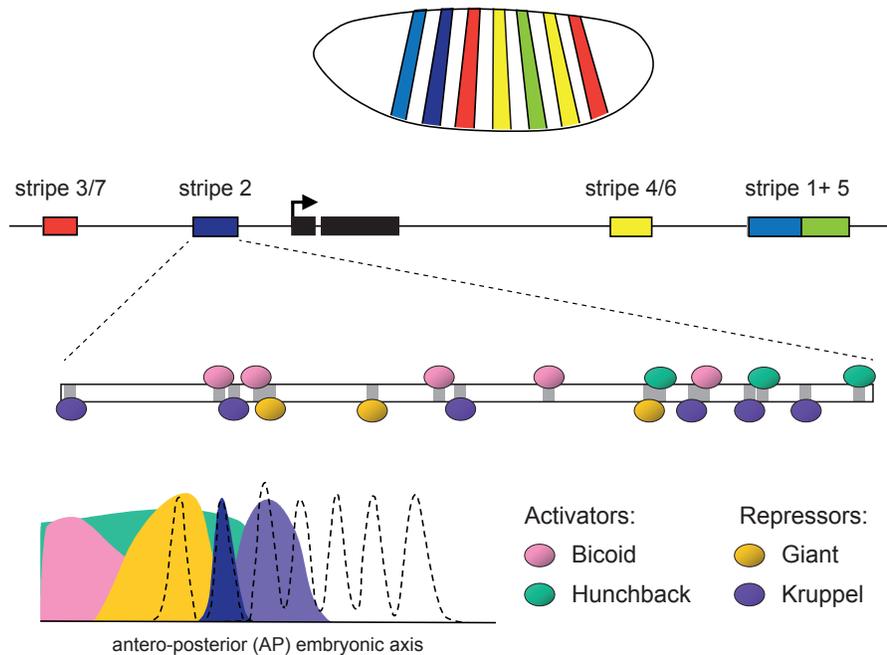


Figure 1.3: Principles of gene regulation in the AP network. Pair-rule gene *even-skipped* is expressed in seven narrow stripes. Locus of *eve* contains five autonomous enhancers in close proximity to the target gene. Each element controls expression in either one or two stripes. As illustrated by stripe 2 enhancer, each element is targeted by multiple activator (Bicoid and Hunchback) and repressor (Giant and Kruppel) transcription factors, often with overlapping binding sites. Concentration gradients of the input TFs along the AP axis are compared to the position of all *eve* stripes (dashed line) as well as the activity pattern of stripe 2 enhancer (dark blue). The enhancer promotes transcription of its target gene only in a specific domain where it receives a net activating input. Figure kindly provided by Prof. Ulrike Gaul.

Overall, the three mechanisms ensure precise definition of positional information in the AP network. Regulation by multiple *cis*-regulatory elements as well as integration of the input from multiple TFs allows a small set of broadly distributed upstream regulators to generate narrower and more defined expression domains of the downstream genes. Additionally, variable affinity of TFBSs enables an enhancer to sense its position along the concentration gradient with higher sensitivity, while cooperativity allows for definition of sharp expression boundaries.

It is important to note that the AP network consists of broadly expressed activators and more spatially restricted repressors [34]. Activator TFs are encoded maternally and are distributed in the embryo either in broad concentration gradients (Bicoid, Hunchback and

Caudal) or ubiquitously (Zelda [8], D-STAT [36]). On the other hand, the downstream zygotic factors act primarily as repressors. Short-range repressors are encoded by gap genes and act independently on individual enhancers, while pair-rule TFs mediate long-range repression that allows for simultaneous silencing of the entire locus [35, 37]. Thus, transcriptional repression appears to play a pervasive role in the initial patterning of the embryo [4].

The AP regulatory network is also characterized by high robustness. The process of axis patterning is highly reproducible, with positions of parasegments showing little variation between individual embryos [38]. Multiple levels of control allow for refinement and adjustment of the initially variable expression of gap and pair-rule genes into a highly stereotypical pattern at the end of cycle 14A. One mechanism involved in maturation of the expression domains is a substantial cross-talk between genes from the same tier of the network, in particular gap and pair-rule genes [39, 40]. Additionally, gap genes are regulated by enhancers with overlapping activities (shadow enhancers), which has been demonstrated to produce sharper and more homogenous gene expression patterns [41].

At the same time, the AP network is characterized by high plasticity. Ectopic expression of regulators, either at modified concentrations or in different regions of the embryo, results in changes of the segmentation layout without interfering with other patterning processes [42, 43]. This indicates high receptiveness of the enhancers to TF input regardless of their position along the AP axis or composition of the regulatory cues.

1.2 Operational principles of axis patterning enhancers

1.2.1 Definition of axis patterning enhancers

In this study, I refer to *cis*-regulatory elements of the AP and DV networks as axis patterning enhancers. I use a separate term in order to highlight their specific features, which are not commonly shared with other developmental regulatory elements. First of all, axis patterning enhancers are targeted by TFs at all positions along the embryonic axis. Second, they simultaneously integrate input from activating and repressing TFs, and very often it is the net prevalence of either regulatory cue that determines the transcriptional output. Finally, the differential regulatory input gives rise to differential transcriptional output along the embryonic axis. Axis patterning enhancers operate at a specific time point of embryogenesis and their differential activity is manifested in space, rather than over developmental time. In this study, axis patterning enhancers are discussed particularly in the context of gap and pair-rule genes.

1.2.2 Discovery of axis patterning enhancers

As the AP regulatory network has been a focus of developmental and genetic studies for decades [44], a considerable number of axis patterning enhancers is already known,

characterized and functionally validated [45]. The enhancers have been identified with a range of different approaches, starting initially from genetic dissection of the regions that are proximal to promoters of key regulators of the network [46, 47, 27]. Fragmentation of the loci with restriction enzymes and subsequent tests with enhancer-reporter assays allowed for identification of minimal genomic regions that drive endogenous expression patterns of target genes. This was accompanied by identification of individual TF binding sites by *in vitro* DNaseI footprinting and their validation by mutagenesis [48].

More recent approaches involved genome-wide identification of candidate CREs based on the local clustering of binding sites for the key regulators of the AP network, as a manifestation of their combinatorial interactions at axis patterning enhancers. The binding sites were predicted from sequence preferences of the TFs [49, 50, 51], additionally accompanied by conservation analysis of site clustering in related *Drosophila* species [48, 52]. In an alternative approach, binding of TFs was inferred experimentally using chromatin immunoprecipitation (ChIP) assays [53, 41].

Recently, regional variation in the level of histone acetylation (H3K27ac), which is a marker of active enhancers [54], has been used for successful *de novo* identification of axis patterning enhancers in the DV network [55]. All of the aforementioned efforts have been also accompanied by a large-scale screen of candidate 2-kb genomic regions for their enhancer activity at different stages of *Drosophila* embryogenesis [56].

1.2.3 Organization of TF binding sites as a determinant of enhancer activity

Spatial activity of axis patterning enhancers in the cellular blastoderm is determined by the composition of regulatory inputs that the elements receive in different regions of the embryo. Multiple efforts in the past focused on deciphering how the identity and arrangement of TF binding sites within the enhancer mediate interpretation of the input and its subsequent computation into a specific transcriptional output [57].

Genetic dissection of native enhancers allowed for identification of minimal sequences conferring the endogenous activity pattern [27] as well as revealed the contribution of individual TFs within the regulatory elements [58, 30]. Simplified synthetic enhancers have been used to systematically examine the interplay between the activating and repressive inputs. By combining a limited set of activator and repressor TFBSs, often from the orthogonal AP and DV networks, the studies highlighted the importance of stoichiometry, affinity and spacing of the binding sites in determination of the final transcriptional output [59, 60, 61].

The explanatory power of the composition and arrangement of TF binding sites has been tested in a number of modelling studies [62, 63, 64]. Activity patterns of AP axis patterning enhancers have been predicted solely as a function of their DNA sequence, binding specificities of key input TFs and their distribution along the AP axis. Position weight matrices (PWMs) of the regulators were used to predict their potential binding sites along the enhancer's sequence (including different affinity ranges), while occupancy of these sites at

different positions along the body axis was computed based on the concentration gradients of the respective TFs. While showing high similarity between the native and predicted activity patterns for certain enhancers, performance of the models was not consistent across all considered *cis*-regulatory elements. The predictive power of TF binding sites was not sufficient to infer all activity patterns, in spite of considering specific functional properties of the AP network such as short-range repression or TF cooperativity.

Remarkably, efforts at reconstituting known axis patterning enhancers have not been successful either. In spite of thorough functional characterization of the minimal enhancer of *even skipped* stripe 2 [27], placing its known TF binding sites in a different sequence context did not allow for recreation of its activity pattern [65]. Additionally, the two tested spacer sequences drove expression of the element in different anterior regions of the embryo and generated different expression levels.

These examples highlight our limitations in understanding the architecture and function of axis patterning enhancers. The inability to fully predict or reconstitute an enhancer's activity might stem either from the fact that we are unable to identify a complete set of input TFs and their contributing binding sites, or have not fully deciphered the rules that govern TF interactions and transformation of the regulatory input into the transcriptional output.

1.2.4 Chromatin context of axis patterning enhancers

The aforementioned efforts at modelling and reconstituting activity patterns of known axis patterning enhancers were based only on the organization of TF binding sites (whether predicted or experimentally validated) within their sequence. These approaches, therefore, did not consider the chromatin context of the regulatory elements, either simplifying them to naked stretches of DNA or assuming that accessibility of the TFBSs to their respective regulators was uniform along the entire length of the enhancer as well as the whole embryonic axis. However, different lines of evidence suggest that the chromatin organization plays a direct role in mediating interactions between TFs or that it varies according to the transcriptional output of the axis patterning enhancers.

Nucleosome destabilization and remodeling are required for the exposure of TF binding sites to their input regulators [66]. Consistently, axis patterning enhancers are characterized by low nucleosome occupancy and high accessibility in the cellular blastoderm [67]. Interestingly though, nucleosome depletion is also observed at these CREs already prior to zygotic genome activation, when the patterning gene regulatory networks are not yet fully in operation [68, 7].

On the other hand, in spite of axis patterning enhancers being globally identified as nucleosome-free regions, there is evidence directly implicating nucleosomes in mediating interaction between input TFs. Ectopic overexpression of the gap repressor Knirps has been demonstrated to increase local nucleosome occupancy of its target *even skipped* enhancers, as manifested by lower sensitivity to MNase digestion and elevated ChIP signal of histone H3 [28]. Importantly, within the entire *eve* locus, only those enhancers that showed sensitivity to Knirps input displayed specific alterations in nucleosome organization.

Additionally, short-range repression has been proposed to be mediated via direct interactions with a nucleosome [69]. Analysis of synthetic enhancers revealed a non-linear distance dependence between the binding sites of repressor and activator TFs [59]. Suppression by short-range repressors was most efficient at the distance of 5 bp (as accounted for by direct competition), decreased at around 30 bp, peaked at 50-60 bp and declined again at larger distances. Teif and Rippe [69] successfully modelled this non-monotonic relationship by assuming local nucleosome stabilization by the repressor and partial unwrapping of the nucleosome. They proposed that at the distance of 6-50 bp activator and repressor TFs display cooperative binding by stabilizing the unwrapped state of the nucleosome and mutually promoting accessibility of their binding sites (collaborative competition, [70]). On the other hand, a partially unwrapped nucleosome can fit between the activator and repressor binding sites when separated by 50-150 bp. With the repressor stabilizing the nucleosome and determining its position, the distance of 50-60 bp corresponds to the core of the nucleosome that is least accessible for activator binding. While accounting for the experimental observations [59, 28], stabilization of a nucleosome by a short-range repressor has not been rigorously tested yet.

It is also important to note that activator and repressor TFs can potentially modify the chromatin environment of their target axis patterning enhancers through their co-regulators, which recruit histone acetyltransferases and deacetylases, respectively [71]. Co-activator CREB-binding protein (CBP) has been demonstrated to acetylate several histone residues, including H3K27 [72, 73]. It is required for the process of body patterning [74], and although it appears to have a predominant role in the DV network [75], it is also recruited by AP regulators [76]. Importantly, CBP has been demonstrated to promote occupancy of its partner TFs at their target enhancers [76]. At the same time, repressors of the AP network primarily interact with CtBP and Groucho co-repressors [77, 78] that both recruit histone deacetylases, in particular Rpd3 [79, 80]. Importantly, Rpd3 mediates deacetylation of H3K27ac and has been shown to be essential for correct progression of the segmentation process [81].

Consistently, it was recently demonstrated that axis patterning enhancers of the DV network display different histone modifications on their flanking nucleosomes in different regions of the cellular blastoderm. Active enhancers, i.e. promoting transcription of their target gene, are characterized by higher levels of H3K27ac as compared to the inactive enhancers [82]. However, the direct causal link between differential interactions with co-regulators and differences in histone acetylation levels still needs to be proven. Additionally, even though suggesting such a link, the authors have not directly demonstrated the effect of histone modifications on local accessibility of enhancer sequences.

Overall, the aforementioned studies suggest and partially demonstrate a role of the local chromatin organization in mediating the regulatory activity of axis patterning enhancers. However, the exact functional relationship and molecular mechanisms are not known at this moment.

1.3 Research questions

1.3.1 Alternative models of chromatin organization at axis patterning enhancers

Traditionally, axis patterning enhancers were analyzed and modeled with the predominant focus on the composition and arrangement of their TF binding sites. How can this view be reconciled with the proposed involvement of nucleosome positioning and chromatin modifications in their regulatory activity? In light of the studies discussed above, I envision two alternative models that relate the chromatin context to the operation of axis patterning enhancers.

Following assumptions of the computational models [62, 63, 64], it can be hypothesized that differential activity of the enhancers along the body axis is purely a reflection of differential TF occupancy, as determined by TF concentration gradients. Taking into consideration global nucleosome depletion of axis patterning enhancers [67], the enhancers would be expected to display equally high accessibility to different combinations of TFs throughout the embryo. TFs would be also able to probe all their potential binding sites along the entire length of the enhancer with the same probability. In this scenario, at cycle 14A when the patterning networks are in full operation, the axis patterning enhancers are expected to be characterized by highly open chromatin structure in all regions of the cellular blastoderm.

In an alternative model that takes into consideration the recent insights into the chromatin organization of axis patterning enhancers [82, 28], a single regulatory element can be hypothesized to display multiple different chromatin states at different regions of the embryo. This heterogeneity could be potentially a reflection of differential regulatory activity of the enhancers along the body axis. The chromatin context may be potentially modulated by differential occupancy of input TFs, possibly through their interactions with nucleosomes or histone-modifying properties of their co-regulators.

The second model requires an active interplay between TFs and chromatin, while the first one assumes the chromatin structure to be essentially invariant along the patterning axes. This study aims, therefore, to distinguish between these two hypotheses and examine regional variation in the chromatin organization of axis patterning enhancers.

1.3.2 Current state of knowledge

To distinguish between these two models, it is necessary to examine whether the chromatin state of axis patterning enhancers varies spatially in the *Drosophila* cellular blastoderm. At the time when this study was started, the only available genome-wide profiles of nucleosome occupancy [7], histone modifications [83] and chromatin accessibility [67] had been acquired from whole blastoderm embryos. As a result, any potential differences along the body axes were convoluted into a single whole-embryo average.

Two studies addressed this spatial heterogeneity by assaying mutant embryos with re-

duced or uniform cell identities. Li and Arnosti [28] performed ectopic overexpression of *Knirps*, the gap repressor, in order to test its effect on nucleosome occupancy and histone acetylation levels at its target enhancers. Koenecke et al. [82] profiled two histone marks H3K27ac and H3K4me in maternal mutants that affected the activity of *Dorsal* and uniformly represented mesodermal and dorsal ectodermal precursor cells. While both studies provided valuable insights into the chromatin context of axis patterning enhancers, they did not offer a global picture of nucleosome organization that would allow for distinction between the two proposed models. In general, assays in a mutant background are informative only for a limited set of regions of interest. Ectopic expression of key regulators might have an unknown effect on the operation of the whole gene regulatory network, making it difficult to relate the mutant states to cell identities from the wild-type embryos.

1.3.3 Outline of the study

The aim of this study was to profile chromatin organization of axis patterning enhancers with increased regional resolution. I focused on the enhancers from the AP network as they define activity domains that vary considerably in size and position within the blastoderm embryo (Figure 1.1). With fairly narrow concentration gradients of AP TFs, the enhancers are also expected to receive highly variable input along the embryonic axis, therefore representing multiple different regulatory states. I used chromatin accessibility as a global metric for the chromatin organization, since accessibility of linker DNA has been demonstrated to be jointly modulated by the positioning and stability of nucleosomes as well as their histone modifications [84]. Taking into consideration dynamic progression of embryogenesis [38] and, in particular, to probe activity of the AP network after global zygotic genome activation, I specifically assayed the stage of cellular blastoderm (stage 5, cycle 14A).

I achieved high spatial resolution by genetically tagging nuclei in well-defined domains of the blastoderm embryo, followed by specific isolation with immunoaffinity purification [85, 86, 87]. I profiled chromatin accessibility of seven domains along the AP axis with ATAC-seq [88]. Fragmentation of native chromatin with a non-specific Tn5 transposase and mapping of short DNA products allows for identification of nucleosome-depleted, highly accessible regions with a single-bp resolution.

During the course of this work, two other published studies examined regional variation in chromatin accessibility of the *Drosophila* blastoderm. Cusanovich et al. [89] performed single-cell ATAC-seq on individual nuclei from dissociated embryos, while Haines and Eisen [90] directly profiled accessibility in cryo-sliced anterior and posterior halves of the cellular blastoderm. Comparison of the experimental approaches as well as insights from these two studies is included in the Discussion.

1.3.4 Summary of the results

I demonstrate that immediately after the onset of zygotic genome activation one quarter of the accessible genome shows significant quantitative variation in its ATAC-seq signal

along the AP axis. Since the most variable regions correspond to the annotated as well as putative axis patterning enhancers, I conclude that differential accessibility is a signature of patterning *cis*-regulatory elements in the *Drosophila* cellular blastoderm. More specifically, I demonstrate that regional differences in accessibility of axis patterning enhancers correlate with their differential regulatory activity. The enhancers display elevated accessibility when they receive a net activating input and promote transcription of their target gene, while their accessibility is consistently reduced when they receive a net repressive input. I conclude that the chromatin context plays an integral role in mediating regulatory activity of axis patterning enhancers. I also discuss potential mechanisms by which accessibility of the enhancers may be modulated by activator and repressor TFs.

Chapter 2

Experimental strategy

I characterized regional variation of chromatin accessibility in the cellular blastoderm of *Drosophila melanogaster* by performing ATAC-seq on specific nuclei populations isolated from well-defined domains along the AP axis. Specific sections of the embryo were genetically tagged with a nuclear envelope marker that was expressed under control of well-characterized axis patterning enhancers. The tagged nuclei were isolated from carefully staged embryos by antibody pull-down (INTACT), followed by limited transposase digestion of native chromatin (ATAC-seq). Short digestion products were subsequently isolated, sequenced and mapped to the reference genome in order to identify highly accessible genomic regions (Figure 2.1).

2.1 Experimental methods

2.1.1 INTACT: genetic tagging and affinity purification of specific cell-types

Genetic tagging and antibody-purification of blastoderm nuclei applied in this study is an adaptation of the previously described method for specific and efficient affinity purification of tagged nuclei from complex tissues, INTACT [85, 86, 87].

The INTACT protocol was originally developed in *Arabidopsis thaliana* [85] and involved affinity purification of a biotinylated nuclear tag with streptavidin beads. Subsequently, the assay was adapted to animal models, and simplified to immunoaffinity purification of the nuclear marker with a specific antibody conjugated to magnetic beads [86, 87]. INTACT has been successfully applied to heterogeneous tissues at different stages of *Drosophila* development, i.e. embryonic mesoderm [87] and adult neurons [86]. The protocol was also proven compatible with subsequent isolations of nuclear RNA (RNA-seq [86]), as well as genome-wide assays of nucleosome occupancy (MNase-seq [87]) and epigenetic chromatin marks (ChIP-seq [86]). Importantly, INTACT was demonstrated in multiple model systems to provide high purity and efficiency of nuclei isolations. For example, Henry et al. [86] report a yield of 15-50% and purity of 99% for antibody-based

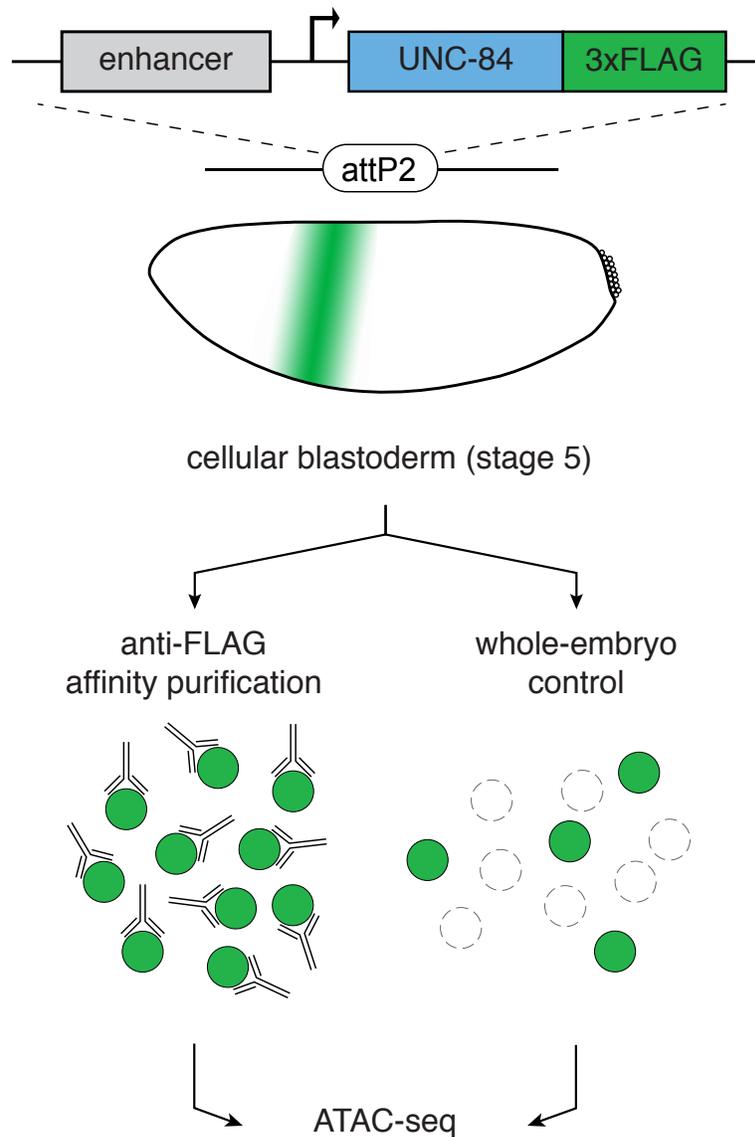


Figure 2.1: Summary of the experimental protocol. Selected AP domains are targeted in the cellular blastoderm by expressing a nuclear tag, UNC84-3xFLAG, under control of well-characterized enhancers of gap and pair-rule genes. All reporter constructs are integrated at the same genomic site (attP2 [91]) to standardize genetic background and integration-site dependent effects. After homogenization of staged embryos, tagged nuclei are affinity-purified with anti-FLAG antibodies, followed by Tn5 transposase fragmentation and ATAC-seq library preparation. Alongside, an ATAC-seq library representing an entire pool of nuclei from homogenized embryos serves as a control (whole-embryo control).

purification of tagged neuronal nuclei. Finally, the simplicity of immunoaffinity purification of INTACT is an advantage over alternative methods for isolation of specifically tagged cell populations, which involve time-consuming FACS sorting and require cross-linking of chromatin: BiTS-ChIP (batch isolation of tissue-specific chromatin [54]) and FANS (fluorescence-activated nuclear sorting [92]). For example, 8 hours of cell sorting yields 40 million nuclei in BiTS-ChIP [93], while as many as 11 million nuclei can be purified within 30 min with INTACT [86].

INTACT presents several advantages in the context of this study. Genetic tagging with a nuclear marker ensures precise definition of spatial coordinates of the tagged AP domains. Combined with the high reported specificity of immunopurification, this should allow for high reproducibility of nuclei isolations from batch collections of embryos. Finally, fast processing times with a short step of antibody pull-down limits perturbation of the native chromatin organization in unfixed nuclei, making INTACT particularly suitable for chromatin accessibility assays.

2.1.2 ATAC-seq: chromatin accessibility assay

Accessibility of DNA in the chromatin context is jointly modulated by nucleosome positioning [94, 95] as well as histone modifications that affect both nucleosome stability and organization of linker DNA [96, 97, 98, 99, 100]. I used accessibility as a global metric for chromatin structure in *Drosophila* embryos and probed it on a genome-wide scale with a recently developed chromatin accessibility assay, ATAC-seq [88].

ATAC-seq relies on the same principle as the alternative long-used assay DNase-seq [101]. Both methods probe sensitivity of chromatin DNA to fragmentation by a small protein that has a comparable size to an average TF. DNase-seq deploys the endonuclease DNaseI, while ATAC-seq relies on an engineered version of Tn5 transposase [102]. The proteins are characterized by minimal sequence specificity [103, 104]. However, due to steric hindrance, both DNaseI and Tn5 transposase are restricted from introducing cuts within DNA that is already associated with other proteins, in particular histones. As a result, they primarily target exposed stretches of naked DNA. This means that during a limiting digest of native chromatin, cleavage frequency of an individual base-pair is determined by how easily its locus can be reached by the protein in the nucleus as well as the locally incurred steric hindrance. Consequently, open chromatin regions are most likely to be represented by short fragmentation products that result from two independent cleavages in close proximity. The fragments are isolated and size-selected during the subsequent step of DNA purification, followed by Illumina sequencing and mapping to the reference genome. DNaseI- and transposase-sensitive regions are characterized by nucleosome-depletion [94, 95] and have been widely demonstrated to co-localize with active regulatory elements of the genome, including promoters, enhancers, silencers, insulators and locus control regions [88, 105, 67].

ATAC-seq uses an engineered version of Tn5 transposase that has been initially designed for rapid and efficient generation of sequencing libraries from isolated genomic DNA [102]. It is associated with standard adaptors used for Illumina sequencing and their integra-

tion directly results in DNA cleavage. As the process of fragmentation (also referred to as "tagmentation") is immediately coupled to the first step of library preparation, the experimental protocol is considerably simplified in comparison to DNase-seq [88]. This substantially reduces the amount of starting material. As a result, ATAC-seq has been already successfully applied to individual *Drosophila* blastoderm embryos [68] as well as single nuclei [89].

2.2 Generation of transgenic strains with tagged AP domains

2.2.1 Selection of nuclear tags

Two nuclear envelope proteins have been combined with an epitope tag in the published INTACT studies of *Drosophila* tissues: RanGAP [87] and UNC-84 [86]. RanGAP (Ran GTPase activating protein) mediates the nucleocytoplasmic transport and localizes to cytoplasmic filaments of the nuclear pore [106, 107]. UNC-84 is a SUN-domain protein of *C. elegans*, that plays a role in the process of nuclear migration [108]. It is embedded in the inner nuclear membrane, with its C-terminal domain localizing to the perinuclear space. As homogenization of tissues in the presence of non-ionic detergents results in removal of the outer nuclear membrane [109], UNC-84 is expected to provide higher efficiency of INTACT purifications. While Henry et al. [86] report a minimal effect of ectopic UNC-84 on the viability of transgenic flies, its overexpression has been observed to induce apoptosis in *Drosophila* larvae (Andrea Ennio Storti, personal communication). As a result, I tested both RanGAP and UNC-84 for their suitability as INTACT tags in *Drosophila* blastoderm (Figure 2.2).

In the published INTACT protocols for *Drosophila* tissues, both RanGAP and UNC-84 are fused to a 3xFLAG epitope that is used subsequently for purification of the tagged nuclei with anti-FLAG antibodies [86, 87]. The fusion proteins also include additional elements, such as a biotin ligase recognition peptide or fluorescent proteins, which aid in quality control and comparisons between streptavidin- and antibody-based purifications. Taking into consideration that *Drosophila* embryogenesis is a very dynamic process with stage 5 lasting only 40 min [110], I aimed to ensure fast translation and folding of the nuclear tag by simplifying it to a C-terminal fusion of 3xFLAG to RanGAP and UNC-84 (Figure 2.2).

2.2.2 Selection of axis patterning enhancers

I selected seven axis patterning enhancers of the AP gene regulatory network as drivers of the nuclear tag in blastoderm embryos. Their activity domains varied in size and position along the AP axis, with enhancers of gap genes (*hunchback*, *Dichaete*, *Kruppel* and *giant*) marking broad domains of the embryo and enhancers of the pair-rule gene *even skipped* defining narrow stripes (Figure 2.3).

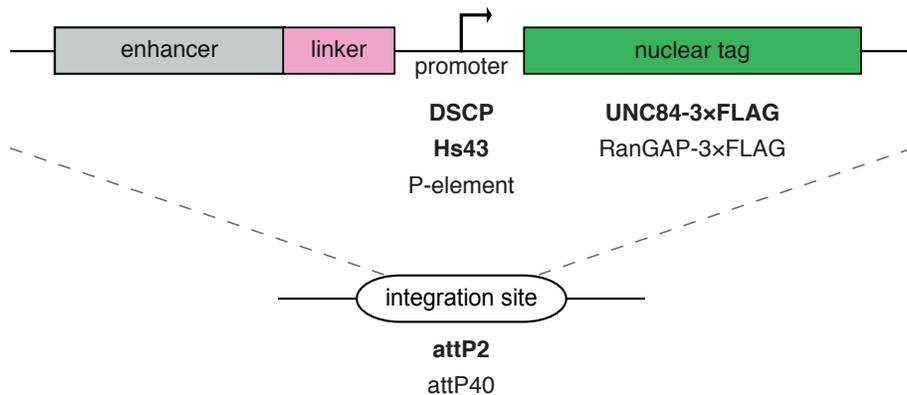


Figure 2.2: Design of the expression cassette. Expression of the nuclear tag was driven by selected axis patterning enhancers that were separated by a 69-bp linker from the basal promoter. Different variants of the promoter, nuclear tag and integration site are listed below the respective elements. In bold: final selection of elements used in the ATAC-seq experiments.

When selecting the drivers, I relied on a curated list of *cis*-regulatory elements used in a thermodynamic model by Segal et al. [64]. It is important to note that genomic coordinates of some elements deviated from their original definitions in the studies that tested and characterized their enhancer activity (Supplementary Material B). Modifications introduced by Segal et al. [64] included: merging of two overlapping elements (*eve_1-ru*), removal of the gene body sequence (*hb_anterior_activ*) and extension of the enhancer to include the flanking predicted TF binding sites (*eve_stripe2*, *Kr_CD1-ru*, *eve_stripe5*). It is unknown, though, to what extent these changes of enhancers' coordinates affect the position and robustness of their activity patterns.

As the global zygotic genome activation occurs only at the beginning of stage 5 of *Drosophila* embryogenesis [111], it was necessary to ensure fast expression of the nuclear tag so that it would reach sufficiently high levels in the assayed cellular blastoderm. In addition to deploying enhancers of early zygotic enhancers, they were combined directly with a basal promoter to reduce any delays in transcription initiation. The expression constructs additionally included a small 69-bp linker (Figure 2.2) to allow for enhancer-promoter looping [112].

While ensuring high expression levels of the nuclear tag for high efficiency of affinity purification, it was also important to control for negative effects of UNC-84 overexpression. For this purpose, I initially deployed three different *Drosophila* basal promoters to assay differences in the promoter-induced transcription levels and also test for promoter compatibility with axis patterning enhancers (Figure 2.2). 1) DSCP (*Drosophila* Synthetic Core Promoter, [91]) is a versatile synthetic promoter with multiple core promoter motifs, designed to provide robust expression with a broad range of enhancers. 2) Hs43 [113] is a minimal version of the *Drosophila* Hsp70 promoter that lacks the heat-shock response

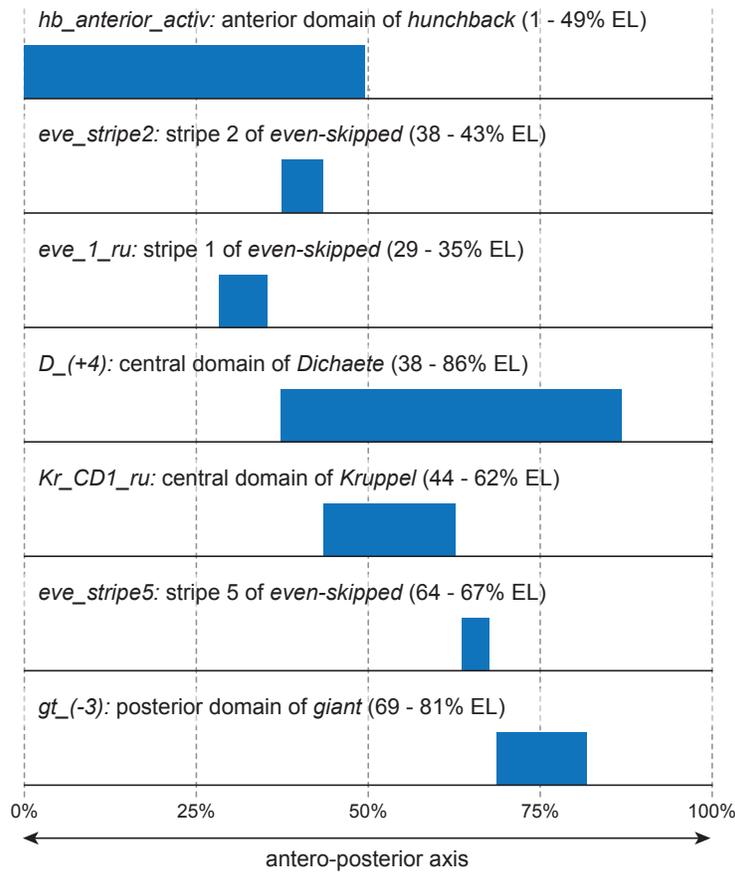


Figure 2.3: Reported activity patterns of the seven enhancers of gap and pair-rule genes used as drivers of the nuclear tag. Indicated corresponding expression domain of the target gene as well as its position along the antero-posterior axis (EL: egg length). 0%: anterior tip. 100%: posterior tip. Gap genes: *hunchback*, *Dichaete*, *Kruppel* and *giant*. Pair-rule genes: *even skipped*.

element [114]. 3) P-element basal promoter [115] contains the promoter and first intron of P-element transposase. All of the considered basal promoters are commonly used in standard expression vectors.

2.2.3 Assembly of expression constructs

All elements of the expression cassette: enhancer, linker, basal promoter and nuclear tag (Figure 2.2) were combined into a minimal vector pBDP [91]. The expression constructs were subsequently integrated with phiC31-mediated recombination into *Drosophila* genome in the wild-type background [116]. I initially tested two insertion sites, attP2 [116] and attP40 [117], which were previously demonstrated to support tissue-specific expression

from a wide range of expression constructs [118].

I used Golden Gate Assembly [119, 120] to efficiently generate expression constructs with different combinations of tested elements (Figure 2.4). The cloning strategy relies on Type II restriction enzymes that cleave DNA outside of their recognition sites, producing 4-nucleotide overhangs that subsequently serve as ligation sites. By combining complementary overhangs, this allows for directional assembly of multiple elements in a single reaction. The enzyme recognition site is independent from the sequence of the element and is eliminated after subcloning. As a result, individual elements of the final construct are separated only by 4-bp ligation sites.

More specifically, I amplified the 69-bp linker and different variants of the enhancer, basal promoter and nuclear tag with PCR primers, whose overhangs contained recognition sites of Type II enzyme (BsaI) as well as element-specific ligation sites (Figure 2.4, Step 1). The sequence of ligation sites that flanked individual elements determined their order in the assembled expression construct. Subsequently, I ligated individual elements into a backbone with the kanamycin resistance gene to generate a set of entry vectors (Figure 2.4, Step 2). For the final reaction of Golden Gate Assembly [119, 120] (Figure 2.4, Step 3), I combined entry vectors and the destination vector pBDP [91] with the BsaI restriction enzyme and DNA ligase. BsaI introduced cuts within the ligation sites, producing 4-nucleotide overhangs. Annealing of complementary overhangs allowed for ligation of the elements in a pre-defined order. With each variant of the same element being flanked by identical ligation sites, the strategy allowed for generation of multiple combinations of expression constructs with the same standard protocol. The pBDP backbone of expression constructs contained the ampicillin resistance gene (Amp^R), *E. coli* origin of replication (ORI), phiC31 attB integration site and marker gene (mini-white⁺). Because entry vectors contained the kanamycin resistance gene, culturing of *E. coli* transformed with the reaction mix in the ampicillin-enriched medium simplified selection of the correctly assembled constructs.

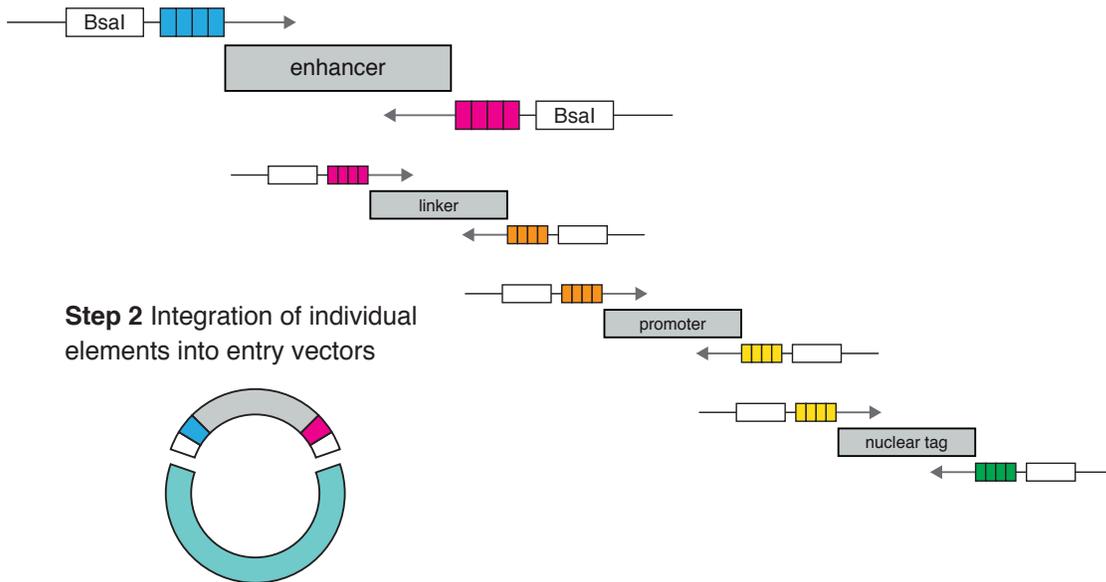
2.3 Validation of transgenic strains

Expression constructs were inserted either in the attP2 (third chromosome [116]) or attP40 (second chromosome [117]) integration sites in the wild-type background. After confirmation of correct integration of the expression cassette, homozygous strains were tested with anti-FLAG immunostaining for expression of the nuclear tag in a correct spatial domain along the AP axis.

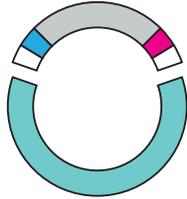
2.3.1 Selection of integration sites, basal promoters and nuclear tags

All generated strains produced detectable levels of the transgenic nuclear tag at stage 5 of embryogenesis. Importantly, expression of the protein was restricted to specific domains

Step 1 Addition of Bsal recognition sites and 4-bp ligation sites to individual elements of the expression cassette



Step 2 Integration of individual elements into entry vectors



Step 3 Golden Gate Assembly of expression constructs

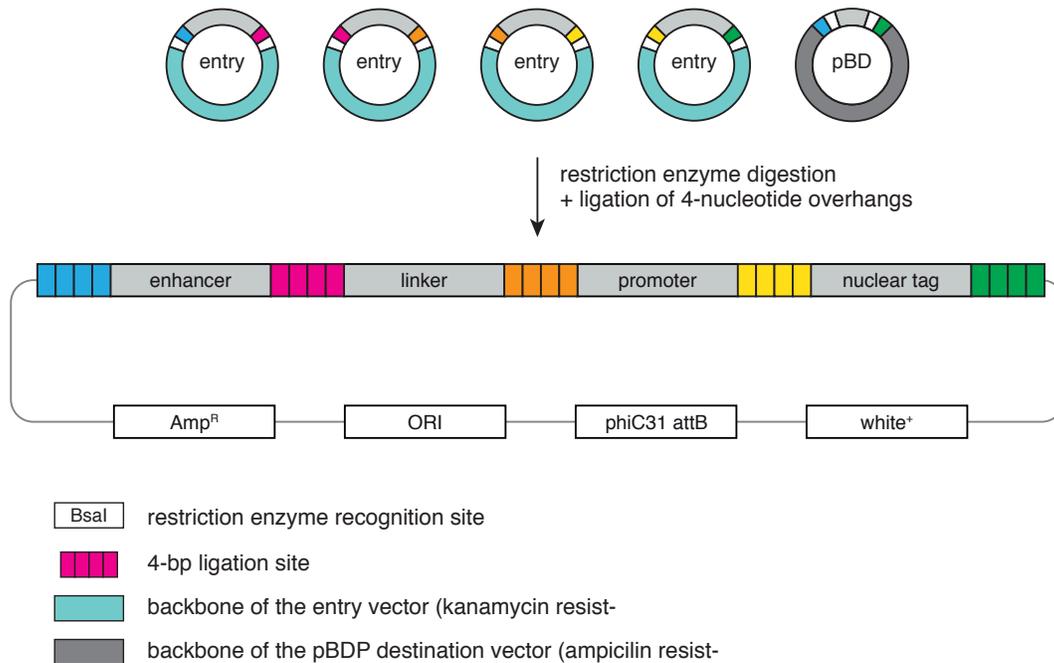


Figure 2.4: Summary of a cloning strategy used for the assembly of expression constructs.

along the AP axis. The anti-FLAG signal localized primarily to the nuclear envelope, with clear exclusion from the nuclei (Figure 2.5).

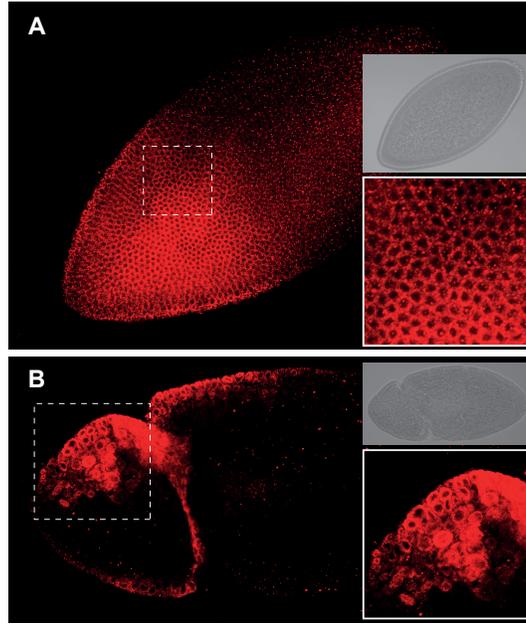


Figure 2.5: Localization of the transgenic tag to the nuclear envelope. Anti-FLAG immunostaining against RanGAP-3xFLAG (driven under control of *hb_anterior_activ* enhancer). Bottom panel represents a magnified view from the field marked with a dashed square. Upper panel shows an image of the embryo in the bright field. A) Stage 5 of embryogenesis. Note that tight packing of nuclei in the cellular blastoderm does not allow for clear discrimination between cytoplasmic and nuclear envelope localization of the anti-FLAG signal. B) Stage 9 of embryogenesis, with less compact organization of nuclei and clear demonstration of RanGAP-3xFLAG localization to the nuclear envelope.

No pronounced difference between expression from the two tested phiC31 landing sites was observed (data not shown), although Pfeiffer et al. [91] report higher transcription levels from the attP2 site as compared to the attP40 site. For this reason, attP2 was selected as an integration site for all transgenic strains.

