An Unexplored Genome Insulating Mechanism in *Caenorhabditis Elegans*

Thesis by

Faisal Alkhaldi

In Partial Fulfillment of the Requirements

For the Degree of

Master of Science

King Abdullah University of Science and Technology

Thuwal, Kingdom of Saudi Arabia

November, 2020

EXAMINATION COMMITTEE PAGE

The thesis of Faisal Alkhaldi is approved by the examination committee.

Committee Chairperson: Christian Froekjaer-Jensen
Committee Members: Simon Krattinger, Jesper Tegnér

ABSTRACT

An Unexplored Genome Insulating Mechanism in *Caenorhabditis Elegans*

Faisal Alkhaldi

Caenorhabditis Elegans genome maintains active H3K36me3 chromatin domains interspersed with repressive H3K27me3 domains on the autosomes' distal ends. The mechanisms stabilizing these domains and the prevention of position-effect variegation remains unknown as no insulator elements have been identified in *C. elegans*. *De-novo* motif discovery applied on *mes-4* binding sites links the H3K36me3-specific methyltransferase to a class of non-coding DNA known as Periodic A_n/T_n Clusters (PATCs). PATCs display characteristics of insulator elements such as local nucleosome depletion and their restriction to genes with specific expression profiles and chromatin marks. Finally, I describe a set of experiments to further investigate the role of PATCs and *mes-4* in the maintenance of stable chromatin domains using a synthetic biology approach.

ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my advisor, Professor Christian Frøkjær-Jensen, who is a brilliant scientist and a very thoughtful person. Thank you for your guidance and enthusiasm.

I would also like to thank everyone at the Synthetic Genome Biology laboratory for their support and comradeship.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

BD	bromodomain
C. elegans	*Caenorhabditis elegans*
CD	chromodomain
ChIP-chip	Chromatin immunoprecipitation followed by microarray hybridization
ChIP-seq	Chromatin immunoprecipitation followed by sequencing
CTCF	CCCTC-binding factor
dMes-4	*Drosophila* maternal-effect sterile gene-4
GFP	Green fluorescent protein
IBP	insulator-binding proteins
MEME	Multiple Em for Motif Elicitation
mes-4	*maternal-effect sterile-4*
PATCs	periodic A_n/T_n clusters
PEV	position-effect variegation
PHD	plant homeodomain
PRC2	Polycomb Repressive Complex 2
RFP	Red fluorescent protein
TADs	topologically associating domains
TF	transcription factors
TPM	transcript per million

LIST OF ILLUSTRATIONS

Introduction

Central to the modern understanding of biology is the inheritance and storage of information in the DNA molecule. The genetic code elegantly deciphers the protein-coding information, and sequence-specific interactions between cis- and trans-elements in non-coding DNA are vital for the intricate control of transcription along with a myriad of functional non-coding RNA molecules. More complex eukaryotic organisms employ elaborate packaging of DNA around histones, and subsequent modification of both histones and DNA plays a significant role in regulating transcription. Mono-, di- and tri-methylation and acetylation of histone H3 at various lysine residues is the most widely studied form of histone modification. Each mark affects the underlying DNA differently. For instance, H3K27me3 is an important marker of long-term repressive chromatin, and H3K4me3 is often restricted to regions of active promoters. An important property of histone modifications that is of substantial relevance to this thesis is their self-propagating nature and is further discussed.

Insulator elements prevent the spreading of heterochromatin

Heterochromatin refers to condensed chromatin states that are unfavorable for active gene transcription. H3K9me3 and H3K9me2 are strongly correlated with areas of such nature. Histone methyltransferase enzymes catalyze the transfer of methyl groups to histone proteins. Usually, they form large complexes with effector proteins and transcription factors to target specific genome locations at different developmental

stages. Proteins in these complexes typically contain a histone-reading domain such as bromodomain (BD), plant homeodomain (PHD), and chromodomain (CD) that are capable of binding to the same histone modification that the complex is adding. This reader-writer theory of histone modification predicts that once a histone modifier is recruited to a target site, it modifies a local region and then start a chain-like reaction of modifying histones along the chromosome (Talbert and Henikoff, 2006). This is illustrated in the phenomena of position-effect variegation (PEV) initially reported in *Drosophila*. PEV refers to the silencing of juxtaposed genes via heterochromatin spreading after the loss of insulator elements separating the two genes. (Yun et al., 2011; Musselman et al., 2013; Elgin and Reuter, 2013).

Insulators, also known as chromatin boundary elements, prevent the spreading of chromatin marks from independently regulated genomic regions. They can also prevent enhancer elements from inadvertently influencing distant genomic sites. Insulators that block the spreading of chromatin marks are referred to as barrier elements. In contrast, insulators that regulate enhancer functions are known as enhancer-blocker elements, although many insulators can mediate both functions. CCCTC-binding factor (CTCF) is one of the best characterized examples of insulators that functions as both a barrier and an enhancer-blocker. CTCF cooperates with cohesin, a chromatin architectural protein, in the three-dimensional formation of self-enclosed chromatin loops known as topologically associating domains (TADs). These domains form a physical barrier that blocks the spreading of repressive chromatin marks and prevents regulatory sites from

influencing distant regions. Insulators are but one of the many parts necessary for the proper regulation of gene expression and the genetic blueprint's successful execution. They allow cells to dictate which genes are to be expressed in a specific tissue, at precise developmental stages, and in response to environmental stimuli without interfering with neighboring genes.

The genomes of *Caenorhabditis elegans* and other related nematodes do not contain homologs of known insulator elements (Heger et al., 2009). Interestingly, *C. elegans* chromatin domains are unique in their establishment compared to other model organisms. Broad domains of H3K27me3 mark the ends of autosomes with H3K36me3 marks separating those domains **(Figure 1.1)**. These chromatin domains are identical in early embryos to that of developed animals in the third larval stage (L3). The establishment of those domains occurs during early embryonic development and they are maintained throughout the animal's lifecycle. Separation of the H3K27me3 and H3K36me3 domains occurs at the boundaries of germline expressed genes that are necessary for embryonic development but are also widely expressed elsewhere (Evans et al., 2016). Since the establishment of such domains occurs during embryonic development, it has been speculated that the active transcription of developmental genes has a regulatory role in defining these domains (Carelli et al., 2017). Laurel et al. (2006) showed the deposition of H3K36me3 marks during embryonic development is dependent solely on a single histone methyltransferase named maternal-effect sterile-4 (MES-4). Maternal-effect sterile is a phenotype in which mutant homozygous

hermaphrodites that are fertile produce sterile progeny; in other words, grandchild-less hermaphrodites. Genes that give rise to such phenotypes are thought to produce components that are maternally supplied to the developing germline but are not necessary for later stages of life. MES-2/3/6 form the *C. elegans* equivalent of the well-conserved Polycomb Repressive Complex 2 (PRC2) responsible for the deposition of repressive "heterochromatic" H3K27 marks. MES-4 participates in the balance between activating and repressive chromatin marks by depositing "euchromatic" H3K36 marks (Ahringer and Gasser, 2018). Additionally, Rechtsteiner et al. (2010) and Kreher et al. (2018) showed that maternal MES-4 associates with germline expressed genes and proposed that MES-4 carries an epigenetic memory of genes expressed in previous generations by deposition of H3K36me3 marks.

Unlike *C. elegans*, chromatin domains in *Drosophila* are distinctly marked by the insulator-binding proteins (IBP) dCTCF and Beaf32 that prevent the spreading of chromatin marks. Drosophila Maternal-effect sterile gene-4 (dMes-4), a homolog of C. elegans *mes-4*, directly interacts with both dCTCF and Beaf32. DMes-4 is recruited by IBP to deposit H3K36me3 marks that prevent spreading of heterochromatin and enable proper regulating of nearby genes. Deposition of H3K36me marks depends on nucleosome depletion induced by IBP binding to insulator elements (Lhoumaud et al., 2014). Multiple myeloma, a relatively rare human plasma cell cancer, is characterized by the overexpression of NSD2, a homolog of *mes-4*. Overexpression of NSD2 leads to the accumulation of H3K36me3 over sequences enriched with the canonical CTCF binding

motif and eventual overexpression of flanking oncogenes. Epigenetic therapies targeting NSD2 and H3K36 are now in development (Luccio, 2015; Popovic et al., 2014; Lhoumaud et al., 2019). Similar effects of H3K36me marks have been observed in yeast (Venkatesh et al., 2012).

