International Core Journal of Engineering Volume 6 Issue 12, 2020
ISSN: 2414-1895 DOI: 10.6919/ICJE.202012_6(12).0036

Zelda is Required for the Transcription and Expression of Class 1 & 2 Gene in Early Drosophila Embryo

Qinyan Feng

Nanjing Foreign Language School, Nanjing, Jiangsu 210008, China.

Abstract

In order to study how genes control development and the functions of several important genes, this work focused on the research of Zelda binding, binding motif of different transcription factors, expression pattern, and conservation of the region upstream the TSS of the gene. Two genes involved in this work are one class 1 gene (Z600) and one patterning gene (twist). This work determined that CAGGTAG site is a binding motif for Zelda, and Zelda is important for the expression of the genes and that highly conserved regions exist around Zelda binding sites, which sets the corner stone on future work to identify more specific role of individual transcription factor in early embryotic development.

Keywords

Zelda binding site, Expression pattern, transcription factors, Binding motifs, conservation.

1. Introduction

Maternal-to-zygotic transition (MZT) in which control on the development is transferred from maternal to zygotic gene is an essential role in the early embryotic development of Drosophila. Zelda is believed to have an important role in the transcription, but how Zelda regulates the expression of Drosophila genes together with other transcription factors is still ambiguous. In this work, we looked at the Zelda binding images, Chip-seq images and compared them with transcription factors' binding situation. Zelda basically binds to TAGteam sites, which are usually conserved. Zelda, together with other important transcription factors, also contributes to the special pattern in gene expression.

2. Data & Result

2.1 Research methods:

Data for this work comes from different websites such as Integrated Genome browser, Jaspar, NCBI, UCSC Genome Browser, Fly express, and Fly base, and data are compared to find correlation and make hypothesis for future work.

2.2 Twist (CG2956)

Here is the snapshot of the gene twist: *twist (twi) encodes a transcription factor required for mesoderm cell fate. The product of twi is essential for gastrulation, the development of mesodermal derivatives, including somatic and visceral muscle, fat body and maintenance of muscle stem cells.* [1]

On integrated genome browser, twist, a class 2 (pattern) gene, is found on the positive strand. There are three peaks in each wildtype Zelda track (orange ones 5^{th} and 6^{th} tracks), all of them located just upstream the gene's coding region. There is one CAGGTAG site right under the second peak, later by using JASPAR to detect all the binding motif, this work helps to reach the conclusion that here really is a strong Zelda binding site, with a relative score of 11.98. Figure 1 shows that of the four

ISSN: 2414-1895 DOI: 10.6919/ICJE.202012_6(12).0036

polymerase tracks, NC 12 shows little expression, the other three tracks are highly expressed, as shown in the peaks in the four blue tracks (1st to 4th track). This phenomenon means that twist start to express from a later stage of embryotic development. In the Zelda mutant track (colored in purple, 7th track), there is only little expression compared with the expression in wild type track, which indicated that Zelda plays an important role in the transcription. (all tracks are one the same scale).

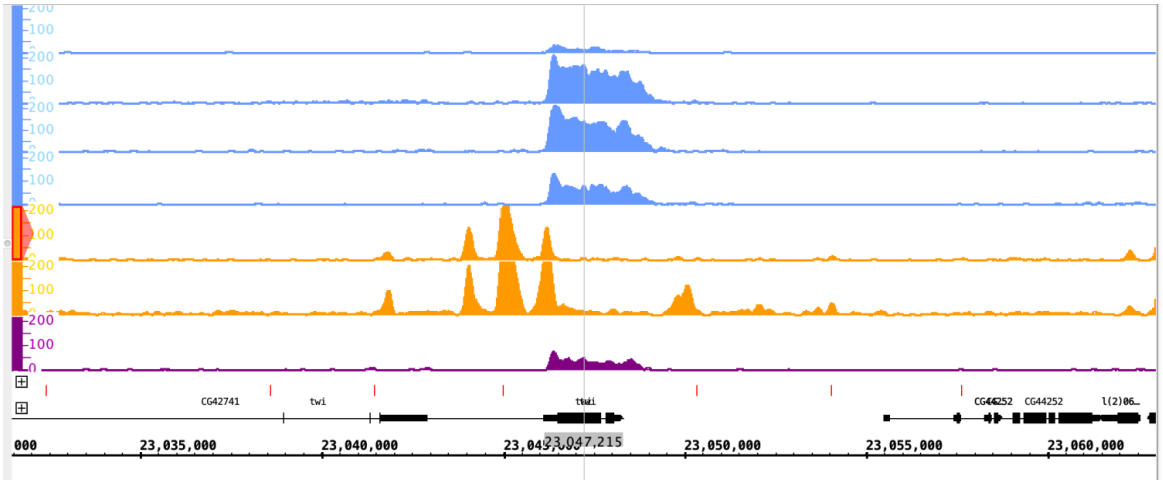

Figure 1 polymerase expression (first 4 blue tracks), wildtype Zelda binding sites (5th and 6th tracks), Zelda mutant binding sites (7th track), TAGteam site (8th track) of the gene twist.[2]

