Proceedings of the 3rd International Conference on Biological Engineering and Medical Science
DOI: 10.54254/2753-8818/24/20231163

Gene co-regulation caused by the special positional relationship in the early stages of Drosophila embryos

Guanhua Zhuang

College of Chemistry and Molecular Sciences, Wuhan University, Wuhan, 10486, China

2020302031076@whu.edu.cn

Abstract. The genetic regulation mechanism of early Drosophila embryo has been a research hotspot in the past few decades. Understanding the early genetic regulation mechanisms of Drosophila embryos plays an important role in understanding the growth and development of other animals. In previous studies, an important Transcription Activator, Zelda (Zld), was reported. Zld is a global activator of early embryonic transcription. During the early development of embryos, all animals undergo the material-to-zygotic Transition (MZT), where Zld plays an important role in degrading maternal mRNA and mediating the expression of over 120 zygotic genes. The absence or mutation of Zld will directly lead to the inability of the embryo to complete the MZT process and result in lethality. TAGteam is a special sequence upstream of many genes in Drosophila, and Zld plays a role as a binding protein of TAGteam. TAGteam is a highly conserved sequence that generally exists in the form of CAGGTAG. Mutations or lack of TAGteam often result in inability to bind to Zld, resulting in the inability of genes to be expressed normally. The two genes involved in this experiment, CG14014 (2L: 5549709.. 550835 [-]) and CG18269 (2L: 5551838.. 5552756 [+]), are two adjacent genes on the Drosophila chromosome. They are located on two different DNA strands and facing in opposite directions, while their current role is not yet clear. There are typical TAGteam sequences upstream of both genes. This article conducts experiments to verify the relationship between these two genes and the regulation of Zld protein, and explores the possibility of their co-regulation by designing experiments.

Keywords: Drosophila Melanogaster, Transcriptional Regulation, Chip-seq, Zelda, Gene-Expression, Transcription

1. Introduction

The development of early animal embryos is regulated by both maternal and zygotic genes, and all animals have to undergo the material-to-zygotic Transition (MZT) [1].

In a 2008 study, Zelda (Zinc-finger early Drosophila activator, or Vielfaltig) was first proposed to be involved as a global transcription factor in mediating early embryonic development in Drosophila embryos and may also dominate the degradation of maternal mRNA. These conclusions have been confirmed in subsequent studies. The experiment indicated that in the case of Zld gene mutation, the important genes for early embryonic development of Drosophila, like genes conducive to sex selection

DOI: 10.54254/2753-8818/24/20231163

could not be effectively expressed. At the same time, the lack of Zld often directly leads to the death of embryos [2-6].

The Zld protein was initially valued by researchers because it is one of the binding proteins of TAGteam. Although TAGteam has three binding proteins, only Zld is currently the most thoroughly studied. TAGteam is an interesting sequence that is often located upstream of an important gene in the embryonic stage of fruit flies. Zld needs to bind to TAGteam to function. Previous experiments have shown that if a mutation occurs in the TAGteam upstream of the gene, the gene cannot be effectively expressed. Among the five TAGteam sequences initially defined, CAGGTAG and Zelda have the strongest specific binding ability, followed by CAGGTAA, which is often used as a marker to judge TAGteam. [7-9] In addition, there are often TATAA sequences near the TAGteam as transcription initiation sites. Experiments in this article also involves these principles.

The function of the two genes involved in this article, CG14014 (2L: 5549709.. 550835 [-]) and CG18269 (2L: 5551838.. 5552756 [+]), in the genome are currently unclear, but they all have distinct TAGteam and TATAA sequences upstream (Figure 1). The two are only about 1000bp apart and are located on two different chains facing different directions, forming a head-to-head juxtaposition positional relationship. Previously, researchers reported a much closer head-to-head juxtaposition of genes with a distance of less than 350 bp between two genes, which has a significant impact on the coordinating relative expression patterns of the two genes [10]. Although there are some differences in the relationship between CG14014 and CG18269 compared to the original text (the distance between the starting sites of CG14014 and CG18269 is about 1000 bp, which is greater than the 350 bp in the original lecture), within 1000 bp, TAGteam may still play a role in promoting gene transcription by binding to Zld. Therefore, it is reasonable to speculate that there is a correlation between CG14014 and CG18269 in expression and to verify it through design experiments.

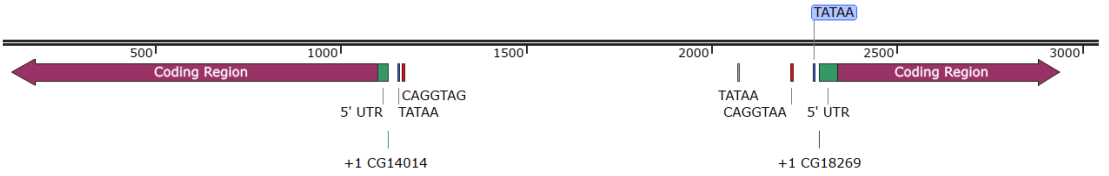

Figure 1. Both CG14014 and CG18269 have CAGGTA and TATAA sequences upstream.