Inspired by these links between MES-4, insulator elements, and germline expressed genes, I have analyzed the binding regions of MES-4 for insulator-like elements. In this thesis, I link the binding of MES-4 to a class of non-coding DNA prevalent in *C. elegans* that have been shown to prevent silencing of transgenes in the germline by preventing position-effect variegation (Frøkjær-Jensen et al., 2016).

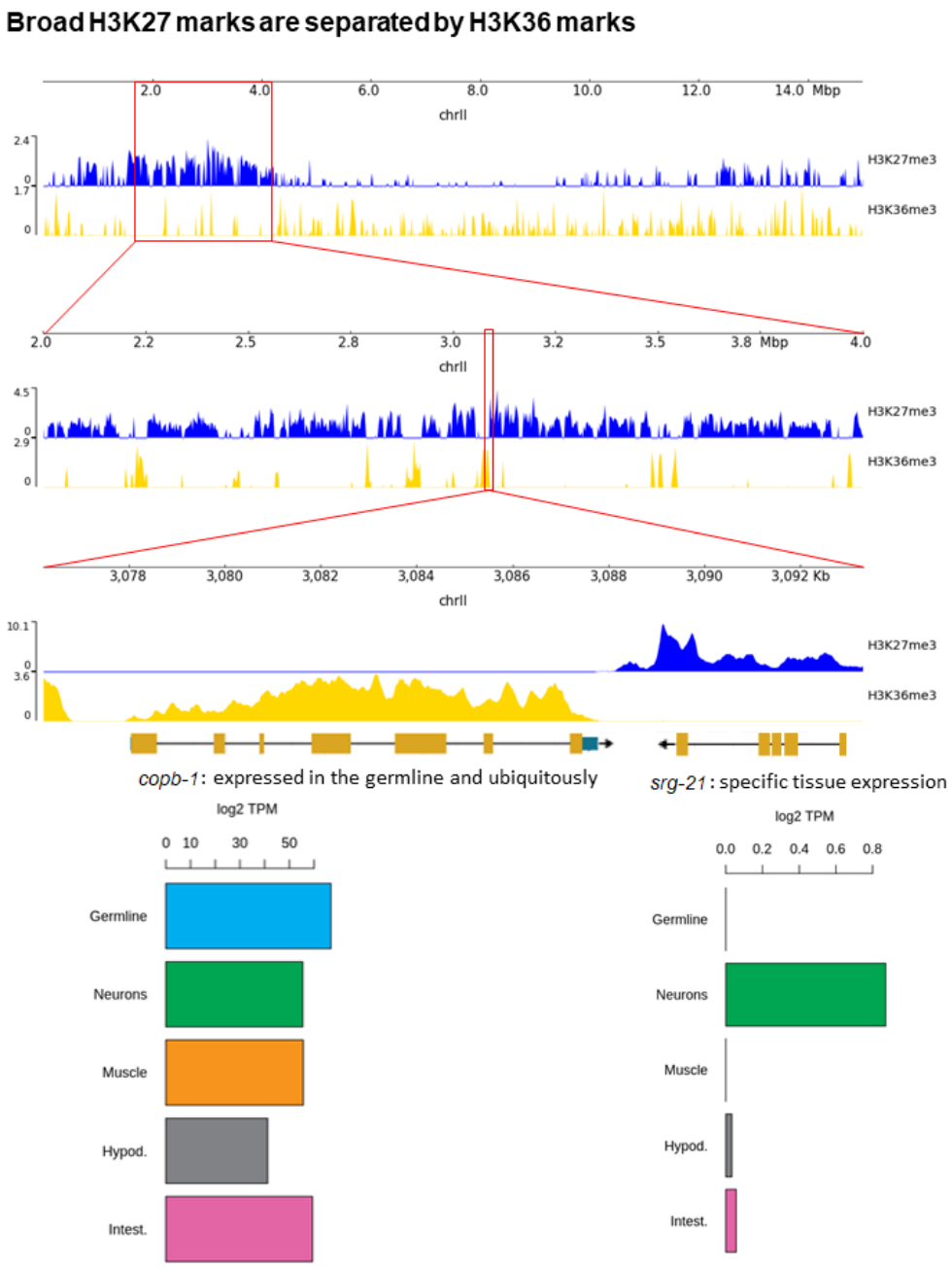

Figure 1.1: H3K27me3 covers the ends of chromosome II

At a chromosomal level, H3K36me3 marks are enriched in the center, and in-between H3K27me3 peaks on the ends. The separation on the ends occurs at the boundaries of ubiquitously expressed genes and tissue-specific genes.

Results

MES-4 binding sites are enriched for novel DNA motif

Recruitment of H3K36 methyltransferases to their target regions depends on transcription in a process known as RNA polymerase II-coupled H3K36 methylation (Jha and Strahl, 2014). However, MES-4 recruitment to genes is at least partially independent of RNA polymerase II and active transcription states (Rechtsteiner et al., 2010). This is in line with the proposed MES-4 function of transmitting an epigenetic memory of genes expressed in the maternal germline to the next generation, even if they were not actively expressed during embryonic development (Rechtsteiner et al., 2010). Although MES-4 possess a plant homeodomain (PHD) predicted to be capable of recognizing H3K36 marks (Laurel et al., 2006), such a mechanism cannot solely explain the recruitment of MES-4 to its target genes. Chromatin marks have to be first added before they are recognized, and ultimately some sequence preference has to guide the modification of histones. PH domains are also known to bind dsDNA with different affinity depending on local nucleosome content and modified chromatin marks (Weaver et al., 2018; Liu et al., 2014).

I obtained early embryo MES-4 ChIP-chip data from the modENCODE Project. Unlike the more commonly used and newer ChIP-seq technology (chromatin immunoprecipitation followed by sequencing), ChIP-chip utilizes binding of immunoprecipitated chromatin to DNA microarrays to infer binding sequences. Using

standard *de novo* motif discovery on binding peaks of MES-4, I identified a highly enriched motif (e-value of 6.3e-039) consisting of three clusters of A nucleotides that are approximately five base pairs apart **(Figure 2.1A)**. Searching JASPAR, CIS-BP, and UniPROBE motif databases using Tomtom motif comparison tool does not identify any match to previously reported motifs, and these findings, therefore, represent a novel motif.

Mapping MES-4 motif occurrences and MES-4 ChIP-chip peaks to gene annotations and tissue-specific expression data shows that the genes bound are most highly expressed in the germline but are also expressed ubiquitously **(Figure2.1B-C)**. There is no standard transcript per million (TPM) threshold for assigning a gene as expressed in a particular tissue. However, if I use an arbitrary threshold of 14, then 45% of all *C. elegans* genes would be assigned as germline expressed compared to 70% and 92% of MES-4 bound genes in ChIP and motif datasets, respectively.

It is important to note that only 25% of genes overlap between the two gene sets, and each group shows a different chromosomal pattern **(Figure 2.2A)**. It has been reported in the literature that sequences retrieved from ChIP experiments are not entirely attributed to valid binding of the target protein to DNA but are biased by different factors such as chromatin accessibility and GC content. Furthermore, binding profiles of the same protein are drastically different between ChIP-chip and ChIP-seq experiments due to underlying systematic and experimental biases, with ChIP-chip being less sensitive and specific (Joshua et al., 2011; Ramachandran et al., 2015). The same

analysis carried out on two transcription factors with highly conserved binding motifs shows that such differences are not specific to MES-4 but may reflect ChIP-chip limitations **(Figure 2.2B)**. Also, technological limitations are not the sole explanation of this recurring phenomenon. Systematic analysis of *Drosophila* transcription factors (TF) binding regions demonstrated that they bind thousands of sites across the genome in a pattern that cannot be attributed to the existence of binding motifs. This adds to the growing evidence that the context of the binding regions might be just as crucial as the binding motifs (Li et al., 2008). Motif discovery algorithms are built with these limitations in mind, and although far from perfect, they can find biologically relevant motifs.