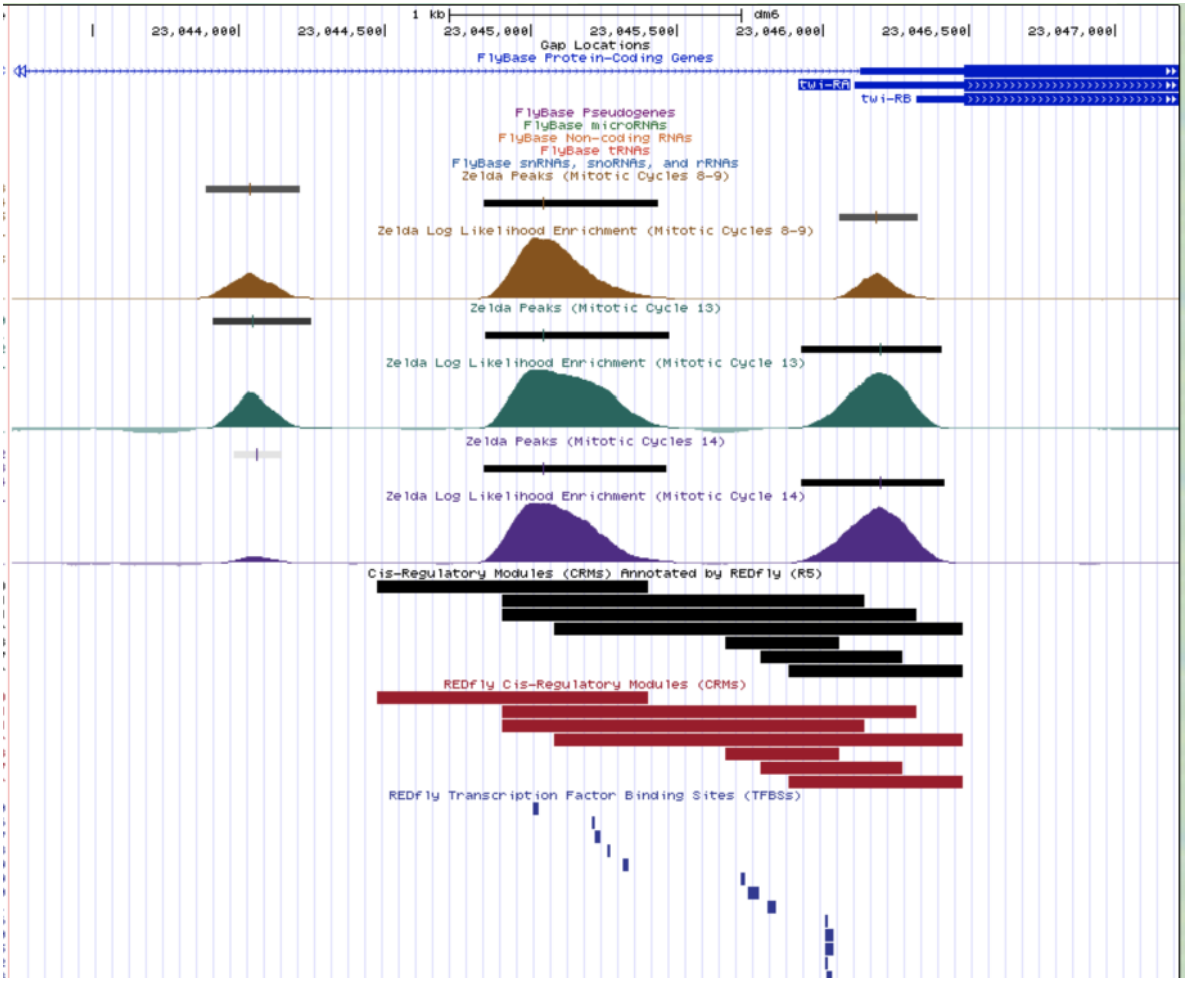

Figure 2 Chip-seq data for twist.[3]

ISSN: 2414-1895 DOI: 10.6919/ICJE.202012_6(12).0036

After having a look at the Zelda binding and expression in IGB, this study entered the gene in UCSC genome browser to learn about the CHIP-seq data, as shown in Figure 2. Three apparent peaks in this image implies three regions that Zelda has a high possibility to attach to, this result matches the three peaks in the IGB image. Selecting one of the representative enhancers, twi_DmD, this work got the overlapping region of the enhancer and one of the peaks. The location of selected region is on chr2R: 23044751-23045498. The exact sequence of this region was got in National Center for Biotechnology Information. Putting the sequence in JASPAR, the study got the binding motif of all transcription factors with a relative profile score threshold of 80%.

