Zelda's Role in Controlling Gene Expression in the Early Embryo

Hao Wang

Abstract:

Much research investigates the influence of Zelda (Zinc-finger early *Drosophila* Embryo activator, Zld) on early embryo development. CAGGTAG and related sequences are common in the early transcribed genes in *Drosophila*. These are all referred to as sites that can activate transcription and are responsible for maternal RNA degradation (as cited in Liang et al., 2008). However, it is still unknown whether other activators can help gene expression. Here, we analyze the effect of Zld on specific genes, such as Halo, by chromatin immunoprecipitation assays. One result is that the gene expression is low in the absence of Zld, which could suggest that Zld plays an important role in the activation of early development of *the Drosophila* embryo.

Keywords: Zinc-finger protein, TAGsites, embryos.

1. Introduction

Many organisms are driven entirely by maternal protein and mRNA deposits during their earliest stages of development, especially the model animal Drosophila. However, in the hours after fertilization, a period known as the mother-to-zygote transition, extensive transcriptional activation of the zygotic genome occurs in organisms. During this transition, Zelda (Zld) plays a key role in transcriptional activation. To understand the function of zld in early embryonic development of Drosophila melanogaster, we use CHIP-chip to observe the tendency of zld binding and transcription of gene Halo in different periods of embryo development or circumstances without zld. Halo is a gene that encodes a protein that acts as a cofactor of the molecular motor kinesin-1 and controls the travel distances of moving lipid droplets. According to a study by Nien et al. [1], they performed Gene Ontolology on the genes associated with Zld-bound sites. They observed that the defects in the gap and pair-rule patterns in embryos lacking maternal expression of zld (hereafter referred to as zld^-) are due to delayed and mis-localized gap repressor activity. They proved the ability of timing of zld and pointed out zld binding to the gap and pair-rule enhancers to activate their expression. However, in this paper, we are more focused on the relationship between sites with the goal of understanding whether the interaction is superimposed, mutual weakening or some sites do not activate gene expression. To explore the relationship between the position of the zld binding site and the start transcription site (TSS), we compared the expression profiles of wild-type and zld^- embryos. In addition, we also used Snapgene to analyze the precise position of sites with the goal of proving that sites are near TSS (The data is from Shelby A. Blythe and Eric F. Wieshcaus). Finally, the strange sites, some of which might be present in 5' untranslated regions or far away from TSS, make people wonder. We found that past theories and ideas can not easily explain it. Therefore, we propose several hypotheses and design experiments to unlock its secret.

2. Results

2.1 Expression of Halo in early embryo

By using CRISPR-cas9 and homology direct repair techniques, we induce clones of mutational zld cells in the embryo during mitosis, which can eliminate maternal zld from the embryo. Embryos of Figures A and B are regular wild type in nucleus cycle 14 (NC 14), and both embryos in Figures C and D are null of zld (zld minus, zld^-). In the embryos in Figures A and B, Halo is expressed ubiquitously in early embryos. Embryos in Figures C and D do not show too much color. Therefore, it exposes that the halo gene does not express in 1-3 hr. It can be explained as they do not have any old transcripts. As a result, without activation of zld in early embryos, the Halo gene cannot be expressed. However, through those pictures, it still cannot reveal other possible regulators that could also control the gene expression in early *drosophila* embryos.