2. Methods

The information about genes comes from Fly Base (https://flybase.org). The techniques used include RNA seq and RNA polymerase II chip sequencing technique. The data is visualized and processed through software such as Snap Gene and IGB (Integrated Gene Browser).

3. Results

3.1. Zld manipulate the transcription of CG14014 in the early stages of fruit fly embryos

Experimental results (A-C) indicated that the expression of CG14014 reaches its peak at early nc 14 and subsequently weaken and that there is no obvious sign of expression of CG14014 in Zld- embryos (D-F).

DOI: 10.54254/2753-8818/24/20231163

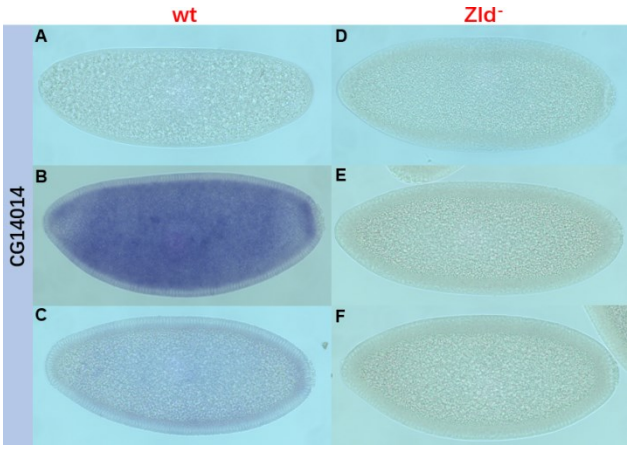

Figure 2. CG14014 expression is downregulated in Zld- mutant embryos. Wild-type (wt A, B, C) and Zld mutated embryos (Zld- D, E, F) were hybridized with RNA probes to visualize to the transcription situation of CG14014 in the early embryos. The pictures of these fruit fly embryos were arranged according to their age, from early to late development.

3.2. Analysis of CG14014 and CG18269 gene region in IGB

RNA polymerase II chip sequencing technique and RNA chip were used to visualize data and results was expressed in different colors with IGB software. (Figure 3) The results indicates that there are two important genes (CG14014 and CG18269) on two different strains and towards opposite directions with a distance of about 1000 bp between them. Both of them have obvious Zld peaks on their upstream. In the normal embryos, CG14014 and CG18269 were expressed normally, while in the Zld- embryos, neither is expressed. The starting and ending sites of transcription of two genes basically match the data on Fly base.

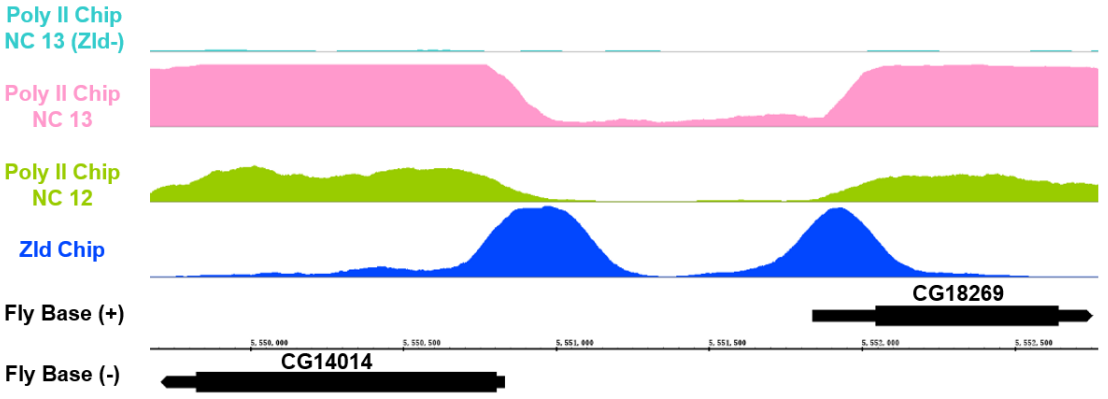

Figure 3. Analysis of Gene CG14014 and CG18269. RNA polymerase II chip was used to measure the expression of CG14014 and CG18269, with different time periods expressed in different colors. RNA seq was used to detect the Zld binding region, presented by blue color. All data is presented and analyzed in IGB.