Coupling of the same enhancer to different basal promoters gave rise to distinct expression levels of the nuclear tag (Figure 2.6). Based on comparisons of the immunostaining signal, embryos with the P-element promoter were characterized by fairly low levels of the transgenic protein, while DSCP provided the strongest expression of the nuclear tag. It is important to note that enhancers of *even skipped* gene are activated downstream of the gap genes in the AP gene regulatory network, allowing less time for transcription and accumulation of the tag in the nuclear envelope. Therefore, in order to ensure sufficient expression of the transgenic protein, enhancers of *even skipped* were coupled to the strongest DSCP

promoter, while enhancers of gap genes were coupled to the weaker Hs43 promoter in the set of strains used in the final ATAC-seq experiments.

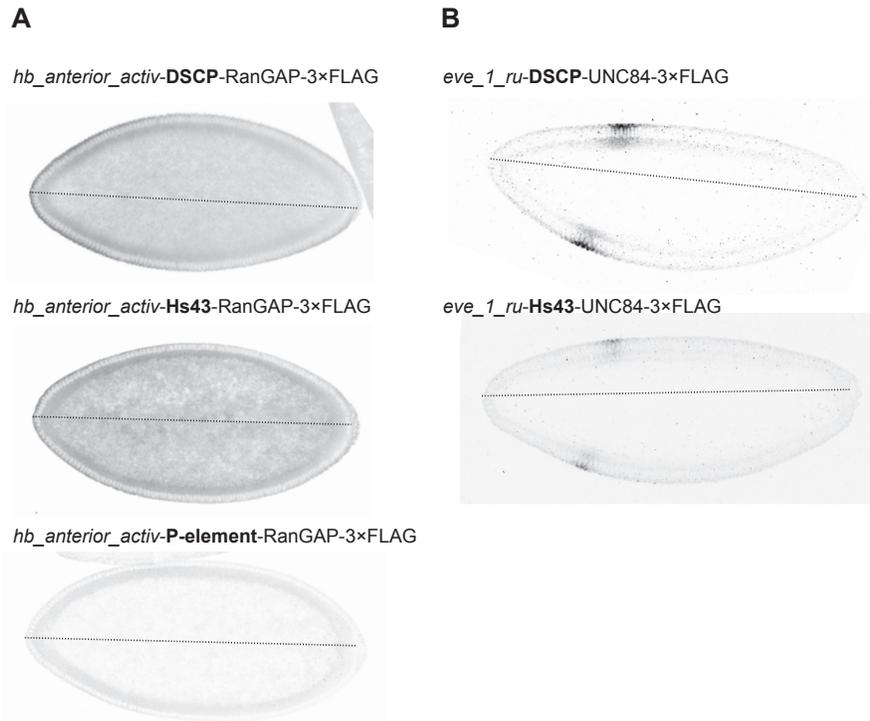


Figure 2.6: Comparison of expression levels driven by different basal promoters. A) Anti-FLAG immunostaining against RanGAP-3xFLAG expressed under control of *hb_anterior_activ* enhancer coupled to DSCP, Hs43 and P-element basal promoters. Images of embryos were acquired with the same settings of the confocal microscope. While the signal is hardly detectable for the P-element promoter, images of the two other strains are already dominated by noise under the same conditions. B) Anti-FLAG immunostaining against UNC84-3xFLAG expressed under control of *eve_1_ru* enhancer coupled to DSCP and Hs43 basal promoters. Stage-5 embryos are positioned with anterior to the left and dorsal side up. Dashed line marks the AP axis.

Importantly, RanGAP-3xFLAG and UNC84-3xFLAG tags were characterized by a different distribution along the AP axis when driven under control of the same enhancer. The RanGAP domains were consistently wider and more diffuse than those of UNC-84 (Figure 2.7). These discrepancies are most likely linked to the fact that the proteins localize to different sub-cellular compartments. RanGAP shuttles between the cytosol and cytoplasmic filaments of the nuclear pore [107] and its mRNA is exported and translated in the cytoplasm. Consequently, it can freely diffuse outside of its transcription domain still at the beginning of stage 5, when the cellularization processes are not yet complete. On the other hand, UNC-84 is targeted to the inner nuclear envelope and is translated at

the endoplasmic reticulum membrane, which restricts its diffusion in the syncytium. As a result, distribution of UNC84-3xFLAG more closely represents the activity domain of its respective enhancer driver (compare with Figure 2.3). Since it allows for definition of more spatially restricted domains of the embryo, UNC84-3xFLAG was chosen as a nuclear marker in the final set of transgenic strains.

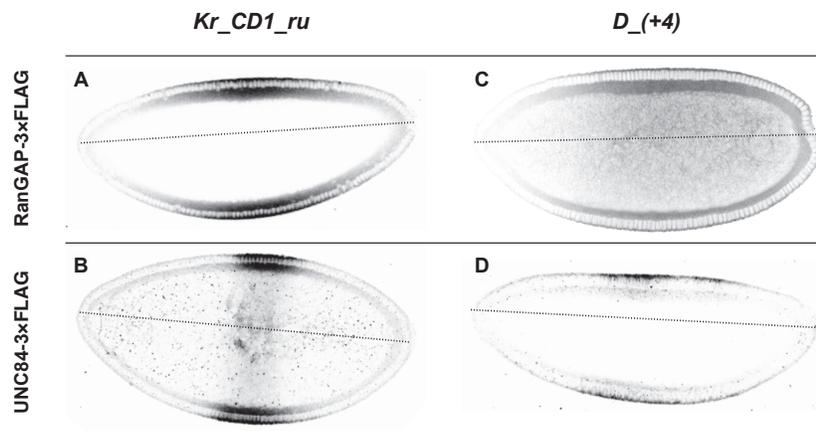


Figure 2.7: RanGAP-3xFLAG is characterized by wider expression domains than UNC84-3xFLAG. Anti-FLAG immunostaining of A) *Kr_CD1_ru*-DSCP-RanGAP-3xFLAG, B) *Kr_CD1_ru*-UNC84-RanGAP-3xFLAG, C) *D_(+4)*-DSCP-RanGAP-3xFLAG, D) *D_(+4)*-Hs43-UNC84-3xFLAG stains. Stage-5 embryos are positioned anterior to the left and dorsal side up. Dashed line marks the AP axis. Immunostaining and imaging was performed by Laura Fischer.

2.3.2 Characterization of tagged domains

Figure 2.8 summarizes seven transgenic strains whose tagged domains were assayed with ATAC-seq in this study. For simplicity, the domains were named D1-D7 according to their position along the AP axis (Figure 2.9). D1 represent the anteriormost domain of *hunchback*, while D7 represents the posteriormost domain of *giant*.

Coordinates of the tagged domains along the AP axis were measured based on anti-FLAG immunostaining of stage 5 embryos (Figure 2.9A). Position of the domains was approximated by projecting the antibody signal onto a line connecting the anterior and posterior tips of the embryo (Figure 2.9C) and was expressed as percent of the axis length (1-100%).

Positions of the tagged domains differ to a varying extent from the reported activity domains of the selected enhancer drivers (Figure 2.9B). In case of domains D3-D6, these differences are rather minimal and might primarily reflect technical variability. Activity domains were measured using the published RNA *in situ* hybridization images [64], while tagged domains have been characterized based on the expression pattern of the protein

Tagged domain	Enhancer	Basal promoter	Nuclear tag	Integration site
D1	<i>hb_anterior_activ</i>	Hs43	UNC84-3×FLAG	attP2
D2	<i>eve_stripe2</i>	DSCP	UNC84-3×FLAG	attP2
D3	<i>eve_1_ru</i>	DSCP	UNC84-3×FLAG	attP2
D4	<i>D_(+4)</i>	Hs43	UNC84-3×FLAG	attP2
D5	<i>Kr_CD1_ru</i>	Hs43	UNC84-3×FLAG	attP2
D6	<i>eve_stripe5</i>	DSCP	UNC84-3×FLAG	attP2
D7	<i>gt_(-3)</i>	Hs43	UNC84-3×FLAG	attP2

Figure 2.8: List of transgenic strains used in the ATAC-seq experiments, including a name of the corresponding tagged domain, composition of the expression cassette and genomic location of the integration site.

tag. Distribution of the transcript and translated protein are likely to differ, especially due to the previously discussed phenomenon of diffusion in the syncytial blastoderm. It is also important to note that both distributions represent a gradient of the signal. As a result, different intensity thresholds as well as different methods for overlaying the signal onto the AP axis could act as additional sources of variability.

On the other hand, the remaining three tagged domains are characterized by much stronger discrepancies. D7 shows strong posterior expansion in comparison to the reported activity domain of *gt_(-3)* enhancer. This is consistent with posterior extension of the endogenous expression domain of *giant* gene at the beginning of cycle 14, which gradually gets sharpened and anteriorly shifted [121]. D1 is characterized by retraction of the nuclear tag from the anterior tip and this pattern is also in agreement with reduced levels of native *hunchback* mRNA in the anteriormost region of the embryo [23]. Finally, UNC84-3xFLAG is expressed in a considerably wider D2 domain than the endogenous stripe 2 of *even skipped*. This is consistent with the fact that the minimal stripe 2 element (*eve_stripe2*) initially promotes transcription in a broad anterior region at the beginning of stage 5 [27].

These three examples illustrate another potential reason for the observed discrepancies between activity domains of the enhancers and positions of the tagged domains. Expression domains of gap and pair-rule genes dynamically evolve before being fully established at the end of stage 5 [3, 38]. At the same time, the nuclear tag appears to be characterized by rather high stability, taking into consideration that it is detectable at later stages of

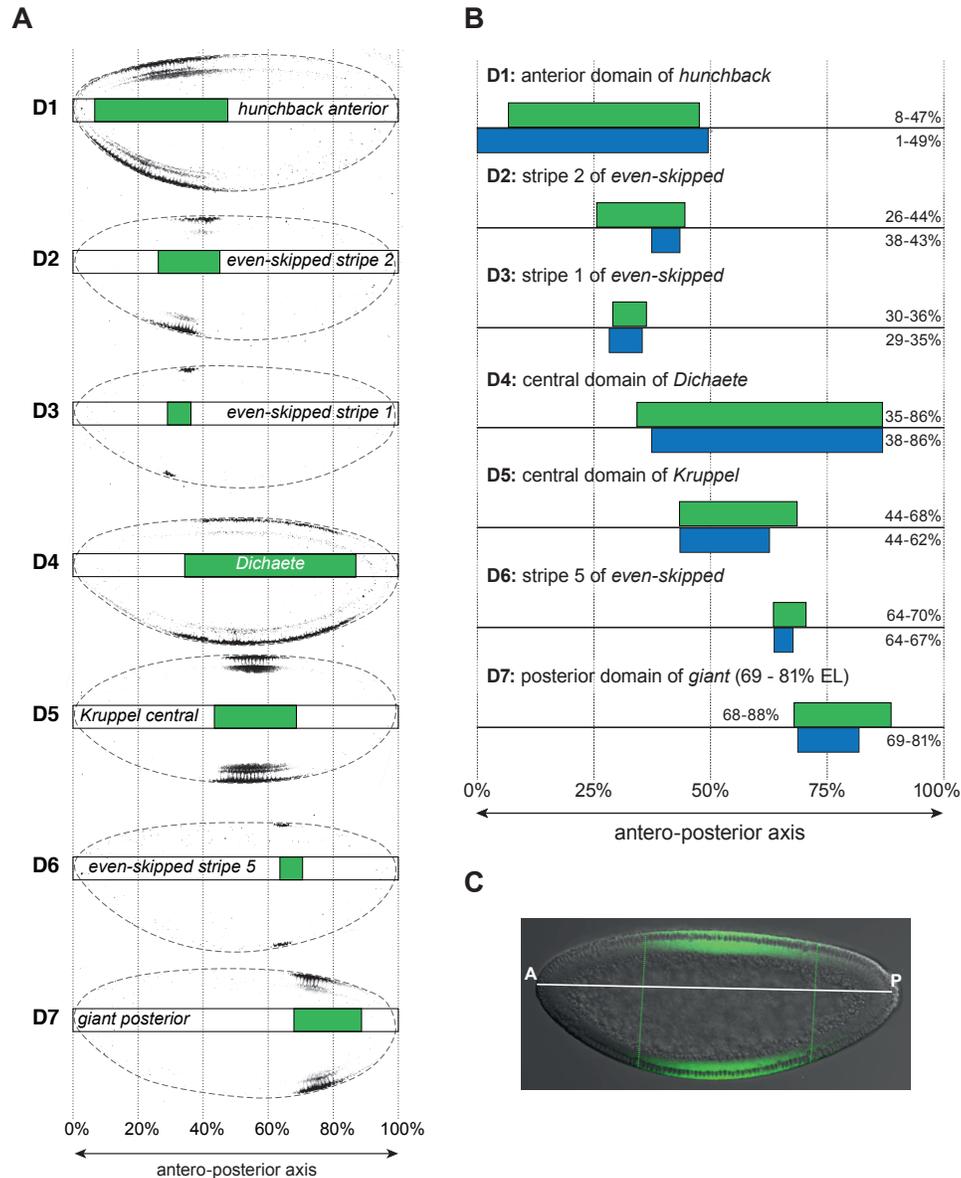


Figure 2.9: Overview of tagged domains (D1-D7). A) Representative images of anti-FLAG immunostaining of UNC84-3xFLAG in the cellular blastoderm of seven transgenic strains. Embryos are positioned with anterior to the left and dorsal side up. Each tagged domain is schematized to indicate its position along the AP axis (green bars). B) Comparison of expression domains of the nuclear tag (green) with reported activity domains of the enhancer drivers (blue, from [64]). Exact positions of the domains along the AP axis are additionally indicated (% of egg length). C) Strategy for measuring expression domains of the nuclear tag along the AP axis. Green lines connecting the anti-FLAG signal between the ventral and dorsal sides are overlaid onto the white line connecting the anterior (A) and posterior (P) tips of the embryo. The image represents cross-section of the blastoderm embryo at its widest point. Grey: interference contrast microscope. Green: anti-FLAG immunostaining of UNC84-3xFLAG.

embryogenesis (Figure 2.5). As a result, the tagged domain likely integrates all positions along the embryonic axis in which the enhancer drives transcription over time.

2.4 Generation of ATAC-seq libraries from tagged domains

2.4.1 Timing of embryonic collections

As chromatin accessibility has been demonstrated to dynamically evolve during *Drosophila* embryogenesis [68, 67], I aimed to specifically target a narrow time point of development. Reduction of the temporal dimension of the assay ensured that any potential differences between accessibility profiles from the tagged domains would primarily stem from regional variation along the AP axis. In this study, I assayed stage-5 blastoderm embryos that represented a time point immediately after zygotic genome activation, when expression patterns of gap and pair-rule genes are already established and the axis patterning enhancers receive a full spectrum of their regulatory input [3, 38]. This stage encompasses progression from the syncytial blastoderm to cellular blastoderm, with gradual invagination of the cellular membrane and separation of embryonic nuclei.

As *Drosophila* females do not always lay eggs immediately after fertilization, batch collections of embryos are prone to contamination with older stages [122]. To ensure high temporal resolution, I performed narrow 20-min collections of embryos. I carefully controlled temperature conditions in order to further reduce variation between independent batches (see Materials and Methods). I also tested different time regimes to obtain highest enrichment of the stage of cellular blastoderm. Overall, with collections at 2:50 - 3:10 h after egg laying, on average 85% of embryos represented the desired stage (Figure 2.10).

2.4.2 Experimental workflow

For each transgenic strain, I performed a series of eight collections from around 4-7 population cages (Figure 2.11). Staggered design allowed for careful control of timing. Embryos were flash-frozen in liquid nitrogen, except for the last collection that was fixed in formaldehyde for subsequent characterization of its temporal profile. On average, 100 μ l of embryos were acquired for each stain.

Embryos were subsequently homogenized, followed by a series of washes with a mild non-ionic detergent to isolate nuclei and remove yolk particles (Figure 2.11). The nuclei were affinity purified with magnetic beads coupled to anti-FLAG antibodies. Short incubation for 30 min at 4°C allowed for preservation of the native chromatin structure. Tagged nuclei, while still bound to the beads, were treated with Tn5 transposase, followed by DNA purification and PCR amplification to complete library preparation [88]. To the best of my knowledge, this is the first demonstration that magnetic beads are compatible with the process of chromatin tagmentation.

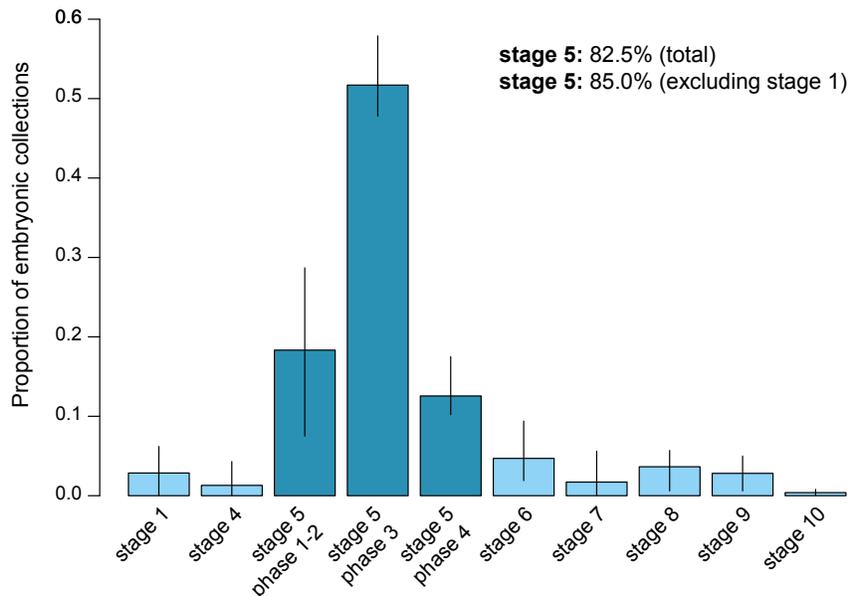


Figure 2.10: Composition of embryonic collections. Proportional representation of different embryonic stages in an average collection of embryos at 2:50 - 3:10 h after egg laying. Stage 5 is split into three subclasses, corresponding to phases defined by Schroeder et al. [51] and representing different morphology of elongating nuclei and invaginating cellular membranes. The measurements are based on seven independent collections ($n > 150$) of four transgenic strains: D1 Rep1, D4 Rep1, D5 Rep1 and D7 Rep1. Vertical bars represent minimum and maximum values. Above: average proportional representation of stage 5 in the collection (for individual collections this value ranged between 73% and 89%). Stage-1 embryos are likely to correspond to unfertilized eggs and, therefore, represent minimal contribution in the ATAC-seq libraries. Embryos were fixed in formaldehyde and examined under interference contrast microscope.

In order to ensure comparability between different ATAC-seq experiments, I performed all tagmentation reactions with the same amount of chromatin and Tn5 transposase. Taking into consideration different sizes of the tagged domains and potential variation in the efficiency of affinity purifications, this required careful quantification of the DNA content of the isolated nuclei (see Materials and Methods). The concentration of Tn5 transposase that offered the best signal-to-noise ratio of the ATAC-seq libraries was determined in a series of titration reactions (data not shown).

2.4.3 Summary of ATAC-seq samples

I assayed seven tagged domains in biological duplicates. Apart from D3 domain, each replicate corresponded to an independent transgenic strain with the same integrated expression construct. Only one strain was obtained for D3, yet the duplicates still represented

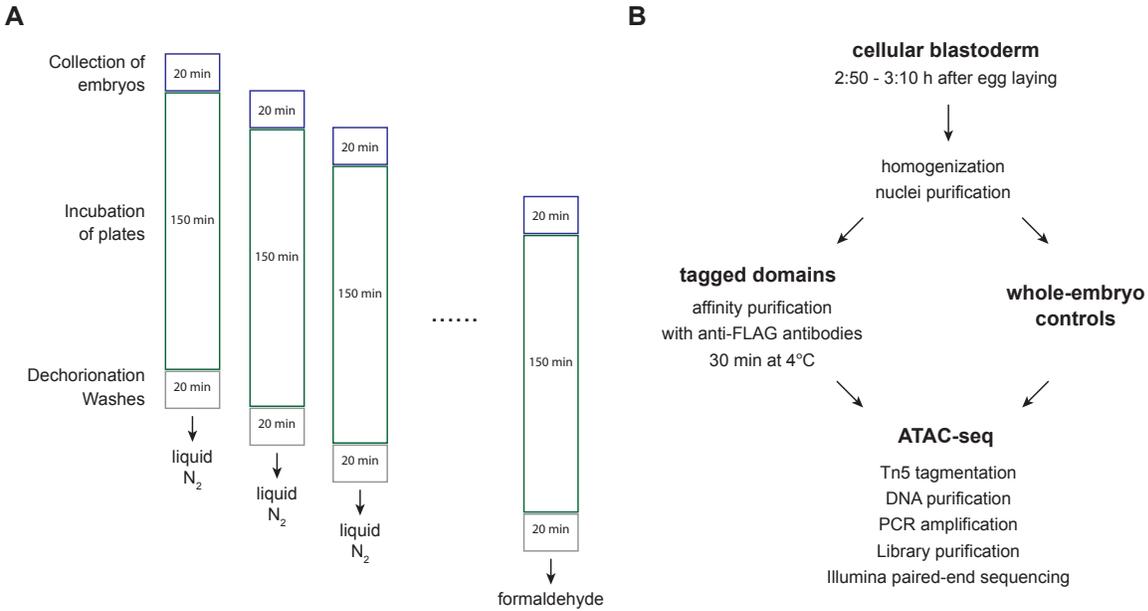


Figure 2.11: Detailed outline of the experimental protocol. A) Staging and collections of embryos. B) Affinity purification of nuclei and generation of ATAC-seq libraries.

independent collections of embryos, affinity purifications and transposase fragmentation reactions.

In addition, I performed a set of control experiments with Tn5 tagmentation on the entire pool of nuclei from staged embryos (whole-embryo controls). I used the same collections of embryos as for affinity-purification experiments on D1, D4, D5 and D7 domains, including biological duplicates. I followed the same experimental protocol, omitting only the step of antibody pull-down and immediately proceeding to Tn5 tagmentation after homogenization and nuclei isolation (Figure 2.11).

Chapter 3

Processing and quality control of ATAC-seq libraries

3.1 Interpretation of ATAC-seq signal

3.1.1 Definitions of ATAC-seq signal

Two definitions of ATAC-seq signal were used in this study: a) total number of Tn5 transposase cuts, and b) coverage of size-selected ATAC-seq fragments (Figure 3.1).

Distribution of transposase cleavages serves as a general measure of sensitivity of each base-pair to Tn5 fragmentation, and as such approximates the level of DNA exposure in the chromatin context. Importantly, Tn5 transposase binds as a dimer and inserts two adaptors separated by 9 bp, corresponding to two cleavage events [88]. As a result, position of the mapped cuts (5' ends of reads) does not exactly correspond to the sites that are directly accessible to and targeted by the transposase. In order to correct for that, Buenrostro et al. [88] recommend shifting the cuts by +4 bp (reads on the plus strand) and -5 bp (reads on the minus strand) in order to more faithfully represent the transposase binding site. The adjustment of cuts is commonly used and also performed in this study.

In addition to providing single bp-resolution of the accessibility landscape, ATAC-seq also offers an insight into the local organization of DNA-bound proteins. This information is encoded in the coverage of ATAC-seq fragments. Each fragment is generated by two independent tagmentation events. Size distribution of ATAC-seq fragments originating from a given genomic interval represents the probability of the two cleavages occurring within a certain distance. This probability is determined by: a) steric hindrance of Tn5 transposase (discussed below), and b) width of the intervening protein-occupied binding site as the transposase cannot access and introduce cuts within protein-associated DNA. As a result, ATAC-seq has been used to estimate the size of proteins bound within open chromatin regions, in particular to distinguish between nucleosome-associated and nucleosome-free regions (Figure 3.2).

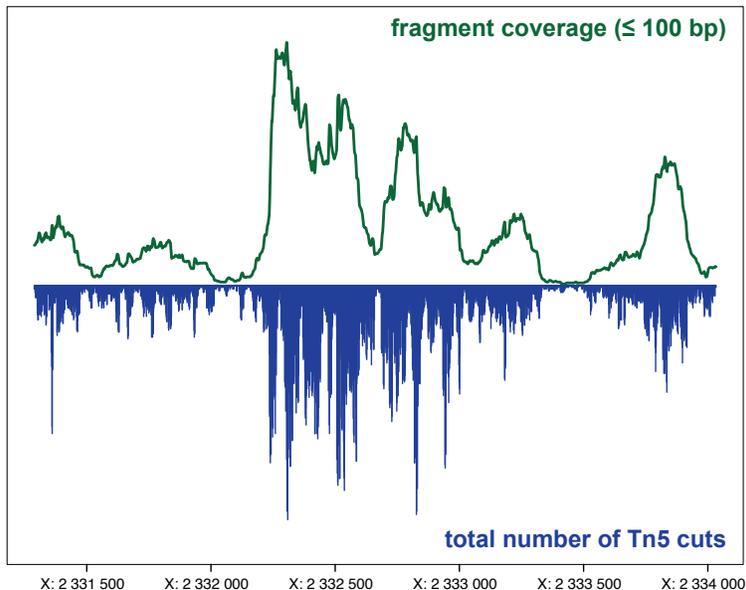


Figure 3.1: Two definitions of ATAC-seq signal. Upper panel (green): coverage of ≤ 100 bp ATAC-seq fragments (originating from nucleosome-free regions). Lower panel (blue): distribution of the total number of Tn5 transposase cuts (without *in silico* size selection). Locus: *gt_-(-10)* enhancer. Reference genome: UCSC dm3. For greater clarity, the profiles are smoothed over a sliding window of 5 bp. Y-axis represents the total number of ATAC-seq fragments / Tn5 cuts mapping to an individual base-pair.

3.1.2 Distinction between nucleosome-bound and nucleosome-free regions

By comparing transposase cut signatures against well-positioned chemically-mapped nucleosomes in *S. cerevisiae*, Schep et al. [123] were able to examine size distribution of ATAC-seq fragments that represent nucleosomal DNA. The study confirmed protection of histone-associated DNA from transposase cleavages, and demonstrated that the most abundant fragments spanning the nucleosome are 143 bp long (152 bp without shifting the cut sites). Importantly, the minimum size of nucleosomal ATAC-seq fragments was demonstrated to be 117 bp (126 bp without shifting the cut sites). First, this shows that the transposase can occasionally introduce cuts within histone-associated DNA, most probably as a result of partial uncoiling (stochastic breathing) of nucleosomal DNA [124]. Second, ATAC-seq fragments below 100 bp can be interpreted with high confidence to represent nucleosome-free regions.

With ATAC-seq performed on human cells, Buenrostro et al. [88] additionally demonstrate anti-correlation between genomic localization of short ≤ 100 bp ATAC-seq fragments that represent nucleosome-free regions and longer 180-247 bp fragments that represent nucleosome-associated regions. Importantly, the inferred nucleosomal intervals co-localize

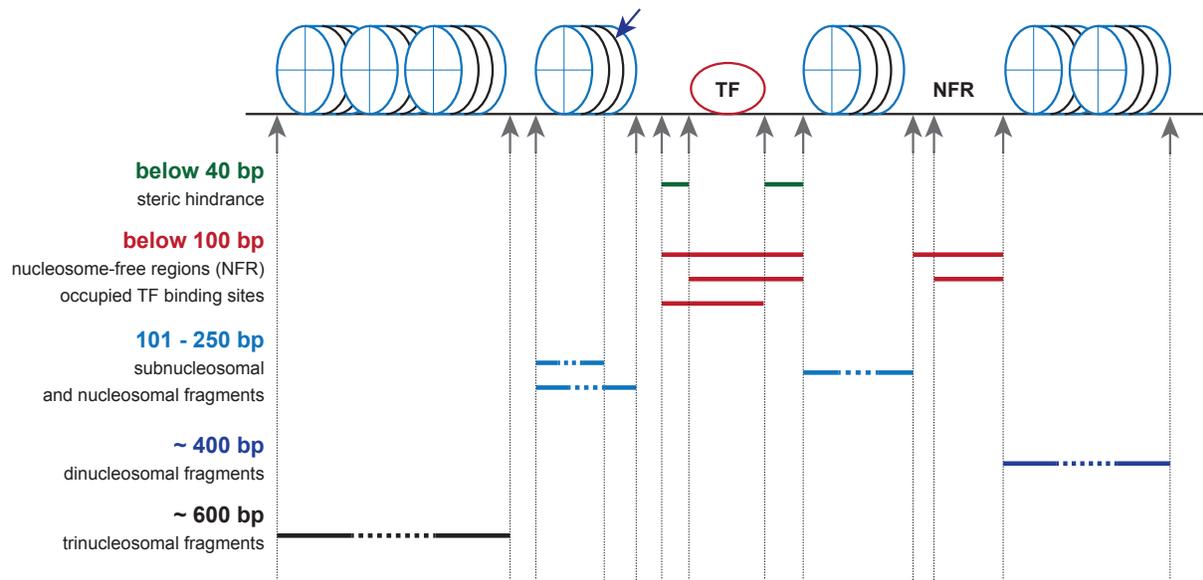


Figure 3.2: Size distribution of ATAC-seq fragments and their position in reference to DNA-bound proteins. Depending on the size and organization of DNA-bound proteins, transposase cleavages produce ATAC-seq fragments of varying size. Fragments below 40 bp (green) that represents cuts within very close proximity are rarely recovered due to steric hindrance of Tn5 transposase. Fragments below 100 bp (red) represent nucleosome-free regions (NFR) that are potentially occupied by TFs and other small proteins. Fragments above 100 bp represent nucleosome-associated regions: subnucleosomal and nucleosomal fragments (below 250 bp, light blue), dinucleosomal fragments (around 400 bp, dark blue) and trinucleosomal fragments (around 600 bp, black). Grey arrows: transposase cuts within exposed DNA. Blue arrow: transposase cut within nucleosomal DNA.

with chromatin immunoprecipitation (ChIP-seq) signal against histone marks. Moreover, the study shows that ≤ 100 bp fragments are enriched within different classes of enhancers, while they are depleted from promoter-flanking and repressed regions of the genome.

In summary, based on *in silico* size selection of mapped fragments, ATAC-seq allows for discrimination between nucleosome-occupied and nucleosome-free regions within broader open chromatin intervals. Since nucleosome remodeling and DNA exposure are an essential prerequisite for binding of DNA-specific proteins [66], in the context of distal *cis*-regulatory elements the nucleosome-free regions are likely to represent sites that are either poised for binding or already occupied by TFs.

3.1.3 Size distribution of ATAC-seq libraries

The different populations of ATAC-seq fragments are also reflected in the size distribution of the sequencing libraries generated in this study. It is possible to distinguish two partially overlapping peaks in the distribution of fragment sizes (Figure 3.3): a major peak with a

summit at 50 bp and a smaller peak centered at 200 bp. Most likely they represent the two classes of nucleosome-free and nucleosome-associated tagmentation products. Additionally, the distribution is characterized by distinct 10-bp periodicity. As it closely resembles the helical pitch of DNA, the periodicity has been previously suggested to result from transposase cleavages within the exposed minor groove of nucleosomal DNA [88]. Finally, the occurrence of fragments below 40 bp rapidly drops in all analyzed ATAC-seq libraries.

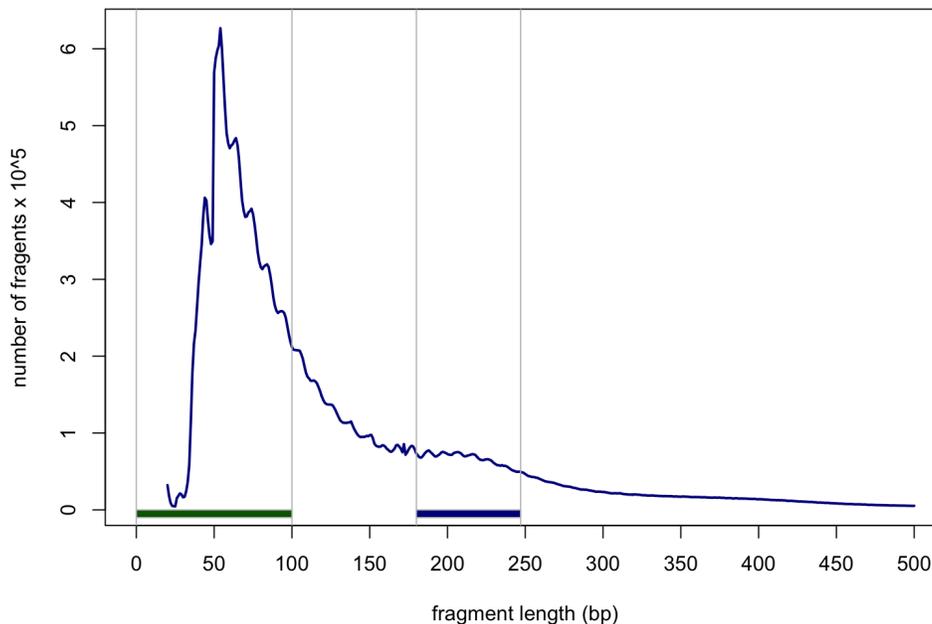


Figure 3.3: Representative size distribution of ATAC-seq libraries from the *Drosophila* blastoderm. Sample: D4 domain Replicate 1. Proportion of fragments ≤ 100 bp: 60.32%. Lower bars mark size thresholds used by Buenrostro et al. [88] to define nucleosome-free fragments (≤ 100 bp, green) and nucleosome-associated fragments (180-247 bp, blue).

This lower threshold of 40 bp has been previously linked to steric hindrance of Tn5 transposase [102]. The protein operates as a dimer and has been reported to occupy a stretch of 35-38 bp, which determines the minimum spacing between the neighbouring tagmentation events. Other causes for the reduced occurrence of very short ATAC-seq fragments could potentially include limited efficiency of DNA purification or their limited mappability to the reference genome. Silica-membrane columns (MinElute, QIAGEN) that are used in the standard ATAC-seq protocol for purification of tagmented DNA allow for efficient recovery of fragments above 70 bp. When considering the additional 60 bp of adaptors ligated by the transposase (Illumina), this corresponds to the minimum size of the tagmentation product being only 10 bp. On the other hand, the theoretical limit on the length of a sequencing read that can be uniquely mapped to the *Drosophila* genome is only 14 bp ($4^{14} = 268\,435\,456$; estimated size of the genome: 180 Mb; [125]). Therefore, steric hindrance of Tn5 transposase appears to be the major factor shaping the lower limit

of the size distribution of ATAC-seq libraries, with the steps of DNA purification and post-sequencing mapping having a negligible effect.

3.1.4 Genomic distribution of size-selected ATAC-seq fragments

In order to test whether ATAC-seq fragments of different size represent different genomic features in my experimental set-up, I examined their distribution over an example locus of *giant* gene. I performed *in silico* size selection and defined five classes of fragments:

- ≤ 100 bp: nucleosome-free fragments according to the definition by Buenrostro et al. [88]
- 180-247 bp: nucleosome-bound fragments according to the definition by Buenrostro et al. [88]
- 40-60 bp: corresponding to the shortest and most abundant population of fragments in the ATAC-seq libraries from the *Drosophila* blastoderm
- 101-125 bp and 125-150 bp: corresponding to the intermediate fragment sizes that were excluded from the analysis by Buenrostro et al. [88]. According to the analysis by Schep et al. [123], fragments below 125 bp are likely to represent nucleosome-free regions, while fragments below 150 bp are likely to represent subnucleosomal and nucleosomal fragments.

Figure 3.4 illustrates differences in the coverage of the five classes of ATAC-seq fragments over 21 kb of the locus of *giant*. The position of highly-accessible regions remains fairly conserved among the different size ranges. However, in comparison to the distribution of the total pool of fragments, size selection allows for recovery of more refined features. For a more quantitative assessment of the similarity between the coverage tracks, I also calculated their signal correlation in pair-wise comparisons (Figure 3.5).

First of all, the analysis confirms low correlation between nucleosome-free ≤ 100 bp and nucleosomal 180-247 bp fragments, with the Pearson correlation coefficient reaching 0.643. This limited correlation is especially evident at the promoter of *giant*, where the nucleosome-free and nucleosomal fragments form distinct non-overlapping peaks. Short fragments pile up into a narrow peak that most probably represents the RNA polymerase II initiation complex (upstream of TSS), while longer fragments form a wider peak that likely corresponds to the well-positioned +1 nucleosome [127].

Second, the shortest pool of 40-60 bp fragments is characterized by a very similar genomic distribution as the ≤ 100 bp fragments (coefficient: 0.965). This is in agreement with the fact that 40-60 bp fragments constitute a majority of tagmentation products from the nucleosome-free pool. However, this shortest pool of fragments that is likely to represent individual TF binding sites [128] forms sharper and more refined coverage peaks (e.g. *gt_-(-10)* enhancer). Importantly, the correlation between 40-60 bp and 180-247 bp fragments is lowest among all pair-wise comparisons (coefficient: 0.557).

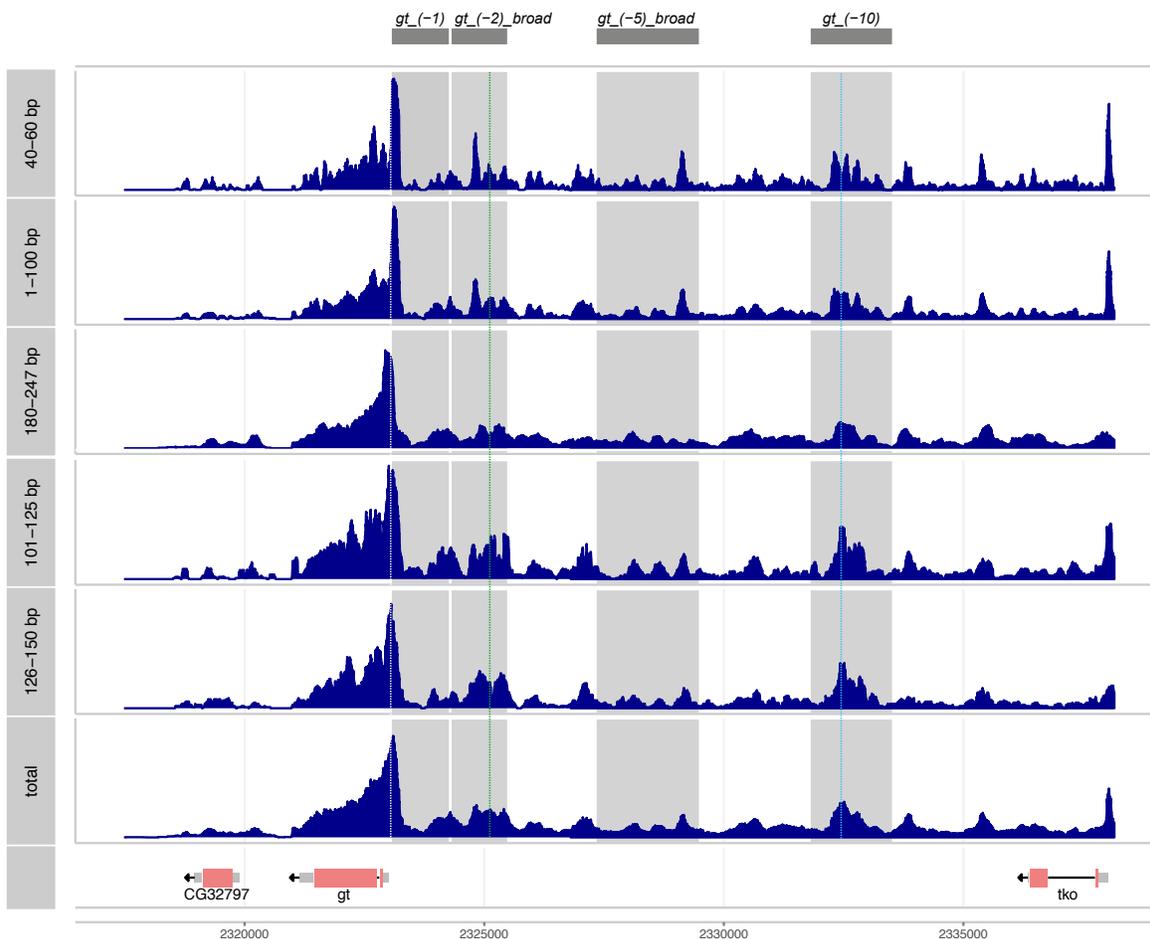


Figure 3.4: Distribution of different classes of size-selected ATAC-seq fragments over the locus of *giant*. Each track represents coverage of ATAC-seq fragments (without normalization), whose size ranges are indicated in the grey panel on the left. Grey bars above indicate position of well-characterized enhancers of *giant* (named as in Supplementary Material D, coordinates from [64]). White dashed line: transcription start site (TSS) of *giant*. Green dashed line: position of a nucleosome-free region within *gt_(-2)_broad*, according to the distribution of ≤ 100 bp fragments. Blue dashed line: position of a nucleosome-occupied region within *gt_(-10)*, according to the distribution of 180-247 bp fragments. Horizontal axis represents genomic coordinates along chromosome X (gene models: FlyBase Release 5.57 [126]).

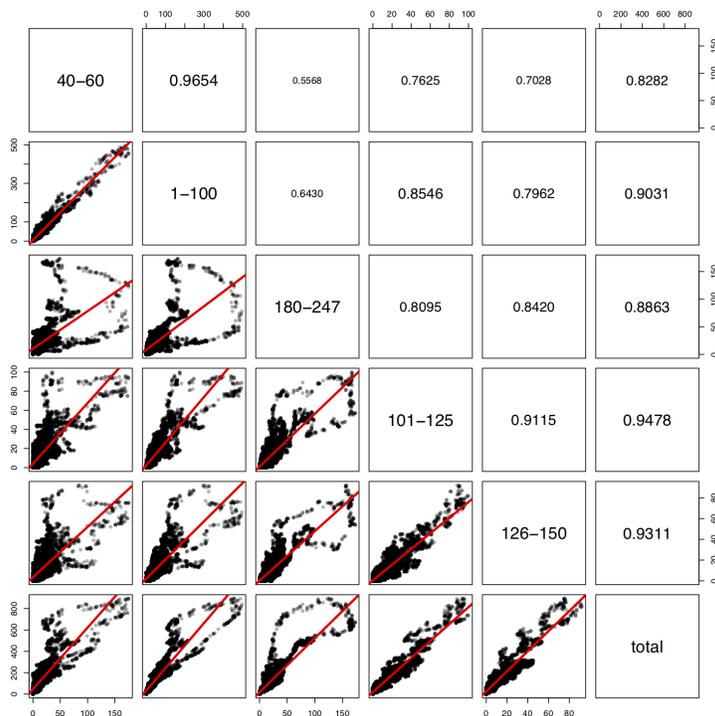


Figure 3.5: Correlation of coverage tracks of different classes of size-selected ATAC-seq fragments over the locus of *giant*. Different size ranges of fragments are indicated on the diagonal. Scatter plots represent pair-wise comparisons of the ATAC-seq coverage signal (without normalization; each data point corresponds to a single base pair). Panels on the right display Pearson correlation coefficients of the corresponding pair-wise comparisons.

Finally, 101-125 bp and 126-150 bp fragments appear to represent a mixed population of genomic regions. For example, within *gt_₋₁₀* enhancer the two pools strongly anti-correlate with the nucleosome-occupied region (marked with a blue line in Figure 3.4). On the other hand, within *gt_₋₂_broad* enhancer, 101-125 bp fragments localize to the nucleosome-free region, while 126-150 bp fragments show strong correlation with the nucleosomal fragments (marked with a green line in Figure 3.4). The mixed nature of these two classes of fragments is also strongly illustrated at the core promoter, where they show an intermediate position between the polymerase II and the +1 nucleosome peak. Based on the correlation analysis, the 101-125 bp size range appears more similar to the nucleosome-free fragments (0.855), while the 126-150 bp size range shows stronger similarity to the nucleosomal fragments (0.842). However, the distinction is not very pronounced. Thus, these two pools represent both subnucleosomal fragments and nucleosome-free regions that are potentially occupied by other big protein complexes. It is understandable therefore that Buenrostro et al. [88] excluded these size ranges from their analysis, in order to allow for unambiguous distinction between nucleosome-free and nucleosomal fragments.