Further complicating the issue for MES-4 is the different pattern of H3K36 marks across the chromosomes. It is possible that the recruitment of MES-4 to the heterochromatic regions of the chromosomal ends is reliant upon the identified motif, but such interaction may not be necessary in the euchromatic central regions.

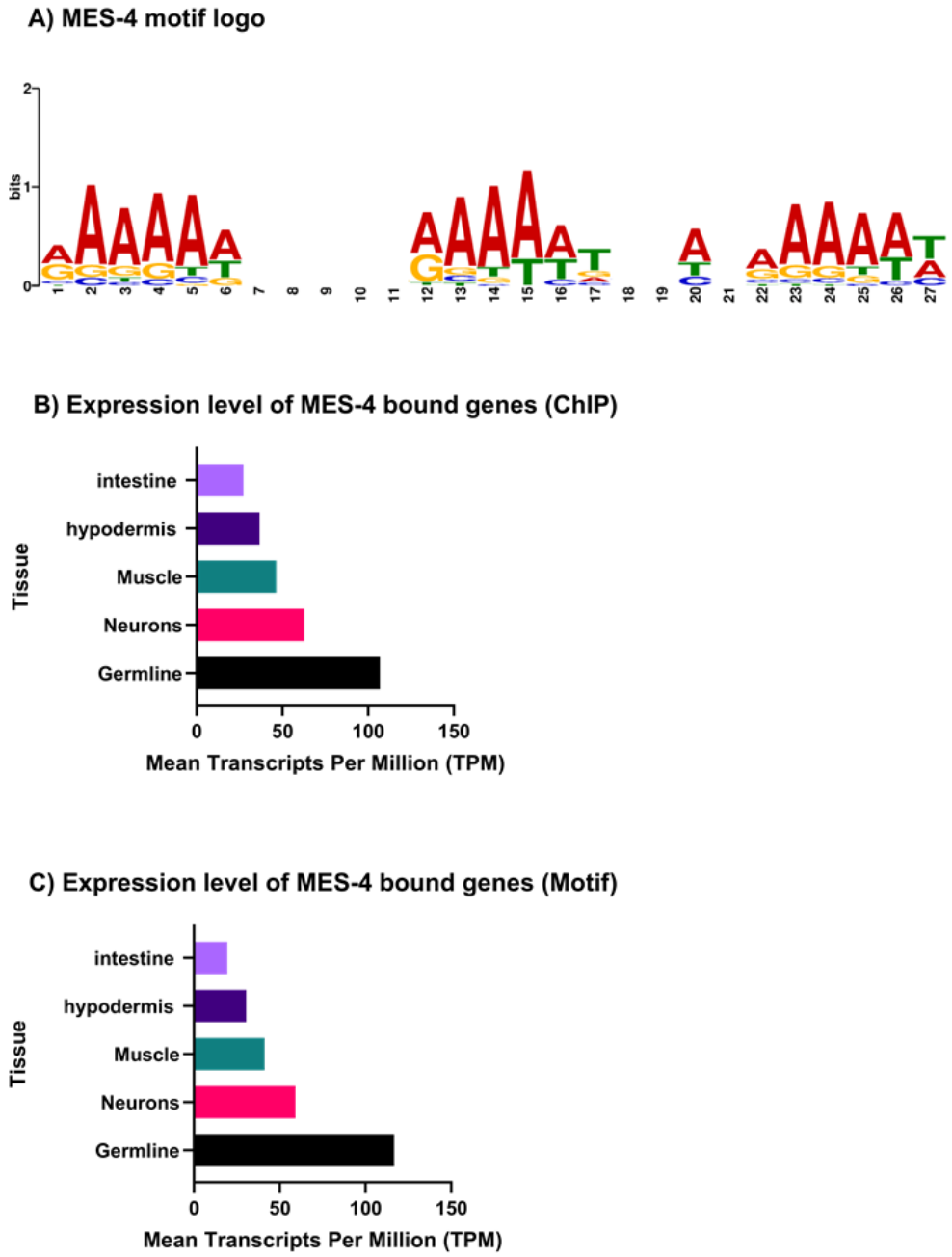

Figure 2.1 MES-4 bound genes are highly expressed in the germline.

A) MES-4 sequence logo. Positions with information content <0.5 have been omitted. Information content is calculated based on Schneider et al. (1986) Shannon's uncertainty method. B-C) We obtained tissue-specific expression data from the *C. elegans* regulatory atlas. Both MES-4 ChIP-chip and motif are enriched on genes that are highly expressed in the germline.

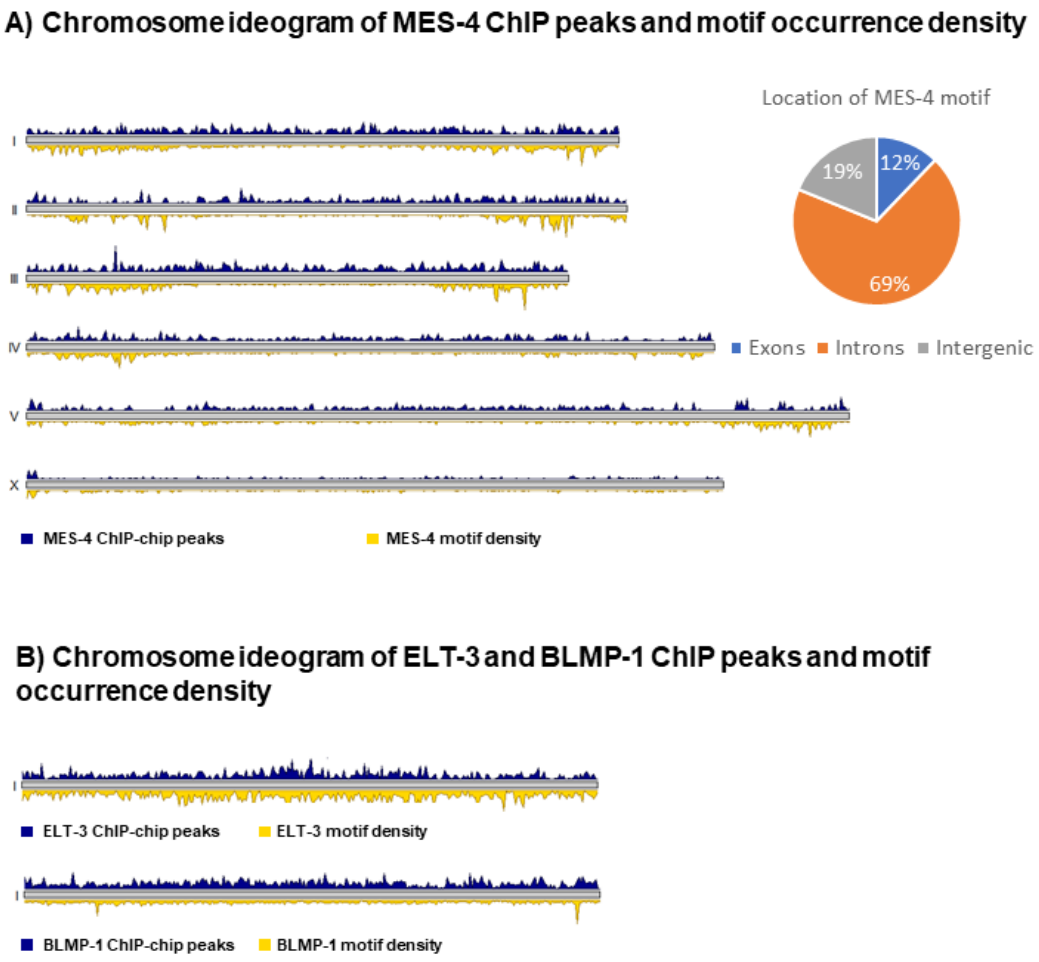

Figure 2.2 MES-4 motif has distinct chromosomal bis and is enriched in introns.