3.3. Analysis of Zld Binding sites on snap gene

In previous researches, researchers reported on the association of Zld with tag team sites and TATAA sequences. Download the gene sequence from Fly base and import it into Snap gene for analysis. CAGGTAG and CAGGTAA are typical TAG team sites, which have strong binding ability with Zld protein, and have been confirmed to have strong correlation with regulating the expression of early zygote genes in embryos. TAG sites generally function within 2000 bp upstream of the gene and there is a strong correlation between the position of TATAA and the position of TAG sites.

DOI: 10.54254/2753-8818/24/20231163

Both CG14014 and CG 18269 has Typical Tagteam on their upstream, CAGGTAG for CG14014 and CAGTAA for CG18269. By mapping the Zld chip obtained from IGB to the position of CAGTAG on the snap gene, it can be seen that the two matched correctly each other.

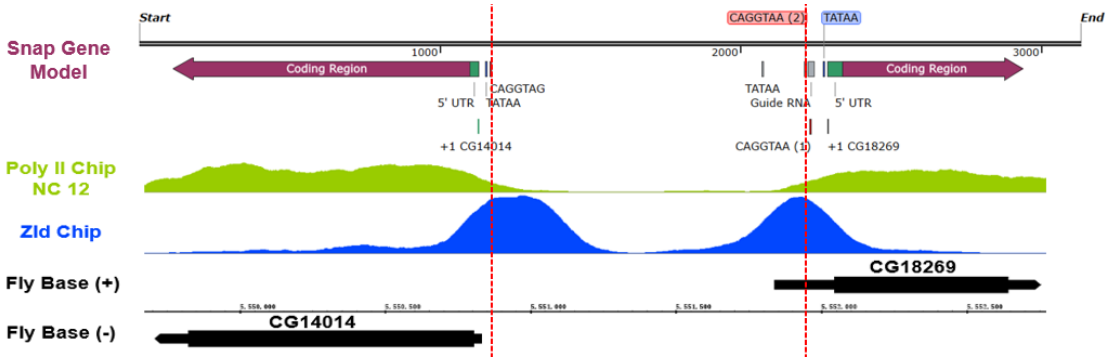

Figure 3. Mapping the Zld chip obtained from IGB to the position of CAGTA on the Snap Gene, it can be seen that the two matched correctly each other. Results were expressed in IGB and Snap Gene.

4. Discussion

As mentioned earlier, there is a very interesting phenomenon in observing Flybase's data graph. It can be seen that at very similar locations (specific locations), approximately 1000 bp apart, located on two different chains, there are genes CG14014 and CG18269 facing opposite directions, forming a typical "head-to-head structure". These two genes have typical TAGteam and TATAA sequences upstream of each other, with specific locations shown in the figure. And in the conclusion section of the experiment, in the case of Zld mutation, both genes cannot be expressed normally, which largely confirms that both genes are regulated by the transcription factor Zld.

Both genes are regulated by the same transcription factor, Zld, and Zld works by binding to the TAG team binding site to regulate transcription. The regulatory mechanisms of two genes that are so similar and their close positions make people curious about whether the expression of these two genes is correlated, whether their expression promotes and suppresses each other, and whether they jointly exercise certain functions.

A common assumption is that if the binding sites of two gene promoters are close, they may interact to enhance their expression, which may be achieved by a single region controlling multiple genes. In order to verify whether the binding of TAGteam and Zld in CG14014 and CG18269 affects their expression, experiments were designed to verify.

The experimental method was to use Crisper-cas9 to knock out the TAGteams of the two end genes separately, making them unable to bind to Zld. Then, RNA polymerase II seq was used to test the transcription of CG14014 and CG18269 genes.

Whether to eliminate the TAG team close to CG14014 and CG18269 or not will divide the experiment into four scenarios (Table 1): retaining both, knocking out the binding sites of CG18269, knocking out the binding sites of CG14014, and knocking out both simultaneously. Knocking out specific position gene sequences is achieved through Crisper-cas9, and the Guide gene and homology Arms used in the design are shown in the Figure4. It is worth noting that there are multiple identical sequences CAGGTAG and CAGGTAA upstream of CG14014 and CG18269, all of which were selected for knockout.

DOI: 10.54254/2753-8818/24/20231163

Figure 4. A. The corresponding positions of guide genes and homology arms in genes are displayed using Snap genes. **B.** Knock out the TAGteam binding site of CG14014 and transform it from CAGGTAG to CTCATAG. **C.** Knock out the TAGteam binding sites of CG18269 and transform it from CAGGTAA to CTCATAA.

Table 1. Experimental Design Plan. "√"means the Zld binding sites were retained, while the "X" means they were knocked out.

	CG14014 Binding Site	**CG18269 Binding Site**
#1	√	√
#2	√	**X**
#3	**X**	√
#4	**X**	**X**

Next, three possible experimental results will be listed and corresponding experimental conclusions will be discussed.