Here is a relation of the gene position and the position of the selected region., the selected region is chr2R: 23044751-23045498, the gene region is chr2R 23046108-23048325, so the selected region is upstream 1357 bp to upstream 610 bp. Supposed the first base in the selected region has the place "1"

There is one strong Zelda binding site in this region, located upstream 1114 bp of the gene with a relative score of 11.98, as shown in Table 1. The motif sequence is CAGGTAG. Later all binding motif from upstream 500 bp to downstream 500 bp are determined, but it results that there is almost no Zelda binding site in -500 bp -+500 bp.

Table 1.

Matrix ID	Name	Score	Relative score	Sequence ID	Start	End	Strand	Predicted sequence
MA1462.1	vfl	11.9849	0.956437454	NT_033778.4:23044751-23045498	243	254	+	GTCCAGGTAGTT
MA1462.1	vfl	8.27959	0.882442505	NT_033778.4:23044751-23045498	425	436	+	CGGCAGGCAAAC
MA1462.1	vfl	6.70526	0.851003244	NT_033778.4:23044751-23045498	320	331	-	GCTTAGGTAATA
MA1462.1	vfl	4.84885	0.813931142	NT_033778.4:23044751-23045498	624	635	-	CATCAGTTAGTT

Table 1 Relative scores, positions, and predicted sequences for Zelda binding sites upstream the gene twist. [4,5]

At the very top of the table, both dorsal binding sites and Zelda binding sites are detected in the study, as you can see in Table 2. Then the expression pattern of twist is gotten in different embryotic stages from FLYEXPRESS. Twist is involved in the establishment and dorsoventral patterning of germ layers in the embryo. [7] Figure 3 shows that from the beginning to the end, the gene mainly express in ventral part. But after stage seven, the bottom left part expresses a little bit inward, forming a small hole. Also, it expresses upward in the right part after stage 7.

Table 2

Matrix ID	Name	Score	Relative score	Sequence ID	Start	End	Strand	Predicted sequence
MA0022.1	dl	12.2329	0.916443773	NT_033778.4:23044751-23045498	369	380	+	TGGTTTTTTCCA
MA1462.1	vfl	11.9849	0.956437454	NT_033778.4:23044751-23045498	243	254	+	GTCCAGGTAGTT
MA0022.1	dl	11.4565	0.896960869	NT_033778.4:23044751-23045498	368	379	+	GTGGTTTTTTCC

Table 2 Relative scores, positions, and predicted sequences for Zelda and Dorsal binding sites upstream the gene twist.[4,5]

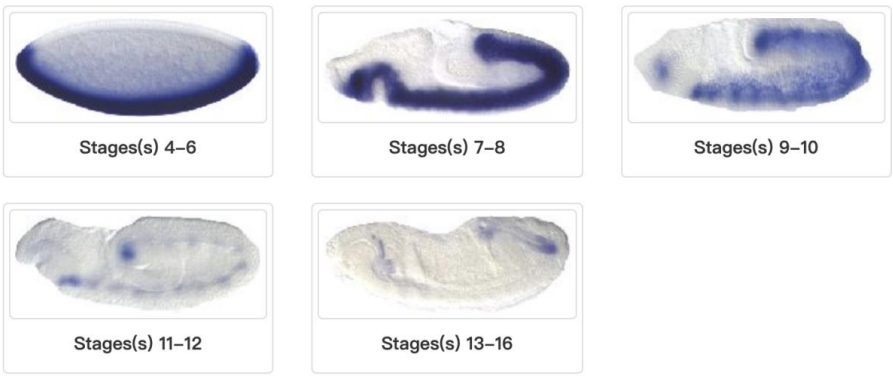

Figure 3 expression pattern for twist.[1]

ISSN: 2414-1895 DOI: 10.6919/ICJE.202012_6(12).0036

Here are the descriptions of the expression pattern**:** A normal blastoderm is formed; at gastrulation, no ventral furrow is visible, but the endoderm invaginates, a cephalic furrow is formed, and the germband elongated. The embryo is twisted or coiled in the egg case, often with posterior side up. [1]

The next thing in this work is to look at the conservation of this particular region on UCSC Genome Browser. The different color in the table represent different conserved region.

As shown in Figure 4, 259-290 bp is a quite conserved region. 3 dorsal binds to this region. One with score 9.0 binds to 263-274(AGGGCAAAACCC), two others with scores 10.4 and 9.1 respectively bind to 279-290 (GGTGGTTTTT).