A) In contrast to MES-4 ChIP-ChIP data, MES-4 motif is concentrated on the autosomal ends. The scarcity of signal on the X chromosome is consistent with observed underrepresentation of germline expressed genes on the X chromosome and its global inactivation in the germline (William et al., 2002). Introns represent 26% of *C. elegans* genome and over-representation of the MES-4 motif in introns suggest a regulatory role for intronic sequences. B) ELT-3 and BLMP-1 are both widely expressed and involved in various regulatory processes. The distribution of conserved ELT-3 and BLMP-1 motifs does not mirror the distribution of their ChIP-chip peaks, highlighting the technology's shortcomings. Motif occurrences are based on position weighted matrix scanning with a p-value of 0.0000005 as threshold and accounting for letter frequency in the genome. Density is calculated based on occurrences in a window of 50000 bp.

MES-4 motif represents a DNA bending A-tract

Although commonly reduced into a sequential arrangement of letters, DNA is a three-dimensional molecule with complex geometry. Similar to proteins, the structure of DNA can be described at different levels of complexity. The primary structure refers to the sequential arrangement of letters in the molecule. The arrangements of those letters, in turn, influence the secondary structure of DNA, which can take multiple forms such as helices and stem-loops. The well-known double-helix secondary structure can give rise to numerous forms of DNA tertiary structures, with B-DNA being the most biologically relevant form. B-DNA refers to a right-handed double helix that completes a turn every 10.5 bases and contains a wide major groove and a narrow minor groove. At this level, the geometry of DNA can be very complex with different bending, rolling, and twisting angles, to name a few parameters. This molecular structure is not an inconsequential coincidence. In some of the human genome's non-coding regions, geometry is under a strong evolutionary selection, more so than the primary DNA sequence (Parker et al., 2009). The genetic triplet code elegantly decodes the information stored in the coding regions of the genome, and it appears likely that non-coding regions also encode genetic information, but we are yet to fully understand the rules describing them.

One of the most studied sequence-dependent geometric changes of DNA is that of A-tracts. These tracts are four to six runs of adenine bases occurring periodically at intervals of ten bases following the helical turn of DNA. A-tracts' peculiar property to

bend the DNA double-helix in vitro long hinted at a biological role for such sequences (Diekmann, 1987, VanWye et al., 1991). Recent advances in molecular modeling, bioimaging, and synthetic biology have proven that A-tracts can function as genetic regulatory elements. Through molecular modeling, Dršata et al. (2014) showed that A-tracts are rigid structures that inhibit the formation of nucleosomes locally but enable long rang looping of DNA. Brunet et al. (2015) used single-molecule tethered particle motion to show the intrinsic flexibility of A-tracts and the importance of tract phasing. A direct biological role for A-tracts has been demonstrated by the drastically reduced expression of transgenes after removal of A-tracts from promoter regions by directly interfering with nucleosome formation (Raveh-Sadka et al., 2012).

The *C. elegans* genome has been reported to contain a class of periodic repeats termed periodic A_n/T_n clusters (PATCs) that are highly enriched in germline expressed genes **(Figure 2.3)** (VanWye et al., 1991; Fire et al., 2006). Atomic force microscopy of PATC-rich introns confirms the flexible nature of such sequences and their classification as A-tracts (Moreno-Herrero et al., 2006; Marin-Gonzalez et al., 2020). Frokjaer-Jensen et al. (2016) showed that synthetic and endogenous PATCs-rich introns could license transgene expression in the germline. The molecular mechanism involved in licensing transgene expression with PATCs, however, remains unknown.

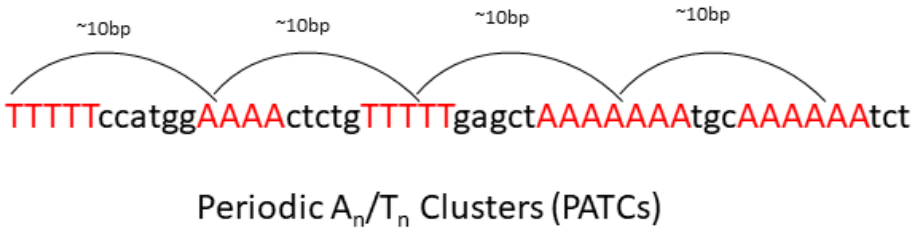

Figure 2.3 Periodic An/Tn Clusters (PATCs)

These periodic clusters cover about 10% of *C. elegans* genome.

Using high-resolution nucleosome maps generated by Valouev et al. (2008) and genome-wide PATCs values, I show that MES-4 motif demarcates regions of low nucleosome occupancy and correlates with the beginning and ending of PATC peaks. *smu-2* introns efficiently license transgene expression (Frokjaer-Jensen et al., 2016) and the association between PATCs and MES-4 motifs is illustrated in **(Figure 2.4A)**. PATC values reflect the degree of perfect periodicity, with higher values associated with longer periodic signals. **(Figure 2.4B)** shows a segment around MES-4 motif the third smu-2 intron. The MES-4 motif is likely a representation of the repeated periodic PATC signal. This is supported by the fact that all MES-4 motifs occur within or partially overlap with PATC peaks. Using a lower threshold of significance to detect a match for MES-4s motif significantly increases the overlap with PATCs and correspond to a less stringent periodicity **(Figure 2.5)**.

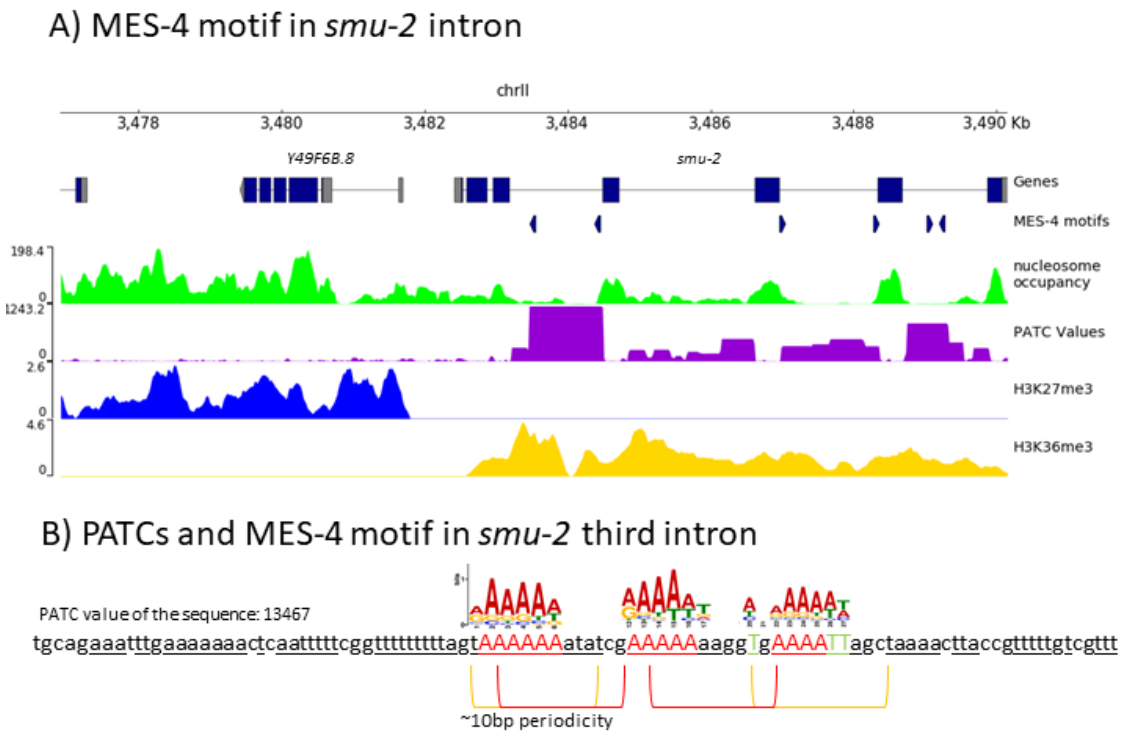

Figure 2.4 MES-4 motif demarcates regions of low nucleosome occupancy.