Based the above analysis, I decided to use the coverage of ≤ 100 bp fragments to

represent nucleosome-free regions that are potentially occupied by TFs.

3.2 Identification of accessible genomic regions

3.2.1 Definition of ATAC-seq peaks

In order to obtain a genome-wide view of accessibility landscapes of the ATAC-seq libraries, the MACS2 algorithm [129] was used to identify genomic regions with significant signal enrichment.

MACS2, originally developed to analyze ChIP-seq data, has been also successfully used for peak calling with ATAC-seq data [90, 55, 130, 123]. In this study, signal enrichment was evaluated based on the distribution of Tn5 transposase cuts. In order to correct for sequencing artifacts and a potential sequence bias of the transposase, a control sample representing tagmentation of purified genomic DNA was used (contributed by Andrea Ennio Storti). The false discovery rate (FDR) threshold was set at 1%. ATAC-seq peaks localizing to heterochromatic regions (2LHet, 2RHet, 3LHet, 3RHet, XHet, YHet) and unmapped intervals (U: unmapped scaffolds) were removed from the analysis. After filtering, the number of peaks identified in individual libraries ranged between 19 000 and 21 000.

3.2.2 Definition of high-confidence ATAC-seq peaks

A set of high-confidence accessible regions was identified by intersecting ATAC-seq peaks from all eight whole-embryo controls. Only peaks mapping to euchromatic regions (chromosomes: 2L, 2R, 3L, 3R, 4 and X) were considered. 17 345 high-confidence peaks covered around 10 Mb, representing 8.3% of the genome. Their median size reached around 400 bp, with the most abundant peaks being 200-250 bp wide (Figure 3.6).

3.3 Reproducibility of ATAC-seq experiments

3.3.1 Comparison between biological replicates

Correlations between replicate ATAC-seq experiments allowed for assessing the reproducibility of the complex multi-step experimental protocol used in this study. The evaluation was based on the signal intensity rather than just the position of accessible regions, as it provides a more reliable measure of the quantitative sensitivity of the assay.

ATAC-seq signal was defined as the total number of Tn5 transposase cuts falling into each of the 17 345 high-confidence accessible regions. Biological replicates of the tagged domains (i.e. originating from independent transgenic strains) showed very high correlation of the signal, with their Pearson correlation coefficients ranging between 0.98 and 0.99 (Figure 3.7A-G). This confirms high reproducibility of the experimental protocol, including the potential sources of technical variability such as: staging and collection of embryos, affinity purification of tagged nuclei and transposase digestion of native chromatin.

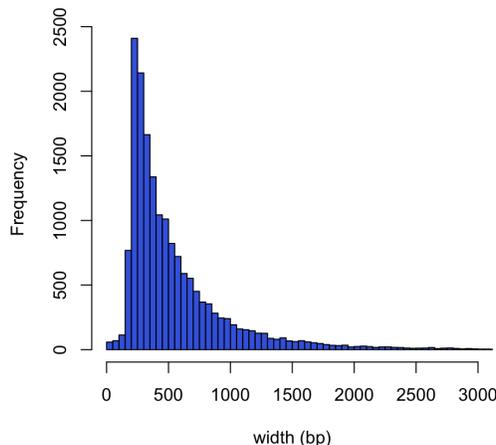


Figure 3.6: Size distribution of high-confidence ATAC-seq peaks. High-confidence peaks represent regions consistently identified as ATAC-seq peaks in all eight whole-embryo control experiments.

Whole-embryo controls were characterized by equally high correlations, both when comparing biological replicates carrying the same expression construct (Figure 3.8) as well as different transgenic strains (Figure 3.7H). High reproducibility between the transgenic strains confirms that expression of the nuclear tag in different regions of the embryo did not affect the genome-wide distribution of ATAC-seq signal.

3.3.2 Comparison with published chromatin accessibility profiles

In order to assess the quality of ATAC-seq profiles obtained in this study, I compared them to the available whole-embryo DNase-seq data from broadly staged blastoderm embryos [67]. I evaluated similarity of the whole-embryo controls to the published DNase-seq profiles both in terms of the extent of overlap between regions with significant signal enrichment as well as the quantitative correlation of the accessibility signal (Figure 3.9).

Among all high-confidence ATAC-seq peaks, 81.6% were identified as accessible regions in the DNase-seq assay, representing 78.8% of the accessible genome (Figure 3.9A). Among all DNase-seq peaks, only 54.1% overlapped ATAC-seq peaks, representing 54.0% of the accessible genome. On the other hand, the two datasets displayed very high quantitative correlation of their signal, defined as the density of either Tn5 transposase or DNaseI cut sites. The Pearson correlation coefficient reached 0.882 when signal intensities were compared over the entire set of 17 345 high-confidence ATAC-seq peaks (Figure 3.9B).

Such high correlation of the accessibility signal is remarkable taking into consideration the technical differences between the two assays, including: a) different timing of embryonic collections (20-min vs. 1-h time window for ATAC-seq and DNase-seq, respectively), b)

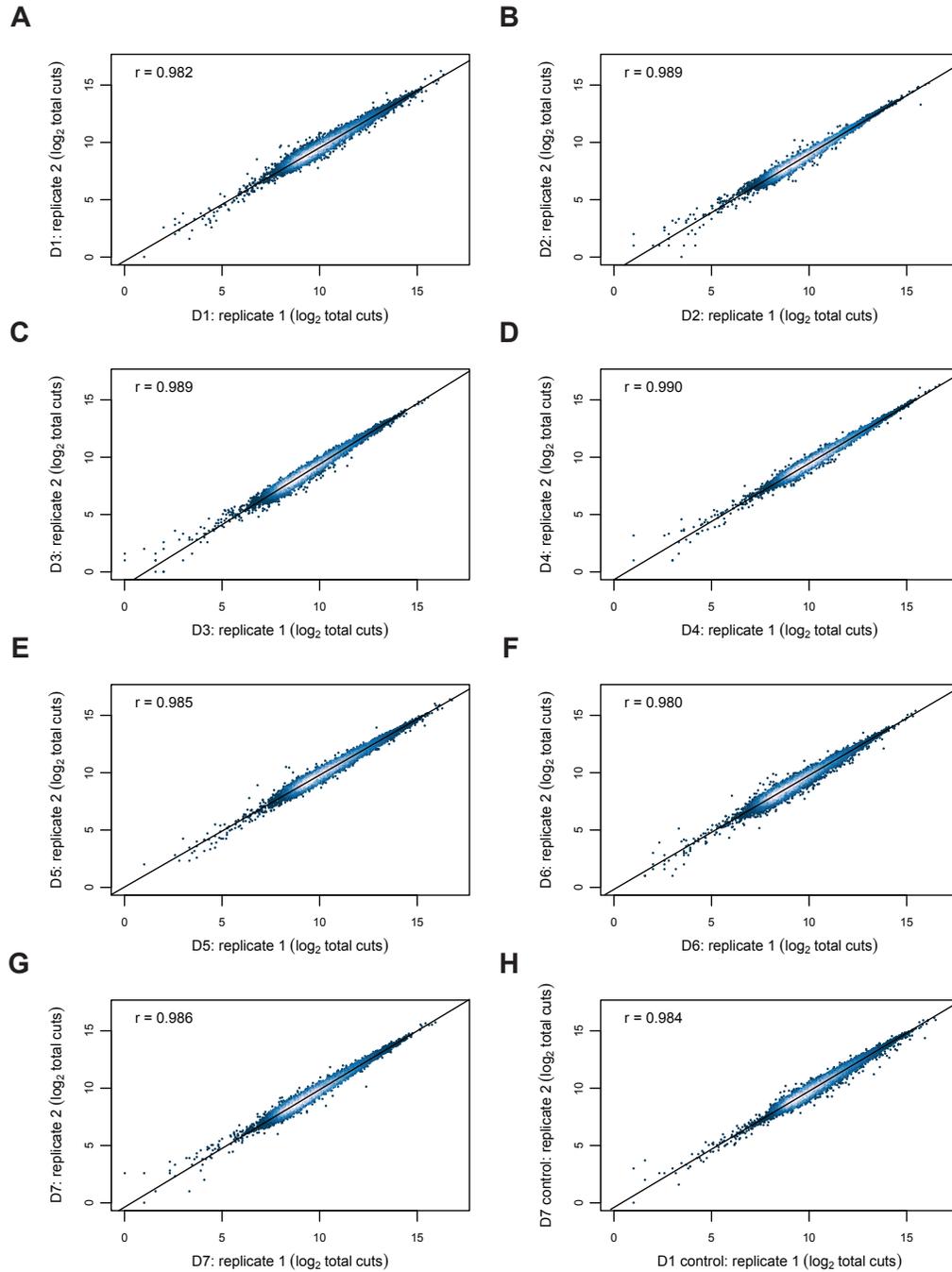


Figure 3.7: Correlation between replicate ATAC-seq experiments on tagged domains. Scatter plots represent pair-wise comparisons between replicate experiments from individual tagged domains (A-G) and a representative comparison between two whole-embryo control experiments from transgenic strains carrying different expression constructs (H). Correlations are evaluated based on \log_2 -transformed ATAC-seq signal (total count of Tn5 transposase cuts) over 17 345 high-confidence ATAC-seq peaks. r = Pearson correlation coefficient.

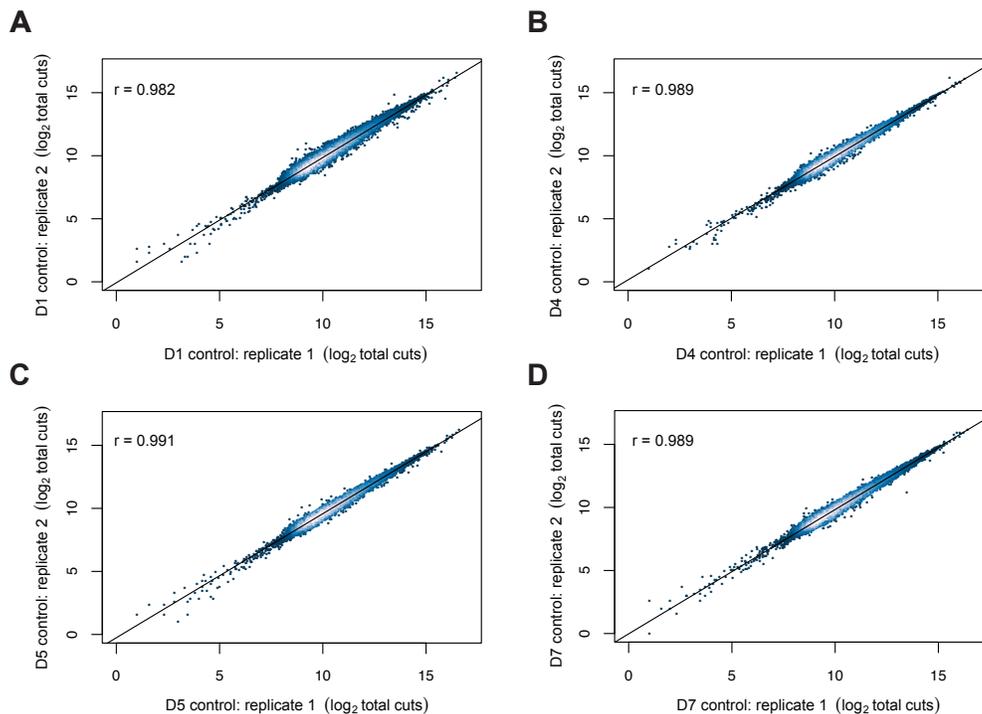


Figure 3.8: Correlation between replicate ATAC-seq experiments on whole-embryos. Scatter plots represent pair-wise comparisons between biological replicates from independent transgenic strains carrying the same expression construct. Correlations are evaluated based on log₂-transformed ATAC-seq signal (total count of Tn5 transposase cuts) over 17 345 high-confidence ATAC-seq peaks. r = Pearson correlation coefficient.

different experimental protocols and c) different enzymatic mechanisms of DNaseI and Tn5 transposase. High similarity between the accessibility profiles proves considerable comparability between the ATAC-seq and DNase-seq assays. The discrepancies regarding the overlap of peaks are likely to reflect the fact that different peak-calling algorithms and different statistical thresholds were used in the two studies to identify highly accessible regions. Thomas et al. [67] applied a customized algorithm, while ATAC-seq peaks were identified with MACS2 with an additional inclusion of a reference sample. Overall, this allows to conclude that, in comparison to mapped cut sites, peak-calling is a less direct measure of accessibility and is more prone to differences in statistical power and sensitivity of the applied algorithms.

DNase-seq was performed on wild-type embryos, while ATAC-seq was performed on transgenic strains with an integrated expression cassette. In addition to demonstrating high quality of the experimental data from this study, similarity to the published DNase-seq profiles confirms that expression of the nuclear tag does not affect the accessibility landscape in the blastoderm embryo.

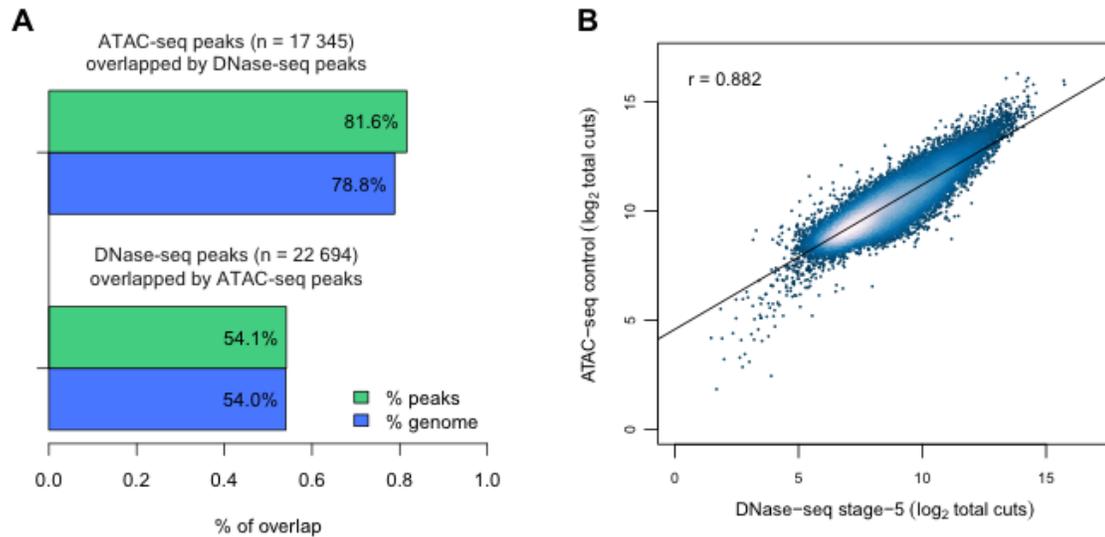


Figure 3.9: High similarity between whole-embryo controls and published DNase-seq accessibility profiles from stage-5 embryos. A) Above: proportion of high-confidence ATAC-seq peaks (identified in all eight whole-embryo controls, $FDR = 1\%$) that co-localize with DNase-seq peaks (intersection from two replicates, $FDR = 5\%$ [67]). Below: proportion of DNase-seq peaks that co-localize with the ATAC-seq peaks. In order to control for different median sizes of ATAC-seq and DNase-seq peaks, both the total proportion of peaks (green) and proportion of their cumulative size (blue) are plotted. B) Scatter plot represents correlation between \log_2 -transformed ATAC-seq and DNase-seq signal intensities over the set of 17 345 high-confidence ATAC-seq peaks. ATAC-seq signal is defined as the normalized count of Tn5 transposase cuts (average over eight whole-embryo controls). DNase-seq signal is defined as the normalized DNaseI tag density (average over two replicates). $r =$ Pearson correlation coefficient.

Chapter 4

Regional differences in chromatin accessibility along the AP axis

4.1 Global variation of accessibility profiles

4.1.1 Examination of individual genomic loci

In order to assess the scale of accessibility variation along the embryonic AP axis, I first examined accessibility profiles of all tagged domains (D1-D7) and whole-embryo controls over selected genomic loci (Figure 4.1). While the position of accessible regions was conserved between the domains, they displayed considerable quantitative differences in the extent of their accessibility signal. Having demonstrated before low technical variability of the experimental protocol and a negligible effect of transgenic tagging on accessibility profiles, I concluded that these differences in ATAC-seq profiles represented true biological variation along the embryonic axis.

4.1.2 Genome-wide assessment of the accessibility variation

In order to check whether these differences in accessibility profiles were also valid on a genome-wide scale, I analyzed variation in the ATAC-seq signal over a complete set of 17 345 high-confidence accessible regions. Principal component analysis (PCA) allowed for simultaneous analysis of variation between the complete set of ATAC-seq samples, including all tagged domains as well as whole-embryo controls (Figure 4.2).

PCA separated the tagged domains based on their accessibility signal into distinct clusters that corresponded to their position along the AP axis: an anterior cluster of D1-D3 domains and a posterior cluster of D4-D7 domains. The analysis also confirmed close similarity between whole-embryo controls, in spite of their different genotype (different integrated expression constructs). Consistent with the controls representing the entire length of the AP axis, their cluster occupied an intermediate position in the PCA, between the anterior and posterior domains (Figure 4.2).

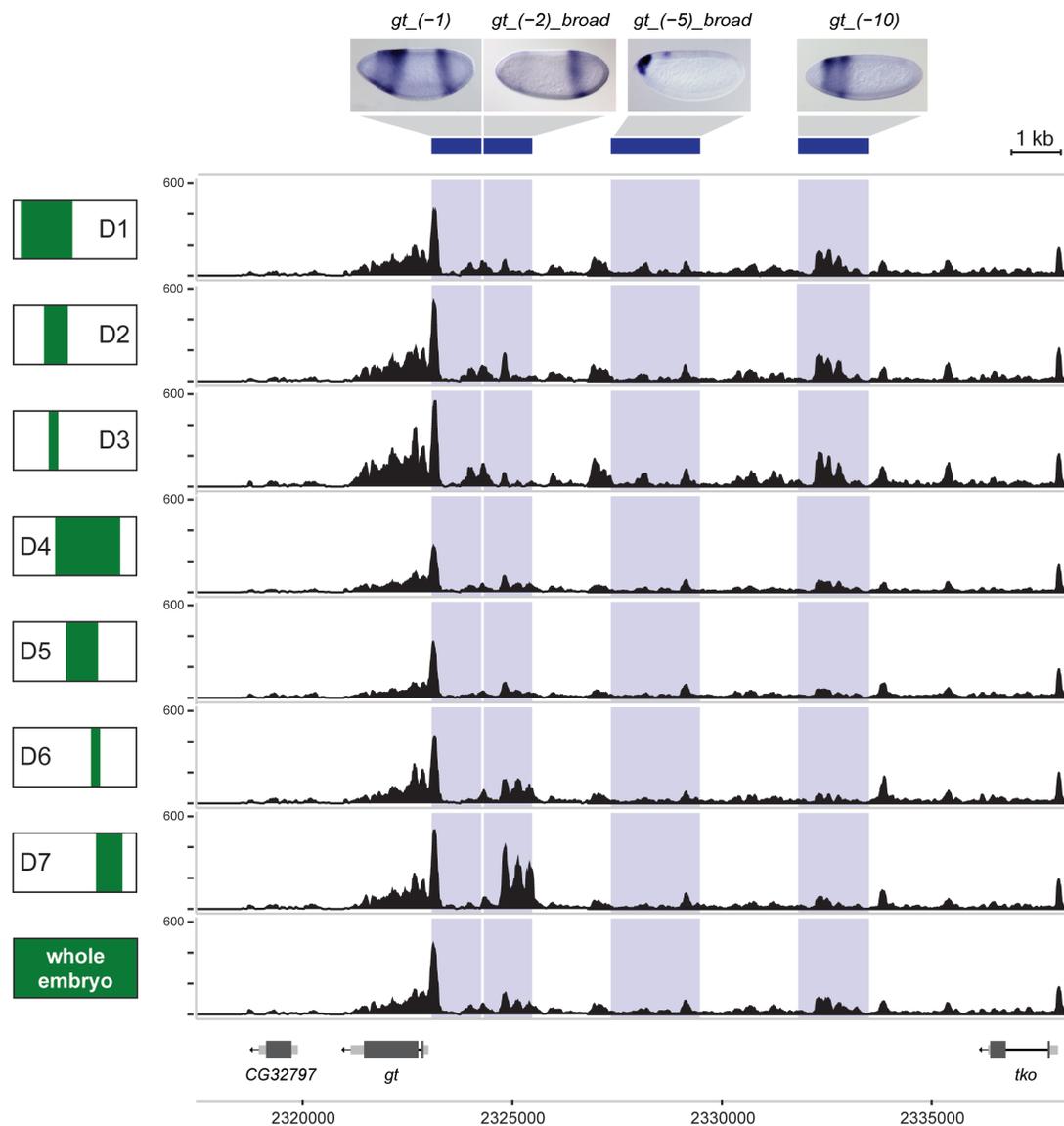


Figure 4.1: Comparison of ATAC-seq accessibility profiles from different tagged domains at the locus of *giant*. Tracks show normalized coverage of ≤ 100 bp ATAC-seq fragments, smoothed over a sliding window of 15 bp. AP positions of the profiled domains are indicated schematically on the left (green shading). Blue bars and underlying shaded regions indicate coordinates of known enhancers of *giant* (names as in Supplementary Material D, coordinates from [64]). Spatial activity of each enhancer in the blastoderm embryo is illustrated above (RNA *in situ* hybridization of a reporter gene, reproduced from [50]). Horizontal axis represents genomic coordinates along chromosome X (gene models: Fly-Base Release 5.57, [126]).

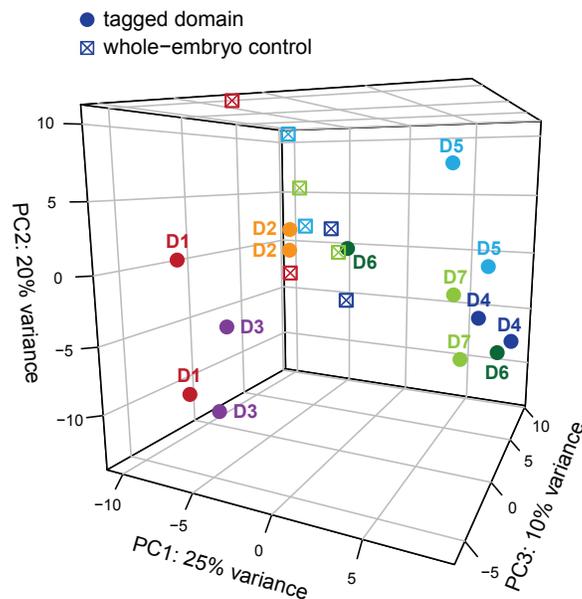


Figure 4.2: Principal component analysis of tagged domains and whole-embryo controls. PCA of genome-wide accessibility variation across individual tagged domains (solid circle) and whole-embryo controls (crossed square). Duplicates are represented as separate data points and color-coded by genotype (D1: red; D2: orange; D3: purple; D4: dark blue; D5: light blue; D6: dark green; D7: light green). PCA is based on accessibility signal (total count of Tn5 transposase cuts) over 17 345 high-confidence ATAC-seq peaks.

The three first components of the PCA analysis captured more than one half of the variance in the ATAC-seq data. Although biological interpretation of the components is not straightforward, PC1 component (25% variance) appeared to represent variation along the AP axis, providing a strong distinction between anterior and posterior domains (Figure 4.3). PC2 component (20% of variance) separated duplicate experiments, possibly representing a batch effect of sequencing and sample processing. PC3 component (10% of variance) separated whole-embryo controls from the tagged domains to a moderate extent.

It is also important to note that one replicate of the D6 domain (Replicate 1) showed strong similarity to whole-embryo controls. This could have potentially resulted from inefficient affinity purification and contamination with untagged nuclei, taking into consideration that D6 represents a narrow posterior stripe of *even skipped*. Clustering with whole-embryo controls was also confirmed in a more specific correlation analysis over a selected set of axis patterning enhancers (Supplementary Material C). For this reason the sample was removed from the subsequent analysis.

The observed differences in ATAC-seq profiles along the AP axis could result from two types of variation: a) global shifts in the genomic position of highly accessible regions

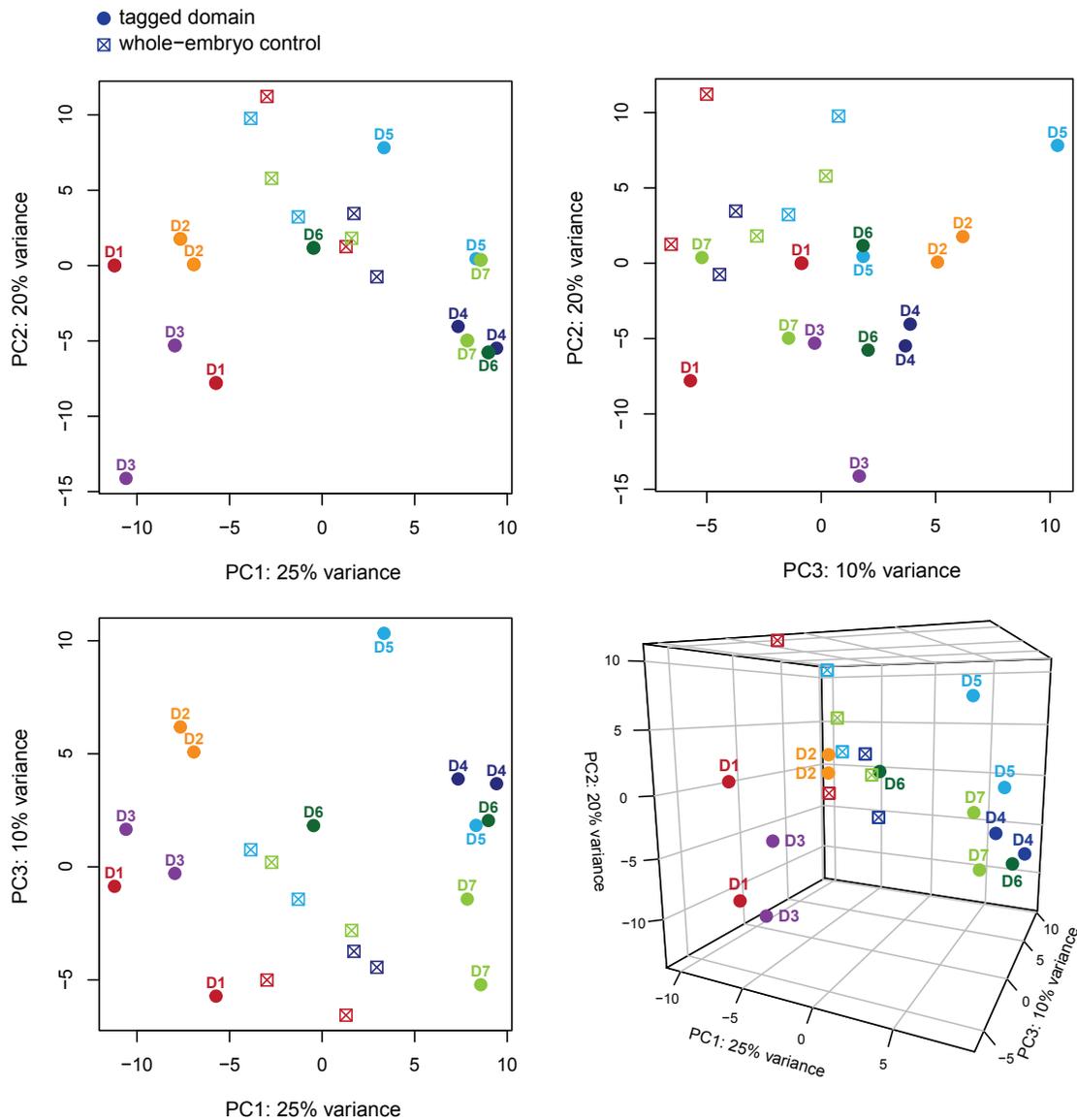


Figure 4.3: Pair-wise representations of PCA components. Panels show distribution of individual tagged domains (solid circle) and whole-embryo controls (crossed square) against two principal components, corresponding to individual planes of the 3D PCA plot (bottom right). Duplicates are represented as separate data points and color-coded by genotype (D1: red; D2: orange; D3: purple; D4: dark blue; D5: light blue; D6: dark green; D7: light green).

between different sections of the embryo or b) quantitative differences in accessibility of the peaks whose genome-wide distribution remained highly conserved along the body axis. In order to distinguish between these two scenarios, I compared the distribution of ATAC-seq peaks between individual tagged domains (Figure 4.4). Around 90% or more of 17 345 high-confidence peaks identified in whole-embryo controls displayed significant signal enrichment in the restricted regions of the blastoderm embryo. Confirming observations from the individual locus of *giant*, this demonstrates that the position of accessible genomic intervals is highly conserved between different regions along the AP axis. As a result, separation of the tagged domains in the PCA analysis primarily reflects quantitative differences in the accessibility signal between the common set of ATAC-seq peaks.

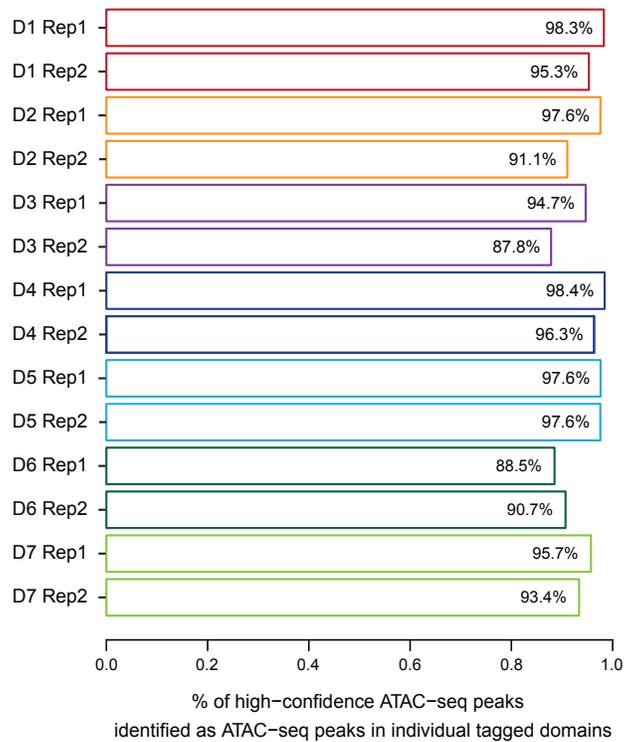


Figure 4.4: Conserved position of highly accessible regions along the AP axis. Bar plot shows the proportion of high-confidence ATAC-seq peaks from whole-embryo controls that are identified as significant peaks in individual tagged domains (minimum overlap = 50 bp). Duplicates are plotted separately.

While duplicates from individual tagged domains were characterized by strong correlation of their signal intensities across the complete set of high-confidence ATAC-seq peaks (Figure 3.7), they differed in the proportion of these peaks that showed significant signal enrichment in the MACS2 analysis (i.e. passing the False Discovery Rate FDR threshold of 1%). This example highlights a certain degree of variability of the peak-calling algorithm, which was performed for each replicate separately, as well as the arbitrariness of statisti-

cal thresholds. Just like the comparison between DNase-seq and ATAC-seq data (section 3.3.2), this demonstrates that signal intensities are a more direct and reliable measure for assessing similarities between the ATAC-seq samples.

Taken together, the results reveal extensive regional variation of chromatin accessibility in blastoderm embryos. While genomic location of accessible intervals is highly conserved, their ATAC-seq signal shows considerable quantitative differences between the assayed domains. Therefore, each position along the AP axis is characterized by a distinct accessibility profile, with the whole-embryo control representing an average across multiple accessibility states.

4.2 Scale of accessibility variation along the AP axis

4.2.1 Identification of differential ATAC-seq peaks

Having established that different regions of the blastoderm embryo are characterized by distinct accessibility profiles, I next aimed to assess what proportion of the accessible genome accounts for these domain-specific differences.

For this purpose, I identified ATAC-seq peaks that displayed significant differences in their accessibility signal in pair-wise comparisons between the tagged domains and whole-embryo controls. I applied the DESeq2 algorithm [131] that was originally developed to identify differential transcripts in RNA-seq experiments, but has been also successfully applied to ChIP-seq, DNase-seq and ATAC-seq data ([82, 132]). The framework of DESeq2 allows for a stringent statistical analysis that is based not only on the presence but also probability of quantitative differences between the compared samples. Additionally, it is applicable to all cases where the genomic signal can be simplified to count data. For this reason, I assessed regional variation in the accessibility signal based on the total number of Tn5 cuts that mapped to each of the 17 345 high-confidence ATAC-seq peaks. I applied a stringent statistical threshold, with False Discovery Rate (FDR) $<1\%$.

I performed a total set of 28 unique pair-wise comparisons between: a) each tagged domain and its corresponding whole-embryo control and b) individual tagged domains (see Materials and Methods). Out of 17 345 high-confidence ATAC-seq peaks that were identified in whole-embryo controls, 4 282 showed significant accessibility changes along the AP axis. This corresponded to 25.8% of the accessible genome displaying significant quantitative variation in the *Drosophila* embryo (Figure 4.5). Importantly, as many as 2 925 of these differential peaks (17.7% of the accessible genome) were supported by at least two independent pair-wise comparisons.

It is important to note that the comparisons solely between the tagged domains and their corresponding whole-embryo controls were not sufficient to identify more than one thousand differential regions (Figure 4.6). If we assume (as has been demonstrated before) that each position along the AP axis is characterized by a unique accessibility profile, this limited sensitivity of detection likely stems from the fact that the whole-embryo sample contains a subset of accessibility states that are represented in the compared domain. This

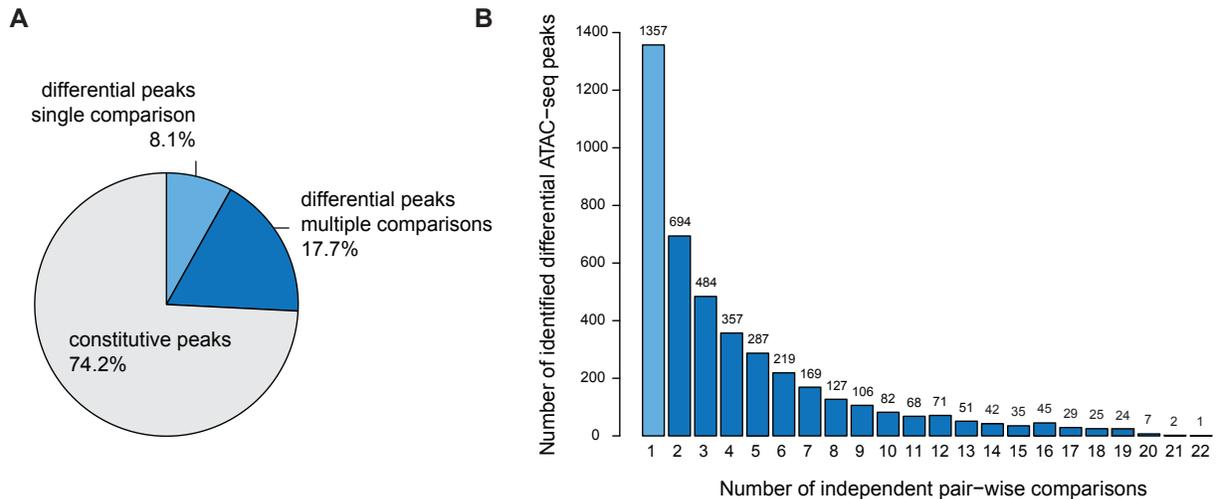


Figure 4.5: One quarter of the accessible genome displays significant regional variation in the *Drosophila* blastoderm. A) Pie chart shows proportional representation of constitutive ATAC-seq peaks that display no significant variation in their accessibility along the AP axis (grey) and differential ATAC-seq peaks that are supported by single pair-wise comparisons (light blue) and multiple pair-wise comparisons (dark blue) in the DESeq2 analysis. The chart shows percentage of the accessible genome (combined size of high-confidence ATAC-seq peaks) that is accounted for by the different classes of peaks. B) Bar plot shows the number of ATAC-seq peaks identified independently in all 28 pair-wise comparisons. Light blue: peaks identified in a single comparison, dark blue: peaks identified in multiple comparisons.

reduces the scale of quantitative differences between their ATAC-seq signal and, consequently, statistical significance of the observed signal fold-change. Likewise, comparison between non-overlapping tagged domains that represent a distinct combination of accessibility profiles, should result in a larger number of ATAC-seq peaks showing a larger magnitude of signal change (Figure 4.6).

In agreement with that, the number of differential peaks identified in individual comparisons between the tagged domains was correlated with their distance along the AP axis (Figure 4.7). DESeq2 consistently identified a larger number of differential ATAC-seq peaks between domains that represented anterior and posterior halves of the embryo (e.g. D1 vs. D4) as opposed to the domains that showed substantial overlap (e.g. D1 and D2). Importantly, Figure 4.7 also clearly illustrates that in any pair-wise comparison, a majority of the ATAC-seq peaks displayed minimal accessibility variation, with their distribution of fold-changes centered at 1 (grey distribution), while only a small proportion of the regions showed significant quantitative differences (blue data points). Overall, this analysis confirms high specificity of the experimental approach and also demonstrates that the measured accessibility is indeed a function of the domain's position along the AP axis.

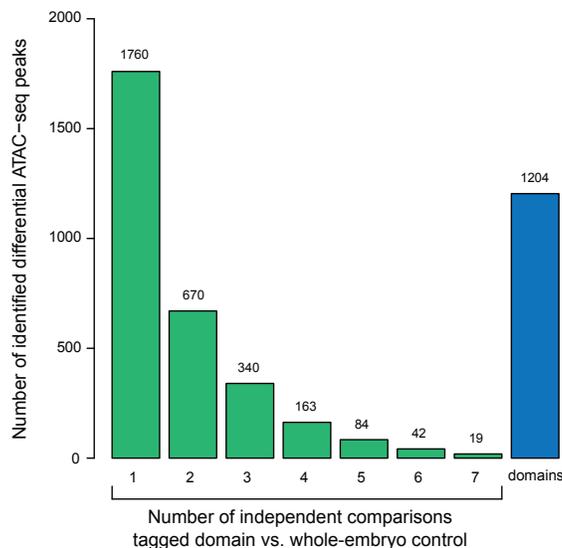


Figure 4.6: Pair-wise comparisons between individual tagged domains provide increased sensitivity of detection. Bar plot shows the number of differential peaks that are supported by multiple pair-wise comparisons between tagged domains and whole-embryo controls (green) as well as those that are uniquely identified in pair-wise comparisons between tagged domains (blue).

4.2.2 Evaluation of differential regions

DESeq2-based identification of differential ATAC-seq peaks relies on the magnitude of signal fold-changes between selected regions of the embryo. However, to what extent does the arbitrary significance threshold of $FDR < 1\%$ represent biologically relevant accessibility differences along the AP axis?

One question worth considering is the extent to which the detection of differential regions was biased by technical variability of the DESeq2 algorithm. More than 2/3 of the differential peaks were identified in multiple independent pair-wise comparisons between different domains (Figure 4.5), providing more confidence that they do not represent technical artifacts. On the other hand, the fact that a subset of ATAC-seq peaks showed significant variation only in a single comparison could likely stem from underrepresentation of certain embryonic regions. Coverage of the AP axis by the seven tagged domains is not uniform in this experimental set-up. Certain sections of the blastoderm are represented independently in multiple domains, e.g. the narrow D3 stripe is included both in the D2 and D1 domains (Figure 2.9). On the other hand, other regions of the embryo are encompassed by only one domain (e.g. anterior 25% of the embryo) or not represented at all (both anterior and posterior termini).

Second, it is feasible that one quarter of the accessible genome shows significant acces-

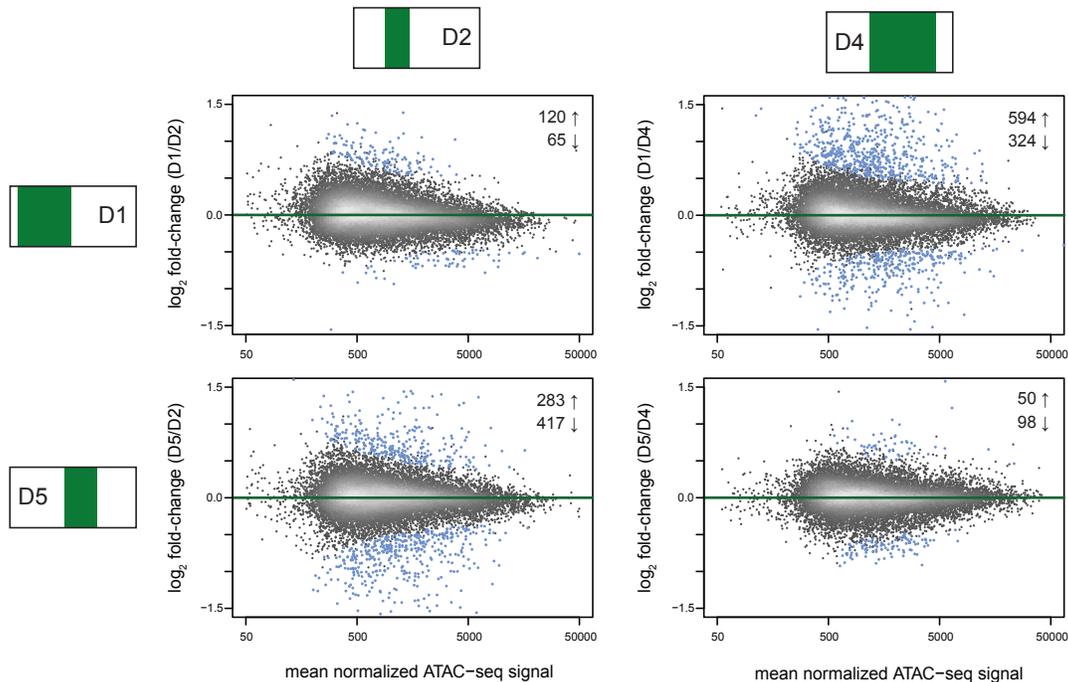


Figure 4.7: Number of differential ATAC-seq peaks is proportional to the distance between compared tagged domains. Example scatter plots show fold-changes of ATAC-seq signal between selected domains (y-axis) against the mean normalized signal intensity (x-axis). Grey: distribution of constitutive peaks. Blue points: individual differential ATAC-seq peaks (FDR < 1%). The number of ATAC-seq peaks showing significant increase (up arrow) and decrease (down arrow) of their signal is indicated in the upper right corner of each plot.

sibility differences along the AP axis? As the functional link between accessibility changes and operation of the differential genomic regions is not fully understood, it is not possible to validate differential ATAC-seq peaks against an independent and biologically-relevant dataset. However, Cusanovich et al. [89] recently performed ATAC-seq profiling of stage 5-8 embryos in a single-cell assay. By comparing accessibility profiles between five big clusters of nuclei, they detected significant regional variation of 24.1% of ATAC-seq peaks, corresponding to as much as 25.0% of the accessible genome. This independent study on a similar number of embryonic regions, yet with very different methodology, confirms my observations and provides more confidence in the significance threshold that I used in the DESeq2 analysis.

On the other hand, Haines and Eisen [90], by comparing the anterior and posterior halves of the cellular blastoderm, identified fewer than 120 ATAC-seq peaks that showed significant quantitative variation, corresponding to only 1.7% of all identified accessible regions. This is in stark contrast with the number of differential peaks identified in this study. For example, a single pair-wise comparison between domains of a similar size, D1 and

D4, identified more than 900 differential peaks. To a certain extent, this discrepancy might reflect differences between algorithms used in the two studies. More generally though, it demonstrates that the number of significantly differential ATAC-seq peaks depends on the number and size of the analyzed domains.

Finally, it is important to note certain limitations of the approach used in this analysis. Since DESeq2 was initially designed to analyse RNA-seq data, it estimates statistical differences between a set of pre-defined genomic regions based on a single count value that represents the total signal intensity. As a result, the algorithm does not detect local accessibility variation within individual ATAC-seq peaks. Secondly, since restricted to the pre-defined coordinates of accessible regions, DESeq2 does not allow for *de novo* identification of differential regions, e.g. based solely on genome-wide comparisons between the accessibility tracks. Finally, definitions of high-confidence ATAC-seq peaks that I used in this study are very stringent, as they only represent regions that were identified by MACS2 as significantly enriched in all eight whole-embryo controls. As a result, this set of peaks is likely to underrepresent regions that show significant accessibility only in a narrow domain of the embryo, and thus would be masked in the whole-embryo average. I envision, therefore, that an even larger number of differential accessible regions could be identified if their definitions were expanded to the entire spectrum of ATAC-seq peaks, i.e. a union of accessible regions identified both in tagged domains and whole-embryo controls.