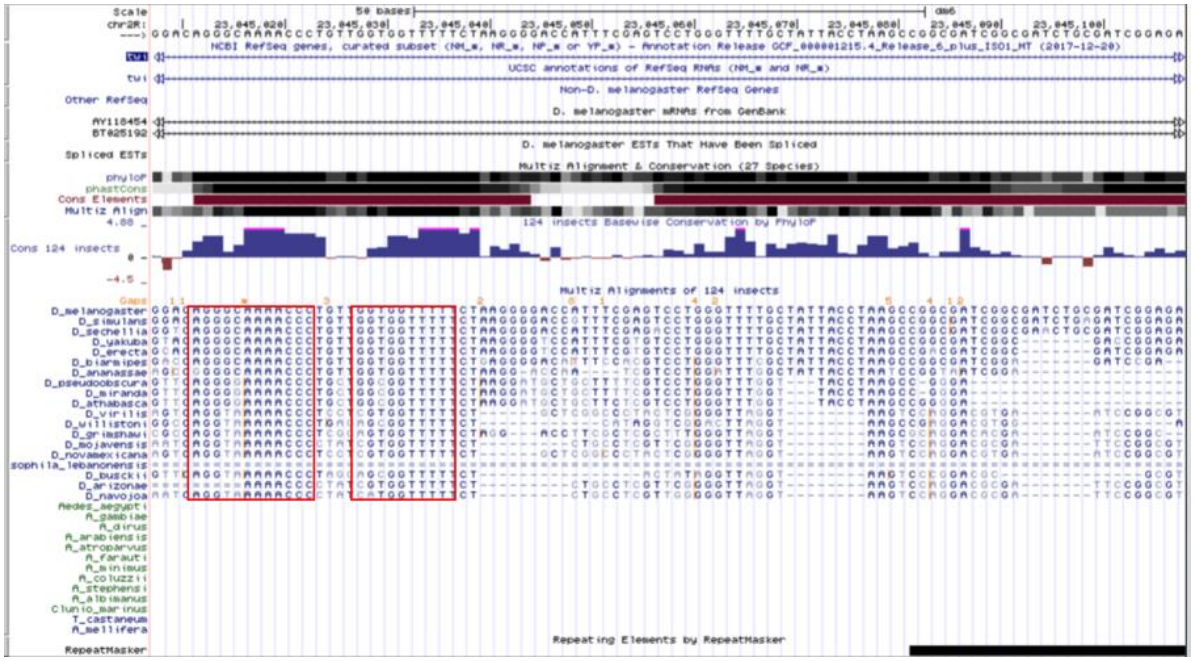

Figure 4. Conservation for twist in the place 259-290, with the first base pair in the selected region called place 1.[8]

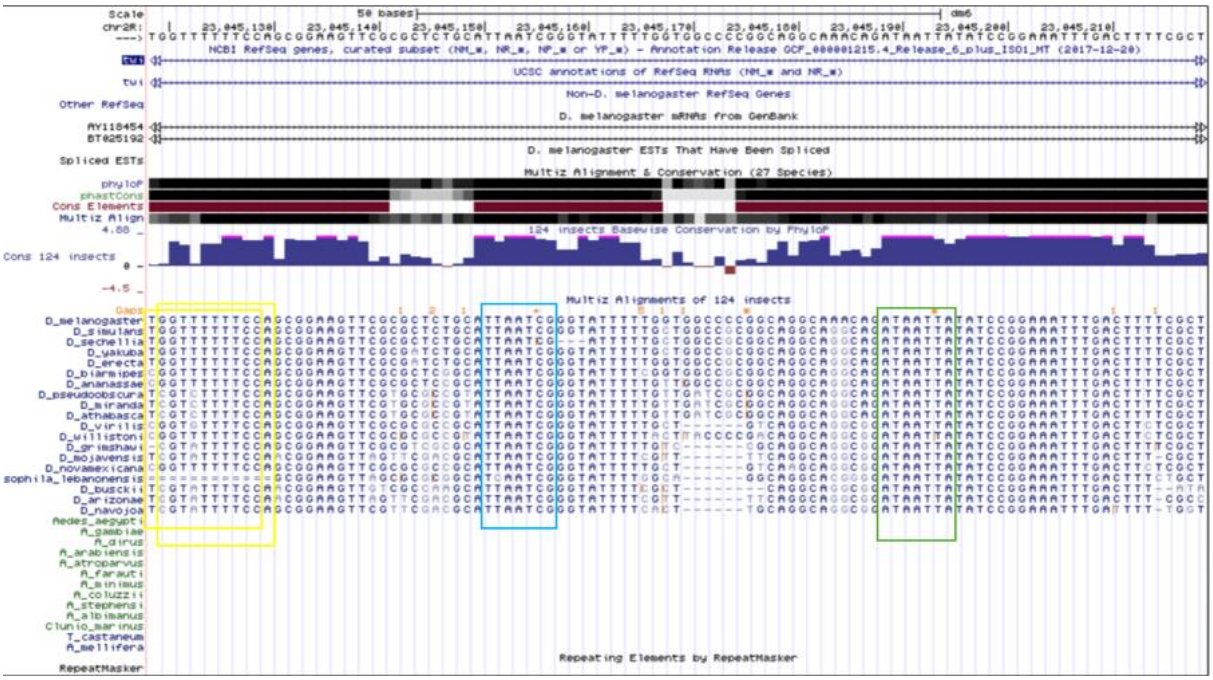

Figure 5. Conservation for twist in the place 368-487, with the first base pair in the selected region called place 1.[8]