A) *smu-2* is a germline expressed gene located at the left end of chromosome II. Introns with MES-4 motifs have lower nucleosome occupancy score. B) A segment of the *smu-2* intron. Colored capital letters match the MES-4 motif, and underlined letters represent the periodic PATC signal.

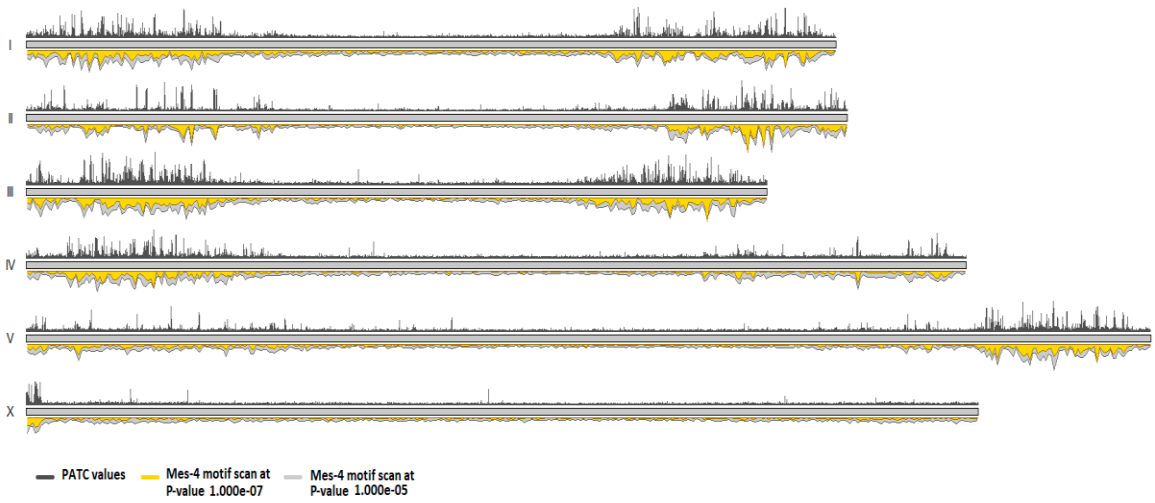

Figure 2.5 MES-4 is highly correlated with PATC density at a chromosomal level. PATC values for every 1bp obtained by the PATC algorithm are compared to the density of MES-4 motifs.

EFL-1 and LIN-53 potentially interact with MES-4

Differential gene expression is achieved, in part, by a physical combinatorial binding of diverse transcription factors and intermediate bridging proteins. The requirement for physical protein-protein interactions requirement often constrains the binding motifs of two interacting proteins to occur at a conserved distance. Whitington et al. (2011) and Felicia et al. (2014) showed that an analysis of an enriched spacing pattern could provide clues on interacting proteins. A similar analysis of MES-4 motif using a spaced motif analysis (SpaMo) algorithm identified EFL-1 as a potential interacting protein. The occurrence pattern of EFL-1 is biased toward the chromosomal ends and is highly enriched in introns, similar to MES-4 motif **(Figure 2.6)**. EFL-1 is a

transcription factor that has been shown to act in both the soma and the germline. In the soma, EFL-1 interacts with LIN-35, as part of the synMuv B pathway, to establish a specific pattern of repressive chromatin in a process that directly antagonizes MES-4. synMuv B mutants are characterized by misexpression of germline-specific genes in somatic cells in a MES-4 dependent manner (Wu et al., 2012). However, during oogenesis and early embryogenesis, EFL-1 is responsible for initiating a germline-specific transcriptional program (Chi and Reinke, 2006).

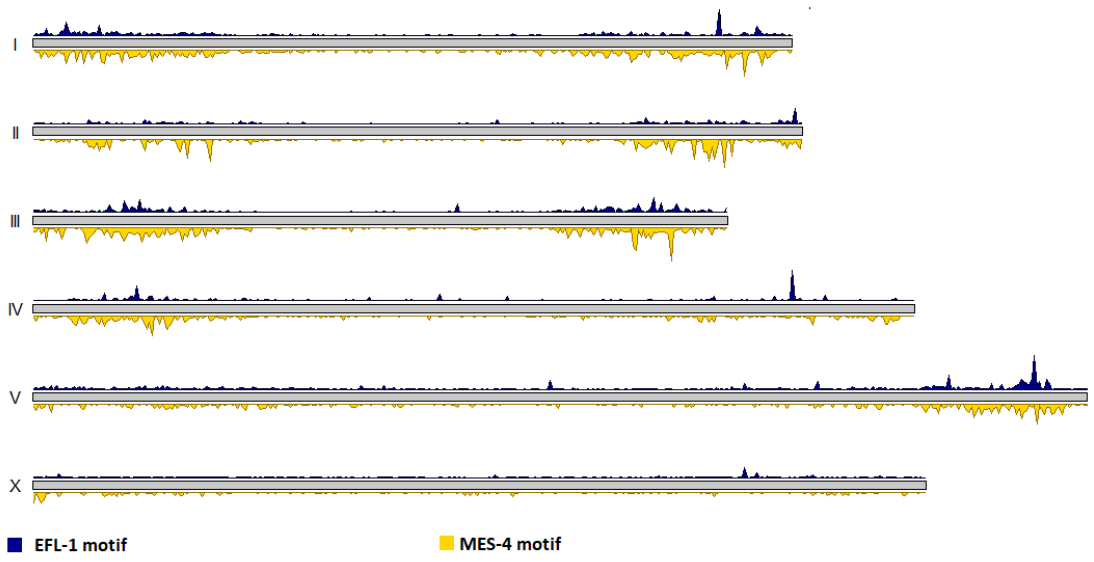

Figure 2.6 EFL-1 and MES-4 motif follow a similar chromosomal bias to MES-4.

EFL-1 motif is absent from the X chromosome and autosomal centers and such chromosomal bias hints toward a common link between EFL-1 and MES-4

Discussion and Future Directions

All identified insulator elements act by enabling the binding of effector proteins to either specific DNA sequences or chromatin marks (Heger et al., 2011). PATCs are an obvious candidate for an insulator. They protect from position-effect variegation, are present between repressive and active chromatin boundaries, and display the nucleosome profile of insulator elements. However, no known protein (except DATIN in yeast) has been shown to bind A-tracts. Whether the insulating properties of PATCs depend on an intrinsic physical property, a binding affinity for MES-4 or an unidentified protein remains a mystery. I have designed two experiments to further our understanding of the association between PATCs and *mes-4* that I plan to carry out as part of my Ph.D. in the Laboratory of Synthetic Genome Biology under Professor Frøkjær-Jensen mentorship:

Can MES-4 motifs protect transgenes from silencing in the germline?