4.3 Characterization of differential ATAC-seq peaks

Accessible genomic regions in stage-5 *Drosophila* embryos have been previously demonstrated to correspond to a wide range of functional genomic elements, such as: transcribed gene bodies, promoters, distal *cis*-regulatory elements, insulators and origins of replication [67]. Taking into consideration that as much as one quarter of the accessible genome displays significant regional variation in the cellular blastoderm, I aimed to test whether the differential ATAC-seq peaks represented any specific class of these functional elements.

For this purpose, I examined genomic localization of differential ATAC-seq peaks as well as their occupancy by different classes of proteins. I contrasted their properties with those of constitutive ATAC-seq peaks that showed no significant accessibility changes in the DESeq2 analysis (Figure 4.5).

Importantly, the differential peaks differed considerably in terms of the scale of accessibility changes that they displayed along the AP axis, ranging from 1.2-fold to 9.3-fold. In order to account for this variability, I binned the ATAC-seq peaks into quarters, based on the maximum reported magnitude of accessibility change among any of the 28 pair-wise comparisons (Figure 4.8).

4.3.1 Genomic localization of differential ATAC-seq peaks

First of all, I compared genomic localization of differential and constitutive ATAC-seq peaks (Figure 4.9). More than one half of constitutive peaks corresponded to promoters

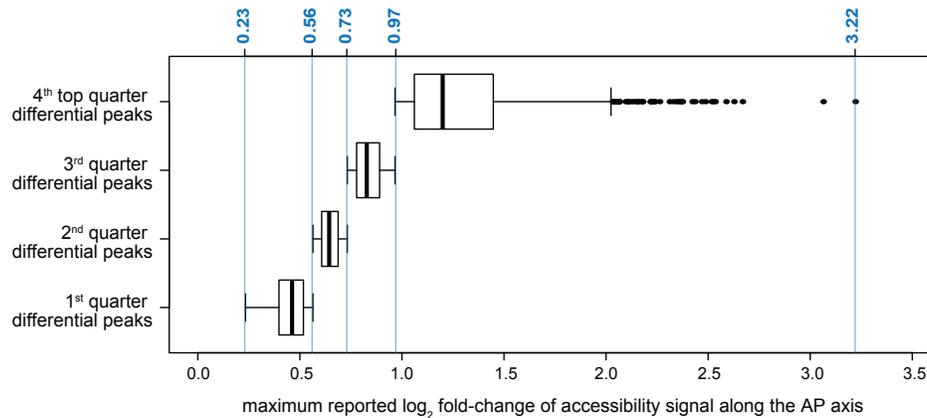


Figure 4.8: Distribution of ATAC-seq signal fold-changes among the entire set of differential peaks. Differential peaks are binned into quarters based on the maximum \log_2 fold-change of their accessibility signal as reported by DESeq2 in any of the 28 pair-wise comparisons. Boxplots represent the distribution of \log_2 fold-changes within each quarter. Blue lines represent the quantiles.

and gene bodies. On the other hand, a vast majority of differential peaks mapped to intergenic and intronic regions. This enrichment was even more pronounced for the top quarter of the peaks that showed the largest magnitude of accessibility changes.

4.3.2 Occupancy by different classes of proteins

Localization of differential ATAC-seq peaks to intronic and intergenic regions suggests that they might represent distal *cis*-regulatory elements. I tested this hypothesis by examining co-localization of differential peaks with binding sites for different classes of proteins (Figure 4.10). For this purpose, I used published chromatin immunoprecipitation (ChIP) data of:

- Origin recognition complex subunit 2 (ORC2). A union of ORC2 ChIP-seq peaks from three *Drosophila* cell lines was used to define genomic positions of origins of replication (ORIs) [133].
- Four insulator proteins: CTCF, BEAF32 and Su(Hw) and CP190 [134]. ChIP-chip experiments were performed as part of the modENCODE project [126].
- Four key TFs of the DV gene regulatory network: Dorsal, Twist, Snail and Mother against dpp [135]. ChIP-chip experiments were performed as part of Berkeley *Drosophila* Transcription Network Project.
- Fourteen TFs of the AP gene regulatory network, including maternal factors: Bicoid and Caudal, gap factors: Giant, Hunchback, Knirps, Kruppel, Hucklebein, Tailless and Dichaete, as well as pair-rule factors: Fushi-tarazu, Hairy, Paired, Runt and

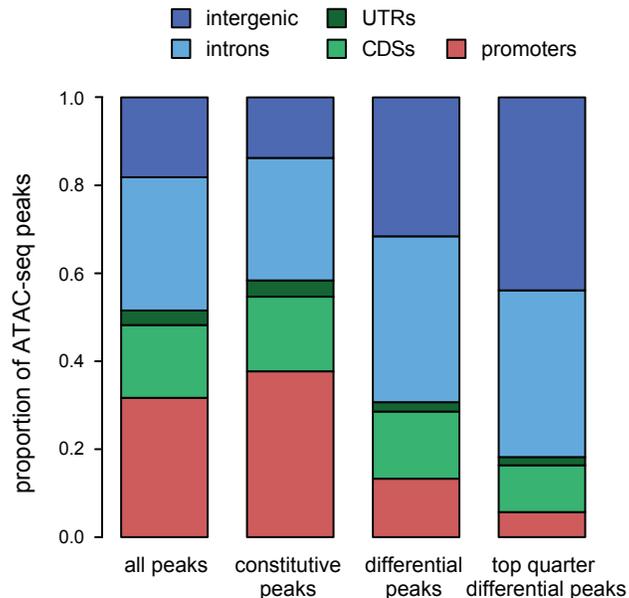


Figure 4.9: Genomic localization of ATAC-seq peaks. Proportional distribution of genomic annotations among different classes of accessible regions: all high-confidence ATAC-seq peaks (all peaks), constitutive peaks, differential peaks and the top quarter of differential peaks (highest values of maximum \log_2 fold-change reported in the DESeq2 analysis). UTR: 5' and 3' untranslated regions. CDS: coding sequence.

Sloppy paired 1 [136, 135]. ChIP-chip experiments were performed as part of Berkeley *Drosophila* Transcription Network Project.

- Six TFs of the AP gene regulatory network, including maternal factors: Bicoid and Caudal, and gap factors: Giant, Hunchback, Knirps, Kruppel [137]. In comparison to ChIP-chip data, ChIP-seq on the same set of TFs was characterized by more localized signal enrichment, allowing for more targeted identification of regions occupied by the AP regulators.

All datasets, apart from ChIP-seq on ORC2, were acquired from broad collections of whole blastoderm embryos, corresponding either to stage 4-5 embryos or the time window of 2-4 h after egg laying.

For this analysis, I exclusively considered differential and constitutive peaks that localized to intronic and intergenic regions. In comparison to constitutive peaks, a significantly smaller proportion of differential peaks corresponded to origins of replication (ORIs) and binding sites of insulator proteins (Figure 4.10). On the other hand, differential peaks were strongly enriched in binding by TFs involved in patterning the embryo both along the antero-posterior and dorso-ventral axes; yet the enrichment was most pronounced for AP TFs. It is important to note that the differences between constitutive and differential peaks were statistically significant for all considered classes of proteins. Additionally, the

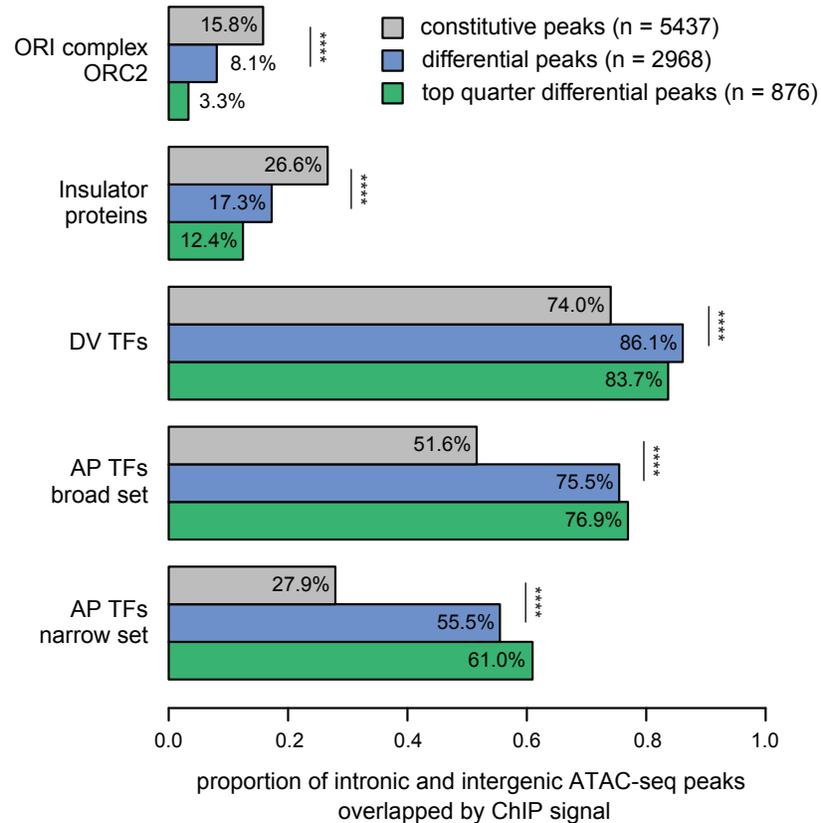


Figure 4.10: Overlap of ATAC-seq peaks with chromatin immunoprecipitation (ChIP) data for different classes of proteins. Bar plot shows the proportion of ATAC-seq peaks that map to intronic and intergenic regions and co-localize with ChIP signal of different classes of proteins. Grey: constitutive peaks, blue: all differential peaks, green: top quarter of differential peaks (number of intervals indicated in the legend). ORI complex: ChIP-seq peaks of ORC2 [133]. Insulator proteins: ChIP-chip peaks of CTCF, BEAF32 and Su(Hw) and CP190 [134]. DV TFs: ChIP-chip peaks of Dorsal, Twist, Snail and Mother against *dpp* [135]. AP TFs (broad set): ChIP-chip peaks of 14 maternal, gap and pair-rule TFs: Bicoid, Caudal, Giant, Hunchback, Knirps, Kruppel, Hucklebein, Tailless, Dichaete, Fushi-tarazu, Hairy, Paired, Runt and Sloppy-paired 1 [136, 135]. AP TFs (narrow set): ChIP-seq peaks of 6 maternal and gap TFs: Bicoid, Caudal, Giant, Hunchback, Knirps, Kruppel [137]. Asterisks indicate significant differences between constitutive and differential peaks ($p < 0.0001$, Fisher's exact test).

respective enrichment or depletion in protein occupancy was even more pronounced for the top quarter of differential peaks that displayed the largest magnitude of accessibility changes.

Overall, localization to intronic and intergenic regions as well as occupancy by patterning TFs suggests that differential peaks represent distal *cis*-regulatory elements that drive patterned gene expression in the early *Drosophila* embryo.

4.3.3 Overlap of ATAC-seq peaks with annotated *cis*-regulatory elements

In order to test whether differential ATAC-seq peaks indeed correspond to axis patterning enhancers of the blastoderm embryo, I examined their overlap with annotated *Drosophila cis*-regulatory elements.

First of all, I considered a broad collection of 2-kb genomic intervals (referred to as Vienna Tiles) that were tested for their enhancer activity during the entire span of *Drosophila* embryogenesis (Kvon et al., 2014). In a large-scale screen, the candidate regions were coupled to a reporter gene in a series of transgenic strains, followed by detailed annotation of their temporal and spatial activity patterns.

Initially, I examined co-localization of differential and constitutive ATAC-seq peaks with the complete set of Vienna Tiles (Figure 4.11). Vienna Tiles that overlapped differential peaks were enriched in elements that displayed enhancer activity at the earliest stage 4-6 of embryogenesis (43%). In contrast, only 27% of Vienna Tiles overlapping constitutive peaks displayed this early enhancer activity, while elements active at stage 4-6 constituted only 16% of all Vienna Tiles. Furthermore, in comparison to differential peaks, a substantially higher proportion of constitutive Vienna Tiles represented enhancers that were active at considerably later developmental stages, up to stage 13-14 (9-11 h after fertilization).

Next, I specifically examined activity patterns of Vienna Tiles that were active at stage 4-6 embryos, thus encompassing the stage of cellular blastoderm (Figure 4.12). In comparison to Vienna Tiles overlapping constitutive ATAC-seq peaks, those that co-localized with differential peaks were enriched in annotation terms referring to patterned expression along the AP axis, e.g. anterior, posterior, gap, A-P stripe and pair-rule. On the other hand, a considerably smaller proportion of differential peaks drove patterned expression along the orthogonal DV axis (e.g. in amnioserosa, ventral ectoderm or mesoderm). Finally, while around 20% of Vienna Tiles with constitutive accessibility produced "ubiquitous" expression patterns, this proportion dropped to 5% for the differential peaks. Overall, 48% of Vienna Tiles active at stage 4-6 displayed differential accessibility in the ATAC-seq assay (Figure 4.13A).

It is important to note that Vienna Tiles were assayed in a large-scale screen that was not followed by any additional validation experiments. As the 2-kb regions were examined outside of their genomic context, this could have led in some cases to ectopic expression of the reporter construct. For this reason, I additionally assayed the overlap of differential ATAC-seq peaks with a closely curated and experimentally verified set of *Drosophila cis*-

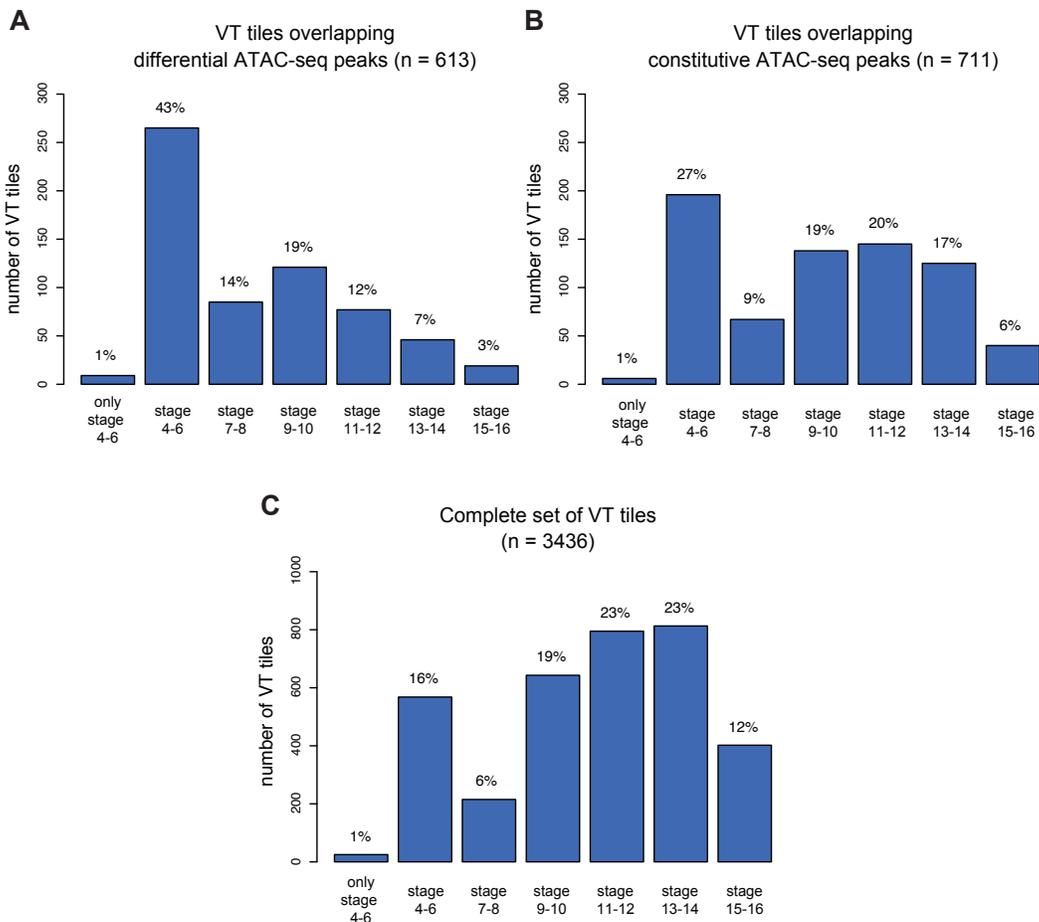


Figure 4.11: Temporal activity of Vienna Tiles overlapping ATAC-seq peaks: A) differential peaks, B) constitutive peaks and C) a complete set of Vienna Tiles (VT tiles). Bar plots show the number of Vienna Tiles active at each stage, with indicated proportional representation among all elements from each category. Since activity of Vienna Tiles usually spans several consecutive stages of embryogenesis, they were annotated based on the earliest stage of embryogenesis at which they drive expression of the reporter construct.

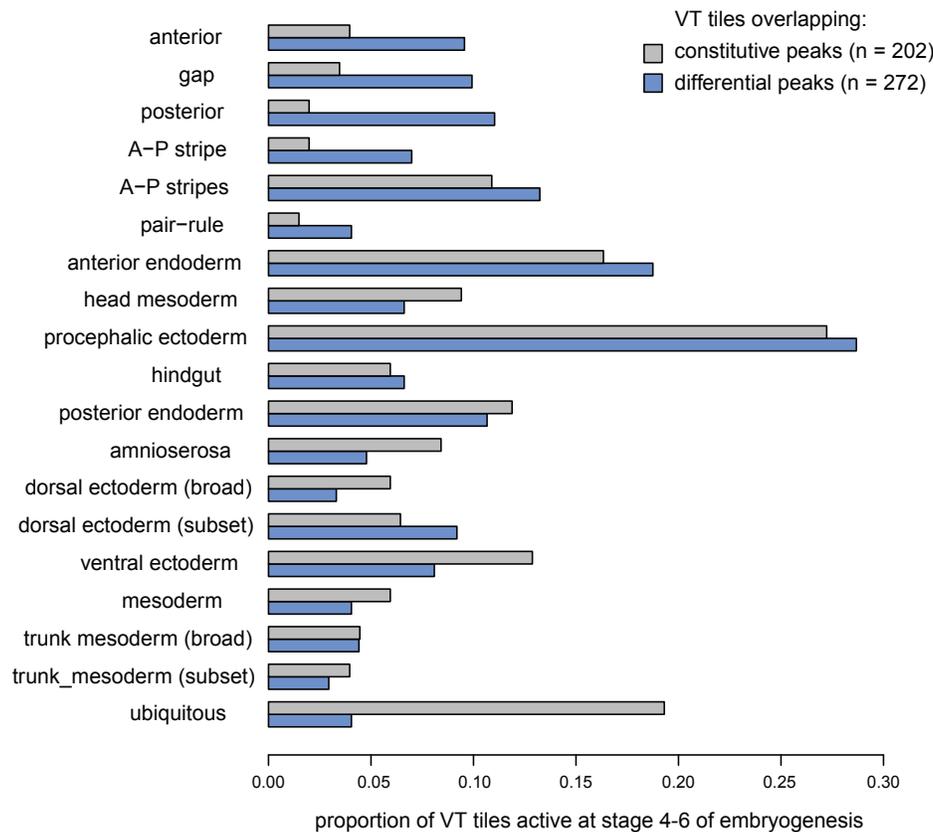


Figure 4.12: Spatial activity of Vienna Tiles overlapping ATAC-seq peaks. Bar plot represents the proportional representation of different annotation terms among Vienna Tiles (VT tiles) that are specifically active in stage 4-6 embryos and overlap either differential peaks (blue) or constitutive peaks (grey). Note that individual Vienna Tiles are often annotated with multiple terms.

regulatory elements collected in the REDfly database [45]. After filtering for CREs that were specifically annotated as active in blastoderm, 77% of those showed substantial overlap with differential ATAC-seq peaks (Figure 4.13A).

Finally, I selected a subset of REDfly CREs representing axis patterning enhancers that specifically drive patterned expression along the AP axis. I only considered elements that met the following criteria: 1) availability of high-quality *in situ* hybridization images, 2) annotated and validated target gene, and 3) non-background accessibility signal in the ATAC-seq assay (indicating an active functional element). Based on the *in situ* images, I selected enhancers whose spatial activity agreed with the expression pattern of their annotated target gene. I excluded elements that produced very weak or spurious expression, as well as those that showed strong modulation along the DV axis. I also excluded *cis*-regulatory elements that overlapped core promoters and gene bodies. Overall, I created a

list of 88 high-confidence enhancers, which I refer to as AP enhancers (genomic coordinates and references to original studies are provided in Supplementary Material D). Importantly, among these carefully curated AP enhancers, 93% were overlapped by differential ATAC-seq peaks (Figure 4.13A).

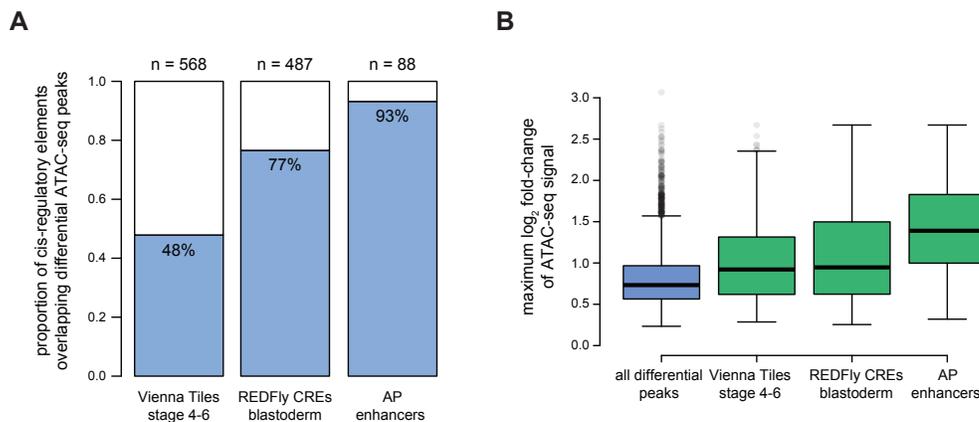


Figure 4.13: Overlap of differential ATAC-seq peaks with annotated *Drosophila* *cis*-regulatory elements. A) Bar plots indicate the proportion of annotated *cis*-regulatory elements that overlap differential ATAC-seq peaks. Total number of CREs from each category is indicated above the bars: Vienna Tiles [56] active specifically at stage 4-6, REDfly CREs [45] active in blastoderm embryos and curated AP enhancers driving patterned expression specifically along the AP axis. B) Box plots show distribution of maximum \log_2 fold-change of ATAC-seq signal reported for all differential peaks (blue) as well as differential peaks overlapping the three categories of annotated CREs (green).

Overall, the proportion of early *Drosophila* enhancers that showed differential accessibility in my assay ranged from 48%, when considering a very broad set of 2-kb genomic elements (Vienna Tiles), to 93%, when considering a restricted set of experimentally validated enhancers driving expression specifically along the AP axis. In all cases, the known *Drosophila* enhancers co-localized with ATAC-seq peaks that displayed a considerably larger magnitude of accessibility variation along the AP axis as compared to the total pool of differential peaks (Figure 4.13B).

4.3.4 Differential accessibility is a signature of axis patterning enhancers

Among all accessible genomic regions in the cellular blastoderm that correspond to different classes of functional elements, regional accessibility variation is primarily displayed by *cis*-regulatory elements that drive patterned gene expression at the earliest stages of embryogenesis.

Constitutive ATAC-seq peaks, which show no significant accessibility variation, mainly localize to promoters and gene bodies. The intergenic constitutive peaks are likely to correspond to insulators and origins of replication. Those that overlap known *cis*-regulatory elements often promote ubiquitous expression patterns. Therefore, constitutive accessible regions are likely to represent functional genomic elements that show little regional variation in their activity in the cellular blastoderm.

In contrast, a majority of differential peaks localize to intronic and intergenic regions. Taking into consideration their enrichment in TF binding and their overlap with known *Drosophila* enhancers, differential accessible regions can be concluded with high confidence to represent distal *cis*-regulatory elements. More importantly, since they are targeted by regulators of the AP and DV gene regulatory networks, promote patterned expression and show enhancer activity specifically in the assayed stage of cellular blastoderm, differential ATAC-seq peaks likely correspond to axis patterning enhancers. Additionally, annotated axis patterning enhancers display a larger magnitude of regional accessibility changes than the total pool of differential peaks identified in the DESeq2 analysis (Figure 4.13B). I propose therefore that differential accessibility is a signature of axis patterning enhancers, which is linked to their differential regulatory activity in the blastoderm embryo.

In this experimental set-up, accessibility variation is displayed most strongly by axis patterning enhancers of the dissected AP axis. This is demonstrated by enrichment of the relevant annotation terms of Vienna Tiles (Figure 4.12) as well as an almost complete representation of the curated set of AP enhancers among the differential ATAC-seq peaks. However, I believe that differential accessibility is not uniquely a feature of the AP network, as a portion of differential peaks co-localizes with *cis*-regulatory elements involved in patterning of the orthogonal DV axis. The AP axis is simply the spatial dimension over which I capture variation in the enhancer's accessibility most clearly. It is important to note, though, that the seven tagged domains also differ in terms of their representation of the presumptive germ layers. For example, D1 domain encompasses the anterior endoderm, while the central D5 domain primarily represents the mesoderm and ectoderm. Thus, sectioning of the embryo along the AP axis also allows for capturing differential activity of the DV network, yet to a limited extent.

Several features of my data appear to be in disagreement with the proposed notion that differential accessibility is a distinct feature of axis patterning enhancers. First of all, a proportion of constitutive peaks is targeted by AP and DV transcription factors, suggesting that they might promote patterned expression without displaying regional accessibility differences. However, sectioning of the embryo into seven domains might not allow for comprehensive representation of accessibility variation that would be detected with a sufficient statistical power in the DESeq2 analysis. It is likely then that the TF-bound constitutive ATAC-seq peaks might be detected as differential regions when assayed with a larger number of spatially-restricted domains. Additionally, ChIP signal over certain constitutive peaks might represent non-specific binding or technical variability of the assay. Indeed, it has been recently demonstrated that not all genomic regions with significant ChIP signal of patterning TFs correspond to functional *cis*-regulatory elements (Fisher et al., 2012).

Second, a small proportion of differential ATAC-seq peaks co-localizes with promot-

ers and gene bodies, suggesting that differential accessibility is not exclusively a feature of distal axis patterning enhancers. Taking into consideration regional modulation of the ATAC-seq signal over the promoter and gene body of *giant* (Figure 4.1), differential accessibility is likely to be additionally linked to differential transcription of the patterned genes.

Third, Vienna Tiles that overlap differential ATAC-seq peaks do not display enhancer activity exclusively in the cellular blastoderm. In fact, one half of the differential elements initiates expression of their reporter gene only at later stages of embryogenesis (Figure 4.11), suggesting that differential accessibility as displayed in the cellular blastoderm is not specific only to elements active at this stage. However, Vienna Tiles represent broad 2-kb intervals and it is likely that outside of their genomic context they might display ectopic temporal activity. Alternatively, this broad temporal profile of differential peaks might result from contamination of my collections with older embryonic stages that would contribute the ATAC-seq signal from enhancers active at later stages (Figure 2.10). Finally, it is possible that certain late enhancers might be characterized by an open chromatin structure immediately after zygotic genome activation and display accessibility modulation that is not linked to their transcriptional activity.

4.3.5 Differential accessibility as a metric for *de novo* identification of enhancers

Taking into consideration that axis patterning enhancers are characterized by strong regional modulation in their accessibility in the cellular blastoderm, I propose that differential accessibility can serve as a metric for *de novo* identification of the early *Drosophila* enhancers.

It is important to note that only one quarter of differential peaks that show the largest magnitude of accessibility changes correspond to the annotated axis patterning enhancers. However, when analysing their co-localization with maternal, gap and pair-rule TFs, both annotated and unannotated differential peaks display similar features. Both classes are dominated by elements bound by the patterning TFs and on average they are targeted by a larger number of different regulators than the constitutive peaks (Figure 4.14). This suggests that at least a portion of differential ATAC-seq peaks might correspond to novel axis patterning enhancers (see Discussion).

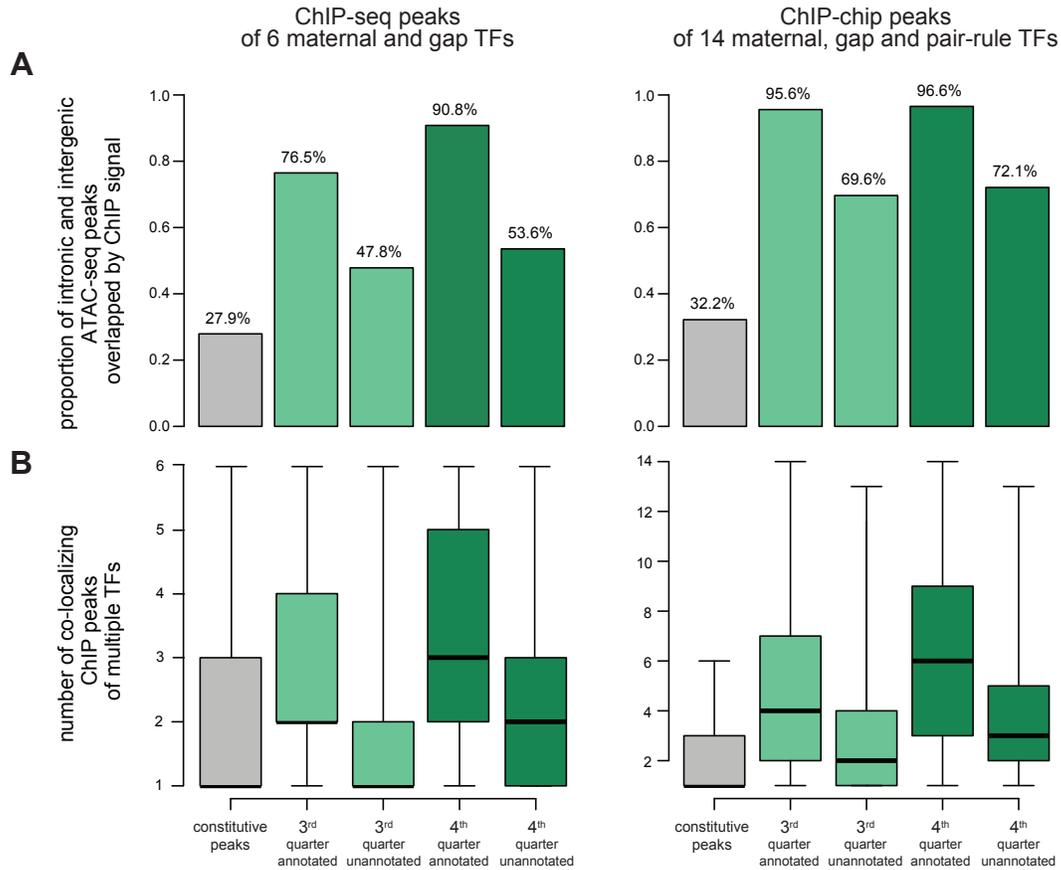


Figure 4.14: Comparison of annotated and unannotated differential peaks. Differential peaks of the 3rd and 4th quarters are divided into those that overlap (annotated) and do not overlap (unannotated) known *cis*-regulatory elements: Vienna Tiles tiles [56], REDfly enhancers [45] and AP enhancers. Only peaks mapping to intronic and intergenic regions are considered. 3rd quarter: 68 annotated and 738 unannotated peaks. 4th: 174 annotated and 702 unannotated peaks. A) Bar plots represent proportion of differential and constitutive ATAC-seq peaks that are overlapped by ChIP-seq peaks of 6 AP TFs and ChIP-chip peaks of 14 AP TFs (as in Figure 4.10). B) Boxplots represents the number of different TFs co-binding within the ATAC-seq peaks (peaks with no overlap were excluded). Note that combinatorial regulation by multiple TFs is a distinguishing feature of axis patterning enhancers.

Chapter 5

Accessibility modulation of individual axis patterning enhancers

5.1 Elevated accessibility of enhancers coincides with domains of their transcriptional activity

5.1.1 Modulation of local accessibility at individual enhancers

So far I demonstrated that axis patterning enhancers are enriched among the ATAC-seq peaks that show the largest regional modulation of their accessibility signal in the cellular blastoderm. I aimed next to examine how these regional accessibility changes are specifically linked to the regulatory activity of the enhancers.

Examination of activity patterns of individual enhancers in reference to the position of the tagged domains revealed a clear trend (Figure 4.1, Figure 5.1). ATAC-seq signal was elevated in the domain that coincided with an expression pattern driven by the enhancer, while it was consistently reduced in the domain from which the expression was excluded. This is illustrated by the *gt₋₁₀* enhancer that regulates the anterior expression domain of *giant* (Figure 5.1). In comparison to the whole-embryo control, its accessibility signal was higher in the anterior of the embryo (IN: D1), while it was lower in the posterior (OUT: D7), where the enhancer did not promote transcription of its target gene. It is important to note that the intermediate value of the ATAC-seq signal in the whole-embryo control is consistent with it representing an average across all accessibility states in the cellular blastoderm.

5.1.2 Relative changes of accessibility signal at active and inactive enhancers

To evaluate the relationship between accessibility and regulatory activity on a larger scale, I examined accessibility changes of all 88 AP enhancers that drive patterned expression

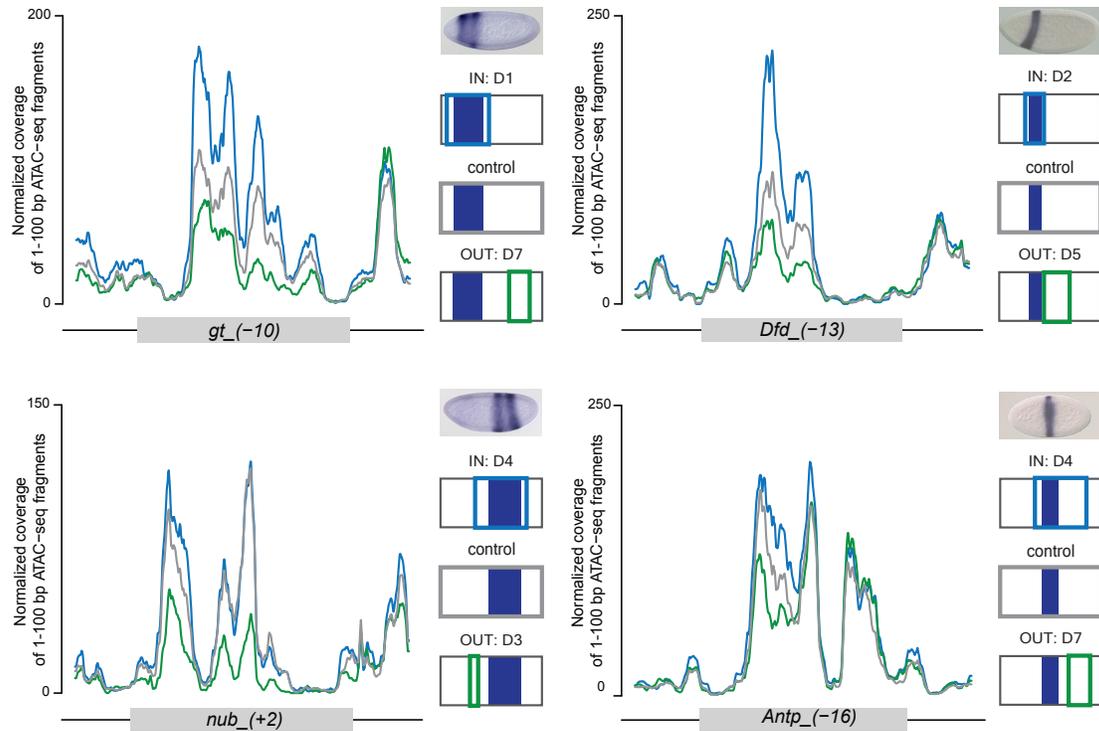


Figure 5.1: Modulation of local accessibility at individual AP enhancers. Coverage of ≤ 100 bp ATAC-seq fragments over four selected AP enhancers. Comparison between a whole-embryo control (control: grey), a tagged domain that encompasses activity pattern of the enhancer (IN: blue) and a tagged domain from which the enhancer's activity is excluded (OUT: green). Activity pattern of the enhancer along the AP axis is indicated schematically in dark blue, with color-coded outlines representing positions of respective domains. ATAC-seq coverage was calculated as mean signal from two replicates. RNA *in situ* hybridization images of a reporter gene: *gt_(-10)* and *nub_(+2)* (reprinted from [50]), *Dfd_(-13)* (reprinted from [138]) and *Antp_(-16)* (reprinted from [63]).

5.1 Elevated accessibility of enhancers coincides with domains of their transcriptional activity

specifically along the AP axis. For each tagged domain, I binned the AP enhancers into three classes: those whose activity pattern was completely included in the tagged domain (IN: active enhancers), those whose activity pattern was completely excluded from the tagged domain (OUT: inactive enhancers) and those that were active both inside and outside of the tagged domain (PARTIAL). I subsequently calculated fold-changes of their ATAC-seq signal between a given domain and the corresponding whole-embryo control.

In comparison to the whole-embryo control, those enhancers that were active inside the assayed tagged domain showed a relative increase of their accessibility, while inactive enhancers displayed a relative decrease of their accessibility signal (Figure 5.2). This trend was consistent and statistically significant across all tested domains and, thus, independent classifications of the AP enhancers.

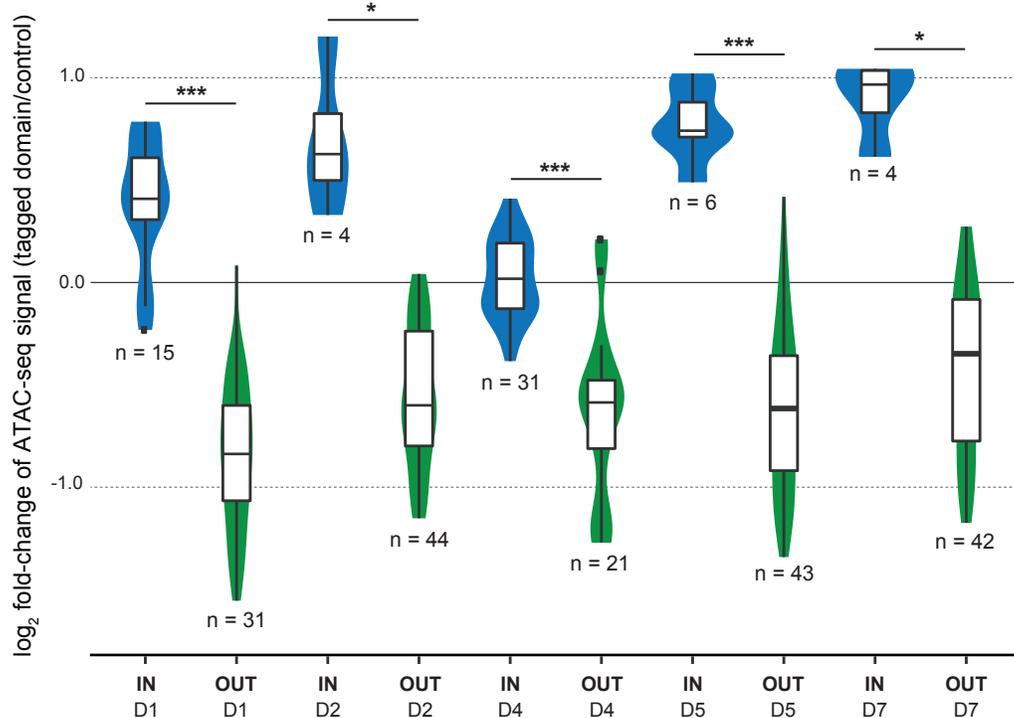


Figure 5.2: Relative changes of accessibility signal at active and inactive enhancers. Box plots show distribution of \log_2 fold-changes of ATAC-seq signal (total count of Tn5 transposase cuts, mean signal from two replicates) between a given tagged domain and its corresponding whole-embryo control, over two classes of AP enhancers. IN (blue): enhancers whose activity patterns are completely included in the tagged domain. OUT (green): enhancers whose activity patterns are completely excluded from the tagged domain. Number of enhancers in each class are indicated below individual plots. Domains D3 and D6 are not presented; due to their limited size, none of the AP enhancers had its activity pattern fully included in the domains. Asterisks indicate significant differences between IN and OUT enhancers (Student's t-test): p-value < 0.05 (*), p-value < 0.0001 (***).

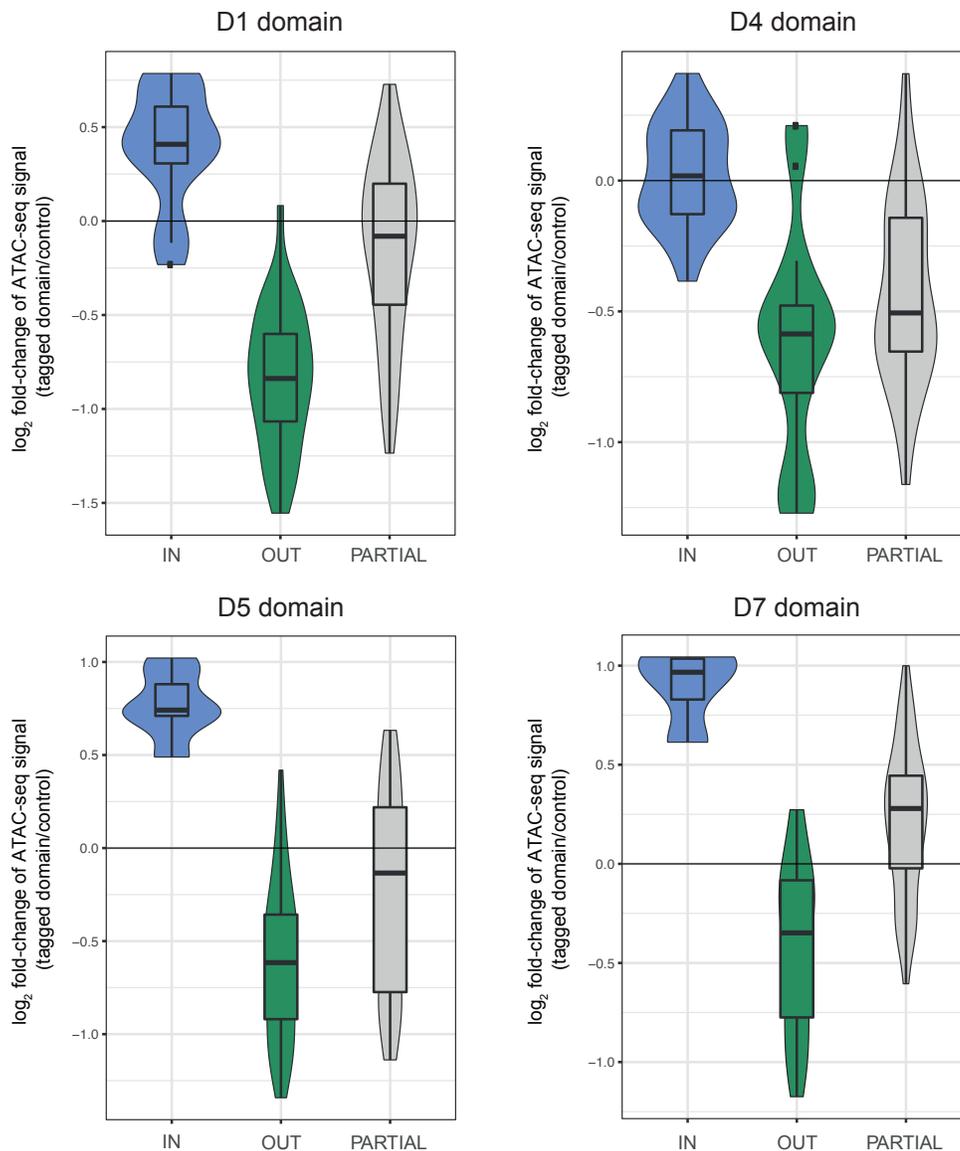


Figure 5.3: Comparison of fold-change accessibility changes at three classes of enhancers. Box plots show distribution of log₂ fold-changes of ATAC-seq signal (total count of Tn5 transposase cuts, mean signal from two replicates) between a given tagged domain and its corresponding whole-embryo control, over three classes of AP enhancers. IN (blue): enhancers whose activity patterns are completely included in the tagged domain. OUT (green): enhancers whose activity patterns are completely excluded from the tagged domain. PARTIAL (grey): enhancers whose activity patterns are present both inside and outside of the tagged domain. Names of the tagged domains are indicated above each plot.