Table 2 Expected Experimental Results A

A	Experiments Results
#1	Both CG14014 and CG18269 were normally expressed
#2	Only CG14014 was normally expressed, while CG18269 wasn't expressed
#3	Only CG18269 was normally expressed, while CG14014 wasn't expressed
#4	Neither CG14014 nor CG18269 could be normally expressed

Experimental conclusion: Knocking out the binding site of one gene will not affect the normal expression of the other gene.

The experimental results indicate that the binding sites of CG14014 and CG18269 do not affect each other, and there is no clear evidence to suggest a correlation between the two.

Table 3 Expected Experimental Results B

B	Experiments Results
#1	Both CG14014 and CG18269 were normally expressed
#2	Only CG14014 was normally expressed, while CG18269 was weakly expressed
#3	Only CG18269 wasn't expressed, while CG14014 was weakly expressed
#4	Neither CG14014 nor CG18269 could be normally expressed

Experimental conclusion: Knocking out the binding site of CG14014 would affect the expression of CG18269, but the opposite was not true.

The experimental results indicate that knocking out the binding site of CG14014 will affect the expression of CG18269, but the opposite is not true.

DOI: 10.54254/2753-8818/24/20231163

Table 4 Expected Experimental Results C

C	Experiments Results
#1	Both CG14014 and CG18269 were normally expressed
#2	Both CG14014 and CG18269 were weakly expressed
#3	Both CG14014 and CG18269 were weakly expressed
#4	Neither CG14014 nor CG18269 could be normally expressed

Knocking out the binding sites of CG14014 and CG18269 with Zld will affect their expression.

The experimental results indicate that the TAG team of CG14014 and CG18269 have a promoting effect on each other's expression.

5. Conclusion

In the past research, the importance of Zelda in regulating the early development of drosophila embryos has been confirmed and has received long-term attention. However, many biological and genetic mechanisms involved in Zelda regulation, such as the role of microRNAs and gene structure, have not yet been clarified. As mentioned in previous literature [10] and this article, the potential impact of gene structural characteristics on the regulatory process of Zelda has not received sufficient researches. This article selects genes from Drosophila embryos that have been confirmed to be regulated by Zelda protein and have significant structural characteristics (close and significant head-to-head structure) for analysis and design of proof experiments. It can be foreseen that the structure of genes plays an important role in the regulation of Drosophila embryonic development, and the complex effect of structure and Zelda protein on Drosophila embryonic development is more helpful for researchers to understand the process of Drosophila embryonic development, and this may help researchers study the embryonic development process of other animals in the future [1].

References

[1] Hamm D.C., Harrison M.M. (2018). Regulatory principles governing the maternal-to- zygotic transition: insights from Drosophila melanogaster. Open Biology, 8 (12) 18013.

[2] Darbo E., Herrmann C., Lecuit T., Thieffry D., van Helden. (2013). Transcriptional and epigenetic signatures of zygotic genome activation during early drosophila embryo genesis. BMC Genomics. 14, 226.

[3] 'Liang, H. L., Nien, C. Y., Liu, H. Y., Metzstein, M. M., Kirov, N., & Rushlow, C. (2008). The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature, 456, 400–403.

[4] Melissa M. Harrison, Michael B. Eisen. (2015). Transcriptional Activation of the Zygotic Genome in Drosophila. Maternal-to-Zygotic Transition, 113, 85-112.

[5] Fu S-B, Nien C-Y, Liang H-L, Rushlow. (2014). Co-activation of microRNAs by Zelda is essential for early Drosophila development. Development, 141, 2108-2118.

[6] Harrison M.M., Li X.Y., Kaplan T., Botchan M.R., Elsen M.B. (2011). Zelda Binding in the Early Drosophila melanogaster Embryo Marks Regions Subsequently Activated at the Maternal-to-Zygotic Transition. Plos Genetics 7(10):e1002266

[7] Nien C-Y, Liang H-L, Butcher S, Sun Y, Fu S, et al. (2011). Temporal Coordination of Gene Networks by Zelda in the Early Drosophila Embryo. Plos Genet 7(10): e1002339.

[8] Ten Bosch, J.R., Benavides J.A., Cline T.W. (2006). The TAGteam DNA motif controls the timing of Drosophila pre-blastoderm transcription. Development. 133(10): 1967-1977.

[9] Satija R., Bradley R.K. (2012). The TAGteam motif facilitates binding of 21 sequence-specific transcription factors in the Drosophila embryo. Genome Research. 22(4) 656-665.

[10] Deron R. Herr*, Greg L. Harris. (2004). Close head-to-head juxtaposition of genes favors their coordinate regulation in Drosophila melanogaster. Department of Biology and Molecular Biology Institute, San Diego State University, San Diego, CA 92182-4614.