Most TF binding motifs are less than ten bp long due to the steric constraints imposed by the width of B-DNA grooves. This makes it unlikely that the MES-4 motif I identified represent a real DNA binding motif. However, A-tracts are known to affect the structure of B-DNA in ways that make it easier for protein to bind DNA. To test the effect of MES-4 motifs, I have designed a green fluorescent protein (GFP) construct with golden gate sites (Engler et al., 2008) that allow the insertion of different introns **(Figure 3.1)**. I designed two sets of introns:

- Endogenous introns enriched with MES-4 motifs and with MES-4 motif removed

- Synthetic introns with MES-4 motifs in a variety of contexts: Low GC content, 0.5 GC content, and artificially designed A-tracts

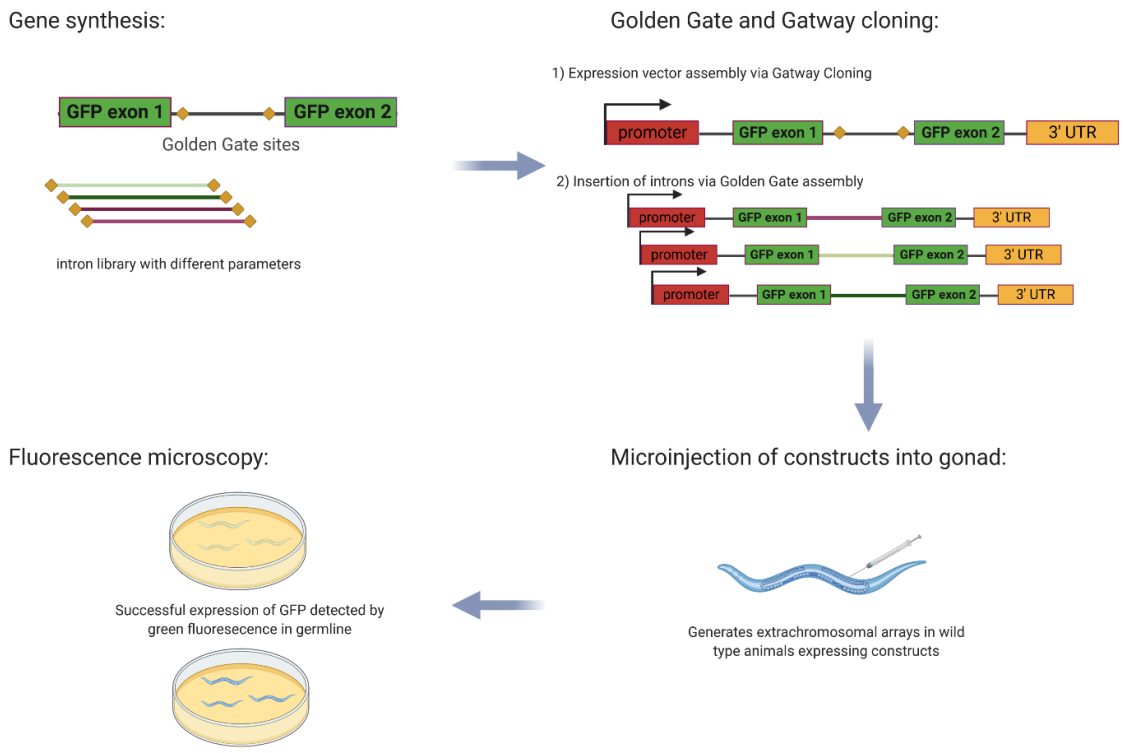

Figure 3.1 Assay to test the effect of different introns

Using gene synthesis, I designed a DNA construct with an intron that contains golden gate sites. These sites can be used to introduce different introns. GFP can be used as a visual marker to detect expression.

Can overexpression of MES-4 lead to spreading of euchromatin?

In multiple myeloma, a chromosomal translocation leads to over-expression of NSD2. It has been shown that spreading of H3K36 marks over CTCF sites and subsequent activation of flanking genes that are usually marked by H3K27 is a major pathway in this progression of the disease (Popovic et al., 2014). I have designed a GFP construct to test if such phenomena can be observed in *C. elegans* and if PATCs/MES-4 motif can affect the degree of spreading **(Figure 3.2)**. The genomic locus shown in **(Figure 1.1)** can also be utilized to test for spreading by tagging *srg-21* with a fluorophore.

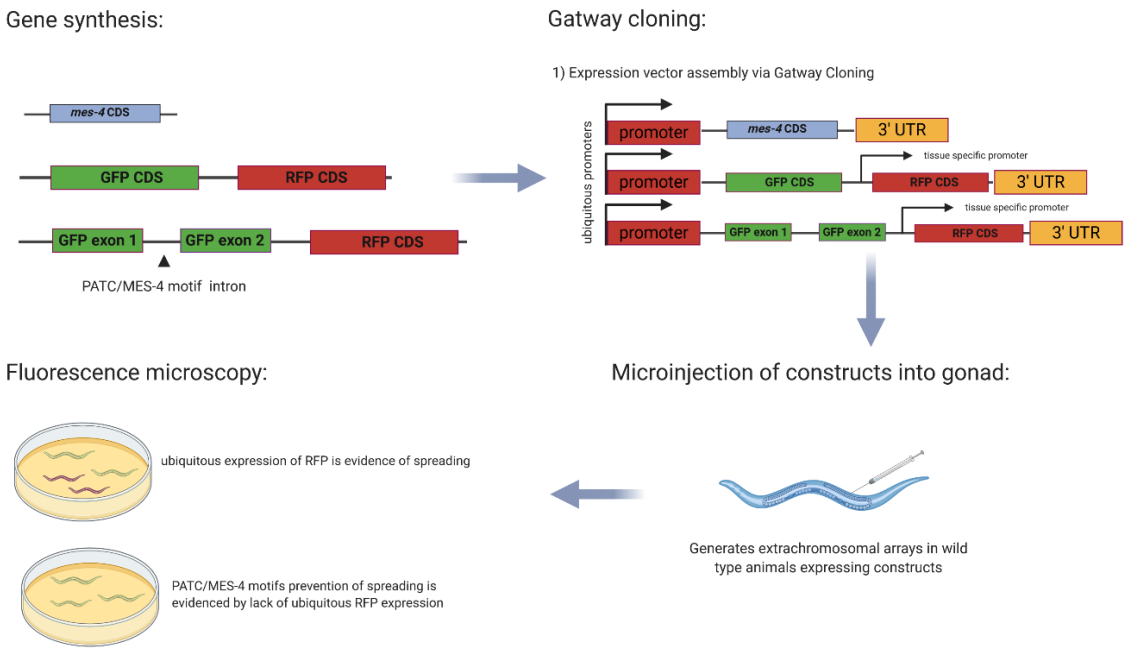

Figure 3.2 Assay to test the effect of mes-4 overexpression

In human cells, overexpression of *mes-4* homolog NSD2 leads to the spreading of "activating" H3K36 marks and subsequent upregulation of gene expression of regions flanked by boundary elements. Using gene synthesis, we can synthesize *mes-4* coding sequence and clone it into an expression vector with a strong promoter. The spreading of H3K36 marks can be visually detected by overexpression or misexpression of red fluorescent protein (RFP).

Methods

De-novo motif discovery. Multiple Em for Motif Elicitation (MEME) program version 5.1.1 (available online at http://meme-suite.org) was used to discover DNA sequence motifs in MES-4 binding sites. The MES-4 chip-chip data was obtained from the modENCODE Project (http://modencode.org). The sequences corresponding to MES-4 binding sites were extracted using bedtools program and the c. elegans genome build (WBcel235) (downloaded from http://ensembl.org).

Position-weight-matrix scan. PWMScan program (available at https://ccg.epfl.ch/pwmscan) was used to scan the c. elegans genome (build WBcel235) using custom position frequency matrix obtained from MEME. Intronic regions coordinates were extracted with the UCSC Table Browser tool (available at http://genome.ucsc.edu) and the corresponding sequences were extracted with bedtools as described above. The intronic sequences were also scanned with PWMScan in order to quantify the occurrences of the motifs in intronic regions.

Gene expression data. Genes expression data was obtained from the Ahringer lab C. elegans regulatory atlas (available at http://ahringerlab.com/RegAtlas). Transcripts Per Million (TPM) values were used to classify genes as germline expressed. The gene set analysis tool was used to obtain the gene expression data and values were further analyzed and visualized using a custom R code.

Modified chromatin marks data. The genomic signals for H3K27me3 and H3K36me3 were obtained from the modENCODE Project (http://modencode.org) and signal peaks were called using Model-based Analysis for ChIP-Seq (MACS) program version 2.2.7 (available form https://github.com/macs3-project/MACS). Correlation between the different chromatin marks and genomic regions was calculated with fisher's exact test using bedtools program.

Nucleosome occupancy data. Data of nucleosome occupancy are available from Nucleosome Explorer++ (http://nucleosome.rutgers.edu). Wig files of predicted nucleosome occupancy were converted to binary BigWig data files and displayed on JBrowse version 1.12.5.