It is important to note that the third class of enhancers whose activity patterns are partially included inside and outside of the tagged domains (PARTIAL) was characterized by a much larger range of their signal fold-changes. In comparison to the whole-embryo control, individual enhancers showed either a relative increase or decrease of their ATAC-seq signal, without following a consistent trend (Figure 5.3).

In summary, I demonstrate that accessibility changes of enhancers are linked to their transcriptional output, with the same enhancer displaying elevated accessibility in the domain of the embryo where it promotes transcription (active) and having reduced accessibility in the domain where it does not induce transcription (inactive).

5.2 Accessibility of axis patterning enhancers is quantitatively correlated with their transcriptional output

Although the AP enhancers were characterized by consistently elevated accessibility in the tagged domain that overlapped their activity patterns, they still displayed a fairly broad distribution of their signal fold-changes relative to the whole-embryo control (Figure 5.2). For example, while ATAC-seq signal of an individual enhancer of *runt*, *run_(+17)*, was comparatively low in all domains from which its activity patterns was excluded (Figure 5.4), its accessibility differed considerably between the remaining two domains that coincided with its activity, D4 and D5. The enhancer was characterized by distinctly elevated signal in the D5 domain, whereas its accessibility in the D4 domain resembled that of the whole-embryo control. This poses a more specific question of interpretation of the ATAC-seq signal measured in individual tagged domains.

5.2.1 Deconvolution of ATAC-seq signal from individual tagged domains

In case of *run_(+17)* enhancer, while both D4 and D5 domains completely overlap the *runt* stripe, they differ considerably in size. As a result, the enhancer is active in a larger portion of the D5 domain as compared to D4, which coincides with its higher accessibility. This suggests that the ATAC-seq signal measured in an individual domain might be a function of the extent of its overlap with an activity pattern of the enhancer. This notion represents a finer assessment than just the binary distinction between enhancers with included (IN) and excluded (OUT) activity patterns.

Similar to the whole-embryo control, each tagged domain consists of a combination of multiple different cell types, often representing a mix of nuclei in which a given axis patterning enhancer is either active or inactive. As a result, the measured ATAC-seq signal could correspond to a weighted average of accessibility states associated with active and inactive enhancers. If the model is correct, we expect to observe a linear correlation between accessibility signal and the proportion of nuclei representing the two states.

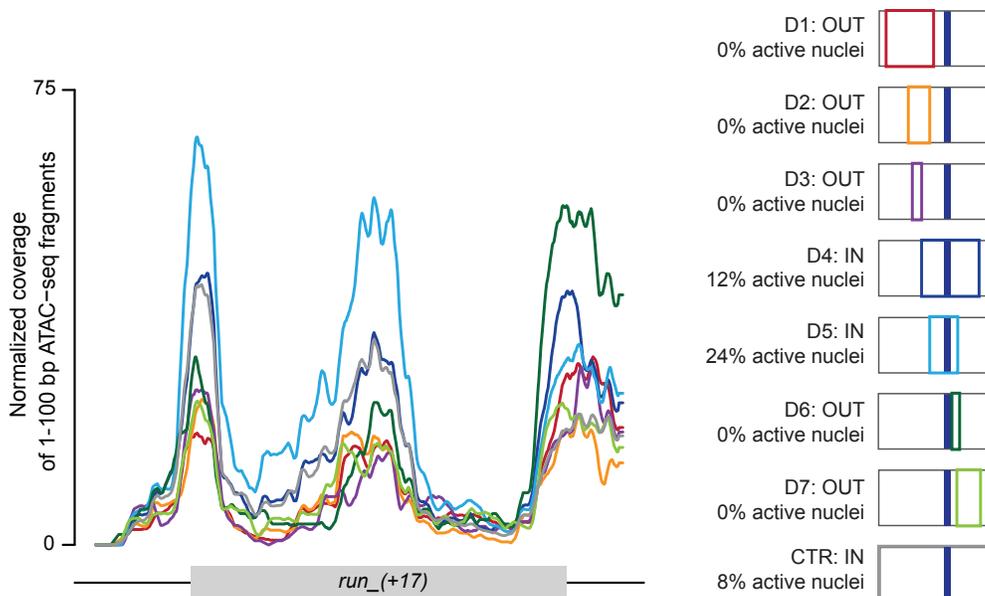


Figure 5.4: Comparison of ATAC-seq signal over *run*₍₊₁₇₎ enhancer between all embryonic domains. Coverage of ≤ 100 bp ATAC-seq fragments over an enhancer of *runt*, *run*₍₊₁₇₎ measured in D1-D7 domains and a whole-embryo control. Domains are color-coded as in the schematic that shows their positions along the AP axis and in respect to the activity pattern of the enhancer (dark blue).

To test this model, I calculated the extent of overlap between each tagged domain and activity patterns of the AP enhancers, in order to evaluate the proportion of enhancers in their active and inactive state. I simplified spatial coordinates of the domains and activity patterns to a single dimension of the AP axis, based on the anti-FLAG staining against the nuclear marker and published RNA *in situ* hybridization images of enhancer-reporter assays, respectively. Importantly though, the *Drosophila* embryo does not have a simple cylindrical shape. In fact, a larger number of nuclei are located in the center of the embryo as compared to its tips. I controlled for these differences by incorporating information on the density of nuclei along the AP axis when calculating overlaps between activity patterns and tagged domains (see Materials and Methods).

5.2.2 Linear correlation between ATAC-seq signal and proportional representation of active enhancers

Subsequently, I compared the ATAC-seq signal measured in each domain against the proportion of nuclei in that domain in which a given enhancer was active. The analysis was performed separately for each AP enhancer and, in addition to the seven data points representing tagged domains, I also included an average signal from the whole-embryo controls

5.2 Accessibility of axis patterning enhancers is quantitatively correlated with their transcriptional output 67

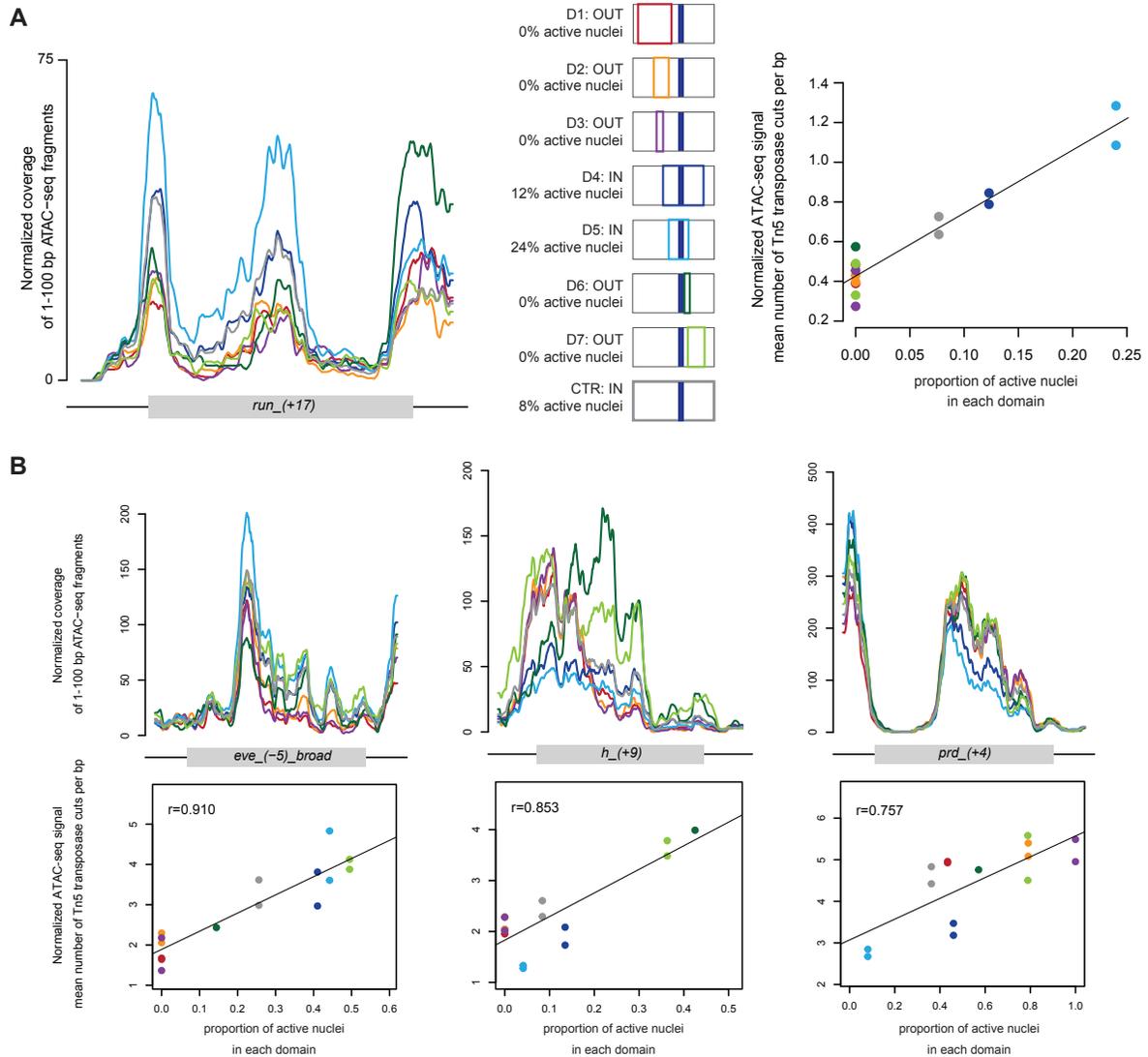


Figure 5.5: Relationship between ATAC-seq signal measured in tagged domains and the proportion of nuclei with active enhancers. A) Left: coverage of ≤ 100 bp ATAC-seq fragments over an enhancer of *run*, *run*₍₊₁₇₎, measured in D1-D7 domains and a whole-embryo control (as in Figure 5.4). Right: total ATAC-seq signal (mean number of Tn5 transposase cuts per bp) plotted against the proportion of a given tagged domain in which the enhancer is active (active nuclei). Each point represents an individual replicate of D1-D7 samples and whole-embryo controls. Coverage tracks and data points are color-coded as in the schematic that shows their positions along the AP axis (middle). D6 Replicate 1 is excluded due to its close similarity to whole-embryo controls. B) Example plots for additional enhancers: *eve*_{(-5)_broad}, *h*₍₊₉₎ and *prd*₍₊₄₎, with an indicated Pearson correlation coefficient (r).

(Figure 5.5).

I observed a strong positive relationship between the ATAC-seq signal and the proportion of nuclei representing an active state. Noticeably, as many as 3/4 of the AP enhancers were characterized by a correlation coefficient above 0.65 (Figure 5.6). For the remaining quarter of elements, I could trace the poor linear relationship to technical difficulties in accurately measuring their activity patterns, i.e. due to additional modulation along the orthogonal DV axis (e.g. *hbn*₋₁) and *gt*₋₅*_broad*) or low quality of available RNA *in situ* hybridization images (e.g. *tara*₋₉, *prd*₋₆ and *Kr*₋₁₁).

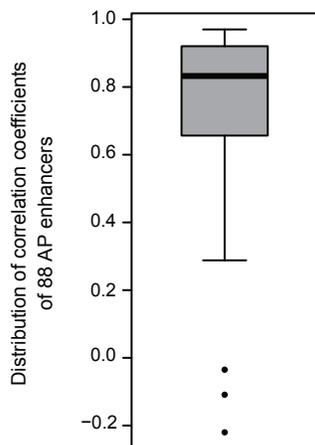


Figure 5.6: Distribution of correlation coefficients. Boxplots represents the distribution of Pearson correlation coefficients across all 88 AP enhancers for the linear relationship between the ATAC-seq signal and proportion of active nuclei in embryonic domains.

5.2.3 Active and inactive axis patterning enhancers display distinct accessibility profiles

In summary, the simple model proved applicable for de-convoluting the ATAC-seq signal from individual domains, potentially offering a more general framework for interpreting genomic signal from complex tissues. As opposed to previous pair-wise comparisons against the whole-embryo control, the new model allows for integration of accessibility signal across multiple independent domains. As a result, I confirmed in a quantitative and systematic fashion the correlation between accessibility of axis patterning enhancers and their regulatory activity. The strong linear relationship suggests that active and inactive enhancers display distinct accessibility states. ATAC-seq signal measured in the tagged domains is simply a proportional contribution of these accessibility signatures. Thus, the model allows for deconvolution of enhancer-specific accessibility profiles in their active (0% active nuclei) and inactive (100% active nuclei) states and the signal fold-change between domains

resulting from different proportional representation of these states.

I conclude that elevated accessibility of axis patterning enhancers coincides with a net activating TF input that promotes transcription of the target gene. Consistently, enhancers display reduced accessibility in regions of the embryo where they receive a net repressive input. It is important to note that here I only consider 'activity' as defined in terms of the transcriptional output. The net repressive input might, in fact, represent either active repression or the absence of activators.

Although I observe a strong relationship for a vast majority of the considered AP enhancers, they still display considerable variation in terms of the strength of their correlations. This could reflect technical inaccuracies in estimating the contribution of active and inactive states in each tagged domain or, alternatively, suggest that accessibility of enhancers is influenced by additional parameters than just the net regulatory input.

5.3 Inactive enhancers exhibit residual accessibility

In the previous section, I demonstrated that axis patterning enhancers are characterized by relatively lower accessibility in their inactive state. To determine the scale of this accessibility reduction, I aimed to test whether inactive enhancers exhibit a closed chromatin structure, comparable to non-functional regions of the genome, or whether they still retain some residual level of accessibility.

For this purpose, I identified a comprehensive set of inaccessible genomic intervals that displayed background sensitivity to transposase digestion, i.e. were not identified as ATAC-seq peaks in any of the tagged domains or whole-embryo controls (see Materials and Methods). I compared accessibility of these background regions to the distribution of accessibility signal from active and inactive AP enhancers. For this analysis, I only considered ATAC-seq signal from domains fully representing the inactive (0% active nuclei) and active (100% active nuclei) states. To correct for varying size of the considered intervals, I expressed their accessibility as a mean number of cleavages introduced by the Tn5 transposase per base pair.

First of all, I could demonstrate that accessibility of inactive AP enhancers was significantly elevated in comparison to the inaccessible portion of the genome (Figure 5.7). Interestingly, while confirming that active enhancers were characterized by a significantly more open chromatin structure than inactive enhancers, the analysis also revealed a rather substantial spread of accessibility signal of individual elements. This is consistent with the earlier observations of selected enhancers (Figure 4.1, Figure 5.1, Figure 5.5), and suggests altogether that different axis patterning enhancers do not share the same level of accessibility. Instead, the accessibility signal in either of the regulatory states appears to be an enhancer-specific signature.

I conclude that accessibility of inactive enhancers, even though significantly reduced, does not decrease to the level of background inaccessible regions. Instead, axis patterning enhancers are characterized by an open chromatin structure in both their active and inactive states, although the absolute level of accessibility is unique to each enhancer.

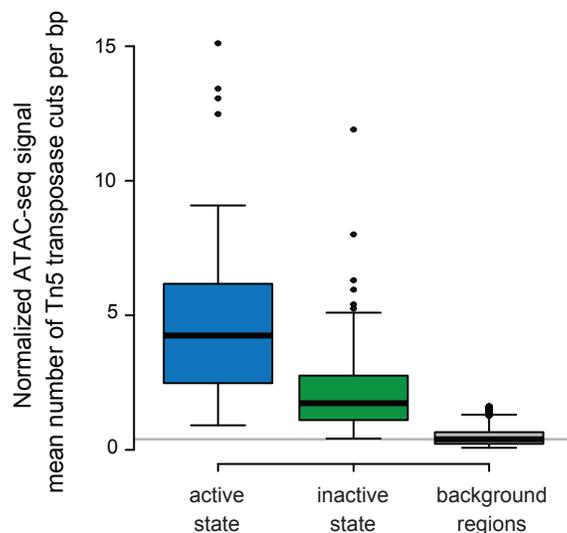


Figure 5.7: Comparison of ATAC-seq signal between background genomic regions and axis patterning enhancers. Box plots show a distribution of ATAC-seq signal (mean number of Tn5 transposase cuts per bp) between active AP enhancers (signal from tagged domains with 100% active nuclei; active state), inactive AP enhancers (signal from tagged domain with 0% active nuclei; inactive state) and 9 309 background regions of the genome (mean signal across all tagged domains; background regions).

5.4 Local accessibility modulation within individual enhancers

It is important to note that in addition to regional modulation along the AP axis, the ATAC-seq analysis also reveals local variation in accessibility signal within individual enhancers. This is most convincingly demonstrated by a non-uniform distribution of short nucleosome-free (≤ 100 bp) ATAC-seq fragments (Figure 5.1). With high probability, troughs of the coverage signal represent nucleosome-occupied intervals, suggesting that not all potential TF binding sites within an enhancer are equally accessible to their target regulators.

Interestingly, the position of nucleosome-free and nucleosome-occupied regions remains fairly conserved across all tagged domains. Although a more thorough examination is required, this suggests the presence of well-positioned nucleosomes within the axis patterning enhancers. Thus, differential regulatory activity of enhancers appears to be accompanied by global changes in their degree of accessibility, preserving the local distribution of nucleosomes.

Chapter 6

Discussion

6.1 Summary of key findings

This study aimed to examine the role of chromatin context in the regulatory activity of axis patterning enhancers, which mediate the establishment of body axes and specification of spatial coordinates during the earliest stages of *Drosophila* embryogenesis. I assayed chromatin accessibility of the enhancers in order to determine whether their chromatin organization is uniform throughout the embryo, and thus invariant to the differential regulatory input, or whether it differs along the body axes, possibly implying an active interplay with the regulatory machinery. The study provided clear evidence in support of the latter model. I demonstrate significant regional variation of chromatin accessibility that is strongly correlated with regulatory activity of the axis patterning enhancers.

Remarkably, the analysis revealed two superimposed properties of chromatin organization of axis patterning enhancers. First, the enhancers are characterized by open chromatin structure along the entire span of the assayed AP axis. This is consistent with the fact that they are targeted by patterning TF throughout the entire embryo and that nucleosome depletion has been demonstrated in multiple paradigms to be essential for recognition of DNA binding sites by different functional proteins [105]. In addition, I observe a fine regional modulation of the enhancers' accessibility, which is correlated with its transcriptional output. Accessibility of axis patterning enhancers is elevated in regions of the embryo where they receive a net activating TF input and promote transcription of the target gene. Their accessibility is comparatively reduced when the enhancers receive a net repressive input, but never decreases to background levels of the inaccessible genome.

Genome-wide comparison of accessibility profiles provided an additional insight into the scale of global accessibility variation in the cellular blastoderm. I demonstrate that chromatin organization of the yet undifferentiated blastoderm is already divergent at the onset of zygotic genome activation. As much as one quarter of the accessible genome displays significant regional variation. Importantly, out of the entire spectrum of functional genomic elements, the most variable regions correspond to known and putative axis patterning enhancers. I conclude that differential accessibility is a signature of patterning

cis-regulatory elements that receive differential regulatory input.

6.2 Advantages and limitations of the experimental strategy

Recently, several experimental approaches have been used to increase spatial resolution of genomic assays in the early *Drosophila* embryo. Cryo-slicing, i.e. sectioning of frozen embryos, has been used to study spatial patterns of gene expression (RNase-seq [139, 140]), regional differences in TF occupancy (ChIP-seq [141]), as well as regional variation in chromatin accessibility profiles along the AP axis (ATAC-seq [90]). Examination of mutagenized embryos with restricted or uniform cell identities has been used to study the distribution of histone modifications along the DV axis (ChIP-seq Koenecke2017) as well as the effect of an individual repressor TF on chromatin organization of its target enhancers [28]. Finally, single-cell assays have been applied to acquire high-resolution profiles of gene expression (RNase-seq [16]) and chromatin accessibility (ATAC-seq [89]) during early *Drosophila* embryogenesis. In this section, I compare and contrast these three experimental strategies with affinity purification of genetically tagged nuclei (INTACT) that I applied in this study.

Genetic tagging allows for precise definition and reproducible isolation of nuclei from specific sections of the embryo. Unlike in the single-cell assay, spatial coordinates of the analyzed nuclei can be directly assigned with high accuracy. This is also in contrast with cryo-slicing, where precision of sectioning depends on the experimenter's skills, while isolation of the exact domain of interest might be problematic without any additional morphological markers.

Genetic tagging also offers high flexibility in terms of the size and location of the analyzed embryonic domains, especially taking into consideration the large number of well-characterized axis patterning enhancers. At the same time, analysis in the mutant background depends highly on the availability of established mutant lines. While cryo-slicing has been used to section the AP axis into narrow 13-25 μm slices (up to 30 slices per embryo) [139, 140], it cannot be easily applied to domains that are modulated along both embryonic axes.

In my experimental strategy, the nuclear tag is expressed in an otherwise wild-type background and does not have any detectable effect neither on the progression of embryogenesis nor on the accessibility landscape of the blastoderm embryo (as demonstrated in section 3.3.2). In contrast, ectopic expression of key regulators in the mutant embryos is likely to alter operation of the patterning gene regulatory networks. As a result, the mutant states might not reliably correspond to cell identities from wild-type embryos.

Even though ATAC-seq has been successfully applied in other studies to single embryos [68] and individual nuclei [89], loss of material during affinity purification requires batch collections of embryos. In spite of narrow 20-min staging, embryonic collections analyzed in this study were nevertheless contaminated with older developmental stages (Figure 2.10).

Furthermore, while INTACT has been reported to offer high purity of isolations [86], contamination with untagged nuclei from other regions of the embryo cannot be fully excluded. In this respect, cryo-slicing offers an important advantage. Hand-sorting of embryos allows for careful selection of the embryonic stage, while manual sectioning reduces contamination with nuclei from other embryonic domains.

In this study, I applied an arbitrary threshold when measuring the boundaries of tagged domains based on their anti-FLAG immunostaining. With a binary distinction between tagged and untagged nuclei, I did not address the fact that the nuclear marker was expressed in a concentration gradient, in line with the graded activity of its enhancer driver. I did not test how the intensity of anti-FLAG staining (or the level of UNC84-3xFLAG expression) translated to the efficiency of INTACT isolations. For this reason, genetic tagging brings some level of uncertainty when estimating positions of the analyzed domains. Additionally, it is possible that the span of the tagged domain might not be uniformly represented in the ATAC-seq sample. For example, a higher amount of UNC84-3xFLAG could result in more efficient purification and, therefore, overrepresentation of nuclei from a peak of the concentration gradient.

6.3 Comparison with other studies examining regional modulation of chromatin accessibility in *Drosophila* embryos

Regional variation in chromatin organization that I report in this study is consistent with two very recent publications that also profiled chromatin accessibility with spatial resolution in the *Drosophila* blastoderm, either in individual nuclei [89] or cryo-sliced anterior and posterior halves of the embryo [90]. Importantly, both studies also demonstrate a correlation between local accessibility and regulatory activity of axis patterning enhancers.

6.3.1 Single-cell ATAC-seq by Cusanovich et al. (2018)

With single-cell ATAC-seq, Cusanovich et al. [89] provided a global and comprehensive view of accessibility variation during the *Drosophila* early development. However, the current methodology did not allow for reconstruction of the exact positional coordinates of the profiled embryonic nuclei. Instead, Cusanovich et al. clustered accessibility profiles from individual nuclei and inferred their developmental identities based on the signal enrichment over a set of tissue-specific regulatory elements.

As a result of clustering, the single-cell assay did not offer a considerably higher spatial resolution than this study on the seven AP domains. For example, global accessibility variation in 2-4 h embryos was evaluated by comparing five big clades of nuclei. The pre-gastrulation stage of the cellular blastoderm was more specifically classified into six clusters, which were mainly located along the AP axis. However, their identities were

assigned only as anterior or posterior, without any additional inference of their size or exact position.

Furthermore, as the study mainly focused on the evolution of accessibility landscapes throughout the first 12 h of *Drosophila* embryogenesis, Cusanovich et al. performed rather broad collections of embryos. The 2-4 h time window that represented the cellular blastoderm additionally contained older gastrulating embryos as well as the earlier syncytial blastoderm that likely represented a stage prior to the global zygotic genome activation.

It is important to note that single-cell ATAC-seq involves tagmentation of individual diploid nuclei, which means that each base-pair can be probed at maximum by four transposase cleavages. This results in a decreased dynamic range of the ATAC-seq signal, which reduces sensitivity of the assay in detecting quantitative modulation of the accessibility profiles. As a consequence, in the analysis by Cusanovich et al., axis patterning enhancers displayed background accessibility levels outside of their activity domains.

Interestingly, Cusanovich et al. fixed the embryos with formaldehyde before homogenization and isolation of individual nuclei. Size distribution of the cross-linked ATAC-seq libraries was dominated by a peak of 200-bp nucleosomal fragments. By contrast, ATAC-seq on native chromatin that was performed in this study allowed for greater recovery of shorter nucleosome-free fragments, which are more informative of DNA accessibility to TF binding (section 3.1).

6.3.2 ATAC-seq on cryo-sliced anterior and posterior domains by Haines and Eisen (2018)

Haines and Eisen [90] focused specifically on accessibility variation along the AP axis by performing ATAC-seq on cryo-sliced anterior and posterior halves of the cellular blastoderm.

With hand-sorting of individual embryos, Haines and Eisen were able to specifically target mitotic cycle 14, corresponding to the stage of cellular blastoderm. However, in comparison to genetic tagging, hand-dissection of embryos led to a lower reproducibility of the replicate experiments, with the Pearson correlation coefficients ranging between 0.80 - 0.88 (when examining 1-kb windows spanning the entire *Drosophila* genome).

Analysis of only two embryonic domains considerably limited the statistical power when evaluating accessibility variation in the cellular blastoderm. Fewer than 120 ATAC-seq peaks showed significant differences in their accessibility signal along the AP axis, corresponding to only 1.7% of all identified accessible regions. Additionally, the analysis of individual regulatory elements was limited to axis patterning enhancers that were exclusively active in the anterior (30 elements) or posterior (9 elements) of the embryo.

6.3.3 Summary

In comparison to the analysis of seven genetically tagged domains, single-cell ATAC-seq does not provide a considerably higher regional resolution. Both studies reveal a surprisingly similar scale of regional variation in the early *Drosophila* embryo, with one quarter

of the accessible genome showing significant signal differences. In contrast, Haines and Eisen conclude that a vast majority of the genome is characterized by an invariant degree of accessibility.

Importantly, Cusanovich et al. rely on *in silico* inference of spatial coordinates of the analysed nuclei (based on tissue-specific markers), while my approach together with that of Haines and Eisen allow for a direct assignment of spatial coordinates to the measured accessibility profiles.

All three studies demonstrate elevated accessibility of axis patterning enhancers in the regions of the embryo where they promote transcription of their target gene. This study together with that of Haines and Eisen show residual accessibility of inactive enhancers, while Cusanovich et al. demonstrate that their accessibility is reduced to the background level of the inaccessible genome. However, this discrepancy is most probably linked to the limited dynamic range of the single-cell ATAC-seq assay. While in case of this study, residual accessibility of inactive enhancers can be potentially accounted for by a small contribution of untagged nuclei or older embryonic stages, manual staging and dissection of embryos reduces the probability of any contamination in the study by Haines and Eisen.

Interestingly, Haines and Eisen detect higher accessibility variation of the anterior enhancers in comparison to the posterior elements. They link this larger magnitude of accessibility changes to the activity of a single TF, the anteriorly deposited Bicoid. However, in this study a vast majority of 88 AP enhancers show a similar and consistent relationship between their ATAC-seq signal and regulatory activity, regardless of their size, position along the AP axis and identity of their target genes.

In this study, comparison of accessibility profiles across a larger number of embryonic domains allowed for more thorough and quantitative analysis of the relationship between enhancers' accessibility and regulatory activity. I was able to demonstrate more systematically the distinction between accessibility states of active and inactive enhancers. In particular, the simple model allowed for disentangling two components that accounted for regional variation of the ATAC-seq signal: enhancer-specific accessibility profile in its active and inactive state and different proportional representation of these states in the tagged domains. With that, I offered a more general framework for interpreting ATAC-seq signal from complex tissues, which can be also applied to other heterogeneous systems.

6.4 Differential accessibility as a signature of axis patterning enhancers

6.4.1 Relationship between activity and accessibility

In this study I demonstrate significant quantitative variation of accessibility profiles in the cellular blastoderm. I show that among a wide range of active genomic elements that display open chromatin structure, the most variable intronic and intergenic regions are enriched in axis patterning enhancers. Additionally, I provide a link between their differential accessibility and differential regulatory activity.

When discussing the relationship between accessibility and activity of *cis*-regulatory elements, it is important to distinguish between two definitions of active enhancers. In a broad definition of being deployed by the cell and receiving regulatory input, active *cis*-regulatory elements have been demonstrated in multiple paradigms to display highly open chromatin structure [67, 142]. This is consistent with eviction and destabilization of nucleosomes being essential for the exposure of binding sites to target TFs, polymerases and other functional proteins [105]. In this sense, accessibility has been used as a predictor of activity in multiple model systems [89, 143]. Consistently, accessibility landscapes show dynamic evolution during embryogenesis, which is a reflection of different *cis*-regulatory elements being deployed to drive different transcriptional programs [144, 67]. According to this broad definition, axis patterning enhancers are active throughout the blastoderm embryo since they are targeted by different combinations of input TFs along the entire body axis. This is consistent with the fact that they display an open chromatin structure in all assayed domains.

However, this study focuses on the less explored relationship between accessibility and activity of enhancers, here defined as their transcriptional output. I demonstrate that accessibility of *cis*-regulatory elements is additionally modulated in accordance with the regulatory input that they receive and its cumulative effect on transcription of their target gene. While general openness of genomic elements has been used as a signature of their temporal deployment by the cell, I propose that differential accessibility is a marker of differential regulatory activity in complex tissues.

This notion of differential accessibility as an indication of differential activity has been already proposed by Pearson et al. [130] in the context of embryonic development of the *Drosophila* central nervous system. The authors demonstrated that elevated accessibility of genomic elements (in comparison to the whole-embryo sample) was a better indicator of midline-specific enhancer activity than just their general openness. Thus, consistent with my observations, Pearson et al. provided a link between cell-type specific transcriptional activity and differential accessibility of *cis*-regulatory elements, and used this relationship to predict novel midline-specific enhancers.

6.4.2 Interpretation of differential accessibility in the context of local chromatin organization

How can quantitative differences in ATAC-seq signal be interpreted in terms of the local change in chromatin organization at axis patterning enhancers?

As discussed in Chapter 3, frequency of Tn5 transposase cleavages approximates the level of DNA exposure in the chromatin context. Due to steric hindrance imposed by histones, the transposase primarily introduces cuts within the linker DNA. Accessibility of individual base-pairs within the linker can be affected by two factors: 1) local occupancy of TFs and other proteins at their binding sites [145] and/or 2) positioning and stability of flanking nucleosomes, including certain histone modifications that affect organization of the linker DNA [98].

6.5 Differential accessibility as a metric for *de novo* discovery of enhancers 77

Since axis patterning enhancers receive varying TF input at different positions along the AP axis, it is plausible that accessibility of the linker DNA could be modulated by the composition and occupancy of bound TFs. As different activators and repressors occupy their specific binding sites, the overall distribution of TFs along the enhancer is expected to vary in each tagged domain. Consequently, different base-pairs are expected to be protected from Tn5 cleavages in each ATAC-seq library. In fact, though, all tagged domains are characterized by very similar coverage profiles of nucleosome-free ATAC-seq fragments, which reveals little modulation in accessibility of individual base-pairs of the linker. Instead, the accessibility signal increases or decreases rather uniformly across longer stretches of DNA between the well-positioned nucleosomes.

In spite of ATAC-seq being used for mapping individual TF binding sites [145], I believe that in this experimental set-up, rather than probing accessibility of individual TF binding sites, I examine the global accessibility of the linker DNA. Quantitative differences in ATAC-seq signal can be interpreted then mainly in terms of the varying stability and positioning of nucleosomes, which affects the overall sensitivity of the linker DNA to transposase cleavages.

6.5 Differential accessibility as a metric for *de novo* discovery of enhancers

One quarter of the accessible genome shows quantitative accessibility differences in the cellular blastoderm, corresponding to around 3 000 intronic and intergenic differential regions. On the other hand, the list of known *cis*-regulatory elements of the early *Drosophila* embryo already comprises several hundred enhancers, i.e. around 500 elements from the REDfly database [45] and around 600 elements represented by Vienna Tiles [56]. How feasible is it that the remaining unannotated differential ATAC-seq peaks represent novel axis patterning enhancers, taking into consideration multiple methods that have been already used for their discovery (section 1.2.2)?

Around 1 400 genes are characterized by patterned expression in the *Drosophila* blastoderm [15]. If we assume that all 3 000 differential peaks represent their regulatory elements, this corresponds to around two enhancers per patterned gene. This is quite a realistic estimate, taking into consideration that the key gap and pair-rule genes are expressed in multiple domains, each regulated by an independent element. Additionally, an increasing number of studies report the presence of partly 'redundant' regulatory elements (shadow enhancers) that simultaneously drive very similar expression patterns of the same target gene, potentially serving as a mechanism of increasing expression robustness [41].

In comparison to annotated differential peaks that overlap known axis patterning enhancers, a smaller proportion of unannotated peaks is bound by AP TFs, and each element is targeted on average by a combination of fewer regulators (Figure 4.14). While this might argue against their correspondence to functional *cis*-regulatory elements, it might also reveal a bias of the previous attempts at identifying axis patterning enhancers.

Systematic genome-wide screens for early *Drosophila* enhancers relied primarily on local clustering of TF binding sites, as evaluated either by PWM-based predictions or co-localization of ChIP signal of multiple regulators [49, 48, 53, 52, 41, 50, 51]. Consequently, tested candidate enhancers were enriched in regions that were combinatorially regulated by multiple different TFs. Additionally, these attempts mainly considered occupancy by key maternal and gap TFs only [49, 41], sometimes being limited to just a single TF such as Bicoid [53, 52]. It is possible then that these studies omitted enhancers regulated by a smaller combination of TFs or TFs from the lower tiers of the AP regulatory network, including pair-rule genes. At the same time, the large scale enhancer-screen of 2-kb Vienna Tiles covered only 13.5% of the non-coding non-repetitive *Drosophila* genome [56].

Therefore, I propose that differential accessibility can complement the previous efforts at discovering axis patterning enhancers in the cellular blastoderm and allow for expansion of the current list of known regulatory elements. Taking into consideration the ongoing attempts at refining PWMs of the key TFs [146] as well as relating their ChIP signal to functionality of the enriched regions [138], I believe that differential accessibility can serve as a more direct measure of differential activity of genomic elements. One advantage of this approach is that accessibility changes along the AP axis (or more generally, in different regions of the embryo) already offer a rough approximation of activity patterns of the candidate elements. Additionally, ATAC-seq provides single base-pair resolution, which might allow for more accurate delineation of their genomic coordinates.

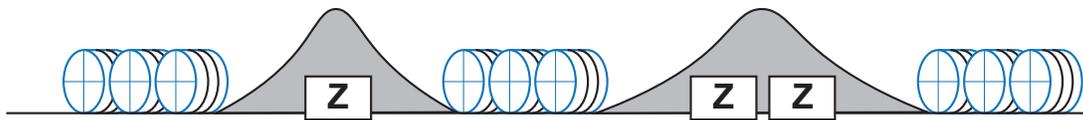
Furthermore, I propose that ATAC-seq data generated in this study can be additionally used to refine coordinates of known axis patterning enhancers. Very often local distribution of the accessibility signal is not in full agreement with the position of annotated *cis*-regulatory elements (Figure 5.1). Taking into consideration that boundaries of these elements were often defined by genetic dissection with restriction enzymes [46, 47, 27] or overlap of broad ChIP peaks [53, 41], it is likely that accessibility is a more direct and accurate approximation of the position of functional genomic regions targeted by input TFs. In fact, it has been recently demonstrated that incorporation of DNase-seq data improved performance of computational models of axis patterning enhancers that previously relied just on the DNA sequence to predict their activity patterns [147].

6.6 Independent mechanisms for establishment and modulation of enhancer's accessibility

This study demonstrated a strong relationship between regional modulation of accessibility and differential activity of axis patterning enhancers. Are these local differences in chromatin organization a cause or consequence of the different transcriptional output of the enhancers? Based on these findings and evidence from other studies, I propose a model in which accessibility of axis patterning enhancers is uniformly established across the entire *Drosophila* embryo prior to zygotic genome activation and is followed by its regional modulation as a result of differential activity of gene regulatory networks.

Prior to zygotic genome activation:

- establishment of the basal level of accessibility by ubiquitous pioneer TFs (Z: Zelda)
- uniform priming of enhancers for the subsequent TF input



During operation of the gene regulatory networks:

- regional modulation of enhancer's accessibility along the AP axis
- accessibility changes correlated with the net regulatory input (A: activators, R: repressors)

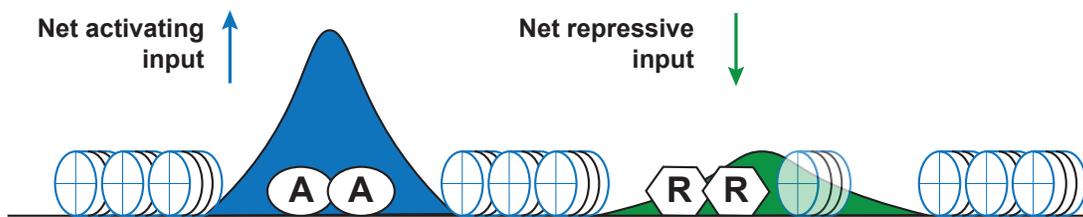


Figure 6.1: Proposed model of chromatin organization at axis patterning enhancers during the earliest stages of *Drosophila* embryogenesis. Global establishment of accessibility prior to zygotic genome activation is followed by finer regional modulation during the operation of gene regulatory networks in the cellular blastoderm.

Axis patterning enhancers have been demonstrated to display highly accessible chromatin structure and reduced nucleosome occupancy already prior to zygotic genome activation [68, 7]. This early opening of enhancers has been shown to be mediated by maternally deposited Zelda, a pioneer TF that either actively destabilizes nucleosomes or prevents their deposition after the initial rapid rounds of genome replication [7, 8]. ChIP peaks of key regulators of the AP and DV patterning networks are most strongly enriched in a recognition motif of Zelda [148] and the protein has been shown to facilitate binding of these TFs to their target enhancers [149, 141, 132]. In fact, Zelda binding sites have been proven necessary for the activity of synthetic enhancers in blastoderm embryos [150].

It appears then that the pioneering activity of Zelda facilitates the access of TFs to their binding sites on axis patterning enhancers. Yet, the uniform distribution of this protein in the embryo falls short of providing spatial cues that could regionally modulate accessibility. Instead, positional information in the *Drosophila* embryo is encoded by concentration gradients of patterning TFs and, as a result, is processed by the enhancers only during operation of the gene regulatory networks after zygotic genome activation.

I propose, therefore, that the basal level of accessibility of axis patterning enhancers is established already prior to zygotic genome activation, possibly by ubiquitous Zelda or other related mechanism, thereby uniformly priming them for the subsequent TF input (Figure 6.1). The basal accessibility would allow enhancers to sample different concentrations and combinations of patterning TFs with no bias towards activating and repressing TFs, consistent with the high plasticity of the blastoderm embryo to ectopic regulatory input [42, 43]. In turn, differential TF occupancy and the resulting differential regulatory activity, would regionally modulate the basal Zelda-dependent enhancer accessibility, to the variation observed in this study.

Therefore, rather than predetermining the activity of axis patterning enhancers, I propose that the reported regional modulations in chromatin accessibility result from differential activity of gene regulatory networks after zygotic genome activation.

6.7 Mechanisms of accessibility modulation at axis patterning enhancers

I demonstrate a strong correlation between accessibility and the net regulatory input received by the axis patterning enhancers. However, at this point it is not clear at what level of regulatory activity this relationship is manifested. Accessibility of enhancers might be directly modulated by their local interactions with TFs or, alternatively, during interactions with the core promoter when the transcriptional output is executed.

One possible scenario is a direct correlation between accessibility of enhancer elements and transcription of their target gene. Although little evidence is available, it cannot be excluded that local accessibility of enhancers might be affected by the frequency of their interactions with the target promoter and the RNA polymerase II. It is also possible that the recently described phenomenon of enhancer transcription [151] could also contribute to the local accessibility variation. Progression of RNA polymerase involves local displacement of nucleosomes, thus transiently increasing DNA exposure [152], while the amount of eRNA (enhancer RNA) has been demonstrated in other model systems to correlate with expression of their nearby protein-coding gene [153].

An alternative scenario involves an active interplay between regulatory machinery and local chromatin organization of axis patterning enhancers. This is supported by various lines of evidence that imply a role of patterning TFs in modulating epigenetic marks and nucleosome stability of their target *cis*-regulatory elements.

The observed correlation between accessibility of axis patterning enhancers and their transcriptional output is highly consistent with a recent study by Koenecke et al. [82], who demonstrated that active and inactive enhancers of the DV network display different histone marks. Active enhancers are characterized by elevated levels of H3K27ac, which is consistent with lysine acetylation being shown in other paradigms to reduce nucleosome stability, while increasing accessibility of its linker DNA [154, 155, 98]. Importantly, Li and Arnosti [28] reported that overexpression of an individual repressor, Knirps, results

in reduction of H4 acetylation over its target enhancer, providing therefore a link between TF occupancy and chromatin states. In fact, key TFs of the patterning networks have been shown to recruit histone acetylases and deacetylases via their co-activators and co-repressors, respectively [74, 75, 81, 80].

Regardless of whether histone modifications are the only mechanisms by which TFs could potentially modulate chromatin accessibility, the repressor Knirps has been demonstrated to increase local nucleosome occupancy [28], while the activator Bicoid has been reported to promote accessibility of its target enhancers [156]. While Haines and Eisen [90] propose that Bicoid, as a maternally deposited activator with a broad anterior concentration gradient, plays a major role in modulating accessibility of enhancers, I speculate that this process is not exclusive to individual regulators. The correlation between regional accessibility and regulatory activity is strong for all considered axis patterning enhancers, regardless of their position along the assayed AP axis and, in consequence, the composition of input TFs.

In light of the aforementioned studies, I propose that activators and repressors of both AP and DV patterning networks share a common mechanism of action, with activators promoting and repressors reducing the accessibility of their target enhancers. As a result, I envision a dynamic interplay between these two opposing effects, with the final chromatin state being determined by the net regulatory input.

6.8 Outlook

6.8.1 Role of local accessibility modulations at axis patterning enhancers

Results of this study raise a question on the functional role of the local accessibility modulations of axis patterning enhancers. I envision several potential mechanisms, which are not necessarily mutually exclusive.

Quenching of activator binding sites. The link between accessibility reduction and the net repressive input could be a manifestation of short-range repression. Stabilization of nucleosomes, either via a direct interaction or recruitment of histone modifying enzymes, has been proposed by Teif and Rippe [69] as a mechanism of action of repressor TFs. This would result in obstruction of the nearby activator binding sites and reduction of the activating input received by the enhancer. This proposed mode of action is consistent with the dominant role of repression in the cellular blastoderm. Maternally deposited activators are broadly distributed in the embryo, while their operation is limited by repressors with more restricted spatial patterns.