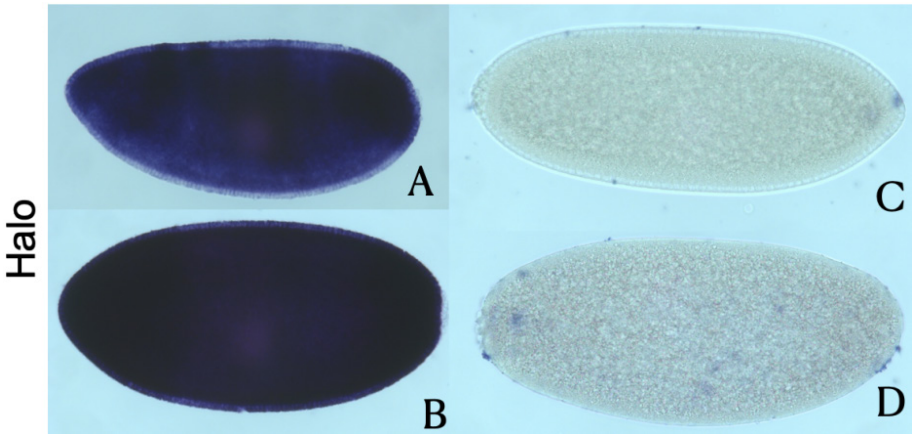

Figure 1. Zld is necessary for the expression of Halo (The picture is from Christine's lab). A-D, Wild-type (A-B), and zld$^-$ (C-D). The shade represents the amount of Halo expressed.

2.2 Zld binds to TAGsites in front of the transcription start site to activate and enhance transcription

To further explore the relationship between the position of the zld binding site relative to the transcription start site(TSS) and its effect on gene expression, we use the method of chromatin immunoprecipitation (see in Method) that indicates the sites are always before TSS and zld can bind with them as an activator. Figure 2 shows browser views of Halo and gene CG18123 region. In the old chip (track 4, figure 2), there is one apparent peak just before the TSS of Halo. This peak can help us identify which sites must be. The polymerase chip (track 1, Figure) also indicates no transcription in zld$^-$; as a result, no zld could bind to sites before genes. After searching the sequence CAGGTAG and other sites, we find that there are two sites upstream of the Halo gene. After we zoom in to examine the distance between the position of the RNA polymerase binding peak (in tracks 2 and 3) and the sites, it has been shown that the increase of gene transcription is just behind the sites. According to a study by ten Bosch et al., as cited in Christine Rushlow et al., the CAGGTAG was the most enriched site (9.5 fold enriched), but the other seven sites are not as strong as CAGGTAG, such as CAGGTAa (4.5 fold), tAGGTAG (3 fold). Therefore, we propose that the closer the sites are to TSS, the more activation they can give to the specific gene. In the upstream, we find some other sites; thanks to we want to observe the highest peak of profiles, the other small peaks caused by that less enriched site are not obvious. However, it is unclear what their meaning of existing is since there is already CAGGTAG, which can enhance transcription to a great extent. In addition, we found that there is a CAGGTAG in the middle of the gene, CG18132; it makes us wonder whether the activation in the previous gene would affect the next one; if it does, we are curious about whether the activation of zld is related to distance (see in Discussion).

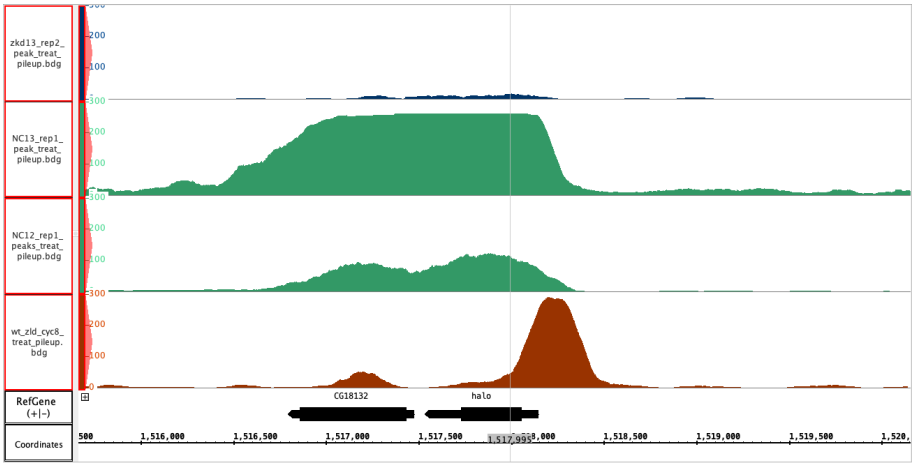

Figure 2. Gene expression is regulated by zld. RNA Polymerase chip for zld- (Track 1), RNA Polymerase chip for wild-type (Track 2 and Track 3). Zld chip (Track 4).