ISSN: 2414-1895 DOI: 10.6919/ICJE.202012_6(12).0036

As shown in Figure 5, 368-487 is a highly conserved region. The yellow part is the binding site of two dorsal with the highest score (12.2/11.5). A few transcription factors bind to the blue part, with a sequence of TTAATCG from 401 to 407. A lot of transcription factors, for example pb, Rx, hbn, ap, zen2, also eve, bind to the green part. Basically, they are two same genes with one on positive strand starting from 439, and another one on negative strand starting from 440, the sequences are ATAATTA or TAATTA for each.

As shown in Figure 6, 530-601 is another highly conserved region. Again, a lot of transcription factors bind to the purple region, including ro, Rx, p, zen, also eve, they are also two same genes with one on positive and one on negative strand. Starting from 545 or 546, the exact sequences are CTAATTA and TAATTA. And there are a few transcription factors bind to the latter dark blue region.

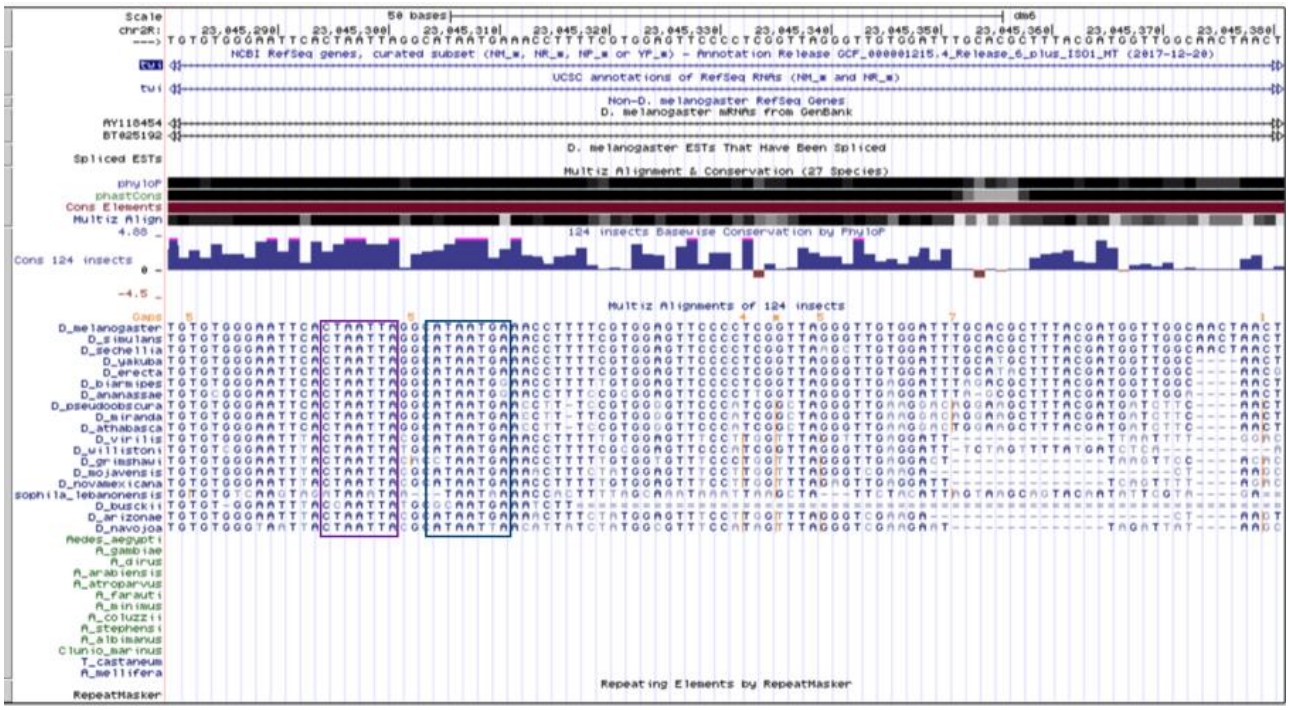

Figure 6. Conservation for twist in the place 530-601, with the first base pair in the selected region called place 1.[8]

Another interesting thing noticed in the study is that huge numbers of transcription factors bind to the pink region, from 134 to 140 with sequence TTAATTA, but it is not a conserved region. In fact, it is not conserved at all.

Other information about twist: Its molecular function is described by: DNA-binding transcription factor activity, RNA polymerase II-specific; protein heterodimerization activity; DNA binding; protein homodimerization activity; RNA polymerase II regulatory region sequence-specific DNA binding. It is involved in the biological process described with 15 unique terms, many of which group under: regulation of transcription, DNA-templated; organic cyclic compound biosynthetic process; signaling; positive regulation of transcription, DNA-templated; gastrulation involving germ band extension. The wild-type allele of twi is involved in the establishment of germ layers. Mutants are embryonic lethals (zygotic), partially dorsalized, and without mesodermal differentiation. [1]

2.3 Z600 (CG17962)

Then the study did similar research on the class 1 gene-Z600, a gene on positive strand.