BIBLIOGRAPHY

Marin-Gonzalez, A., Pastrana, C. L., Bocanegra, R., Martín-González, A., Vilhena, J. G., Pérez, R., Ibarra, B., Aicart-Ramos, C., & Moreno-Herrero, F. (2020). Understanding the paradoxical mechanical response of in-phase A-tracts at different force regimes. Nucleic Acids Research, 48(9), 5024–5036. https://doi.org/10.1093/nar/gkaa225

Lhoumaud, P., Badri, S., Rodriguez-Hernaez, J., Sakellaropoulos, T., Sethia, G., Kloetgen, A., Cornwell, M. I., Bhattacharyya, S., Ay, F., Bonneau, R., Tsirigos, A., & Skok, J. A. (2019). NSD2 overexpression drives clustered chromatin and transcriptional changes in a subset of insulated domains. Nature Communications, 10(1), 1–18. https://doi.org/10.1038/s41467-019-12811-4

Ahringer, J., & Gasser, S. M. (2018). Repressive chromatin in caenorhabditis elegans: Establishment, composition, and function. Genetics, 208(2), 491–511. https://doi.org/10.1534/genetics.117.300386

Kreher, J., Takasaki, T., Cockrum, C., Sidoli, S., Garcia, B. A., Jensen, O. N., & Strome, S. (2018). Distinct roles of two histone methyltransferases in transmitting h3k36me3-based epigenetic memory across generations in caenorhabditis elegans. Genetics, 210(3), 969–982. https://doi.org/10.1534/genetics.118.301353

Weaver, T. M., Morrison, E. A., & Musselman, C. A. (2018). Reading more than Histones: The prevalence of nucleic acid binding among reader domains. In Molecules (Vol. 23, Issue 10). MDPI AG. https://doi.org/10.3390/molecules23102614

Huang, C., & Zhu, B. (2018). Roles of H3K36-specific histone methyltransferases in transcription: antagonizing silencing and safeguarding transcription fidelity. Biophysics Reports, 4(4), 170–177. https://doi.org/10.1007/s41048-018-0063-1

Gates, L. A., Foulds, C. E., & O'Malley, B. W. (2017). Histone Marks in the 'Driver's Seat': Functional Roles in Steering the Transcription Cycle. In Trends in Biochemical Sciences (Vol. 42, Issue 12, pp. 977–989). Elsevier Ltd. https://doi.org/10.1016/j.tibs.2017.10.004

Frøkjær-Jensen, C., Jain, N., Hansen, L., Davis, M. W., Li, Y., Zhao, D., Rebora, K., Millet, J. R. R. M., Liu, X., Kim, S. K., Dupuy, D., Jorgensen, E. M., & Fire, A. Z. (2016). An Abundant Class of Non-coding DNA Can Prevent Stochastic Gene Silencing in the C. elegans Germline. Cell, 166(2), 343–357. https://doi.org/10.1016/j.cell.2016.05.072

Luccio, E. di. (2015). Inhibition of Nuclear Receptor Binding SET Domain 2/ Multiple Myeloma SET Domain by LEM-06 Implication for Epigenetic Cancer Therapies. Journal of Cancer Prevention, 20(2), 113–120. https://doi.org/10.15430/jcp.2015.20.2.113

Brunet, A., Chevalier, S., Destainville, N., Manghi, M., Rousseau, P., Salhi, M., Salome, L., & Tardin, C. (2015). Probing a label-free local bend in DNA by single molecule tethered particle motion. Nucleic Acids Research, 43(11), e72. https://doi.org/10.1093/nar/gkv201

Ramachandran, P., Palidwor, G. A., & Perkins, T. J. (2015). BIDCHIPS: Bias decomposition and removal from ChIP-seq data clarifies true binding signal and its functional correlates Medicine. Epigenetics and Chromatin, 8(1), 33. https://doi.org/10.1186/s13072-015-0028-2

Popovic, R., Martinez-Garcia, E., Giannopoulou, E. G., Zhang, Q., Zhang, Q., Ezponda, T., Shah, M. Y., Zheng, Y., Will, C. M., Small, E. C., Hua, Y., Bulic, M., Jiang, Y., Carrara, M., Calogero, R. A., Kath, W. L., Kelleher, N. L., Wang, J. P., Elemento, O., & Licht, J. D. (2014). Histone Methyltransferase MMSET/NSD2 Alters EZH2 Binding and Reprograms the Myeloma Epigenome through Global and Focal Changes in H3K36 and H3K27 Methylation. PLoS Genetics, 10(9). https://doi.org/10.1371/journal.pgen.1004566

Popovic, R., Martinez-Garcia, E., Giannopoulou, E. G., Zhang, Q., Zhang, Q., Ezponda, T., Shah, M. Y., Zheng, Y., Will, C. M., Small, E. C., Hua, Y., Bulic, M., Jiang, Y., Carrara, M., Calogero, R. A., Kath, W. L., Kelleher, N. L., Wang, J. P., Elemento, O., & Licht, J. D.

(2014). Histone Methyltransferase MMSET/NSD2 Alters EZH2 Binding and Reprograms the Myeloma Epigenome through Global and Focal Changes in H3K36 and H3K27 Methylation. PLoS Genetics, 10(9). https://doi.org/10.1371/journal.pgen.1004566

Dršata, T., Špačková, N., Jurečka, P., Zgarbová, M., Šponer, J., & Lankaš, F. (2014). Mechanical properties of symmetric and asymmetric DNA A-tracts: Implications for looping and nucleosome positioning. Nucleic Acids Research, 42(11), 7383–7394. https://doi.org/10.1093/nar/gku338

Liu, Z., Li, F., Ruan, K., Zhang, J., Mei, Y., Wu, J., & Shi, Y. (2014). Structural and functional insights into the human börjeson-forssman- lehmann syndrome-associated protein PHF6. Journal of Biological Chemistry, 289(14), 10069–10083. https://doi.org/10.1074/jbc.M113.535351

Perez, P. J., Clauvelin, N., Grosner, M. A., Colasanti, A. V., & Olson, W. K. (2014). What controls DNA looping? International Journal of Molecular Sciences, 15(9), 15090–15108. https://doi.org/10.3390/ijms150915090

Ng, F. S. L., Schütte, J., Ruau, D., Diamanti, E., Hannah, R., Kinston, S. J., & Göttgens, B. (2014). Constrained transcription factor spacing is prevalent and important for transcriptional control of mouse blood cells. Nucleic Acids Research, 42(22), 13513–13524. https://doi.org/10.1093/nar/gku1254

Jha, D. K., & Strahl, B. D. (2014). An RNA polymerase II-coupled function for histone H3K36 methylation in checkpoint activation and DSB repair. Nature Communications, 5(1), 1–13. https://doi.org/10.1038/ncomms4965

Lhoumaud, P., Hennion, M., Gamot, A., Cuddapah, S., Queille, S., Liang, J., Micas, G., Morillon, P., Urbach, S., Bouchez, O., Severac, D., Emberly, E., Zhao, K., & Cuvier, O. (2014). Insulators recruit histone methyltransferase d M es4 to regulate chromatin of flanking genes. The EMBO Journal, 33(14), 1599–1613. https://doi.org/10.15252/embj.201385965

Elgin, S. C. R., & Reuter, G. (2013). Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. Cold Spring Harbor Perspectives in Biology, 5(8). https://doi.org/10.1101/cshperspect.a017780

Wu, X., Shi, Z., Cui, M., Han, M., & Ruvkun, G. (2012). Repression of Germline RNAi Pathways in Somatic Cells by Retinoblastoma Pathway Chromatin Complexes. PLoS Genetics, 8(3), e1002542. https://doi.org/10.1371/journal.pgen.1002542

Musselman, C. A., Lalonde, M. E., Côté, J., & Kutateladze, T. G. (2012). Perceiving the epigenetic landscape through histone readers. In Nature Structural and Molecular Biology (Vol. 19, Issue 12, pp. 1218–1227). NIH Public Access. https://doi.org/10.1038/nsmb.2436