2.3 Knocking out two TAGsites differentially to observe their influence on gene expression

To investigate the relationship between the two TAGsites, we design three experiments to help determine their property of them. By looking at the position of TAGsites in Figure 3, we found that the two TAGsites appear to only serve Halo because there is no other gene close to them, and we attempt to examine the separate effect. Consequently, we designed three experiments to observe gene expression under different site activations. In the experiment, CRISPR-cas9 cutting and HDR will be included (see Method). After we examined the profiles from Flybase, Snapgene, and Integrated Genome Browser (data from Nien et al. and Shelby A.Blythe, Eric F. Wieshcaus), we designed TAGsite mutation which can eliminate the factor of zld. The third group is a negative control group, and in this group, it is necessary to knock out both sites to exclude their ability to activate. In the first and second groups, we plan to knock out the sites upstream to see their single effect. Even though we have not carried out this experiment, we predict that the activation of zld could be passed down (see Discussion). By examining Figure 2, we observed the zld binding peak before Halo is extremely higher than in gene CG18132, and the TAGsite of CG18132 is located in the coding region rather than before the gene. We found that the polymerase binding peak of CG18132 is similar to that of Halo, so the prediction is that TAGsites could be set in between the gene, and the activation can be passed down if the previous TAGsites are close to the gene (see in Discussion).

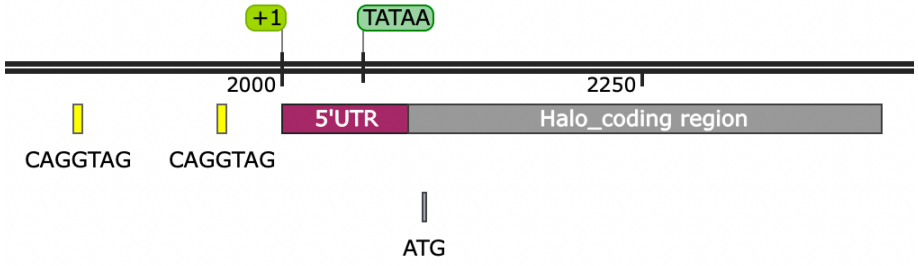

Figure 3. The specific region of different DNA sequences in gene Halo.

+1 is the beginning of the Halo gene. The length of the 5' untranslated region is 87bp after +1. The length of the halo coding region is 330bp. After searching for sites, there are two CAGGTAG before and close to TSS. The graph comes from the snap gene, and by looking at the nucleotides view, we determine the position of UTR and coding region.

3. Discussion

All of those data, zld-chip from Nien et al., embryos picture from Christine lab, help us understand the function of zld by examining the relationship between expression profiles and the nucleotides view. It also inspires our prediction about how zld could affect gene expression regarding presence, location, or intensity. Our integrated results demonstrate that the gene transcription is strongly enriched just after the presence of zld, which aligns with global ideas. Since Halo is a ubiquitous gene, it strongly suggests that other ubiquitous genes might also be controlled by zld. Another prediction is that zld activation weakens with distance, or there might be other promoters or repressors to influence the gene expression. In addition, we also found another gene that would be worthy of study when we attempt to figure out the relationship between TAGsites and CG18123.

3.1 The coordination of TAGsites

The transcription only begins when zld exists, and by zld chip, there is a peak just near TSS. We found two sites, both are CAGGTAG. In Nien et al. paper, they strongly suggest that CAGGTAG is the most enriched site. However, we are wondering why there are two of them. It is not known whether both of them can function. It might be explained as Halo plays an important function in embryos, so it needs to be activated intensively and punctually. Normally, sites are within 2kb of TSS, which can strengthen one gene's transcription and expression. Nevertheless, we still found some sites in untranslated or coding regions. We designed the experiment to knock out certain sites so that we could focus on the effect of a single site. What's more, we also found other less enriched sites in front of TSS, and we found that the old binding peak is pretty low in Figure 2. It is hard to observe their peak if we zoom out to look at the peak at CAGGTAG. Consequently, we suggest they are not useless, but they can slightly activate gene expression and might be responsible for the transition between gene and gene, TAGsites and TAGsites.