Here is a snapshot of the gene Z600: Z600 (Z600) encodes a mitotic inhibitor that binds to the hydrophobic patch of Cyclins. It contributes to the switch in cell cycle mode in blastoderm embryos, delaying mitosis in the mesoderm anlage to avoid an interference of mitosis and morphogenesis. [9]

ISSN: 2414-1895 DOI: 10.6919/ICJE.202012_6(12).0036

First by using Integrated Genome Browser, one obvious peak is detected in each of the wildtype Zelda track (5th and 6th tracks) just upstream the transcription start site. Under the Zelda peak, there is one CAGGTAG site, located 126 bp upstream TSS. Later by using JASPAR, this work further identify this region as a strong Zelda binding site with a relative sore of 13.59. Taking a look at four polymerase tracks shown in Figure 7, the gene has almost no expression in NC12_rep1 track (1st track), but in later stages, which are NC13_rep1, NC14E_rep1, and NC14M_rep 1 (2nd to 4th track), the gene is highly expressed. So, the gene start to express from a later stage of embryotic development. There is barely any expression in Zelda mutant track (7th track). In addition, there is a TATAA box exactly 30 bp upstream the transcription start site.

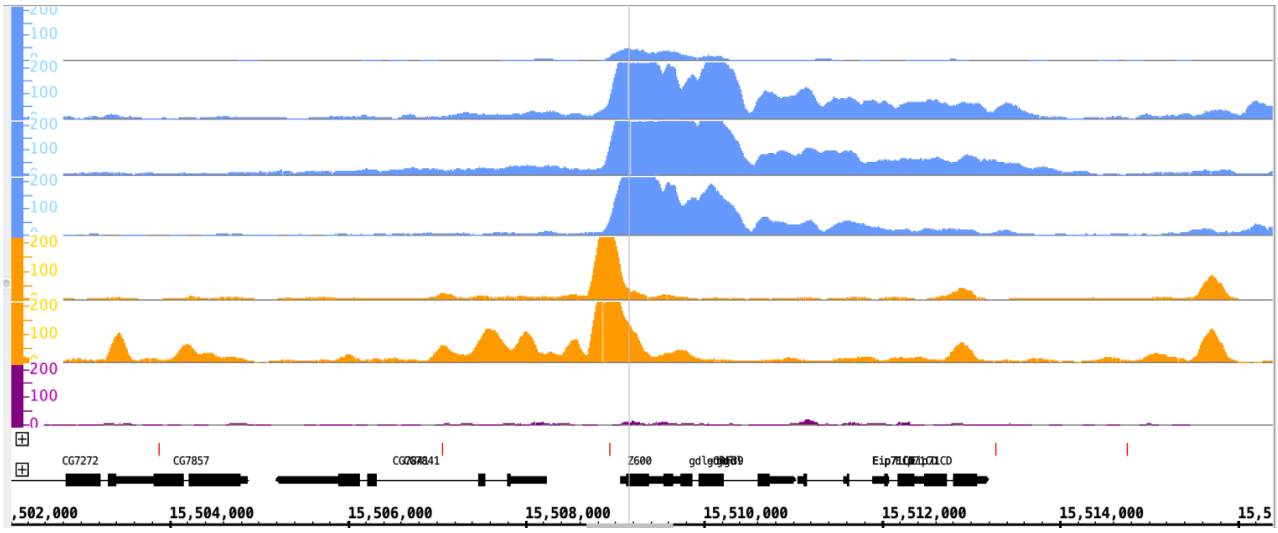

Figure 7. Polymerase expression, wildtype Zelda binding sites, Zelda mutant binding sites, TAGteam site of the gene Z600.[2]

Entering Z600 into UCSC Genome browser, CHIP-seq data is found for the gene, as shown in Figure 8. There is one Zelda binding peak in both IGB and CHIP-seq data. This work chose a region from 300 bp upstream to the end of the gene to look at the exact sequence. Put the position of the gene on the chromosome into NCBI, and then exact sequence of this region was extracted.