Venkatesh, S., Smolle, M., Li, H., Gogol, M. M., Saint, M., Kumar, S., Natarajan, K., & Workman, J. L. (2012). Set2 methylation of histone H3 lysine 36 suppresses histone

exchange on transcribed genes. Nature, 489(7416), 452–455. https://doi.org/10.1038/nature11326

Raveh-Sadka, T., Levo, M., Shabi, U., Shany, B., Keren, L., Lotan-Pompan, M., Zeevi, D., Sharon, E., Weinberger, A., & Segal, E. (2012). Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. Nature Genetics, 44(7), 743–750. https://doi.org/10.1038/ng.2305

Parker, S. C. J., & Tullius, T. D. (2011). DNA shape, genetic codes, and evolution. In Current Opinion in Structural Biology (Vol. 21, Issue 3, pp. 342–347). NIH Public Access. https://doi.org/10.1016/j.sbi.2011.03.002

Yun, M., Wu, J., Workman, J. L., & Li, B. (2011). Readers of histone modifications. In Cell Research (Vol. 21, Issue 4, pp. 564–578). Nature Publishing Group. https://doi.org/10.1038/cr.2011.42

Ho, J. W. K., Bishop, E., Karchenko, P. V., Nègre, N., White, K. P., & Park, P. J. (2011). ChIP-chip versus ChIP-seq: Lessons for experimental design and data analysis. BMC Genomics, 12. https://doi.org/10.1186/1471-2164-12-134

Whitington, T., Frith, M. C., Johnson, J., & Bailey, T. L. (2011). Inferring transcription factor complexes from ChIP-seq data. Nucleic Acids Research, 39(15), e98–e98. https://doi.org/10.1093/nar/gkr341

Rohs, R., Jin, X., West, S. M., Joshi, R., Honig, B., & Mann, R. S. (2010). Origins of specificity in protein-DNA recognition. In Annual Review of Biochemistry (Vol. 79, pp. 233–269). NIH Public Access. https://doi.org/10.1146/annurev-biochem-060408-091030

Rechtsteiner, A., Ercan, S., Takasaki, T., Phippen, T. M., Egelhofer, T. A., Wang, W., Kimura, H., Lieb, J. D., & Strome, S. (2010). The Histone H3K36 Methyltransferase MES-4 Acts Epigenetically to Transmit the Memory of Germline Gene Expression to Progeny. PLoS Genetics, 6(9), e1001091. https://doi.org/10.1371/journal.pgen.1001091

Parker, S. C. J., Hansen, L., Abaan, H. O., Tullius, T. D., & Margulies, E. H. (2009). Local DNA topography correlates with functional noncoding regions of the human genome. Science, 324(5925), 389–392. https://doi.org/10.1126/science.1169050

Heger, P., Marin, B., & Schierenberg, E. (2009). Loss of the insulator protein CTCF during nematode evolution. BMC Molecular Biology, 10, 84. https://doi.org/10.1186/1471-2199-10-84

Engler, C., Kandzia, R., & Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. PLoS ONE, 3(11). https://doi.org/10.1371/journal.pone.0003647

Li, X. Y., MacArthur, S., Bourgon, R., Nix, D., Pollard, D. A., Iyer, V. N., Hechmer, A., Simirenko, L., Stapleton, M., Luengo Hendriks, C. L., Hou, C. C., Ogawa, N., Inwood, W., Sementchenko, V., Beaton, A., Weiszmann, R., Celniker, S. E., Knowles, D. W., Gingeras, T., … Biggin, M. D. (2008). Transcription factors bind thousands of active and inactive

regions in the Drosophila blastoderm. PLoS Biology, 6(2), 0365–0388. https://doi.org/10.1371/journal.pbio.0060027

Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., Zeng, K., Malek, J. A., Costa, G., McKernan, K., Sidow, A., Fire, A., & Johnson, S. M. (2008). A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Research, 18(7), 1051–1063. https://doi.org/10.1101/gr.076463.108

Talbert, P. B., & Henikoff, S. (2006). Spreading of silent chromatin: Inaction at a distance. In Nature Reviews Genetics (Vol. 7, Issue 10, pp. 793–803). Nature Publishing Group. https://doi.org/10.1038/nrg1920

Cui, M., Kim, E. B., & Han, M. (2006). Diverse Chromatin Remodeling Genes Antagonize the Rb-Involved SynMuv Pathways in C. elegans. PLoS Genetics, 2(5), e74. https://doi.org/10.1371/journal.pgen.0020074

Moreno-Herrero, F., Seidel, R., Johnson, S. M., Fire, A., & Dekker, N. H. (2006). Structural analysis of hyperperiodic DNA from Caenorhabditis elegans. Nucleic Acids Research, 34(10), 3057–3066. https://doi.org/10.1093/nar/gkl397

Bender, L. B., Suh, J., Caroll, C. R., Fong, Y., Fingerman, I. M., Briggs, S. D., Cao, R., Zhang, Y., Reinke, V., & Strome, S. (2006). MES-4: An autosome-associated histone methyltransferase that participates in silencing the X chromosomes in the C. elegans germ line. Development, 133(19), 3907–3917. https://doi.org/10.1242/dev.02584

Fire, A., Alcazar, R., & Tan, F. (2006). Unusual DNA structures associated with germline genetic activity in Caenorhabditis elegans. Genetics, 173(3), 1259–1273. https://doi.org/10.1534/genetics.106.057364

Kim, J. K., Gabel, H. W., Kamath, R. S., Tewari, M., Pasquinelli, A., Rual, J. F., Kennedy, S., Dybbs, M., Bertin, N., Kaplan, J. M., Vidal, M., & Ruvkun, G. (2005). Functional genomic analysis of RNA interference in C. elegans. Science, 308(5725), 1164–1167. https://doi.org/10.1126/science.1109267

Reinke, V., Gil, I. S., Ward, S., & Kazmer, K. (2004). Genome-wide germline-enriched and sex-biased expression profiles in Caenorhabditis elegans. Development, 131(2), 311–323. https://doi.org/10.1242/dev.00914

Kelly, W. G., Schaner, C. E., Dernburg, A. F., Lee, M. H., Kim, S. K., Villeneuve, A. M., & Reinke, V. (2002). X-chromosome silencing in the germline of C. elegans. In Development (Vol. 129, Issue 2, pp. 479–492). NIH Public Access. /pmc/articles/PMC4066729/?report=abstract

Dudley, N. R., Labbé, J. C., & Goldstein, B. (2002). Using RNA interference to identify genes required for RNA interference. Proceedings of the National Academy of Sciences of the United States of America, 99(7), 4191–4196. https://doi.org/10.1073/pnas.062605199

Fong, Y., Bender, L., Wang, W., & Strome, S. (2002). Regulation of the different chromatin states of autosomes and X chromosomes in the germ line of C. elegans. Science, 296(5576), 2235–2238. https://doi.org/10.1126/science.1070790

McKinney, K., & Prives, C. (2002). Efficient Specific DNA Binding by p53 Requires both Its Central and C-Terminal Domains as Revealed by Studies with High-Mobility Group 1 Protein. Molecular and Cellular Biology, 22(19), 6797–6808. https://doi.org/10.1128/mcb.22.19.6797-6808.2002

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). DNA-Binding Motifs in Gene Regulatory Proteins. https://www.ncbi.nlm.nih.gov/books/NBK26806/

Vanwye, J. D., Bronson, E. C., & Anderson, J. N. (1991). Species-specific patterns of DNA bending and sequence. Nucleic Acids Research, 19(19), 5253–5261. https://doi.org/10.1093/nar/19.19.5253

Kemphues, K. J., Priess, J. R., Morton, D. G., & Cheng, N. (1988). Identification of genes required for cytoplasmic localization in early C. elegans embryos. Cell, 52(3), 311–320. https://doi.org/10.1016/S0092-8674(88)80024-2

Diekmann, S. (1987). Temperature and salt dependence of the gel migration anomaly of curved DNA fragments. Nucleic Acids Research, 15(1), 247–265. https://doi.org/10.1093/nar/15.1.247