3.2 Zld activation could be passed down

To further understand the coordination of zld, we need

to confirm whether the activation of zld can be passed down. Since we saw similar RNA polymerase binding peaks in Halo and CG18132 but no TAGsites before CG18132, we predict TAGsites can serve several genes, and activation decays with distance. Therefore, we design the experiment to investigate the influence of knocking out TAGsites before gene Halo on gene CG18132. As mentioned in the results, we need to mutate the sites to make them lose function. We found the guide RNA using chop-chop, which could be put into a guide vector after complementary base pairing. Then, we also chose a 1kb homologous arm on both sides of the sites, which was later mutated in vitro and put into a donor vector. Finally, we put the guide and donor vectors into the embryos during mitosis. We must repeat the experiment to ensure it grows in the embryo to succeed. As a result, we could get the different versions of the RNA polymerase binding profiles of CG18132. However, though we do not get the complete data, we could still predict the transmission characteristics of old activation.

4. Conclusion

Zinc-finger protein plays a significant role in gene expression in embryos. However, the interaction between them is not clear. If this experiment can be completed, we can continue to use zld to enhance or weaken the expression of some genes to help us treat some diseases.

5. Method

5.1 Fly strain

The halo strain was used to obtain wild-type embryos, and the mutated zld allele was used to obtain zld^- embryos.

5.2 ChIP-chip

The chromatin was immunoprecipitated by purified anti-zld antibody RNA polymerase antibody. Binding efficiency was visualized using the Integrated Genome Browser [4], and a median filter was applied to the ChIP/input ratio for the visualization. And the higher peak indicates a higher ChIP/input ratio in the region.

5.3 CRISPR-cas9 editing

The CRISPR-Cas system is a natural immune system of prokaryotes. After being invaded by a virus, some bacteria can store a small piece of the viral gene in their DNA in a storage space called CRISPR. When the virus is invaded again, the bacteria can recognize the virus from the stored fragments, cutting the DNA of the virus and rendering it useless. As a result, we can design the distinctive guide vector and insert it into the embryos. To find a guide RNA that includes CAGGTAG, we used the website Chopchop (http://chopchop.cbu.uib.no/). This guide vector can identify the specific sequences of nucleotides and cut. Therefore, the Donor vector replicates CAGGTAG to CTCATAG, called knock out.

5.4 Homology direct repair

Homology direct repair happens after donor vectors are inserted into embryos. The donor vector contains 2kb homologous arms and mutated DNA sequences with 20bp. During mitosis, crossing over will replace these mutated DNA sequences into embryos.

References

1. Chung-Yi Nien, Hsiao-Lan Liang, et al. (2011) Temporal Coordination of Gene Networks by Zelda in the Early Drosophila Embryo.

2. Hsiao-Lan Liang, Chung-Yi Nien, et al. (2008) The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila.

3. Yujia Sun, Chung-Yi Nien, et al. (2023) Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation.

4. Nicol JW, Helt GA, Blanchard SG, Raja A, Loraine AE (2009) The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics 25: 2730–2731.

5. Ten Bosch, J. R., Benavides, J. A. & Cline, T. W. The TAGteam DNA motif controls the timing of Drosophila pre-blastoderm transcription. Development 133, 1967–1977 (2006).

6. Li, X. et al. Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm. PLoS Biol. 6, 365–388 (2008).

7. Jiang, J., Rushlow, C. A., Zhou, Q., Small, S. & Levine, M. Individual Dorsal morphogen binding sites mediate activation and repression in the Drosophila embryo. EMBO J. 11, 3147–3154 (1992).

8. Lecuit, T., Samanta, R. &Wieschaus, E.slam encodes a developmental regulator of polarized membrane growth during cleavage of the Drosophila embryo. Dev. Cell 2, 425–436 (2002).

9. Staudt, N., Fellert, S., Chung, H., Ja¨ckle, H. & Vorbru¨ggen, G. Mutations of the Drosophila zinc finger-encoding gene vielfa¨ltig impair mitotic cell divisions and cause improper chromosome segregation. Mol. Biol. Cell 17, 2356–2365 (2006).