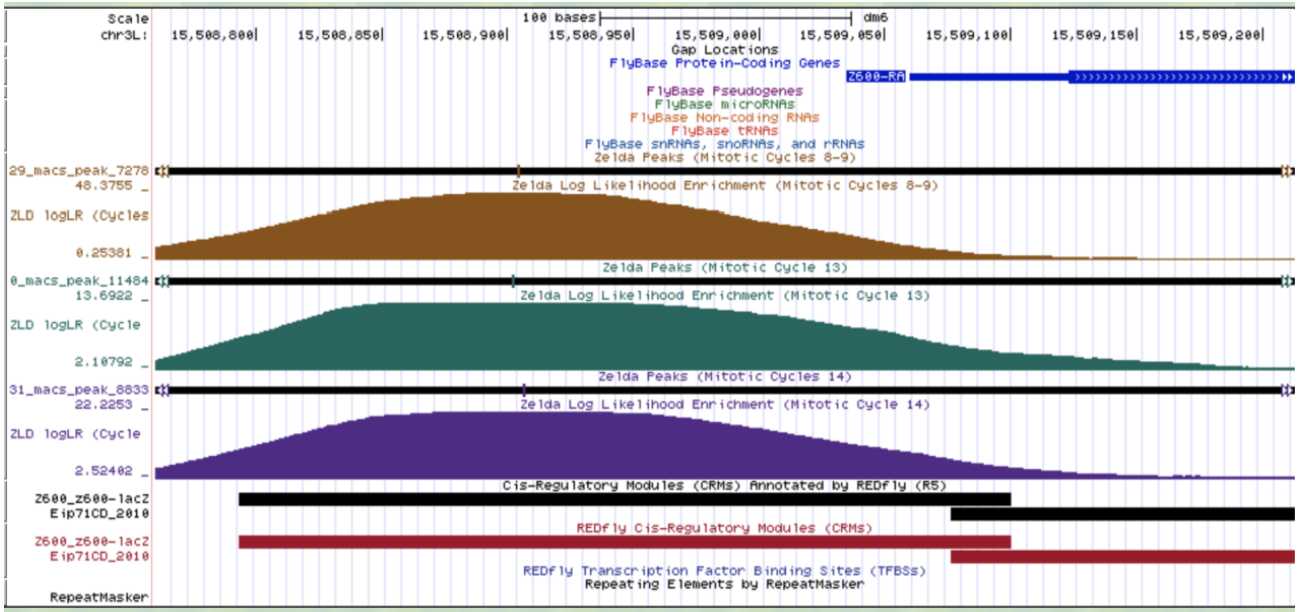

Figure 8. Chip-seq data for Z600.[3]

ISSN: 2414-1895 DOI: 10.6919/ICJE.202012_6(12).0036

Table 3. shows that from the Jaspar result, the study can find two strong Zelda binding sites upstream the TSS. The binding motif is CAGGTAG.

Table 3.

Matrix ID	Name	Score	Relative score	Sequence ID	Start	End	Strand	Predicted sequence
MA1462.1	vfl	13.5856	0.988402510357	NT_037436.4:15508761-15509513	171	182	+	GAGCAGGTAGCA
MA1462.1	vfl	13.2606	0.981911601706	NT_037436.4:15508761-15509513	98	109	-	GAGCAGGTAGTA
MA1462.1	vfl	5.53333	0.827600168435	NT_037436.4:15508761-15509513	333	344	-	GATCAGGAAGTT
MA1462.1	vfl	5.20706	0.821084634992	NT_037436.4:15508761-15509513	494	505	-	TGGCTGGTACCA
MA1462.1	vfl	4.62302	0.809421329633	NT_037436.4:15508761-15509513	466	477	-	GTCCAGGGAGTA

Table 3 Relative scores, positions, and predicted sequences for Zelda binding sites upstream the gene Z600.[4,6]

Then the conservation in the region upstream TSS was observed in UCSC Genome Browser. Suppose that the upstream 300 bp is called place '1'in this situation: it is not conserved from 0 to 100, it is quite conserved from 100 to 200, as shown in Figure 9. Two specific regions that I want to talk about are 134/135-140/141 and 162-170.