Detection of concentration gradients. Hannon et al. [156] proposed that Bicoid, by affecting accessibility of its enhancers in a concentration-dependent manner, provides information on their position along the embryonic axis. Thus, the concentration-dependent effect of activator and repressor TFs on the local organization of enhancers could serve as a sensitive mechanisms of sensing TF dosage and conveying specific positional information.

Integration of regulatory input. Opposing effects of activators and repressors on the local chromatin accessibility, especially if executed in a concentration-dependent manner, could allow for integration of the complex regulatory input into a simple read-out of the chromatin organization. It is not clear though how the variable accessibility of enhancers would affect their communication with the promoter and the basal transcriptional machinery. Whether this would be mediated simply by exclusion of the activating input or indirect interactions with the repressive histone marks still remains to be tested.

6.8.2 Implications for the future studies on axis patterning enhancers

This study provides multiple insights that might guide future work on deciphering the regulatory logics of axis patterning enhancers.

First of all, the results suggest an interplay between the regulatory machinery and the immediate chromatin context of enhancers. In addition to investigating the composition and organization of TF binding sites, consideration of TF interactions with nucleosomes and histone modifying enzymes could provide additional mechanical insights into integration and interpretation of the regulatory input by axis patterning enhancers.

Secondly, this study also demonstrates that accessibility of axis patterning enhancers is not uniform along their entire length and suggests the presence of well-positioned nucleosomes inside these elements. This could account for the contribution of the background sequence between TF binding sites, which would possibly affect the positioning and stability of nucleosomes. Additionally, potential obstruction of certain predicted TFBSs by nucleosomes could inform modelling efforts and allow for more specific selection of binding sites that contribute to the activity patterns of axis patterning enhancers.

Chapter 7

Materials and Methods

7.1 Generation of transgenic strains

7.1.1 Elements of the expression cassette

UNC-84 coding sequence was amplified from pMUH_unc84_tdTFlag (a gift from Sean Eddy, Addgene plasmid #46024 [86]). Internal BsaI recognition sites were removed by synonymous substitutions (G1989A and T2088C), using QuickChange Multi Site-Directed Mutagenesis Kit (Agilent Technologies). RanGAP coding sequence was amplified from cDNA collection of *Drosophila* Genomics Resource Center (LD16356 clone). UNC84-3xFLAG and RanGAP-3xFLAG nuclear tags were generated by C-terminal fusion of the coding sequence to a 25-aa peptide linker (HL4 from [157]), followed by 3xFLAG tag (Sigma) and SV40 polyadenylation signal. RanGAP-3xFLAG nuclear tag was a gift from Andrea Ennio Storti. Enhancers of gap and pair-rule genes were amplified from genomic DNA, according to their definitions from [64]. 69-bp DNA linker represents a standard multiple cloning site (a gift from Myrto Deligiannaki). Sequences of DSCP [91], Hs43 [113] and P-element [115] promoters were a gift from Katja Frhauf.

7.1.2 Entry vectors

Enhancers, DNA linker, promoters and nuclear tags were flanked by BsaI recognition sites and specific 4-bp ligation sites in an additional round of PCR amplification. A-overhangs were added to blunt ends of PCR amplicons by combining 8 μ l of the purified PCR product with 1 μ l of GoTaq[®] G2 Hot Start Polymerase (Promega), 2 mM of dATP and 0.75 mM of MgCl₂ in the total volume of 20 μ l, followed by incubation at 74°C for 20 min. The polymerase was pre-activated by heating to 95°C for 2 min. A-tailed PCR products were directly integrated into the pCR[®]8/GW/TOPO[®] backbone (pCR[®]8/GW/TOPO[®] TA Cloning Kit, Invitrogen). 8 μ l of DNA was combined with 2 μ l of the salt solution and 1 μ l of the vector, followed by 15-min incubation at room temperature and bacterial transformation. Correct integration of the elements in the entry vectors was confirmed by

Sanger sequencing.

All PCR amplification reactions were performed with iProof™ High-Fidelity DNA Polymerase (Bio-Rad), according to the manufacturer's protocol. DNA sequences of the elements of expression constructs as well as their corresponding PCR primers are available in Supplementary Material A.

7.1.3 Assembly of expression constructs

pBDP plasmid (a gift from Gerald Rubin, Addgene plasmid #17566 [91]) was modified to serve as a destination vector for Golden Gate Assembly with BsaI Type IIs restriction enzyme [119, 120]. First, an internal BsaI restriction site in the coding sequence of AmpR gene was removed by synonymous substitution (T720C), using QuickChange Multi Site-Directed Mutagenesis Kit (Agilent Technologies). BsaI site within phiC31 attB was retained to avoid interference with the integraton process. Second, a cassette with BsaI recognition sites and 4-bp ligation sites flanking the DNA linker was inserted in the multiple cloning site of the pBDP plasmid, using FseI and KpnI restriction enzymes. The modified pBDP plasmid was sequenced to confirm its correct assembly.

For the final Golden Gate Assembly reaction, 80 fmol of entry vectors with the enhancer, DNA linker, promoter and nuclear tag were combined with 80 fmol of the pBDP plasmid, 2 µl of T4 DNA ligase (3 U/ml, Promega M180B) and 2 µl of BsaI (10 U/ml, NEB) in the final volume of 20 µl. Correct assembly of the expression constructs was validated by Sanger sequencing.

7.1.4 Site-specific genomic integration

Expression constructs were inserted into attP2 [116] and attP40 [117] integration sites in the wild-type background. Homozygous lines representing independent insertions of the same construct were used as biological duplicates, except for the *eve_stripe1_DSCP_UNC84-3xFLAG* construct (D3 domain) for which only one transgenic line was obtained.

7.2 Characterization of tagged domains

7.2.1 Immunostaining of transgenic embryos

Specific localization of the nuclear tag in transgenic embryos was validated with anti-FLAG antibody staining. In short, embryos were collected in population cages (broad window: 0-4 h AEL), dechorionated in 50% bleach (2 min), fixed for 20 min in 4% formaldehyde-heptane and devitellinised with methanol. Tagged domains were visualized by anti-FLAG immunostaining according to standard procedures [158]. Primary antibody: monoclonal anti-FLAG M2 (Sigma-Aldrich: F1804) in 1:500 dilution. Secondary antibody: Alexa Fluor 647-conjugated goat anti-mouse antibody (Jackson ImmunoResearch: 115-605-166)

in 1:500 dilution. The embryos were imaged under LSM 710 Carl Zeiss laser-scanning confocal microscope.

7.2.2 Definition of positions of tagged domains along the AP axis

Positions of tagged domains were measured by projection of the anti-FLAG signal onto the axis connecting the anteriormost and posteriormost tips of the embryo, and were expressed as percent of the axis length (1-100%).

7.3 Preparation of ATAC-seq libraries

7.3.1 Collection and staging of embryos

Transgenic strains were expanded into population cages. After pre-clearing, embryos were collected on yeasted grape juice plates for 20 min and aged for additional 150 min (25°C incubator), followed by harvesting and dechoriation (50% bleach, 2 min) for 20 min (25°C temperature-controlled room). The embryos were flash-frozen in liquid nitrogen exactly 190 min after the beginning of collections (2:50 - 3:10 h after egg laying, AEL). Around 100 μ l of embryos were collected per transgenic strain and stored at -80°C until nuclei isolation. Additionally, each final collection was fixed with formaldehyde and examined under the differential interference contrast (DIC) microscope to confirm that it represented the cellular blastoderm (stage 5, mitotic cycle 14).

7.3.2 Affinity-purification of tagged nuclei

Staged embryos were homogenized on ice by 10 strokes of Dounce homogenizer (tight pestle) in Buffer A (15 mM Tris-HCl pH = 7.4, 60 mM KCl, 15 mM NaCl, 0.1 mM EGTA, 5 mM MgCl₂ and EDTA-free protease inhibitor (Roche: 04693132001)). Homogenate was filtered through 10- μ m nylon net filter (Merck Millipore) and centrifuged at 500 g for 7 min at 4°C. The nuclei pellet was resuspended in Buffer A + 0.5% NP-40 (Tergitol™, Sigma), incubated on ice for 3 min and centrifuged at 500 g for 7 min at 4°C. Purified nuclei were resuspended in 500 μ l of Buffer A. 50 μ l of Dynabeads protein G (ThermoFisher: 10009D) were adsorbed to 1 μ l of monoclonal anti-FLAG M2 antibody (Sigma- Aldrich: F1804) in 200 μ l PBS + 0.02% Tween-20 for 1 h at 4°C. Magnetic beads were washed once with PBS + 0.02% Tween-20, resuspended in 100 μ l of Buffer A and combined with the nuclei suspension. After 30-min incubation with rotation at 4°C, the beads with tagged nuclei were bound to a magnetic stand, washed three times with Buffer A and resuspended in 700 μ l of Buffer A.

Whole-embryo controls were generated from the same collections of staged embryos as isolations of tagged domains D1, D4, D5 and D7 (including duplicates). After homoge-

nization and nuclei purification, the nuclear pellet was resuspended directly in 700 μ l of Buffer A, excluding incubation with magnetic beads.

7.3.3 ATAC-seq on purified nuclei

In order to estimate the final yield of nuclei isolations, 25 μ l of nuclei suspension (+ beads) was combined with 10 μ l of 10 mg/ml proteinase K and 1.25 μ l of 10% SDS. Following incubation at 55°C for 20 min and vortexing at maximum speed for 5 min, concentration of genomic DNA was measured with a Qubit 2.0 fluorometer (Invitrogen). Nuclei suspension corresponding to 340 ng of genomic DNA was used for each ATAC-seq reaction.

Fragmentation and amplification of ATAC-seq libraries were performed according to the standard protocol [159]. An appropriate volume of nuclei suspension (+beads) was pelleted at 500 g for 7 min at 4°C. The pellet was resuspended in Nextera Tagment DNA Buffer with 6 μ l of Nextera Tn5 transposase (Illumina) in the final volume of 25 μ l. After incubation at 37°C for 30 min, the reaction was terminated by addition of Buffer PB (Qiagen). Magnetic beads were bound to the magnetic stand and the supernatant was purified with Qiagen MinElute PCR Purification Kit, followed by elution in 10 μ l Elution Buffer (10 mM Tris buffer, pH = 8). 5 μ l of eluted DNA was combined with custom Nextera PCR primers (from [88]) and NEBNext Q5 Hot Start HiFi PCR Master Mix (New England Biolabs) in the total volume of 50 μ l, followed by amplification for 12 cycles (PCR program according to [159]). Purification was performed with Agencourt AMPure XP beads (Beckman Coulter: A63881), using double size selection (left ratio: 2.0x, right ratio: 0.5x) according to the manufacturer's protocol. Libraries were eluted in 30 μ l of 10 mM Tris-HCl, pH = 8.0 and paired-end sequencing was performed on Illumina HiSeq 1500.

7.3.4 ATAC-seq on genomic DNA

In order to control for potential sequence bias of Tn5 transposase, a reference ATAC-seq library was prepared from purified genomic DNA (gDNA). gDNA was isolated from larval tissues of a wild-type *Drosophila melanogaster* strain, using a high-salt extraction method [160] with minor modifications: a step of vortexing was omitted to reduce shearing of gDNA and a final step of RNase digestion was included (with 50 ng/ μ l of Ambion RNaseA, 30-min incubation at 37°C). 12 ng of gDNA was combined with 2.5 μ l of Tn5 transposase in Nextera Tagment DNA Buffer (final volume: 50 μ l), followed by incubation at 37°C for 30 min. The reaction was directly terminated by purification with Qiagen MinElute PCR Purification Kit and elution in 11 μ l of nuclease-free water. 10 μ l of eluted DNA was combined with custom Nextera PCR primers (from [88]) and NEBNext Q5 Hot Start HiFi PCR Master Mix (New England Biolabs) in the total volume of 50 μ l, followed by amplification for 12 cycles (according to [159]). The library was subsequently purified and processed as other ATAC-seq libraries. ATAC-seq on genomic DNA was performed by Andrea Ennio Storti.

7.4 Data analysis

7.4.1 ATAC-seq data processing

After demultiplexing and removal of adaptor sequences, ATAC-seq reads were aligned to a reference genome (UCSC: dm3) using Bowtie2 v.2.2.9 [161], with the following settings: `bowtie2 -local -very-sensitive-local`. Mapped reads were filtered for mapping quality ($q > 10$) and proper pairing (read paired & read mapped in proper pair) using SAMtools v.1.3.1 [162], with the following settings: `samtools view -f 0x3 - q 10` (performed by Roberto Cortini).

As a general measure of sensitivity to Tn5 transposase fragmentation, ATAC-seq signal was defined as a number of transposase cuts mapping to each bp (or a total count of cuts mapping to a selected genomic interval). The cuts were defined as 5' ends of ATAC-seq reads, with additional shifting by +4 bp and -5 bp for reads mapping to the plus and minus strands, respectively [88]. Total pool of ATAC-seq fragments was considered, with no prior size selection. For visualization purposes, ATAC-seq signal was alternatively defined as coverage of short digestion products (≤ 100 bp, with *in silico* size selection).

7.4.2 Peak calling

ATAC-seq peaks were called with MACS2 v.2.1.1.2 [129], based on the enrichment of Tn5 transposase cuts (smoothed by +/- 100 bp extension). ATAC-seq library from genomic DNA was used as a common control sample. Peak calling was performed separately for each replicate, with the q-value cut-off = 0.01 and the following non-default parameters: `macs2 callpeak -keep-dup all -q 0.01 -nomodel -shift -100 -extsize 200 -f BAM -g dm -B` (performed by Roberto Cortini).

A set of 17 345 high-confidence accessible regions was identified by intersection of ATAC-seq peaks from all eight whole-embryo controls. Only peaks mapping to chromosomes X, 2L, 2R, 3L, 3R and 4 were considered.

7.4.3 Normalization of ATAC-seq signal

To normalize for differences in sequencing depth between the samples, their ATAC-seq signal was divided by scaling factors representing a median of signal ratios over a complete set of high-confidence ATAC-seq peaks (calculated with the function `estimateSizeFactorsForMatrix`, from DESeq2 v.1.20.0 [131]). Size factors were calculated separately for ATAC-seq signal expressed as the total number of cuts and coverage of ≤ 100 bp fragments, in order to control for differences in size distribution of the sequencing libraries.

7.4.4 Correlation of ATAC-seq profiles with published DNase-seq data

Published DNase-seq accessibility data from whole stage-5 embryos [67] was downloaded from the UCSC Genome Browser, dm3 release [163]: coordinates of DNase-seq peaks (intersection of peaks from two replicates, FDR = 5%) and wig tracks with normalized DNaseI tag density (both replicates). In order to control for the different size and genomic distribution of DNase-seq and high-confidence ATAC-seq peaks, first a union of these accessible regions was defined, followed by identification of peaks from each dataset that overlapped the union (minimum overlap = 1 bp). Correlation of accessibility signal was calculated over a complete set of high-confidence ATAC-seq peaks. DNaseI tag density and the total count of Tn5 transposase cuts was calculated separately for each replicate, followed by estimation of a Pearson correlation coefficient between mean ATAC-seq signal from eight whole-embryo controls and mean DNase-seq signal from both duplicates.

7.4.5 Principal component analysis

Principal component analysis (PCA) was performed with DESeq2 package v.1.20.0 [131] on the normalized and rlog-transformed total count of Tn5 transposase cuts, over a complete set of high-confidence ATAC-seq peaks. Function plotPCA was customized in order to include all input genomic intervals.

7.4.6 Identification of differential ATAC-seq peaks

High-confidence ATAC-seq peaks that showed significant differences in the total count of Tn5 transposase cuts between different tagged domains were identified with DESeq2 package v.1.20.0 [131], using default parameters and an adjusted p-value below 0.01. A total of 28 unique pair-wise comparisons was performed:

- between domains D1, D4, D5, D7 and their corresponding whole-embryo controls with the same genotype (multiple-factor design: domain/control + transgenic line A/line B)
- between domains D2, D3, D6 and all eight whole-embryo controls (single-factor design: domain/control)
- between individual domains (single-factor design: domain A/ domain B).

As D6 replicate 1 showed strong clustering with whole-embryo controls (Figure 4.2, Supplementary Material C), it was excluded from the analysis.

Constitutive ATAC-seq peaks were defined as a complementary subset of high-confidence ATAC-seq peaks that showed no significant difference in their accessibility signal in any of the pair-wise comparisons (adjusted p-value ≥ 0.01). Differential peaks were divided into quarters based on the maximum absolute value of \log_2 fold-change reported among all 28 pair-wise comparisons.

7.4.7 Genomic annotations of ATAC-seq peaks

Drosophila gene models were downloaded from FlyBase release 5.57 [126]. Promoters were defined as intervals 100 bp upstream and downstream of transcription start sites (TSS). Peaks were tested consecutively for their overlap with promoters, CDSs and introns, with a minimum overlap of 100 bp. Remaining peaks were tested for their overlap with UTR regions (5' and 3' UTRs), with a minimum overlap of 1 bp. The remaining peaks, without any gene annotations, were assigned as "intergenic" peaks.

7.4.8 Overlap of ATAC-seq peak with published ChIP data

ATAC-seq peaks were filtered to include only the "intronic" and "intergenic" annotations and tested for their overlap with ChIP peaks from published datasets (minimum overlap = 50 bp). Coordinates of ORIs (origins of replication) were downloaded from FlyBase release 5.57 [126] and represented a union of ChIP-seq peaks of ORC2 (origin recognition complex subunit 2) from three *Drosophila* cell lines [133]. Coordinates of ChIP-chip peaks of insulator proteins from 2-4 h embryos were downloaded from modENCODE [134], with the following IDs: BEAF-32 (5130), CP190 (5131), CTCF (5057) and Su(Hw) (5066).

Definitions of ChIP-chip peaks of TFs from the AP and DV patterning networks in stage 4-5 embryos [136, 135] were downloaded from the UCSC Genome Browser (dm3) as wig files representing signal intensity above FDR = 1% (single replicate, indicated as "best antibody"): Bicoid AB2, Caudal AB1, Giant AB2, Hunchback AB1, Knirps AB2, Kruppel AB2, Hucklebein AB1, Tailless AB1, Dichaete AB1, Fushi-tarazu AB3, Hairy AB2, Paired AB1, Runt AB1, Sloppy-paired1 AB1, Dorsal AB3, Mothers against dpp AB2, Snail AB2 and Twist AB2. In order to represent every bp of ChIP-chip peaks, intervals from wig files were extended by +/- 25 bp.

Definitions of ChIP-seq peaks of TFs from the AP patterning network in stage 4-5 embryos [137] were downloaded from the NCBI Gene Expression Omnibus repository with the following accession numbers (single replicate): Bicoid AB1 (GSM511083), Caudal AB1 (GSM511087), Giant AB2 (GSM511086), Hunchback AB1 (GSM511081), Knirps AB2 (GSM511088) and Kruppel AB2 (GSM511085). For greater specificity, ChIP-seq peaks were additionally filtered based on their overlap with peak summits from ChIP-chip experiments generated with the same antibody [136]. Definitions of ChIP-chip summits were downloaded from Berkeley *Drosophila* Transcription Network Project (primary peaks defined by the symmetric null test at FDR = 1%).

7.4.9 Overlap of ATAC-seq peaks with annotated *cis*-regulatory elements

Differential and constitutive peaks were tested for their overlap with three categories of annotated *cis*-regulatory elements (minimum overlap = 100 bp). Vienna Tiles [56] were filtered with the following parameters: `verification_status = correct`, `positive = 1` (reported

as active at any stage of embryogenesis). Temporal activity was assigned based on the earliest stage of embryogenesis at which a VT tile was reported as active.

REDfly enhancers were downloaded from the REDfly database v.5.4.2 [45] with the following criteria: CRM (data type), expression + only (restrictions), blastoderm embryo (ontology/expression term).

AP enhancers represent a carefully selected set of REDfly enhancers that drive patterned expression specifically along the AP axis (Supplementary Material D). For greater clarity, names of AP enhancers were standardized to represent their distance in kb to the closest TSS of a target gene. Overlapping AP enhancers that represent the same regulatory element are indicated as "broad" and "narrow".

7.4.10 Overlap of tagged domains with activity patterns of AP enhancers

Activity patterns of AP enhancers were measured using published *in situ* hybridization images of reporter constructs (references listed in Supplementary Material D) and defined as a position of the signal along the axis connecting the anteriormost and posteriormost tips of the embryo. Activity patterns of the enhancers, expressed as percent of the axis length (1-100%), are provided in Supplementary Material B. The proportion of a domain being overlapped by the activity pattern of an AP enhancer, i.e. the proportion of active nuclei, was additionally scaled based on the density of blastoderm nuclei along the AP axis. The number of nuclei within each percent of the axis was calculated using the embryo model from *Drosophila* Virtual Expression eXplorer (DVEX.org [16]).

7.4.11 Identification of background regions

Background regions were defined by subtraction of the union of ATAC-seq peaks that were called in all samples (tagged domains and whole-embryo controls) from the entire *Drosophila* genome. Only intervals 1-50 kb long were considered. After normalization, background ATAC-seq signal was defined as an average across all tagged domains. Intervals from the top and bottom deciles were removed before comparison with ATAC-seq signal of active and inactive enhancers.

Chapter 8

Appendix

8.1 Supplementary Material A: DNA sequences

- A list of DNA sequences of elements used for generation of expression constructs: enhancer drivers, DNA linker, basal promoters and nuclear tags.
- A list of PCR primers used for amplification of the UNC-84 coding sequence from the pMUH_unc84_tdTFlag plasmid as well as amplification of enhancer elements from genomic DNA.
- A list of PCR primers used for addition of Golden Gate Assembly cloning sequences to enhancer elements, DNA linker, basal promoters and nuclear tags.
- DNA sequence of the cloning cassette integrated in the pBDP plasmid.

1. Enhancer drivers

D1 domain: *hb_anterior_actv*

AACAATTGCAACAGGCATTAGTTTATATATCGCTCAGGTAGACGGATGCACGCGTCAAGGGATTAGATGGGCAGAGGTGACGGGA
AGTCAGGTACAGGTCGCGGATCGGTCGGGAATCGAGGATCACGGATCGCGGATTGAGGATTGCGCTCTTGATCCATTCTGGATTA
GAGCAGAAACAAAAATTTATGCGCACTTGGATTTGGATGATCCGGGAGCTTAGCGGATGGCCAGCTTAGCAGCGAGCTGCGAATT
TTCCACCGGTTTTCTATGGGGATTACGTTGGTTCAGGAGTTCGACAGCAGGAGTAGGCAGCTAGCGTGGGCAGTTTCGTAGTTAATA
ATAAAAAGTAAAAAGGATTGCGGGACTTAACTAAATTAACGGATCAGAAGTGCCTACACCTGCGGGAAAACTCTAAGGACCAACT
AAACTATATGCATAAATATGTGCAGTATAAATATTACACACCCATTTGAAAAACATTTTCTGACAACAATTTCCGCCAGACATT
TCACTTTGATTTGCGTAGTTTTCCTAATAAATTCGCAATAAAATGCTTGTGCTTATATTTTTCCATTTCCAATTTACACT
GAAAAATGTGCAGTTGCTGCATTTTGGCTAATTGTTGTGCTTTCAAGTAAATATTATTA AAAACGCAAAACGGGAAAAAGGG
GCATTTACGGAAATATTATATGGGAGGATGGTGTGCTGCTA

D2 domain: *eve_stripe2*

AATATAACCAATAATTTGAAGTAACTGGCAGGAGCGAGGTATCCTTCTGGTTACCCGGTACTGCATAACAATGGAACCCGAAC
CGTAACTGGGACAGATCGAAAAGCTGGCCTGGTTTCTCGCTGTGTGCCGTGTAATCCGTTTTGCCATCAGCGAGATTATTAGT
CAATTGCAGTTGCAGCGTTTCGCTTTCGTCCTCGTTTCACTTTCGAGTTAGACTTTATTCAGCATCTTGAAACAATCGTCGCAGT
TTGGTAAACAGCTGTGCCATACTTTCATTTAGACGGAAATCGAGGGACCTGGACTATAATCGCAACAACGAGACCGGGTTGCCAAG
TCAGGGCATTCCGCCGATCTAGCCATCGCCATCTTCGCGGGCGTTTGTGTTGTTGTTGCTGGGATTAGCCAAGGGCTTGAAGT
GGATCCCAATCCCGATCCCTAGCCGATCCCAATCCCAATCCCAATCCCTTGTCTTTTTCATTAGAAAGTCATAAAAACACATAA
TAATGATGTCAAGGGATTAGGGGCGCGCAGGTCCAGGCAACGCAATTAACGGACTAGCGAAGTGGGTTATTTTTTGCGCCGAC
TTAGCCCTGATCCGCGAGCTTAACCCGTTTTGAGCCGGGACAGAGTAGTTGTGGGTGGACCCACGA

D3 domain: *eve_1_ru*

GGCCTAATCACTTCCCTGAAATGCATAAATGTGCCGCGGCTTTTGATACGCTCCTGGCGGAGAGGGAGATGAGGAAAGGATGCAC
GGGAACCGCAGCCAAGTGGCAGTCGAGATTGGCAAAATCCGCCAGCGGACAATGCCAGAGAATGGGCAACAAGTAGCGGCGCAATT
AGCAATCCATATCATGCTTTTATGGCCGGCCAACCTTTCGCGCGCATCTCAGTTTCATCCGAAGCGGGACCAAGTCCAGGTTCAAG
TCGAGGTCAGTACCCCTGCTATCCCGTCAACCCCTTTAGGGCGATAATCCTTCTAAATGTTTGCATTAATTTTCAGGGCGTGGAC
GGATTAGGGCGTGTGGCTGGGCGGAACCCGCAGCAGAAACCGCCGAGGACACTGCACCGACTGACCTGCAGCCTACAGATCTCT
GATCTTCCATCTCTAATCCTTTCCGATTTGCAACTGACTTTCGCACTGGTCCGCCCTAATCCTTCCGCCGAGAAAGCGGGCAGA
GTCGCGAGGTTACTGGCCCGGGGTAATGGGATTATCTGCGATTACCCAGATGATCCGCAGAAAGTCAATCTGGTTTCAAGGGCTAA
TTGTCAGCGAAGTCAACTAAATCCAATCCTTTCGCGCCCTTCTGTATTATTGTTTGGTTTTCGTTTGGTTTGGAGATTCTGGC
AATTAAGTTGCCGTTTTGATGCGCGGGGGCGGTGCATCAAACTCTTTCGGCATACCTGTCTGCACAAATGCTGAATTCGCA
TCCCATGGATACCCAGATATTCAGATATCCAAGGCCGCAAG

D4 domain: *D_(+4)*

CGGGGAGATCCTTAAATACGCAGAAAACCTACCAGAGTTACAAAACAGAGAACAGCTTGACCCCTTTGCAAAACGATCTTGGAAGA
GTCCCGCCAGCAAGTAAATGTTTATGGCAGCCACAACCCGAACGATAAAACATAAACTGAACCTTCCGCGTAAGAAAGGCATTTGG
AAATTGCACGAAAATGTGAAAAATGAGAACAGCCGACGAGGTCGTCACAATGATTTCTACTAGCAAAATGATACCCGTTATTAG
CTAATCCGCTCTGGTCTTTGTGCGTAATCCGGCGGATTAGCGCGTGCATCTTTCGCTAAAAAAGTGCACAAAAACAGACGGATT
TTTGGGAAAACCTGAAGGAAATCAATAGTCTGGCTAATCCGGGTCCAAGTTACCACAGGTAGCCAAAACACTGCACGAAAAGAG
CAACAACAACCCAAACAGACAGCGCTAGAAGCATAAAAACTCCTGCGACATTTTCGACACTCCTTTCTGCGGACCGGACGCTAATC
CCCGTCAATAAGTCTGTAATAAAATCGAAGCAACCCGGGAGGCTGCAAAATATATTTGTATACAGAGCTGACCAAAATGGGACGAAA
GGCGTGGCTTTTGGGAGTAGGAGGCGTAGGCGAGATGGGTCTGTAACCGTTCTCGACAATCCTTAAAAACCAACAGGTTGAAAT
GGAATTTCAATTAATCTGTCATCAATTTTTTTGGATGAAATGTTTAACTATGGATCAGGAAGGCGTGAGCGGGCTTATGTTTTA
AGCTTGGCTTAGTCGGATGGATGGATTGGGCGTTTTAGCGAAAAGTTAATCACTTCCGTTTTGGGTTCCGCAAGGCAACATCCGATA
TTTGATCATAGATTGTTTTAATGGCTGATTATGGTTGTAACAGTTTTATGATCAGAATGTTATGATGCGAAAAATGGCATAAAA
TTTAACCTGTAAGATTCAAAGTAGTTACGAAATATAATGTAATAAAGTAGAATAATAAAAATACAATATGTTACTTTCAGCAGCGG
AATCTTGAGCCGCTATTAATCTAAAGGATGTCAAATAAAAATGCAATTTTAAAAATGCGGATGCGGTTCTTTTAAATGATATTTA
TTATTTACAAAATCTGTAATAGTTTCTAGATTTGAAAAATAGTTATCCAATCAAATACTATAAAAAACCTTTTCTTACAATAC
AAATGATTTGTTGAACCCCTTGAAGCCTTTACCGCTGTGCACACAAGTAATGTTCCAAAGAGCTAAGCCAGAAGATCCAATTT
TTTATATAATTAATCAATATCAGTCAATGCTGGGTAAATATTTGTTTGCAGAGCAAACGGAGAATGGAGAAACAAGGCATGGAA
AAAAGACCAATAACAATAACGCATAATAACAACAATAATCAATATTTGATATTTTAAACACATTTGACATACAACGTTGAAAC
GAAGGGAGCGGGCAGAGGGCATGAATAGACTCCCAAAGACTTGAAAACGGCAATCGAAGAGTCAGAGTTCCGGGGGAAAGACAA
TGCAAGTGGATTTAATAATGACCGCAATTAATCTTATTATTATACGAAAAGAGCTAAAAATCAAACGGAAAAACACCGCTTG
ATTTCTTTCGCCAGGTTTACGTTACACTCCGAATAGTGAGGTTTACAGGGCCAGGCCCCCGTTGATCCCGAAAAAAAAC
ACAAATGTGACCCCATCCAAGAGC

D5 domain: *Kr_CD1_ru*

AAATCGGGATCCTAAGTAACTATAATCCAGGCTTAATCACTGGATCAATAACTAAGTAGCATTTCCTGGGATGGAATATGAAG
 TTACCTGCATATGCCTACCGATCCGAAACTGCTTAACTTAATCGACATGCATGATCATAAAAAGCAATTTGCTACAATTTAT
 ATTTTTTTGCTTTTCTTCTTTAAGCATCTGGGATCTGGATCAGAAAAGAAAAGTGTAAACGCTACCTTCAGAAACGGATTAA
 ATTTTTTTCAGACAAATAATCCAGCCTTAAGCATGGTGATTAAGCTTGATCCCCACCAAGGGCGTAATATTGACGGATTTTCT
 TAAATCCCTCTGTTAATCTCCGGCTTAGAGCGGACGCGTTTTTTCGCGACTCCGCTGCATTGTTTTTTTTTTCAGTTTCTTCA
 ATTCGCAAGAAGGCAGGCCTATGGACCGAATGAGGATCATAATATGGAATTCCTAAATAACTAAGAAGGGCAGTCGGCATAGT
 ATTGATCTACCTGTAAGCGTGGGTCTATCTTTGCCCTCGCATTGAGACTCTCTAGTCACAGGTAGACCAGCCTTGAGTTCGT
 CGGCAATTAGAAGTCAAAATTTCTTAAAAACAACAAAAATGTCAAAGTAAAAACAATGCAAAAAATATGTGTAAGTAACTA
 AATCCGGCTTAGGATTTCTTGCGTCATAAACATGACTAGGGAGCCATTAATAATTTGATAATTTGCATGTCACCTGTGCACCGTAAA
 CGAAGCCAACACATCGGCTGAAACCCAGCGTCATTTATGCTATGCTATCTCACCTTTTGCATAATTTTTTAAATTTATTGACTT
 GTCTTTTTTACGAATGCAAGGAAGTACTTGTGTACTTCCGTTTTTCAAAGTCCGATTTTATTATCTACCTGCAATTTATAATGA
 CAACCAAGTTTTAACATTTCCAAAACGTGGTTTTTTGAGTTCAATAAGCTTCATTTTTTAAAGGCTTTTTTGAACCAATTAAC
 ATTTGCGGTACGTTTTTTGTAGAAAATTTTCCATAGCTGCATAAAGGTACAAGTTTCATTTATTTGATGGGAAAAATGAATTAGAGA
 CCAGCACGATCTTAAACAGAAACAAAAAGGCGTGGATATTTGTTATATGTACAAAAACAACCATGGCATATACGTATGTATTGTAT
 ATCCTAATTCGGCCTTCCCTTTATTTAATCACTTGCATAAATTTTCAACTATACAGATCAGCAATCAAAGTTCTTGATCCAATCG
 CTATAGTAACAGATGCGCAAGCTGATAATTTATTGAAAATTTTTTCTTAGGATTTCTCCATATTAAGCATTTCCTTGCTGTG
 ACCTCGATTCAAAAGAAAGCACTGATTTAATGCAGGTAATCGTCAAAAAATAA

D6 domain: *eve_stripe5*

GGGCGGGTGCATCAAATCCTTTCCGCATACCTGTCTGCACAAATGCTGAATTCGCATCCCATGGATACCCAGATATTCAGATA
 TCCCAAGGCCGCAAAAGTCAACAAGTCGGCAGCAAATTTCCCTTTGTCCGGCGATGTGTTTTTTTTTAGCCATAACTCGCTGCAT
 GTTTTGGCCCAAGTTTTTCTTCTGCCAAATTCGGGATGATGCGGGGATTATGCGCTGATTGCGTGAATTTGGACATCCTGC
 GAGGCCCGAGGAACCTTCTGCTAAATCCTTTCATCCGCTACAGAACCCCTTTGTGTCCCGTTCCGCGGGAGTCTTTGACGGGT
 CCTTCGACTATTCGCTTACAGCAGCTTGCCTAAAATTTTATAACCTACGAGCGGCTCTTCCGCGAATCCCTGGCATTATCCTT
 TTTACCTTTGCCAATCCGTTGGCTAAAAACGGCTTCGACTTCCGCTAACTGCTGGACAACAAGACAAAAACGGCGAAAGG
 ACGGCGATTTCCAGGTAGCATTCGGAATTCGCTCAAACCTAAAGGACCGGTTATATAACGGGTTTATATGGCCAGAATCTCTGCAT
 CTCCACGACCGCCAGAAGCTGCGTAAAACTGCAGGCTCTGTTTTGATTTCTGCAACTTCAGTTAATTGCCGGGATGGCCAGCAA
 TTGCCGGCAATTATAAACAGCGCAGATGTGACTCAGCTTCCATATCTAATCTATATCTCATGCCGAAAATCTAGGGTGGGGAG
 CGGAGGGCGGGGTGCGTGGGTGACTTGCCTGCCA

D7 domain: *gt_l(-3)*

CTGCTCGTGTGTTGCCCTCCTCCTTAAACATCCTTGATTCTACGTCCTTTATCCCTTTGGCTTCCGCTGCGGCGTGGGATGTGGGA
 TCCTCGACAAGTTGTTTCACTTTGTTAACTGTTTGTCTAACGGTCGCTCGACGCCAACGGAAGTGGGTAGCGGTACAAAGTT
 GGCTATCATCCATGGGAAATGTTATGCTAGATACTTACTGCAGCTAATTCGAGGACATAACGTAGTTTAGATTTGTACGCTGAA
 ATGTTATCAAATCCTATCAACAATACCTACATATATACTATCTTTCATTCATCTCTTTTTTACAACCTGCCATTTCAGGGGGATT
 GGGTGAATTTGCAATAAAAAAGAGGGATCGATCGATGGTAATCAGTAAAAAACCGCGAAGCTTATCCGCTAATCCGTGCGCTAATCC
 AACCGTAAAAATGGTGGATTAACATAAACTCGGACCGTCACGTCGCGCTCATTGAGTCGATCCCTTACCTGCGGTTACCGGGCC
 CGAACCCGAGCCCTTCAGATCCTGGACATACGTATAGCCAGCCCAATCGATCCCGAGCAGATGCCATAAAAAGTCCGCGCCGA
 AGTCGAAGCCAAAGAGTTGCCAAGTTCGGTGTAAAAACAGATTCAGCGAAAAAAAACGTAAAAAACAGCAGGCCATAAAACCA
 TATATTCATATACTCGGACTGCTGTTCTTCGCCATAATGAGCTCCGAACCAAGATGCAATATTTGTGCAATCCTTTGAGTTGT
 GACCTTGGGCGGTTTTTTTACGGAATCATCTACCTGATGGCCGAACGGGTTGCCACCCACGGTATGCGCTGTATTAGGCTGTTA
 TGCTGTTGCCATAAAAAACATCCAAAGTTTTATTGCATAATGGACGCTTCGTGCGAAAAGGCTCATTCCGAGTCCGCGGGCCCTA
 AAACCGCTGCAGTTTTTTAAGGAGCCATAAACTGATATTTCCGACACCGCGGTCCTTAAAAAAAAGGGTTCTCTCCTCCGAC
 CATTGTGGCTGCGATTTTTTATATTGGTCCGCTGCTAGGTGACAAAAGAAAATCAGCGACTAATTGACTGAAAGGGGTTTTGCCTTA
 TATGACCCGTCCACCCGTCCAAAAAGGGTGGAGGTGGTTTTGATTGGATTCGTGCTTGGTTTTCCGGGAATTGCCGTTAACT
 CCTTTCGCGGCCCTTTGTCT

2. Linker

TCCAGTGTGGTGAATTTCTGCAGATATCCAGCACAGTGGCGGCGCTCGAGTCTAGAGGGCCCTTCGAA

3. Basal promoters**DSCP**

GAGCTCGCCCGGGGATCGAGCGCAGCGGTATAAAAGGGCGCGGGTGGCTGAGAGCATCAGTTGTGAATGAATGTTTCGAGCCGAG
 CAGACGTGCCGCTGCCTTCGTTAATATCCTTTGAATAAGCCAACTTTGAATCAACAAGACGCATACCAAAC

Hs43

AAGAGCGCCGGAGTATAAATAGAGGCGCTTCGTCTACGGAGCGACAATTCATTCAAACAAGCAAAGTGAACACGTCGCTAAGCG
 AAAGCTAAGCAAATAACAAGCGCAGCTGAACAAGCTAAACAATCTGCAG

P-element promoter

AGCCGAAGCTTACCGAAGTATACACTTAAATTCAGTGCACGTTTGCTTGTGAGAGGAAAGGTTGTGCGGACGAATTTTTTTT
TGAAAACATTAACCTTACGTGG

4. Nuclear tags

Legend: **coding sequence of the nuclear-envelope protein** HL4 peptide linker 3×FLAG STOP codon SV40
polyadenylation signal

Unc84-3xFLAG

ATGGCTCCCGCAACGGAAGCCGACAACAACCTTCGACACCCATGAATGGAAATCGGAATTCGCATCCACACGCTCTGGACGCAATT
CTCCAAACATTTTTGCAAAAGTTCCCGGAAGCTTTCCTGACTCCACCAGTTCGAAACGCCAGATCGCCACGCTTACCGAAGA
AGAGCTGGATGCTTTGACAGGCGACTTACCATACGCAACCAACTACACATACGCATACAGCAAAATCTACGATCCATCCTTGCCG
GACCACTGGGAAGTGCCAAACCTTGGTGGTACTACTTCAGGATCACTCTCTGAGCAGGAGCAGTGGTCCAGCGCCAGTCTCAGCA
GACAGCTTCTCTATATCCTCCGTTTCCCGTCTACCTTGTCTTCAGTTCATCACCTACATTTTGAAGCTTTCTACCACGTCAT
CAAGATCACTAGCTTACCATCTGGGACTACCTGTTGATTTGGTGAACCTCGGAAAACCTCGTTACTACGCCATCCAAAGATCAT
CGTCGCCGTACAGCTCTCATTCGCAACCGGCAAGAGCCATTCTCCACTAAGGCTGCTCGTCTATTCGTCGATTCCTTTGAGATCC
TTGTCTACGTCGTCTTACTCCTTACAGAAATGCTCAAGAAGTAACAATGGCGTGGAAACAGTACAGTACCGTTCGATCAAGGA
TCAATTGGAAAAATGAGAGAGCTAGCAGAAATGACGACAAGATCCCAAACATTGGAAAGAAGCCGCAAGTTTGTATGGATTATCGAAA
TCACCAGCACGCCGAGCAGCTCCAGCCTTTGTGAAGACTAGTACAATTACCAGAATCACTGCCAAGGTGCTCTCGAGCTCTCCAT
TCGGAGAAGGAACGTCGAAAATAAACCCTGACTGTTGTGACTACTAGAACAGTGAAGCAACGCTCAGTTACCCCAAGATTCCG
CCAAACCGTGCCTACTGTAAGCTATAACTCGAGCACTCGATACTCCGGAACCGAAATCGACACACCCTCTCCACATATGGA
CTTCGAAGCCGAGGACTGAGTCATCTGAATCTCCTGAACCACTTTTGACATGGTTCATGCTGCTGCAACTTCCACGCTTTGT
TCCCACAAGAACTTACAATTATCAATACGAAGAAGCGACAGGAAATAAGATTAAAATGCATTCATTTGGCTAGGTTACTTGTAT
ATTGTTCCCGTTCTTTGCGGCACGACATGTATGGTATACGTTCTACGATTTATGGAAAGAGTGCCATACATGAAGTACCAATAT
CAGCAAGCGCAATGGAGACTATTTCATGTTCAGAGATATCAACGAACCGGCACCAAGTTCATCAGATGTTTCATGATGCTGTTGGTG
TTTCTTGAGAAATTCGAATTCGCCGATTTCTTGGAGCTCATTCTGAGCAACAATCGTTGAAGCGCATCAAGTGGTATTTGCAATGTT
CAAAGGAGGAATTTGTCGAGACAGTTTCTATTTTGGAGGACTATTTGCTGGTCTACCAGATAAGAAATCATCAAAGTTTCTCGTGG
TGTCAAAATCTCGGTTACTTCTGGCTTCTCTTCCGCACTTTCTCCTTGGATTCTTGACATCGACAACACAGCAATAAGAG
TTAAAGAAATTAACCAAGATAAGAAATGCATTAAGAAGTCGGAAGGATCCCTCCAGCTGTGCCAATCTGGATTTTCAAGTGCAAA
TCAGGTTAAACATTACACATGGATGGTGAAGGAATTTGTTGTAGATATTGCATTTGACACGTACAACATATGAAAGTTCGACGATT
GGTAGACTTGGCACTACTCCAGTTATGCTTGGGACTGATTGCAAGCGGATGCGCGCTGTTGGAAATGGCTTAAATCTGTGC
TCTCATCGAGTTTTCGATTCATCGATTTTGTGCTGAAAGCTATTTTACTATGGCTCAGATGGGTTCTTGTGTCAGCAACAAGT
TATCGGAACCTTTTCAACGGTTGCTACGAAACCTTTGACAAACGGATGCACAGCAATTTGTTGGCCATACAAAGAGCTTTCATCTAC
AATGCTTCAAATGCTGTTTACAACTTTTTCTCAACTATCTTTGCCGGCTCTTAACTTTTCTACTTCTTCCAAAACCTCCATTC
TTTCTCTTCAAGTCATTTGGCAGCGGAATCACTAACATTTTATAACTTCATTTATGCACCAATCGCTGGAGTGTTCAACTT
TGCTGGTGATAACTACATGATTTTCTTCAATGAGGTAGCGGAGCTTTTGGAAAAGTGTACAACCTCCGTTGGTTTCCGTGCTCAA
ACTGTAATTAACGATGATTTCTTCTCATTGCTTACCATTTCAGTTTGTGCACCTCGTTCGATTCGCAATCAGCCAAATATGCTC
CAGAAGATGTTGTTCAAGTATTCCAATTCACAAGCTATTACCCCAACTCCGGATGTGGAGCGTATTGTTGAAGAGCCACTGAG
AAAAGTCCAGATGTGGAGGACGAAGAACTAGTGATAATCCCGCCCCCGCACCTAACCTATCCAGTCCAGCGCCAACCTCCG
GCCCCAGTAATATCCATCAGACTAACGTTGTTGAGACTGTTGACAAAGATGCCATCATTAAGGAGGTAACGGAGAAGCTTCCGG
CCGAGTTGTCCGCCAATTCAGCAAGAGCTTAGCGCAAAGTTGAGCAAAACTACAACAATTAATGAGCAACTGAAAATGGA
AAACACCAACATTCATATGATAAGAATCATTTGGAAGCTATCATCCGTCAAAATGATCTACGAGTATGACACGGATAAAACCTGGG
AAAGTTGACTATGCCCTGGAGAGCTCAGGTGGAGCTGTTGTGTCAACAAGATGCTCGGAGACGTACAAAAGCTACACAAGGCTGG
AAAAGTTTGGGATATCCCAATCTACTATTTCCATTAATCTCCTCAAGAGTTGCTATTACAGAGAAATCCAAATCCCTGTTTCTGG
GGAATGCTGGTGTCTCAAAGAATCCCGTGGCTACATTTGCTGTGAGCTGTCTCATTTTATGATGTTTCTAGCATCAGCTATGAG
CACATTTGATCAGAAGTTGCTCCAGAAGGAAACCGTTCGAGTGTCCAAAGGGAGTCCCGTTTGGGCTTACAAGCAGATTGACG
ACCTGAACTCGAGAGTTTGGATTGGCGACTACACTTATGATCTTGATGGCCCGCCACTTCAATTTCTTCCAAAGCACAACCC
CGATTTTCTGTCAAGTTTGTGGAGCTCAGGTGACAAGCAATACGGAGCTCCGTTACATGTCTTACCAGCTTCTGTTTTCAT
GGAAAAGTTGTTCAAGTTCTGGCCGAGGCCGCCGCAAGGAGGCCGCCGCAAGGAGGCCGCCGCAAGG
CCGCCGCCGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGTGA GTTTAAA
CCCGCTGATCAGCTCAGCTGTGCTTCTAAGATC CAGACATGATAAGATACATTTGATGAGTTTGGACAAAACACAACCTAGAATG
CAGTGAATAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTAAGCTGCAATAAACAAGTTAACA
ACAACAATTCATTCATTTTATGTTTCAGGTTTCAGGGGGAGGTGTGGGAGGTTTTTTT

RanGAP-3xFLAG

ATGTCCACCTTTAACTTCGCCAGCATGGCCGCCAACTGGGCCAGGAGCAGGGAATATCATTCGAGAACAAGGTGCTTTCCTGGA
 ACACAGCTGCCGATGTCCAGGATGTGGTGGATGCCCTTAACAAACAGACCACCGTGCATATCTGAATCTGGACGGGAACACACT
 GGGCGTTGAGGCCGCCAAGGCGATTGGTGGGGTCTGAAGCGTCATCCAGAGTTTCGGAAGGCGCTGTGGAAGAATGTTTACT
 GGTGCTCTCATATCGGAGATTCCGGAGGCACTCAAGCACCTGGGAGCCGCGCTAATTGTCGCGGGCGCCAACTGACAGTCCCTGG
 ATCTCAGCGACAATGCCTTAGGACCGAATGGCATGCGAGGCTTAGAGGAGTTACTGCGATCCCCGGTCTGCTACTCGCTGCAGGA
 GCTGCTGCTGTGCAATTGTGGCCTTGGTCCGAGGGCGGTATGCTGTCCCGGGCTCTGATCGATCTGCATGCCAATGCCAAC
 AAGGCGGGCTTCCCGCTCCAGCTGCGTGTGTTTCATAGGTTCCGCGCAATCGTCTCGAGGATGCCGGTGTCTACGGAAATGGCAACCG
 CATTCCAAACCCCAAGACCTTCGAGGAGATTGTTCTGGAGCAAACTCCATTACATCGAAGGCGTCGAGGCCCTTGCCGAATC
 CTTCAGCATAATCCTCATCTACGAGTGTAAACATGAACGACAATACTCTAAAGTCCGAGGGAGCTGAAAAATAGCTGAGGCT
 CTCCCTTCTTGCCACTGCTGCGTGAAATGAGCTTGGAGACTGCCGTATCAAACTAATGGCGCTACCCTTCGGTGGGCTC
 TGGAGAGAGGAAACGAACGACTGGAAGTTATCGACTTAGGTTTAAACGAAATCAACAGCGACGGCGGCTTGGTGTGGTGAATGC
 TATGGGAAACAAGCCCAAGCTACGCATCTTGAATCTAGATGGCAATAGCTTTGGAGAAGAAGGCAGCGAGAAGATAATCAGCGAG
 ATGAGTAAGTTGCCAATGCTGCCCACTGCAACCGTTTTCAGCACCAGGAAGAGGAGGATTTGGAAGATGAATACAGGCTGACA
 AGCAGGACGCAGATTACGAAGAGGAAGAGGAAGTACACGAGCACGCCAACGATACTACCAGAAGAAGCAGATGAGGATAGCGAGGG
 CGACGAGGACGACGAGGAAGAGCGAGGAGACGAGGAGTACAGCAACCTCGCGGAGGAGACTGCCATGTCACTACGAATGCCTAC
 ACGACCAAGCTTTTAAACGACACAACCAACTCGATGGCCAGCGAACTTTTGGCGTCCGGAACAAGCAGATCAGCCAAAAATGCA
 CTCCAGAGAAGTTCTGTTTGGAGCCAGAAACCTGCTCCAGGAAGATTTCGATTCGCTAGATATGGATAACAACCTTGAGGCTTT
 GCAGTCGATTGTCAACCAATTCACCGGCGACAACCAATTTGCTACTGCTCGTCTCACCACCTTGAAGTGCAGCGCATTTGTGCGAA
 TCCTCGAAAGCTGCGTTGGATCTGGCCGCTCCTTTGTACCAGCCACCTTTGACTATGCCATCAAGACAAGCAGGAGACAGTG
 TACTCAACTATGTACTGATGCAGCTCCGTTTGTGGCCCTGCAAGGAGGTATTCATTCGGACTACGATGTCAAGAACTGTGATT
 TGCTCTTCGCGAGGCTCTCAAGCAACCACTTTGCCAAGCACAACATTAAGAATTCCCTTAAGACTTTCCTGGAGGGTGGCGGAG
 TCGCTGGCCGAGGCCCGCCCAAGGAGGCCGCGCCAGGAGGCCGCCCAAGGAGGCCCGCCCAAGGCCCGCCGCTGACTAC
 AAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGTGA GTTTAAACCCGCTGATCAGCCT
 CGACTGTGCCTTCTAAGATC CAGACATGATAAGATACATTGATGAGTTGGACAACCAACAACCTAGAATGCAGTGAAAAAATGC
 TTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTCATT
 ATTTTATGTTTCAGGTTTCAGGGGAGGTGTGGGAGGTTTTTTT

5. PCR primers**UNC-84 coding sequence**

Forward primer: ATGGCTCCCGCAACGGAAG

Reverse primer: AACTTGAACAACCTTTCCATGAACAC

***hb_{anterior}* *actv* enhancer**

Forward primer: AACAAATTGCAACAGGCATTAGT

Reverse primer: TAGCACAGCACCATCCTCCC

***eve_{stripe2}* enhancer**

Forward primer: AATATAACCCAATAATTTGAAGTAACT

Reverse primer: TCGTGGGGTCCACCCACAACCTA

***eve₁* *ru* enhancer**

Forward primer: GGCTAATCACTTCCCTGAA

Reverse primer: CTTTGGCGCCTTGGGATATC

***D₍₊₄₎* enhancer**

Forward primer: CGGGGAGATCCTTAAATACGCAGAA

Reverse primer: GCTCTTTGGATGGGGTTCACATT

***Kr_{CD1}* *ru* enhancer**

Forward primer: AAATCGGGATCCTAAGTTAACT

Reverse primer: TTATTTTTTTGACGATTACCTGCA

***eve_stripe5* enhancer**

Forward primer: GGGCGGGTGCATCAAATCCTTTCCGG

Reverse primer: TGGCAGGCAAGTCACCCACGCACC

***gt(-3)* enhancer**

Forward primer: CTGCTCGTGTGTTGCCCTCCTCCTT

Reverse primer: GACAAAGGGCCGCGAAAGGAGTTA

6. PCR primers with overhangs for Golden Gate Assembly

Legend: PCR primer overhang BsaI recognition site 4-bp ligation site

***hb_anterior_actv* enhancer**

Forward primer: GACTAGGGTCTCAGATGAACAATTGCAACAGGCATTAGT

Reverse primer: GTAGTCGGTCTCTACCTTAGCACAGCACCATCCTCCC

***eve_stripe2* enhancer**

Forward primer: GACTAGGGTCTCAGATGAATATAACCCAATAATTGAAGTAACT

Reverse primer: GTAGTCGGTCTCTACCTTCGTGGGGTCCACCCACAATA

***eve_1_ru* enhancer**

Forward primer: GACTAGGGTCTCAGATGGCCTAATCACTTCCCTGAA

Reverse primer: GTAGTCGGTCTCTACCTCTTTGCGGCCTTGGGATATC

***D(+4)* enhancer**

Forward primer: GACTAGGGTCTCAGATGCGGGGAGATCCTTAAATACGCAGAA

Reverse primer: GTAGTCGGTCTCTACCTGCTCTTTGGATGGGGTACATT

***Kr_CD1_ru* enhancer**

Forward primer: GACTAGGGTCTCAGATGAAATCGGGATCCTAAGTAACT

Reverse primer: GTAGTCGGTCTCTACCTTTATTTTTTTGACGATTACCTGCA

***eve_stripe5* enhancer**

Forward primer: GACTAGGGTCTCAGATGGGCGGGTGCATCAAATCCTTTCCGG

Reverse primer: GTAGTCGGTCTCTACCTTGGCAGGCAAGTCACCCACGCACC

***gt(-3)* enhancer**

Forward primer: GACTAGGGTCTCAGATGCTGCTCGTGTGTTGCCCTCCTCCTT

Reverse primer: GTAGTCGGTCTCTACCTGACAAAGGGCCGCGAAAGGAGTTA

Linker

Forward primer: GACTAGGGTCTCAGGTTCCAGTGTGGTGAATTCTG

Reverse primer: GTAGTCGGTCTCTTACC TTCGAAGGGCCCTCTAGACT

DSCP promoter

Forward primer: GACTAGGGTCTCAGGTA GAGCTCGCCCGGGGATCGAGCGC

Reverse primer: GTAGTCGGTCTCTTCA GTTTGGTATGCGTCTGTGATTC

Hs43 promoter

Forward primer: GACTAGGGTCTCAGGTA AAGAGCGCCGGAGTATAAAT

Reverse primer: GTAGTCGGTCTCTTCA CTGCAGATGTTTAGCTTGTTC

P-element promoter

Forward primer: GACTAGGGTCTCAGGTAAGCCGAAGCTTACCGAAGTA

Reverse primer: GTAGTCGGTCTCTTCAACCACGTAAGGGTTAATGTTTTCA

Unc84-3xFLAG

Forward primer: GACTAGGGTCTCATGAAATGGCTCCCGCAACGGAAG

Reverse primer: GTAGTCGGTCTCTCCAGAACTTGAACAACCTTTCCATGAACAC

RanGAP-3xFLAG

Forward primer: GACTAGGGTCTCATGAAATGTCCACCTTTAACTTCGCC

Reverse primer: GTAGTCGGTCTCTCCAGAAAAACCTCCACACCTCC

7. Modification of the pBDP plasmid

Sequence of the cassette introduced in the MCS of the pBDP plasmid.

Legend: FseI KpnI BsaI recognition site 4-bp ligation site

TAAGCA GGCCGGCC GATGA GAGACC TCCAGTGTGGTGGAAATCTGCAGATATCCAGCACAGTGGCGGCCGCTCGAGTCTAGAGG
 GCCCTTCGAA GGTCTCT CTGG GGTACC TAAGCA

8.2 Supplementary Material B: Enhancer drivers

Comparison of genomic coordinates of the seven enhancers used as drivers of the nuclear tag. The table lists:

- A name of the corresponding tagged domain.
- A name of the enhancer from Segal et al. (2008) [64].
- Genomic coordinates of the enhancer as defined in Segal et al. (2008) [64].
- Corresponding name of the enhancer from the REDfly database v.5.4.2 [45].
- A reference to the original study that characterized and validated the enhancer.
- Genomic coordinates of the enhancer as reported in the original study.
- A summary of changes introduced by Segal et al. (2008) [64]. Personal communication with Ulrich Unnerstall.

References to the original studies: Schröder et al., 1988 [164]; Small et al., 1992 [27]; Fujioka et al., 1999 [46]; Schroeder et al., 2004 [50]; Hoch et al., 1990 [47].

Tagged domain	Enhancer name Segal et al. (2008)	Genomic coordinates Segal et al. (2008)	Enhancer name from REDfly	Reference of the original study	Genomic coordinates original study	Changes in reference to the original definition
D1	<i>hb_anterior_actv</i>	3R:4520323..4521043	hb_1.2	Schröder et al., 1988	3R:4519886..4521119	Removal of 3'UTR and first intron
D2	<i>eve_stripe2</i>	2R:5865217..5865879	eve_stripe_2	Small et al., 1992	2R:5865267..5865750	Extension of the enhancer
D3	<i>eve_1_ru</i>	2R:5873440..5874247	eve_stripe1-construct1 eve_stripe1-construct2	Fujioka et al., 1999	2R:5873439..5874233	Merging of two elements
D4	<i>D_(+4)</i>	3L:14166136..14167860	D_(+5)	Schroeder et al., 2004	3L:14166136..14167860	Original
D5	<i>Kr_CDI_ru</i>	2R:21110136..21111549	Kr_CDI	Hoch et al., 1990	2R:21110142..21111300	Extension of the enhancer
D6	<i>eve_stripe5</i>	2R:5874147..5874946	eve_stripe5	Fujioka et al., 1999	2R:5874221..5874834	Extension of the enhancer
D7	<i>gt_(-3)</i>	X:2324294..2325502	gt_-3_construct	Schroeder et al., 2004	X:2324294..2325502	Original

8.3 Supplementary Material C: Evaluation of D6 Replicate 1

D6 Rep1 clusters with whole-embryo controls in the global analysis of accessibility variation across the entire set of high-confidence ATAC-seq peaks (Figure 4.2).

D6 Rep1 is also characterized by strong similarity to the control samples when accessibility across the selected set of 88 AP enhancers is considered (Figure 8.1). D6 Rep1 is included in the cluster of controls, while the other tagged domains (including D6 Rep 2) display limited correlation with the whole-embryo samples.

The two replicates of D6 produce different results in the DESeq2 analysis. Figure 8.2 presents an example of differential analysis between D4 and D6 domains. Inclusion of a single D6 Rep2 gives rise to an expected distribution of signal fold-changes and estimated p-values of the ATAC-seq peaks, taking into consideration that a majority of regions are constitutive in this comparison [131]. \log_2 fold-changes are symmetrically distributed around 0 and the majority of regions is characterized by the adjusted p-value of 1 (Figure 8.2A). In contrast, differential analysis with D6 Rep2 produces a non-symmetric distribution of \log_2 fold-changes, while adjusted p-values show a fairly broad spread (Figure 8.2B). This suggests a lack of strong distinction between constitutive and differential peaks in this comparison, which could be caused by technical variability of D6 or contamination with untagged nuclei.

For these reasons, D6 Rep1 was excluded from the subsequent analysis, i.e. identification of differential peaks and examination of accessibility signal over individual enhancers.

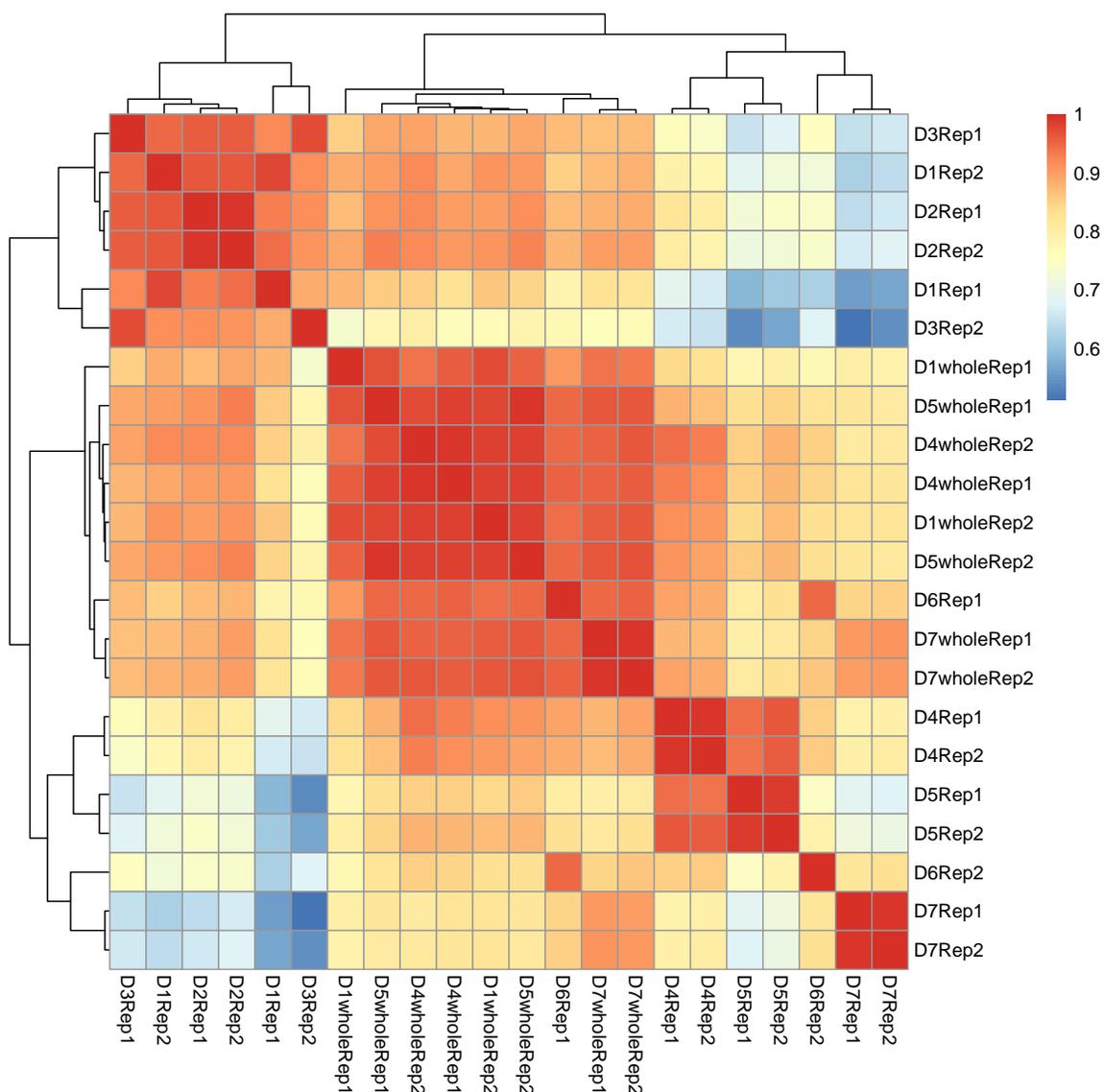


Figure 8.1: Correlation matrix of all ATAC-seq samples across 88 AP enhancers. Correlations are based on normalized \log_2 -transformed ATAC-seq signal, defined as the total number of Tn5 cuts. Heatmap: Pearson correlation coefficient. Note clustering of whole-embryo samples and separation of anterior and posterior tagged domains. Unlike D6 Rep2, D6 Rep1 is included in the cluster of control samples. A majority of tagged domains cluster with their biological replicates, apart from D1 and D3 domains.

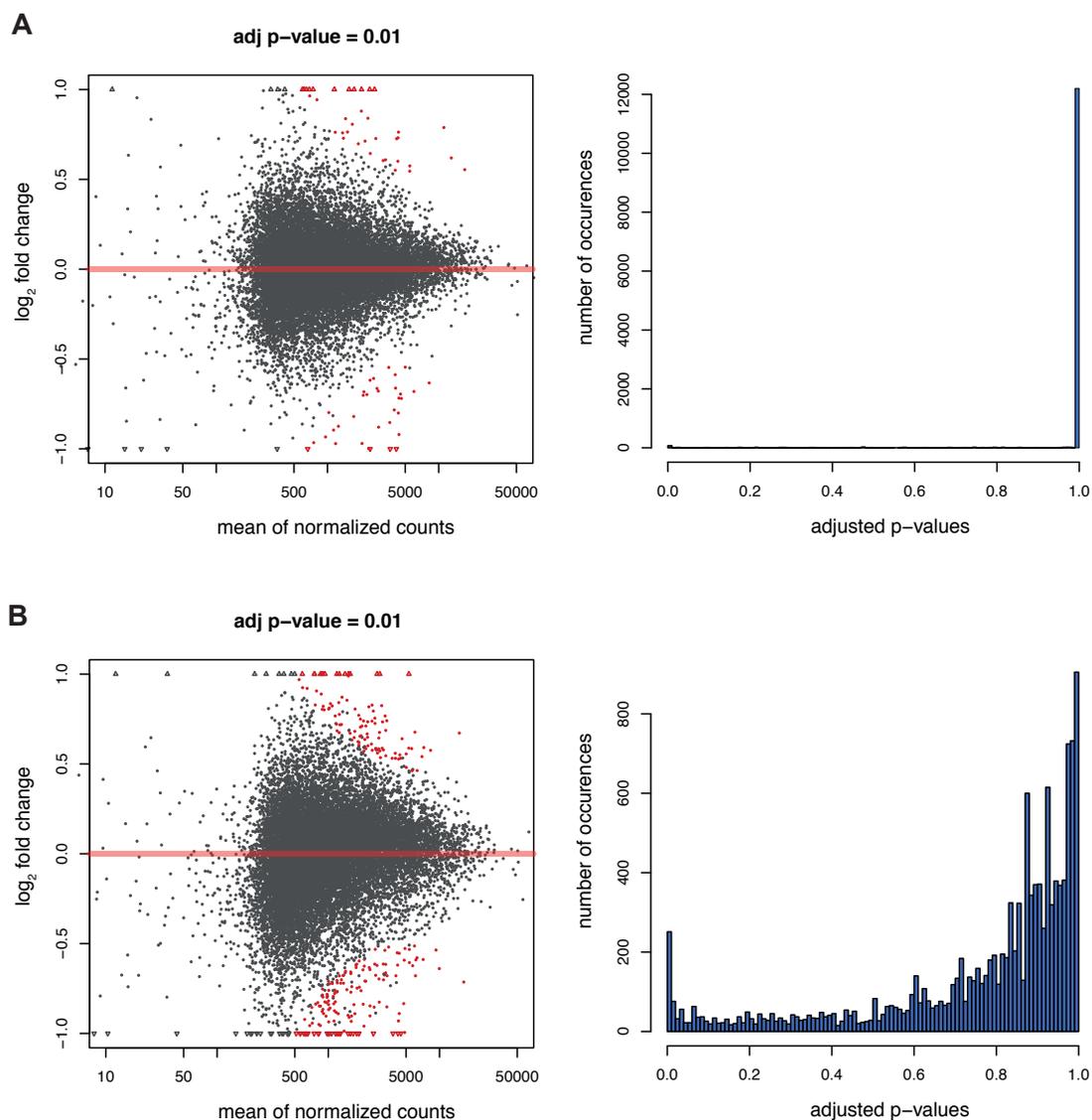


Figure 8.2: DESeq2 analysis with replicates of D6. Differential analysis between D4 and D6 domains. A) Analysis performed with two replicates of D4 and D6 Rep 2. B) Analysis performed with two replicates of D4 and D6 Rep1. Left: distribution of ATAC-seq signal \log_2 fold-changes vs. mean accessibility signal. ATAC-seq peaks that show no significant quantitative changes are marked in grey, while differential peaks are marked in red. Triangles mark regions beyond the range of the y-axis. Right: distribution of adjusted p-values across all considered ATAC-seq peaks.

8.4 Supplementary Material D: AP enhancers

- **Name:** name of the AP enhancer used in this study
- **chr:** chromosome name
- **start:** start position (reference genome: UCSC dm3)
- **end:** start position (reference genome: UCSC dm3)
- **REDfly name:** REDfly identifier of the enhancer (REDfly v. 5.4.3 [45])
- **Source:** reference to the original study that identified and characterized the enhancer. Source of the RNA *in situ* hybridization image used for estimation of the activity pattern
- **Target gene:** name of the enhancer's target gene
- **AP position:** position of the enhancer's activity domain along the AP axis

References to the original studies: Berman et al. 2004 [48], Chen et al. 2012 [53], Fisher et al. 2012 [138], Fujioka et al. 1999 [46], Gao and Finkelstein 1998 [165], Hader et al. 2000 [166], Hoch et al. 1990 [47], Howard & Struhl 1990 [167], Jacob et al. 1991 [168], Kazemian et al. 2010 [63], Klingler et al. 1996 [169], Kvon et al. 2014 [56], La Rosee et al. 1997 [170], Liaw and Lengyel 1993 [171], Margolis et al. 1995 [172], Ochoa-Espinosa et al. 2005 [52], Pankratz et al. 1992 [173], Perry et al. 2011 [41], Riddihough & Ish-Horowicz 1991 [174], Rudolph et al. 1997 [175], Schröder et al. 1988 [164], Schroeder et al. 2004 [50], Schroeder et al. 2011 [51], Small et al. 1992 [27], Small et al. 1996 [34], Wimmer et al. 1995 [176].

Name	chr	start	end	REDfly name	Source	Target gene	AP position
Abd-B_(+11)	3R	12745522	12747731	VT42848	Kvon et al. 2014	Abd-B	77-94%
Abd-B_(+30)	3R	12726430	12728642	VT42837	Kvon et al. 2014	Abd-B	71-91%
Antp_(-16)	3R	2774228	2775839	Antp_4	Kazemian et al. 2010	Antp	41-57%
Blimp-1_(+7)	3L	5616300	5617332	Cluster-8331	Fisher et al. 2012	Blimp-1	56-63%
Blimp-1_(+9)	3L	5614588	5615641	Cluster-8307	Fisher et al. 2012	Blimp-1	24-91%
bowl_(+3)	2L	3774540	3777027	Cluster-8198	Fisher et al. 2012	bowl	20-56%
btd_(+3)	X	9584455	9585527	Btd_Ss-Bg	Wimmer et al. 1995	btd	19-39%
cad_(+2)	2L	20768045	20769545	ChIP_AHD2	Fisher et al. 2012	cad	35-100%
D_(+4)	3L	14166136	14167860	D_(+5)	Schroeder et al. 2004	D	39-86%
Dfd_(-13)	3R	2630027	2631527	Scr_ChIP-miRNA9	Fisher et al. 2012	Dfd	30-41%
ems_(-9)	3R	9736403	9737444	Unspecified HC_18	Chen et al. 2012	ems	15-28%
ems_(+22)	3R	9705258	9706352	Unspecified HC_25	Chen et al. 2012	ems	17-98%
eve_(-5)_broad	2R	5870741	5872141	eve_construct20	Fujioka et al. 1999	eve	54-80%
eve_(-5)_narrow	2R	5871404	5871999	eve_stripe_4_6	Fujioka et al. 1999	eve	54-80%
eve_(-7)	2R	5873439	5874233	eve_stripe1- construct1/2	Fujioka et al. 1999	eve	26-36%
eve_(-8)_broad	2R	5874141	5875341	eve_construct30	Fujioka et al. 1999	eve	62-71%
eve_(-8)_narrow	2R	5874221	5874834	eve_stripe5	Fujioka et al. 1999	eve	62-71%
eve_(+1)	2R	5865267	5865750	eve_stripe_2	Small et al. 1992	eve	36-44%
eve_(+4)	2R	5863006	5863516	eve_stripe_3+7	Small et al. 1996	eve	45-86%
fkh_(-2)	3R	24411719	24413426	fkh_-2_construct	Schroeder et al. 2004	fkh	89-100%
fkh_(-4)	3R	24413511	24415247	fkh_distal_-4	Perry et al. 2011	fkh	87-100%
ftz_(-3)	3R	2692616	2694360	ftz_(+3)	Schroeder et al.	fkh	30-73%

					2011		
ftz ₍₋₄₎	3R	2693265	2694696	ftz _{CE8024}	Berman et al. 2004	fkx	33-76%
ftz ₍₊₇₎	3R	2681761	2683378	ftz ₍₋₇₎	Schroeder et al. 2011	fkx	53-88%
gt ₍₋₁₎	X	2323048	2324286	gt _{-1_construct}	Schroeder et al. 2004	gt	16-80%
gt ₍₋₁₀₎	X	2331789	2333533	gt _{-10_construct}	Schroeder et al. 2004	gt	13-41%
gt _{(-2)_broad}	X	2324294	2325502	gt _{-3_construct}	Schroeder et al. 2004	gt	64-82%
gt _{(-2)_narrow}	X	2324617	2325714	gt _{CE8001}	Berman et al. 2004	gt	63-83%
gt _{(-5)_broad}	X	2327322	2329503	gt _{-6_construct}	Schroeder et al. 2004	gt	1-14%
gt _{(-5)_narrow}	X	2327975	2328778	gt _{gt1}	Ochoa-Espinosa et al. 2005	gt	1-16%
h ₍₋₁₂₎	3L	8680591	8681060	h _{stripe0}	Ochoa-Espinosa et al. 2005	h	1-16%
h ₍₊₁₀₎	3L	8658178	8659109	h _{h7_element}	La Rosee et al. 1997	h	74-86%
h _{(+11)_broad}	3L	8656630	8658374	h _{stripe_3+4_ET22}	Howard & Struhl 1990	h	46-59%
h _{(+11)_narrow}	3L	8657463	8657938	h _{stripe3_ET38}	Howard & Struhl 1990	h	46-59%
h ₍₊₅₎	3L	8662070	8664900	h _{betah1(2)5}	Riddihough & Ish- Horowicz 1991	h	23-64%
h ₍₊₉₎	3L	8659411	8660491	h _{stripe_6+2}	Howard & Struhl 1990	h	68-74%
hb _{(-2)_broad}	3R	4524484	4526003	hb _{distal_nonminimal}	Perry et al. 2011	hb	5-55%
hb _{(-2)_narrow}	3R	4524620	4525479	Unspecified _{HC_01}	Chen et al. 2012	hb	6-50%
hb ₍₋₄₎	3R	4526522	4527945	hb _{upstream_enhancer (HZ1.4)}	Margolis et al. 1995	hb	38-92%
hb _{(-4)_downstream}	3R	4526520	4526862	hb _{HZ340}	Margolis et al. 1995	hb	41-92%
hb _{(-4)_upstream}	3R	4526861	4527388	hb _{HZ526}	Margolis et al. 1995	hb	78-92%
hb _{(+0)_broad}	3R	4519856	4520626	hb _{proximal}	Perry et al. 2011	hb	1-53%
hb _{(+0)_narrow}	3R	4519886	4521119	hb _{1.2}	Schröder et al. 1988	hb	1-53%
hbn ₍₋₁₎	2R	16849286	16850294	Unspecified _{HC_14}	Chen et al. 2012	hbn	3-19%

hkb_(-2)	3R	173892	174474	hkb_0.6kbRIRV	Hader et al. 2000	hkb	1-99%
hkb_(+6)	3R	165647	167219	hkb_distal_+6kb	Perry et al. 2011	hkb	1-15%
kni_(-2)_broad	3L	20689644	20690670	kni_distal	Perry et al. 2011	kni	52-78%
kni_(-2)_narrow	3L	20689643	20690512	kni_KD	Pankratz et al. 1992	kni	55-73%
kni_(+0)	3L	20687687	20688345	kni_proximal_minimal	Perry et al. 2011	kni	48-78%
Kr_(-11)	2R	21124126	21125171	Kr_Cluster-8297	Fisher et al. 2012	Kr	10-28%
Kr_(+0)	2R	21113281	21115489	Kr_Kr/E	Jacob et al. 1991	Kr	11-28%
Kr_(+1)	2R	21112355	21113940	Kr_proximal	Perry et al. 2011	Kr	37-67%
Kr_(+2)	2R	21111575	21113281	Kr_NcS1.7HZ	Hoch et al. 1990	Kr	36-67%
Kr_(+3)	2R	21110142	21111300	Kr_CD1	Hoch et al. 1990	Kr	45-68%
noc_(+1)	2L	14489159	14489648	Unspecified_HC_34	Chen et al. 2012	noc	7-19%
nub_(+2)	2L	12615792	12617776	nub_-2_construct	Schroeder et al. 2004	nub	49-80%
oc_(-1)	X	8537082	8538914	oc_+7_construct	Schroeder et al. 2004	oc	1-28%
oc_(-4)	X	8547936	8549796	oc_otd_early_enhancer	Gao and Finkelstein 1998	oc	3-28%
oc_(+0)	X	8536987	8537940	oc_intronic_distal	Perry et al. 2011	oc	8-34%
odd_(-3)_broad	2L	3608812	3610461	odd_-3_construct	Schroeder et al. 2004	odd	50-78%
odd_(-3)_narrow	2L	3608835	3610207	odd_CE8010	Berman et al. 2004	odd	50-85%
odd_(-4)	2L	3610420	3611803	odd_-5_construct	Schroeder et al. 2004	odd	35-81%
pdm2_(-1)	2L	12678898	12680520	pdm2_+1_construct	Schroeder et al. 2004	pdm2	50-80%
prd_(-3)	2L	12088746	12089847	Unspecified_HC_03	Chen et al. 2012	prd	25-35%
prd_(-6)	2L	12091700	12092777	prd_Cluster-8520	Fisher et al. 2012	prd	21-83%
prd_(+4)	2L	12080488	12081888	prd_O-E	Ochoa-Espinosa et al. 2005	prd	28-83%
rib_(+2)	2R	15160750	15161413	Unspecified_HC_11	Chen et al. 2012	rib	1-12%

run_(+10)	X	20554215	20556618	run_stripe3+7	Klingler et al. 1996	run	45-96%
run_(+14)	X	20551039	20552649	run_stripe1+7	Klingler et al. 1996	run	28-37%
run_(+17)	X	20548261	20549257	run_(-17)	Schroeder et al. 2011	run	57-62%
run_(+19)	X	20583284	20585544	run_(+19)	Schroeder et al. 2011	run	19-88%
run_(+30)	X	20594595	20597303	run_(+30)	Schroeder et al. 2011	run	40-93%
run_(+31)	X	20533075	20535598	run_(-31)	Schroeder et al. 2011	run	82-92%
run_(+41)	X	20523501	20524783	run_(-41)	Schroeder et al. 2011	run	71-83%
run_(+42)	X	20522461	20523683	run_(-42)	Schroeder et al. 2011	run	71-81%
run_(+9)	X	20555735	20556596	run_(-10)	Schroeder et al. 2011	run	48-59%
slp1_(-4)	2L	3828818	3829610	slp1_slp_B	Ochoa-Espinosa et al. 2005	slp1	16-30%
slp1_(+1)	2L	3824596	3824967	slp1_slp_A	Ochoa-Espinosa et al. 2005	slp1	11-34%
slp2_(+3)	2L	3833484	3834592	Unspecified_HC_02	Chen et al. 2012	slp2	13-34%
SoxN_(-7)	2L	8831698	8833198	SoxN_ChIP-50	Fisher et al. 2012	SoxN	6-31%
tara_(-9)	3R	12076498	12077998	tara_ChIP-37	Fisher et al. 2012	tara	1-41%
tll_(+1)	3R	26675738	26678501	tll_P2	Liaw and Lengyel 1993	tll	1-100%
tll_(+2)	3R	26675091	26676127	Unspecified_HC_07	Chen et al. 2012	tll	2-100%
tll_(+3)	3R	26675104	26675738	tll_K10	Rudolph et al. 1997	tll	1-100%
tll_(+4)	3R	26673281	26675739	tll_K2	Rudolph et al. 1997	tll	77-100%
Tollo_(+12)	3L	15216686	15217594	Unspecified_HC_23	Chen et al. 2012	Tollo	18-32%
tsh_(-49)	2L	21876631	21877660	Unspecified_HC_17	Chen et al. 2012	tsh	37-77%
Ubx_(-10)	3R	12570248	12571305	Ubx_Cluster-8166	Fisher et al. 2012	Ubx	51-88%

Bibliography

- [1] M. Levine and E. Davidson, “Gene regulatory networks for development,” *Proceedings of the National Academy of Sciences*, vol. 102, no. 14, pp. 4936–4942, 2005.
- [2] A. Stathopoulos and M. Levine, “Genomic regulatory networks and animal development,” *Developmental Cell*, vol. 9, no. 4, pp. 449–462, 2005.
- [3] J. Jaeger, “The gap gene network,” *Cellular and Molecular Life Sciences*, vol. 68, no. 2, pp. 243–274, 2011.
- [4] A. Nasiadka, B. H. Dietrich, and H. M. Krause, “Anterior posterior patterning in the *Drosophila* embryo,” *Advances in Developmental Biology and Biochemistry*, vol. 12, pp. 155–204, 2002.
- [5] K. Weigmann, R. Klapper, T. Strasser, C. Rickert, G. Technau, H. Jäckle, W. Janing, and C. Klämbt, “FlyMove - A new way to look at development of *Drosophila*,” *Trends in Genetics*, vol. 19, no. 6, pp. 310–311, 2003.
- [6] B. A. Edgar and P. H. O’Farrell, “Genetic control of cell division patterns in the *Drosophila* embryo,” *Cell*, vol. 57, no. 1, pp. 177–187, 1989.
- [7] X. Y. Li, M. M. Harrison, J. E. Villalta, T. Kaplan, and M. B. Eisen, “Establishment of regions of genomic activity during the *Drosophila* maternal to zygotic transition,” *eLife*, vol. 3, pp. 1–20, 2014.
- [8] H. L. Liang, C. Y. Nien, H. Y. Liu, M. M. Metzstein, N. Kirov, and C. Rushlow, “The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*,” *Nature*, vol. 456, no. 7220, pp. 400–403, 2008.
- [9] K. V. Anderson and C. Nüsslein-Volhard, “Information for the dorsal-ventral pattern of the *Drosophila* embryo is stored as maternal mRNA,” *Nature*, vol. 311, no. 5983, pp. 223–227, 1984.
- [10] H. G. Frohnhof and C. Nüsslein-Volhard, “Maternal genes required for the anterior localization of bicoid activity in the embryo of *Drosophila*,” *Genes & Development*, vol. 1, no. 8, pp. 880–890, 1987.

- [11] C. Nüsslein-Volhard, H. G. Frohnhofer, and R. Lehmann, "Determination of antero-posterior polarity in *Drosophila*," *Science*, vol. 238, no. 4834, pp. 1675–1681, 1987.
- [12] T. Schüpbach and E. Wieschaus, "Maternal-effect mutations altering the anterior-posterior pattern of the *Drosophila* embryo," *Roux's Archives of Developmental Biology*, vol. 195, no. 5, pp. 302–317, 1986.
- [13] Z. Ali-Murthy, S. E. Lott, M. B. Eisen, and T. B. Kornberg, "An Essential Role for Zygotic Expression in the Pre-Cellular *Drosophila* Embryo," *PLoS Genetics*, vol. 9, no. 4, p. e1003428, 2013.
- [14] S. De Renzis, O. Elemento, S. Tavazoie, and E. F. Wieschaus, "Unmasking activation of the zygotic genome using chromosomal deletions in the *Drosophila* embryo," *PLoS Biology*, vol. 5, no. 5, pp. 1036–1051, 2007.
- [15] E. Lécuyer, H. Yoshida, N. Parthasarathy, C. Alm, T. Babak, T. Cerovina, T. R. Hughes, P. Tomancak, and H. M. Krause, "Global Analysis of mRNA Localization Reveals a Prominent Role in Organizing Cellular Architecture and Function," *Cell*, vol. 131, no. 1, pp. 174–187, 2007.
- [16] N. Karaiskos, P. Wahle, J. Alles, A. Boltengagen, S. Ayoub, C. Kipar, C. Kocks, N. Rajewsky, and R. P. Zinzen, "The *Drosophila* embryo at single-cell transcriptome resolution," *Science*, vol. 358, no. 6360, pp. 194–199, 2017.
- [17] C. Nüsslein-Volhard, "Determination of the Embryonic Axes of *Drosophila*," *Development*, vol. 1, pp. 1–10, 1991.
- [18] B. Moussian and S. Roth, "Dorsoventral axis formation in the *Drosophila* embryo - Shaping and transducing a morphogen gradient," *Current Biology*, vol. 15, no. 21, pp. 887–899, 2005.
- [19] S. B. Carroll, G. M. Winslow, V. J. Twombly, and M. P. Scott, "Genes that control dorsoventral polarity affect gene expression along the anteroposterior axis of the *Drosophila* embryo," *Development*, vol. 99, pp. 327–332, 1987.
- [20] S. V. E. Keränen, C. Fowlkes, C. Luengo, D. Sudar, D. K. Knowles, J. Malik, and M. D. Biggin, "Three-dimensional morphology and gene expression in the *Drosophila* blastoderm at cellular resolution II: dynamics," *Genome Biology*, vol. 7, no. 12, p. R124, 2006.
- [21] D. Papatsenko, Y. Goltsev, and M. Levine, "Organization of developmental enhancers in the *Drosophila* embryo," *Nucleic Acids Research*, vol. 37, no. 17, pp. 5665–5677, 2009.
- [22] J. Zeitlinger, R. P. Zinzen, A. Stark, M. Kellis, H. Zhang, R. A. Young, and M. Levine, "Whole-genome ChIPchip analysis of Dorsal, Twist, and Snail suggests integration

- of diverse patterning processes in the *Drosophila* embryo,” *Genes & Development*, no. 510, pp. 385–390, 2007.
- [23] P. Tomancak, A. Beaton, R. Weiszmann, E. Kwan, S. Shu, S. E. Lewis, S. Richards, M. Ashburner, V. Hartenstein, S. E. Celniker, and G. M. Rubin, “Systematic determination of patterns of gene expression during *Drosophila* embryogenesis,” *Genome biology*, vol. 3, no. 12, p. RESEARCH0088, 2002.
- [24] M. Mallo and C. R. Alonso, “The regulation of Hox gene expression during animal development,” *Development*, vol. 140, no. 19, pp. 3951–3963, 2013.
- [25] I. Gray, P. Szymanski, and M. Levine, “Short-range repression permits multiple enhancers to function autonomously within a complex promoter,” *Genes and Development*, vol. 8, no. 15, pp. 1829–1838, 1994.
- [26] S. Small, D. N. Arnosti, and M. Levine, “Spacing ensures autonomous expression of different stripe enhancers in the even-skipped promoter,” *Development*, vol. 119, no. 3, pp. 762–772, 1993.
- [27] S. Small, A. Blair, and M. Levine, “Regulation of even-skipped stripe 2 in the *Drosophila* embryo,” *The EMBO journal*, vol. 11, no. 11, pp. 4047–57, 1992.
- [28] L. M. Li and D. N. Arnosti, “Long- and short-range transcriptional repressors induce distinct chromatin states on repressed genes,” *Current Biology*, vol. 21, no. 5, pp. 406–412, 2011.
- [29] G. Struhl, K. Struhl, and P. M. Macdonald, “The gradient morphogen bicoid is a concentration-dependent transcriptional activator,” *Cell*, vol. 57, no. 7, pp. 1259–1273, 1989.
- [30] W. Driever, G. Thoma, and C. Nüsslein-Volhard, “Determination of spatial domains of zygotic gene expression in the *Drosophila* embryo by the affinity of binding sites for the bicoid morphogen,” *Nature*, vol. 340, no. 6232, pp. 363–367, 1989.
- [31] A. P. Lifanov, V. J. Makeev, A. G. Nazina, and D. A. Papatsenko, “Homotypic Regulatory Clusters in *Drosophila*,” *Genome Research*, vol. 13, no. 4, pp. 579–588, 2003.
- [32] D. S. Burz, R. Rivera-Pomar, H. Jäckle, and S. D. Hanes, “Cooperative DNA-binding by Bicoid provides a mechanism for threshold-dependent gene activation in the *Drosophila* embryo,” *EMBO Journal*, vol. 17, no. 20, pp. 5998–6009, 1998.
- [33] D. Lebrecht, M. Foehr, E. Smith, F. J. P. Lopes, C. E. Vanario-Alonso, J. Reinitz, D. S. Burz, and S. D. Hanes, “Bicoid cooperative DNA binding is critical for embryonic patterning in *Drosophila*,” *Proceedings of the National Academy of Sciences*, vol. 102, no. 37, pp. 13176–13181, 2005.

- [34] S. Small, A. Blair, and M. Levine, "Regulation of two pair-rule stripes by a single enhancer in the *Drosophila* embryo," *Developmental Biology*, vol. 175, no. 2, pp. 314–324, 1996.
- [35] H. N. Cai, D. N. Arnosti, and M. Levine, "Long-range repression in the *Drosophila* embryo.," *Proceedings of the National Academy of Sciences*, vol. 93, no. 18, pp. 9309–9314, 1996.
- [36] R. Yan, S. Small, C. Desplan, C. R. Dearolf, and J. E. Darnell, "Identification of a Stat gene that functions in *Drosophila* development," *Cell*, vol. 84, no. 3, pp. 421–430, 1996.
- [37] C. A. Martinez and D. N. Arnosti, "Spreading of a corepressor linked to action of long-range repressor hairy.," *Molecular and Cellular Biology*, vol. 28, no. 8, pp. 2792–802, 2008.
- [38] S. Surkova, D. Kosman, K. Kozlov, Manu, E. Myasnikova, A. A. Samsonova, A. Spirov, C. E. Vanario-Alonso, M. Samsonova, and J. Reinitz, "Characterization of the *Drosophila* segment determination morphome," *Developmental Biology*, vol. 313, no. 2, pp. 844–862, 2008.
- [39] J. Jaeger, S. Surkova, M. Blagov, H. Janssens, D. Kosman, K. N. Kozlov, Manu, E. Myasnikova, C. E. Vanario-Alonso, M. Samsonova, D. H. Sharp, and J. Reinitz, "Dynamic control of positional information in the early *Drosophila* embryo," *Nature*, vol. 430, no. 6997, pp. 368–371, 2004.
- [40] Manu, S. Surkova, A. V. Spirov, V. V. Gursky, H. Janssens, A. R. Kim, O. Radulescu, C. E. Vanario-Alonso, D. H. Sharp, M. Samsonova, and J. Reinitz, "Canalization of gene expression in the *Drosophila* blastoderm by gap gene cross regulation," *PLoS Biology*, vol. 7, no. 3, pp. 0591–0603, 2009.
- [41] M. W. Perry, A. N. Boettiger, and M. Levine, "Multiple enhancers ensure precision of gap gene-expression patterns in the *Drosophila* embryo," *Proceedings of the National Academy of Sciences*, vol. 108, no. 33, pp. 13570–13575, 2011.
- [42] W. Driever and C. Nüsslein-Volhard, "The bicoid protein determines position in the *Drosophila* embryo in a concentration-dependent manner," *Cell*, vol. 54, no. 1, pp. 95–104, 1988.
- [43] W. Driever, V. Siegel, and C. Nüsslein-Volhard, "Autonomous determination of anterior structures in the early *Drosophila* embryo by the bicoid morphogen," *Development*, vol. 109, no. 4, pp. 811–20, 1990.
- [44] C. Nüsslein-Volhard and E. Wieschaus, "Mutations affecting segment number and polarity in *Drosophila*," *Nature*, vol. 287, no. 5785, pp. 795–801, 1980.

- [45] S. M. Gallo, D. T. Gerrard, D. Miner, M. Simich, B. Des Soye, C. M. Bergman, and M. S. Halfon, “REDfly v3.0: Toward a comprehensive database of transcriptional regulatory elements in *Drosophila*,” *Nucleic Acids Research*, vol. 39, no. SUPPL. 1, pp. 118–123, 2011.
- [46] M. Fujioka, Y. Emi-sarker, G. L. Yusibova, T. Goto, and J. B. Jaynes, “Analysis of an even-skipped rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients,” *Development*, vol. 2538, pp. 2527–2538, 1999.
- [47] M. Hoch, C. Schröder, E. Seifert, and H. Jäckle, “cis-acting control elements for Krüppel expression in the *Drosophila* embryo,” *The EMBO journal*, vol. 9, no. 8, pp. 2587–95, 1990.
- [48] B. P. Berman, B. D. Pfeiffer, T. R. Lavery, S. L. Salzberg, G. M. Rubin, M. B. Eisen, and S. E. Celniker, “Computational identification of developmental enhancers: conservation and function of transcription factor binding-site clusters in *Drosophila melanogaster* and *Drosophila pseudoobscura*,” *Genome biology*, vol. 5, no. 9, p. R61, 2004.
- [49] B. P. Berman, Y. Nibu, B. D. Pfeiffer, P. Tomancak, S. E. Celniker, M. Levine, G. M. Rubin, and M. B. Eisen, “Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the *Drosophila* genome,” *Proceedings of the National Academy of Sciences*, vol. 99, no. 2, pp. 757–762, 2002.
- [50] M. D. Schroeder, M. Pearce, J. Fak, H. Fan, U. Unnerstall, E. Emberly, N. Rajewsky, E. D. Siggia, and U. Gaul, “Transcriptional control in the segmentation gene network of *Drosophila*,” *PLoS Biology*, vol. 2, no. 9, p. e271, 2004.
- [51] M. D. Schroeder, C. Greer, and U. Gaul, “How to make stripes: deciphering the transition from non-periodic to periodic patterns in *Drosophila* segmentation,” *Development*, vol. 138, no. 14, pp. 3067–3078, 2011.
- [52] A. Ochoa-Espinosa, G. Yucel, L. Kaplan, A. Pare, N. Pura, A. Oberstein, D. Papatzenko, and S. Small, “The role of binding site cluster strength in Bicoid-dependent patterning in *Drosophila*,” *Proc Natl Acad Sci U S A*, vol. 102, no. 14, pp. 4960–4965, 2005.
- [53] H. Chen, Z. Xu, C. Mei, D. Yu, and S. Small, “A system of repressor gradients spatially organizes the boundaries of Bicoid-dependent target genes,” *Cell*, vol. 149, no. 3, pp. 618–629, 2012.
- [54] S. Bonn, R. P. Zinzen, C. Girardot, E. H. Gustafson, A. Perez-Gonzalez, N. Delhomme, Y. Ghavi-Helm, B. Wilczynski, A. Riddell, and E. E. M. Furlong, “Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development,” *Nature Genetics*, vol. 44, no. 2, pp. 148–156, 2012.

- [55] N. Koenecke, J. Johnston, B. Gaertner, M. Natarajan, and J. Zeitlinger, “Genome-wide identification of *Drosophila* dorso-ventral enhancers by differential histone acetylation analysis,” *Genome Biology*, vol. 17, no. 1, pp. 1–19, 2016.
- [56] E. Z. Kvon, T. Kazmar, G. Stampfel, J. O. Yáñez-Cuna, M. Pagani, K. Schernhuber, B. J. Dickson, and A. Stark, “Genome-scale functional characterization of *Drosophila* developmental enhancers in vivo,” *Nature*, vol. 512, no. 1, pp. 91–95, 2014.
- [57] M. Levine, “Transcriptional enhancers in animal development and evolution,” *Current Biology*, vol. 20, no. 17, pp. R754–R763, 2010.
- [58] D. N. Arnosti, S. Barolo, M. Levine, and S. Small, “The eve stripe 2 enhancer employs multiple modes of transcriptional synergy,” *Development*, vol. 122, no. 1, pp. 205–14, 1996.
- [59] W. D. Fakhouri, A. Ay, R. Sayal, J. Dresch, E. Dayringer, and D. N. Arnosti, “Deciphering a transcriptional regulatory code: Modeling short-range repression in the *Drosophila* embryo,” *Molecular Systems Biology*, vol. 6, no. 341, pp. 1–14, 2010.
- [60] M. M. Kulkarni and D. N. Arnosti, “Information display by transcriptional enhancers,” *Development*, vol. 130, no. 26, pp. 6569–6575, 2003.
- [61] M. M. Kulkarni and D. N. Arnosti, “Logic of Short-Range Transcriptional Repression in,” *Society*, vol. 25, no. 9, pp. 3411–3420, 2005.
- [62] X. He, M. A. H. Samee, C. Blatti, and S. Sinha, “Thermodynamics-based models of transcriptional regulation by enhancers: The roles of synergistic activation, cooperative binding and short-range repression,” *PLoS Computational Biology*, vol. 6, no. 9, p. e1000935, 2010.
- [63] M. Kazemian, C. Blatti, A. Richards, M. McCutchan, N. Wakabayashi-Ito, A. S. Hammonds, S. E. Celniker, S. Kumar, S. A. Wolfe, M. H. Brodsky, and S. Sinha, “Quantitative analysis of the *Drosophila* segmentation regulatory network using pattern generating potentials,” *PLoS Biology*, vol. 8, no. 8, pp. 51–52, 2010.
- [64] E. Segal, T. Raveh-Sadka, M. Schroeder, U. Unnerstall, and U. Gaul, “Predicting expression patterns from regulatory sequence in *Drosophila* segmentation,” *Nature*, vol. 451, no. 7178, pp. 535–540, 2008.
- [65] B. J. Vincent, J. Estrada, and A. H. DePace, “The appeasement of Doug: a synthetic approach to enhancer biology,” *Integrative Biology*, vol. 8, no. 4, pp. 475–484, 2016.
- [66] M. J. Guertin and J. T. Lis, “Mechanisms by which transcription factors gain access to target sequence elements in chromatin,” *Current Opinion in Genetics and Development*, vol. 23, no. 2, pp. 116–123, 2013.

- [67] S. Thomas, X. Y. Li, P. J. Sabo, R. Sandstrom, R. E. Thurman, T. K. Canfield, E. Giste, W. Fisher, A. Hammonds, S. E. Celniker, M. D. Biggin, and J. A. Stamatoyannopoulos, “Dynamic reprogramming of chromatin accessibility during *Drosophila* embryo development,” *Genome Biology*, vol. 12, no. 5, p. R43, 2011.
- [68] S. A. Blythe and E. F. Wieschaus, “Establishment and maintenance of heritable chromatin structure during early *drosophila* embryogenesis,” *eLife*, vol. 5, pp. 1–21, 2016.
- [69] V. B. Teif and K. Rippe, “Nucleosome mediated crosstalk between transcription factors at eukaryotic enhancers,” *Physical Biology*, vol. 8, no. 4, pp. 4–8, 2011.
- [70] J. A. Miller and J. Widom, “Collaborative competition mechanism for gene activation in vivo,” *Molecular and Cellular biology*, vol. 23, no. 5, pp. 1623–32, 2003.
- [71] M. Mannervik, “Control of *Drosophila* embryo patterning by transcriptional co-regulators,” *Experimental Cell Research*, vol. 321, no. 1, pp. 47–57, 2014.
- [72] W. H. W. Ludlam, M. M. H. Taylor, K. K. G. Tanner, J. J. M. Denu, R. H. R. Goodman, and S. S. M. Smolik, “The acetyltransferase activity of CBP is required for wingless activation and H4 acetylation in *Drosophila melanogaster*,” *Molecular and Cellular Biology*, vol. 22, no. 1, p. 3832, 2002.
- [73] F. Tie, R. Banerjee, C. A. Stratton, J. Prasad-Sinha, V. Stepanik, A. Zlobin, M. O. Diaz, P. C. Scacheri, and P. J. Harte, “CBP-mediated acetylation of histone H3 lysine 27 antagonizes *Drosophila* Polycomb silencing,” *Development*, vol. 136, no. 18, pp. 3131–3141, 2009.
- [74] H. Akimaru, D. X. Hou, and S. Ishii, “*Drosophila* CBP is required for dorsal-dependent twist gene expression,” *Nature Genetics*, vol. 17, no. 2, pp. 211–214, 1997.
- [75] P. H. Holmqvist, A. Boija, P. Philip, F. Crona, P. Stenberg, and M. Mannervik, “Preferential genome targeting of the CBP co-activator by Rel and Smad proteins in early *Drosophila melanogaster* embryos,” *PLoS Genetics*, vol. 8, no. 6, p. e1002769, 2012.
- [76] D. Fu, Y. Wen, and J. Ma, “The co-activator CREB-binding protein participates in enhancer-dependent activities of Bicoid,” *Journal of Biological Chemistry*, vol. 279, no. 47, pp. 48725–48733, 2004.
- [77] Y. Nibu, H. Zhang, and M. Levine, “Interaction of short-range repressors with *Drosophila* CtBP in the embryo,” *Science*, vol. 280, no. 5360, pp. 101–104, 1998.
- [78] Z. Paroush, R. L. Finley, T. Kidd, S. M. Wainwright, P. W. Ingham, R. Brent, and D. Ish-Horowicz, “Groucho is required for *Drosophila* neurogenesis, segmentation, and sex determination and interacts directly with hairy-related bHLH proteins,” *Cell*, vol. 79, no. 5, pp. 805–815, 1994.

- [79] G. Chen, J. Fernandez, S. Mische, and A. J. Courey, “A functional interaction between the histone deacetylase Rpd3 and the corepressor Groucho in *Drosophila* development,” *Genes and Development*, vol. 13, no. 17, pp. 2218–2230, 1999.
- [80] P. Struffi and D. N. Arnosti, “Functional interaction between the *Drosophila* knirps short range transcriptional repressor and RPD3 histone deacetylase,” *Journal of Biological Chemistry*, vol. 280, no. 49, pp. 40757–40765, 2005.
- [81] M. Mannervik and M. Levine, “The Rpd3 histone deacetylase is required for segmentation of the *Drosophila* embryo,” *Proceedings of the National Academy of Sciences*, vol. 96, no. 12, pp. 6797–801, 1999.
- [82] N. Koenecke, J. Johnston, Q. He, S. Meier, and J. Zeitlinger, “*Drosophila* poised enhancers are generated during tissue patterning with the help of repression,” *Genome Research*, vol. 27, no. 1, pp. 64–74, 2017.
- [83] S. Roy, J. Ernst, P. V. Kharchenko, P. Kheradpour, N. Negre, M. L. Eaton, J. M. Landolin, C. A. Bristow, L. Ma, M. F. Lin, S. Washietl, B. I. Arshinoff, F. Ay, P. E. Meyer, N. Robine, N. L. Washington, L. Di Stefano, E. Berezikov, C. D. Brown, R. Candeias, J. W. Carlson, A. Carr, I. Jungreis, D. Marbach, R. Sealfon, M. Y. Tolstorukov, S. Will, A. A. Alekseyenko, C. Artieri, B. W. Booth, A. N. Brooks, Q. Dai, C. A. Davis, M. O. Duff, X. Feng, A. A. Gorchakov, T. Gu, J. G. Henikoff, P. Kapranov, R. Li, H. K. MacAlpine, J. Malone, A. Minoda, J. Nordman, K. Okamura, M. Perry, S. K. Powell, N. C. Riddle, A. Sakai, A. Samsonova, J. E. Sandler, Y. B. Schwartz, N. Sher, R. Spokony, D. Sturgill, M. van Baren, K. H. Wan, L. Yang, C. Yu, E. Feingold, P. Good, M. Guyer, R. Lowdon, K. Ahmad, J. Andrews, B. Berger, S. E. Brenner, M. R. Brent, L. Cherbas, S. C. Elgin, T. R. Gingeras, R. Grossman, R. A. Hoskins, T. C. Kaufman, W. Kent, M. I. Kuroda, T. Orr-Weaver, N. Perrimon, V. Pirrotta, J. W. Posakony, B. Ren, S. Russell, P. Cherbas, B. R. Graveley, S. Lewis, G. Micklem, B. Oliver, P. J. Park, S. E. Celniker, S. Henikoff, G. H. Karpen, E. C. Lai, D. M. MacAlpine, L. D. Stein, K. P. White, M. Kellis, B. Booth, A. A. Samsonova, M. van Baren, T. R. Gingeras, R. A. Hoskins, C. L. Comstock, A. Dobin, J. Drenkow, S. Dudoit, J. Dumais, D. Fagegaltier, S. Ghosh, K. D. Hansen, S. Jha, L. Langton, W. Lin, D. Miller, A. E. Tenney, H. Wang, A. T. Willingham, C. Zaleski, D. Zhang, D. Acevedo, E. P. Bishop, S. E. Gadel, Y. L. Jung, C. D. Kennedy, O. K. Lee, D. Linder-Basso, S. E. Marchetti, G. Shanower, N. Nègre, R. L. Grossman, R. Auburn, H. J. Bellen, J. Chen, M. H. Domanus, D. Hanley, E. Heinz, Z. Li, F. Meyer, S. W. Miller, C. A. Morrison, D. A. Scheftner, L. Senderowicz, P. K. Shah, S. Suchy, F. Tian, K. J. Venken, R. White, J. Wilkening, J. Zieba, J. T. Nordman, T. L. Orr-Weaver, L. C. DeNapoli, Q. Ding, T. Eng, H. Kashevsky, S. Li, J. A. Prinz, Q. Dai, G. J. Hannon, M. Hirst, M. Marra, M. Rooks, Y. Zhao, T. D. Bryson, M. D. Perry, W. J. Kent, S. E. Lewis, G. Barber, A. Chateigner, H. Clawson, S. Contrino, F. Guillier, A. S. Hinrichs, E. T. Kephart, P. Lloyd, R. Lyne, S. McKay, R. A. Moore, C. Mungall, K. M. Rutherford, P. Ruzanov, R. Smith, E. O. Stinson, Z. Zha, C. G.

- Artieri, J. H. Malone, L. Jiang, N. Mattiuzzo, E. A. Feingold, P. J. Good, M. S. Guyer, and R. F. Lowdon, "Identification of functional elements and regulatory circuits by *Drosophila* modENCODE," *Science*, vol. 330, no. 6012, pp. 1787–1797, 2010.
- [84] O. Bell, V. K. Tiwari, N. H. Thomä, and D. Schübeler, "Determinants and dynamics of genome accessibility," *Nature Reviews Genetics*, vol. 12, no. 8, pp. 554–564, 2011.
- [85] R. B. Deal and S. Henikoff, "A simple method for gene expression and chromatin profiling of individual cell types within a tissue," *Developmental Cell*, vol. 18, no. 6, pp. 1030–1040, 2010.
- [86] G. L. Henry, F. P. Davis, S. Picard, and S. R. Eddy, "Cell type-specific genomics of *Drosophila* neurons," *Nucleic Acids Research*, vol. 40, no. 19, pp. 9691–9704, 2012.
- [87] F. A. Steiner, P. B. Talbert, S. Kasinathan, R. B. Deal, and S. Henikoff, "Cell-type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling," *Genome Research*, vol. 22, no. 4, pp. 766–777, 2012.
- [88] J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, and W. J. Greenleaf, "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position," *Nature Methods*, vol. 10, no. 12, pp. 1213–1218, 2013.
- [89] D. A. Cusanovich, J. P. Reddington, D. A. Garfield, R. M. Daza, D. Aghamirzaie, R. Marco-Ferrerres, H. A. Pliner, L. Christiansen, X. Qiu, F. J. Steemers, C. Trapnell, J. Shendure, and E. E. M. Furlong, "The cis-regulatory dynamics of embryonic development at single-cell resolution," *Nature*, vol. 555, no. 7697, pp. 538–542, 2018.
- [90] J. E. Haines and M. B. Eisen, "Patterns of chromatin accessibility along the anterior-posterior axis in the early *Drosophila* embryo," *PLOS Genetics*, vol. 14, no. 5, p. e1007367, 2018.
- [91] B. D. Pfeiffer, A. Jenett, A. S. Hammonds, T.-T. B. Ngo, S. Misra, C. Murphy, A. Scully, J. W. Carlson, K. H. Wan, T. R. Laverty, C. Mungall, R. Svirskas, J. T. Kadonaga, C. Q. Doe, M. B. Eisen, S. E. Celniker, and G. M. Rubin, "Tools for neuroanatomy and neurogenetics in *Drosophila*," *Proceedings of the National Academy of Sciences*, vol. 105, no. 28, pp. 9715–9720, 2008.
- [92] S. Haenni, Z. Ji, M. Hoque, N. Rust, H. Sharpe, R. Eberhard, C. Browne, M. O. Hengartner, J. Mellor, B. Tian, and A. Furger, "Analysis of *C. elegans* intestinal gene expression and polyadenylation by fluorescence-activated nuclei sorting and 3-end-seq," *Nucleic Acids Research*, vol. 40, no. 13, pp. 6304–6318, 2012.
- [93] S. Bonn, R. P. Zinzen, A. Perez-Gonzalez, A. Riddell, A. C. Gavin, and E. E. Furlong, "Cell type-specific chromatin immunoprecipitation from multicellular complex samples using BiTS-ChIP," *Nature Protocols*, vol. 7, no. 5, pp. 978–994, 2012.

- [94] J. Vierstra, H. Wang, S. John, R. Sandstrom, and J. A. Stamatoyannopoulos, “Coupling transcription factor occupancy to nucleosome architecture with DNase-FLASH,” *Nature Methods*, vol. 11, no. 1, pp. 66–72, 2014.
- [95] C. Wu, “The 5’ ends of drosophila heat shock genes in chromatin are hypersensitive to DNase I,” *Nature*, vol. 286, no. 5776, pp. 854–860, 1980.
- [96] J. D. Anderson, P. T. Lowary, and J. Widom, “Effects of histone acetylation on the equilibrium accessibility of nucleosomal DNA target sites,” *Journal of Molecular Biology*, vol. 307, no. 4, pp. 977–985, 2001.
- [97] A. Gansen, K. Tóth, N. Schwarz, and J. Langowski, “Opposing roles of H3-and H4-acetylation in the regulation of nucleosome structure - a FRET study,” *Nucleic Acids Research*, vol. 43, no. 3, pp. 1433–1443, 2015.
- [98] L. N. Mishra, S. Peppenella, R. Rogge, J. C. Hansen, and J. J. Hayes, “Acetylation Mimics Within a Single Nucleosome Alter Local DNA Accessibility in Compacted Nucleosome Arrays,” *Scientific Reports*, vol. 6, no. 2016, p. 34808, 2016.
- [99] M. Shogren-Knaak, H. Ishii, J. M. Sun, M. J. Pazin, J. R. Davie, and C. L. Peterson, “Histone H4-K16 acetylation controls chromatin structure and protein interactions,” *Science*, vol. 311, no. 5762, pp. 844–847, 2006.
- [100] K. Tóth, N. Brun, and J. Langowski, “Chromatin compaction at the mononucleosome level,” *Biochemistry*, vol. 45, no. 6, pp. 1591–1598, 2006.
- [101] A. P. Boyle, S. Davis, H. P. Shulha, P. Meltzer, E. H. Margulies, Z. Weng, T. S. Furey, and G. E. Crawford, “High-Resolution Mapping and Characterization of Open Chromatin across the Genome,” *Cell*, vol. 132, no. 2, pp. 311–322, 2008.
- [102] A. Adey, H. G. Morrison, Asan, X. Xun, J. O. Kitzman, E. H. Turner, B. Stackhouse, A. P. MacKenzie, N. C. Caruccio, X. Zhang, and J. Shendure, “Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition,” *Genome Biology*, vol. 11, no. 12, 2010.
- [103] H. Koohy, T. A. Down, and T. J. Hubbard, “Chromatin Accessibility Data Sets Show Bias Due to Sequence Specificity of the DNase I Enzyme,” *PLoS ONE*, vol. 8, no. 7, p. e69853, 2013.
- [104] P. Madrigal, “On Accounting for Sequence-Specific Bias in Genome-Wide Chromatin Accessibility Experiments: Recent Advances and Contradictions,” *Frontiers in Bioengineering and Biotechnology*, vol. 3, no. September, pp. 1–4, 2015.
- [105] P. N. Cockerill, “Structure and function of active chromatin and DNase i hypersensitive sites,” *FEBS Journal*, vol. 278, no. 13, pp. 2182–2210, 2011.

- [106] D. Görlich and U. Kutay, “Transport between the cell nucleus and the cytoplasm,” *Annual review of cell and developmental biology*, vol. 15, pp. 607–60, 1999.
- [107] A. Kusano, C. Staber, and B. Ganetzky, “Nuclear Mislocalization of Enzymatically Active RanGAP Causes Segregation Distortion in *Drosophila*,” *Developmental Cell*, vol. 1, no. 3, pp. 351–361, 2001.
- [108] C. R. Bone, E. C. Tapley, M. Gorjanacz, and D. A. Starr, “The *Caenorhabditis elegans* SUN protein UNC-84 interacts with lamin to transfer forces from the cytoplasm to the nucleoskeleton during nuclear migration,” *Molecular Biology of the Cell*, vol. 25, no. 18, pp. 2853–2865, 2014.
- [109] L. Klyszejko-Stefanowicz and L. S. Hnilica, “Isolation, fractionation, and analysis of non-histone chromosomal proteins,” in *Progress In Nonhistone Protein Research: Volume III*, ch. Chapter I, pp. 1–68, CRC Press, 2017.
- [110] V. E. Foe, G. M. Odell, and B. a. Edgar, “Timetable of *Drosophila* Early Development,” in *The Development of Drosophila melanogaster*, pp. 149–300, 1993.
- [111] F. Pilot, J.-M. Philippe, C. Lemmers, J.-P. Chauvin, and T. Lecuit, “Developmental control of nuclear morphogenesis and anchoring by charleston, identified in a functional genomic screen of *Drosophila* cellularisation,” *Development*, vol. 133, no. 4, pp. 711–723, 2006.
- [112] Z. Hu and W.-W. Tee, “Enhancers and chromatin structures: regulatory hubs in gene expression and diseases,” *Bioscience Reports*, vol. 37, no. 2, p. BSR20160183, 2017.
- [113] C. Thummel and V. Pirrotta, “New pCaSpeR P element vectors,” *Dros. Info. Serv.*, vol. 71, p. 150, 1992.
- [114] J. Amin, R. Mestril, P. Schiller, M. Dreano, and R. Voellmy, “Organization of the *Drosophila melanogaster* hsp70 heat shock regulation unit,” *Molecular and Cellular biology*, vol. 7, no. 3, pp. 1055–1062, 1987.
- [115] P. Rørth, “Gal4 in the *Drosophila* female germline,” *Mechanisms of Development*, vol. 78, no. 1-2, pp. 113–118, 1998.
- [116] A. C. Groth, M. Fish, R. Nusse, and M. P. Calos, “Construction of Transgenic *Drosophila* by Using the Site-Specific Integrase from Phage ϕ C31,” *Genetics*, vol. 166, no. 4, pp. 1775–1782, 2004.
- [117] M. Markstein, C. Pitsouli, C. Villalta, S. E. Celniker, and N. Perrimon, “Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes,” *Nature Genetics*, vol. 40, no. 4, pp. 476–483, 2008.

- [118] B. D. Pfeiffer, T. T. B. Ngo, K. L. Hibbard, C. Murphy, A. Jenett, J. W. Truman, and G. M. Rubin, “Refinement of tools for targeted gene expression in *Drosophila*,” *Genetics*, vol. 186, no. 2, pp. 735–755, 2010.
- [119] C. Engler, R. Kandzia, and S. Marillonnet, “A one pot, one step, precision cloning method with high throughput capability,” *PLoS ONE*, vol. 3, no. 11, p. e3647, 2008.
- [120] C. Engler, R. Gruetzner, R. Kandzia, and S. Marillonnet, “Golden gate shuffling: a one-pot DNA shuffling method based on type II restriction enzymes,” *PloS one*, vol. 4, no. 5, p. e5553, 2009.
- [121] E. D. Eldon and V. Pirrotta, “Interactions of the *Drosophila* gap gene giant with maternal and zygotic pattern-forming genes,” *Development*, vol. 111, pp. 367–378, 1991.
- [122] M. M. Harrison, X. Y. Li, T. Kaplan, M. R. Botchan, and M. B. Eisen, “Zelda binding in the early *Drosophila melanogaster* embryo marks regions subsequently activated at the maternal-to-zygotic transition,” *PLoS Genetics*, vol. 7, no. 10, 2011.
- [123] A. N. Schep, J. D. Buenrostro, S. K. Denny, K. Schwartz, G. Sherlock, and W. J. Greenleaf, “Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions,” *Genome Research*, vol. 25, no. 11, pp. 1757–1770, 2015.
- [124] J. D. Anderson, A. Thåström, and J. Widom, “Spontaneous Access of Proteins to Buried Nucleosomal DNA Target Sites Occurs via a Mechanism That Is Distinct from Nucleosome Translocation,” *Molecular and Cellular Biology*, vol. 22, no. 20, pp. 7147–7157, 2002.
- [125] M. D. Adams, “The Genome Sequence of *Drosophila melanogaster*,” *Science*, vol. 287, no. 5461, pp. 2185–2195, 2000.
- [126] L. S. Gramates, S. J. Marygold, G. Dos Santos, J. M. Urbano, G. Antonazzo, B. B. Matthews, A. J. Rey, C. J. Tabone, M. A. Crosby, D. B. Emmert, K. Falls, J. L. Goodman, Y. Hu, L. Ponting, A. J. Schroeder, V. B. Strelets, J. Thurmond, P. Zhou, N. Perrimon, S. R. Gelbart, C. Extavour, K. Broll, M. Zytkevich, N. H. Brown, H. Attrill, M. Costa, S. Fexova, T. Jones, A. Larkin, G. H. Millburn, N. Staudt, T. Kaufman, G. B. Grumblin, R. Cripps, M. Werner-Washburne, and P. Baker, “FlyBase at 25: Looking to the future,” *Nucleic Acids Research*, vol. 45, no. D1, pp. D663–D671, 2017.
- [127] T. N. Mavrich, C. Jiang, I. P. Ioshikhes, X. Li, B. J. Venters, S. J. Zanton, L. P. Tomsho, J. Qi, R. L. Glaser, S. C. Schuster, D. S. Gilmour, I. Albert, and B. F. Pugh, “Nucleosome organization in the *Drosophila* genome,” *Nature*, vol. 453, no. 7193, pp. 358–362, 2008.

- [128] A. J. Stewart and J. B. Plotkin, “Why transcription factor binding sites are ten nucleotides long,” *Genetics*, vol. 192, no. 3, pp. 973–985, 2012.
- [129] Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nussbaum, R. M. Myers, M. Brown, W. Li, and X. S. Shirley, “Model-based Analysis of ChIP-Seq (MACS),” *Genome Biology*, vol. 9, no. 9, p. R137, 2008.
- [130] J. C. Pearson, D. J. McKay, J. D. Lieb, and S. T. Crews, “Chromatin profiling of *Drosophila* CNS subpopulations identifies active transcriptional enhancers,” *Development*, vol. 143, no. 20, pp. 3723–3732, 2016.
- [131] M. I. Love, W. Huber, and S. Anders, “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2,” *Genome Biology*, vol. 15, no. 12, pp. 1–21, 2014.
- [132] Y. Sun, C. Y. Nien, K. Chen, H. Y. Liu, J. Johnston, J. Zeitlinger, and C. Rushlow, “Zelda overcomes the high intrinsic nucleosome barrier at enhancers during *Drosophila* zygotic genome activation,” *Genome Research*, vol. 25, no. 11, pp. 1703–1714, 2015.
- [133] M. L. Eaton, J. A. Prinz, H. K. MacAlpine, G. Tretyakov, P. V. Kharchenko, and D. M. MacAlpine, “Chromatin signatures of the *Drosophila* replication program,” *Genome Research*, vol. 21, no. 2, pp. 164–174, 2011.
- [134] S. E. Celniker, L. A. Dillon, M. B. Gerstein, K. C. Gunsalus, S. Henikoff, G. H. Karpen, M. Kellis, E. C. Lai, J. D. Lieb, D. M. MacAlpine, G. Micklem, F. Piano, M. Snyder, L. Stein, K. P. White, and R. H. Waterston, “Unlocking the secrets of the genome,” *Nature*, vol. 459, no. 7249, pp. 927–930, 2009.
- [135] S. MacArthur, X. Y. Li, J. Li, J. B. Brown, H. C. Cheng, L. Zeng, B. P. Grondona, A. Hechmer, L. Simirenko, S. V. Keränen, D. W. Knowles, M. Stapleton, P. Bickel, M. D. Biggin, and M. B. Eisen, “Developmental roles of 21 *Drosophila* transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions,” *Genome Biology*, vol. 10, no. 7, 2009.
- [136] X. Y. Li, S. MacArthur, R. Bourgon, D. Nix, D. A. Pollard, V. N. Iyer, A. Hechmer, L. Simirenko, M. Stapleton, C. L. Luengo Hendriks, C. C. Hou, N. Ogawa, W. Inwood, V. Sementchenko, A. Beaton, R. Weizmann, S. E. Celniker, D. W. Knowles, T. Gingeras, T. P. Speed, M. B. Eisen, and M. D. Biggin, “Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm,” *PLoS Biology*, vol. 6, no. 2, pp. 0365–0388, 2008.
- [137] R. K. Bradley, X. Y. Li, C. Trapnell, S. Davidson, L. Pachter, H. C. Chu, L. A. Tonkin, M. D. Biggin, and M. B. Eisen, “Binding site turnover produces pervasive quantitative changes in transcription factor binding between closely related *drosophila* species,” *PLoS Biology*, vol. 8, no. 3, p. e1000343, 2010.

- [138] W. W. Fisher, J. J. Li, A. S. Hammonds, J. B. Brown, B. D. Pfeiffer, R. Weiszmann, S. MacArthur, S. Thomas, J. A. Stamatoyannopoulos, M. B. Eisen, P. J. Bickel, M. D. Biggin, and S. E. Celniker, “DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in *Drosophila*,” *Proceedings of the National Academy of Sciences*, vol. 109, no. 52, pp. 21330–21335, 2012.
- [139] P. A. Combs and M. B. Eisen, “Sequencing mRNA from Cryo-Sliced *Drosophila* Embryos to Determine Genome-Wide Spatial Patterns of Gene Expression,” *PLoS ONE*, vol. 8, no. 8, pp. 2–8, 2013.
- [140] P. A. Combs and M. B. Eisen, “Genome-wide measurement of spatial expression in patterning mutants of *Drosophila melanogaster*,” *F1000Research*, vol. 6, p. 41, 2017.
- [141] M. Mir, A. Reimer, J. E. Haines, X. Y. Li, M. Stadler, H. Garcia, M. B. Eisen, and X. Darzacq, “Dense Bicoid hubs accentuate binding along the morphogen gradient,” *Genes and Development*, vol. 31, no. 17, pp. 1784–1794, 2017.
- [142] R. E. Thurman, E. Rynes, R. Humbert, J. Vierstra, M. T. Maurano, E. Haugen, N. C. Sheffield, A. B. Stergachis, H. Wang, B. Vernot, K. Garg, S. John, R. Sandstrom, D. Bates, L. Boatman, T. K. Canfield, M. Diegel, D. Dunn, A. K. Ebersol, T. Frum, E. Giste, A. K. Johnson, E. M. Johnson, T. Kutuyavin, B. Lajoie, B. K. Lee, K. Lee, D. London, D. Lotakis, S. Neph, F. Neri, E. D. Nguyen, H. Qu, A. P. Reynolds, V. Roach, A. Safi, M. E. Sanchez, A. Sanyal, A. Shafer, J. M. Simon, L. Song, S. Vong, M. Weaver, Y. Yan, Z. Zhang, Z. Zhang, B. Lenhard, M. Tewari, M. O. Dorschner, R. S. Hansen, P. A. Navas, G. Stamatoyannopoulos, V. R. Iyer, J. D. Lieb, S. R. Sunyaev, J. M. Akey, P. J. Sabo, R. Kaul, T. S. Furey, J. Dekker, G. E. Crawford, and J. A. Stamatoyannopoulos, “The accessible chromatin landscape of the human genome,” *Nature*, vol. 489, no. 7414, pp. 75–82, 2012.
- [143] S. Fu, Q. Wang, J. E. Moore, M. J. Purcaro, H. E. Pratt, K. Fan, C. Gu, C. Jiang, R. Zhu, A. Kundaje, A. Lu, and Z. Weng, “Differential analysis of chromatin accessibility and histone modifications for predicting mouse developmental enhancers,” *Nucleic Acids Research*, pp. 1–18, 2018.
- [144] D. J. McKay and J. D. Lieb, “A Common Set of DNA Regulatory Elements Shapes *Drosophila* Appendages,” *Developmental Cell*, vol. 27, no. 3, pp. 306–318, 2013.
- [145] J. R. Hesselberth, X. Chen, Z. Zhang, P. J. Sabo, R. Sandstrom, A. P. Reynolds, R. E. Thurman, S. Neph, M. S. Kuehn, W. S. Noble, S. Fields, and J. A. Stamatoyannopoulos, “Global mapping of protein-DNA interactions in vivo by digital genomic footprinting,” *Nature Methods*, vol. 6, no. 4, pp. 283–289, 2009.
- [146] C. Jung, P. Bandilla, M. Von Reutern, M. Schnepf, S. Rieder, U. Unnerstall, and U. Gaul, “True equilibrium measurement of transcription factor-DNA binding affinities using automated polarization microscopy,” *Nature Communications*, vol. 9, no. 1, p. 1605, 2018.

- [147] P. C. Peng, M. A. Hassan Samee, and S. Sinha, “Incorporating chromatin accessibility data into sequence-to-expression modeling,” *Biophysical Journal*, vol. 108, no. 5, pp. 1257–1267, 2015.
- [148] R. Satija and R. K. Bradley, “The TAGteam motif facilitates binding of 21 sequence-specific transcription factors in the *Drosophila* embryo,” *Genome Research*, vol. 22, no. 4, pp. 656–665, 2012.
- [149] S. M. Foo, Y. Sun, B. Lim, R. Ziukaite, K. O’Brien, C. Y. Nien, N. Kirov, S. Y. Shvartsman, and C. A. Rushlow, “Zelda potentiates morphogen activity by increasing chromatin accessibility,” *Current Biology*, vol. 24, no. 12, pp. 1341–1346, 2014.
- [150] J. Crocker, A. Tsai, and D. L. Stern, “A Fully Synthetic Transcriptional Platform for a Multicellular Eukaryote,” *Cell Reports*, vol. 18, no. 1, pp. 287–296, 2017.
- [151] F. Liu, “Enhancer-derived RNA: A Primer,” *Genomics, Proteomics and Bioinformatics*, vol. 15, no. 3, pp. 196–200, 2017.
- [152] J. L. Workman, “Nucleosome displacement in transcription,” *Genes and Development*, vol. 20, no. 15, pp. 2009–2017, 2006.
- [153] N. E. Hott, J. A. Heward, B. Roux, E. Tsitsiou, P. S. Fenwick, L. Lenzi, I. Goodhead, C. Hertz-Fowler, A. Heger, N. Hall, L. E. Donnelly, D. Sims, and M. A. Lindsay, “Long non-coding RNAs and enhancer RNAs regulate the lipopolysaccharide-induced inflammatory response in human monocytes,” *Nature Communications*, vol. 5, 2014.
- [154] D. Y. Lee, J. J. Hayes, D. Pruss, and A. P. Wolffe, “A Positive Role for Histone Acetylation in Transcription Factor Access to Nucleosomal DNA,” *Cell*, vol. 72, pp. 73–84, 1993.
- [155] Z. Li and H. Kono, “Distinct Roles of Histone H3 and H2A Tails in Nucleosome Stability,” *Scientific Reports*, vol. 6, no. March, pp. 1–12, 2016.
- [156] C. E. Hannon, S. A. Blythe, and E. F. Wieschaus, “Concentration dependent chromatin states induced by the bicoid morphogen gradient,” *eLife*, vol. 6, pp. 1–29, 2017.
- [157] R. Arai, H. Ueda, A. Kitayama, N. Kamiya, and T. Nagamune, “Design of the linkers which effectively separate domains of a bifunctional fusion protein,” *Protein Engineering Design and Selection*, vol. 14, no. 8, pp. 529–532, 2001.
- [158] H. A. Müller, “Immunolabeling of embryos,” *Methods in molecular biology*, vol. 420, pp. 207–218, 2008.
- [159] J. D. Buenrostro, B. Wu, H. Y. Chang, and W. J. Greenleaf, “ATAC-seq: A method for assaying chromatin accessibility genome-wide,” *Current Protocols in Molecular Biology*, vol. 2015, pp. 21–29, 2015.

- [160] S. Aljanabi, “Universal and rapid salt-extraction of high quality genomic DNA for PCR- based techniques,” *Nucleic Acids Research*, vol. 25, no. 22, pp. 4692–4693, 1997.
- [161] B. Langmead and S. L. Salzberg, “Fast gapped-read alignment with Bowtie 2,” *Nature Methods*, vol. 9, no. 4, pp. 357–359, 2012.
- [162] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin, “The Sequence Alignment/Map format and SAMtools,” *Bioinformatics*, vol. 25, no. 16, pp. 2078–2079, 2009.
- [163] W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, and D. Haussler, “The Human Genome Browser at UCSC,” *Genome Research*, vol. 12, no. 6, pp. 996–1006, 2002.
- [164] C. Schröder, D. Tautz, E. Seifert, and H. Jäckle, “Differential regulation of the two transcripts from the *Drosophila* gap segmentation gene *hunchback*,” *The EMBO journal*, vol. 7, no. 9, pp. T2881–2887, 1988.
- [165] Q. Gao and R. Finkelstein, “Targeting gene expression to the head: the *Drosophila* orthodenticle gene is a direct target of the Bicoid morphogen,” *Development*, vol. 125, no. 21, pp. 4185–93, 1998.
- [166] T. Häder, D. Wainwright, T. Shandala, R. Saint, H. Taubert, G. Brönner, and H. Jäckle, “Receptor tyrosine kinase signaling regulates different modes of Groucho-dependent control of Dorsal,” *Current biology*, vol. 10, no. 1, pp. 51–54, 2000.
- [167] K. R. Howard and G. Struhl, “Decoding positional information: regulation of the pair-rule gene *hairy*,” *Development*, vol. 110, no. 4, pp. 1223–31, 1990.
- [168] Y. Jacob, S. Sather, J. R. Martin, and R. Olo, “Analysis of Krüppel control elements reveals that localized expression results from the interaction of multiple subelements,” *Proceedings of the National Academy of Sciences*, vol. 88, no. 13, pp. 5912–5916, 1991.
- [169] M. Klingler, J. Soong, B. Butler, and J. P. Gergen, “Disperse versus compact elements for the regulation of runt stripes in *Drosophila*,” *Developmental Biology*, vol. 177, no. 1, pp. 73–84, 1996.
- [170] A. La Rosée, T. Häder, H. Taubert, R. Rivera-Pomar, and H. Jäckle, “Mechanism and Bicoid-dependent control of hairy stripe 7 expression in the posterior region of the *Drosophila* embryo,” *EMBO Journal*, vol. 16, no. 14, pp. 4403–4411, 1997.
- [171] G. J. Liaw and J. A. Lengyel, “Control of tailless expression by bicoid, dorsal and synergistically interacting terminal system regulatory elements,” *Mechanisms of Development*, vol. 40, no. 1-2, pp. 47–61, 1993.

- [172] J. S. Margolis, M. L. Borowsky, E. Steingrímsson, C. W. Shim, J. A. Lengyel, and J. W. Posakony, "Posterior stripe expression of hunchback is driven from two promoters by a common enhancer element," *Development*, vol. 121, no. 9, pp. 3067–3077, 1995.
- [173] M. J. Pankratz, M. Busch, M. Hoch, E. Seifert, and H. Jäckle, "Spatial control of the gap gene knirps in the Drosophila embryo by posterior morphogen system," *Science*, vol. 255, no. 5047, pp. 986–989, 1992.
- [174] G. Riddihough and D. Ish-Horowicz, "Individual stripe regulatory elements in the Drosophila hairy promoter respond to maternal, gap, and pair-rule genes," *Genes and Development*, vol. 5, no. 5, pp. 840–854, 1991.
- [175] K. M. Rudolph, G.-J. Liaw, A. Daniel, P. Green, A. J. Courey, V. Hartenstein, and J. A. Lengyel, "Complex regulatory region mediating tailless expression in early embryonic patterning and brain development," *Development*, vol. 124, no. 21, pp. 4297–4308, 1997.
- [176] E. A. Wimmer, M. Simpson-Brose, S. M. Cohen, C. Desplan, and H. Jäckle, "Trans- and cis-acting requirements for blastodermal expression of the head gap gene buttonhead," *Mechanisms of Development*, vol. 53, no. 2, pp. 235–245, 1995.