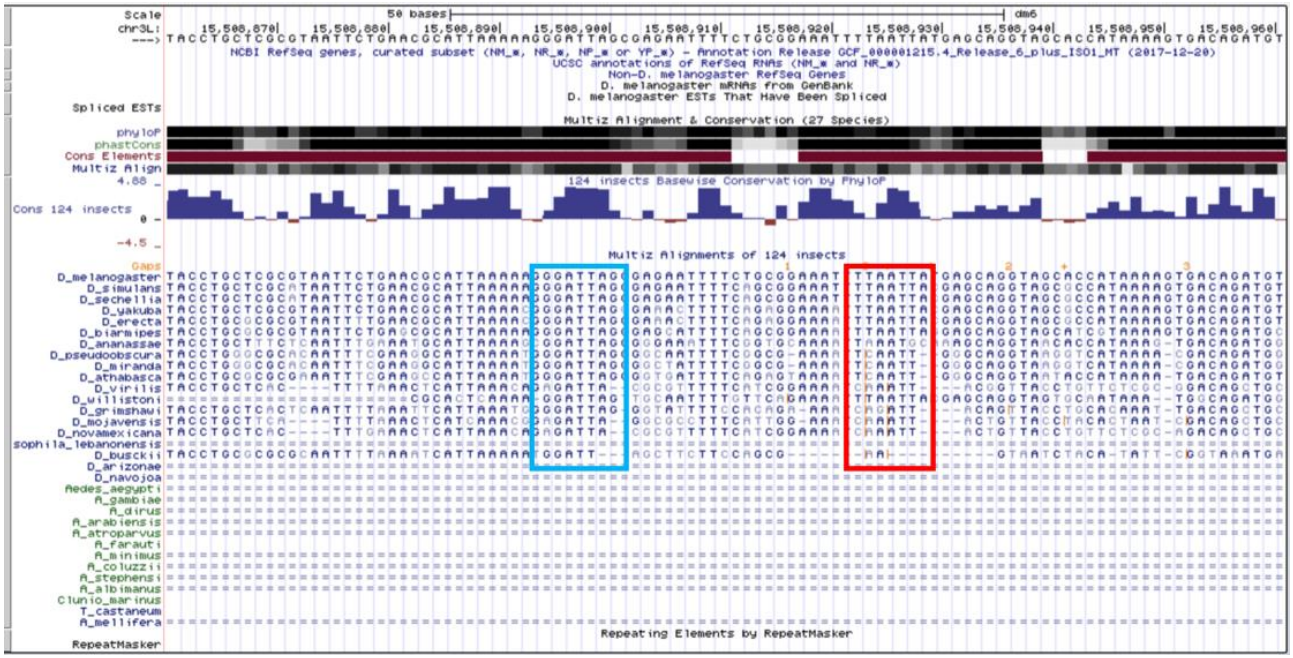

Figure 9. Conservation for Z600 in the place 101-200, with the first base pair in the selected region called place 1.[10]

162-170, which is upstream 138 to upstream 130 bp, is a conserved region with huge numbers of transcription factors binding to it. So, all the red box in this table are transcription factors that bind to 162-170. These TFs include al, ubx, repo, and so on.

Also, there are a few transcription factors binding to 134-140 or 135-141, which is about upstream 166 to upstream 159 bp. They are highlighted in blue in the table, which is a relatively small number compared with the red region.

TATAAA box, located 30 bp upstream TSS, is also a highly conserved region, A TATA box is a DNA sequence that indicates where a genetic sequence can be read and decoded. It is a type of promoter sequence, which can tell other molecules where transcription begins. [11]

Other information about Z600**:** Its molecular function is described by: protein binding; cyclin binding. It is involved in the biological process described with: negative regulation of G2/M transition of

ISSN: 2414-1895 DOI: 10.6919/ICJE.202012_6(12).0036

mitotic cell cycle; gastrulation; negative regulation of mitotic nuclear division; ventral furrow formation. [9]

3. Conclusion

To make a conclusion of the research on the gene twist: CAGGTAG site is a binding motif for Zelda, and Zelda is important for the expression of the gene twist. It is involved in the establishment and dorsoventral patterning of germ layers in the embryo, as regulated by Dorsal and Zelda. Twist has Zelda binding sites upstream TSS, in the region where chip-seq peak and enhancer overlaps. It has highly conserved region in the region where chip-seq peak and enhancer overlaps, the conserved regions are usually binding sites of several transcription factors.

To make a conclusion for the research on the gene Z600: CAGGTAG site is a binding motif for Zelda, and Zelda is important for the expression of the gene Z600. It has Zelda binding sites upstream TSS. Z600 has TATAA box 30 bp upstream TSS, which is important for the transcription. It has highly conserved region upstream the TSS, and some of the conserved region are strong binding sites of TF

Comparing class I gene, which is represented by Z600, and patterning gene, which is represented by twist, this work finds the similarities between them are that they both have strong Zelda binding sites upstream TSS, Zelda plays an important role in the expression of both genes. They both have highly conserved regions upstream TSS, with some of these regions acting as strong binding motif for several TFs. The differences are that Zelda and Dorsal work together to contribute to the expression pattern we see in the gene twist, while there is no clear expression pattern for class 1 gene, Z600. Z600 has a TATAA box 30 bp upstream TSS, but twist doesn't have.

To make a hypothesis for the regulation of other transcription factors, this work reveals the possibility that Ap is an important transcription factor for the expression of twist, because it appears in almost every highly conserved region. Its binding motif maybe like TAATTA, because though there can be little difference for their binding motif, these 6 consecutive base pair never change, as indicated in Table 4. This work proposes that ap also contribute to the ventral expression of the gene. In the future, more research should be done on how different transcription factors collaborate to contribute to the specific expression pattern..

Table 4.

Matrix ID	Name	Score	Relative score	Sequence ID	Start	End	Strand	Predicted sequence
MA0209.1	ap	10.8014	1.000000007	NT_033778.4:23044751-23045498	545	551	+	CTAATTA
MA0209.1	ap	10.8014	1.000000007	NT_033778.4:23044751-23045498	546	552	-	CTAATTA
MA0209.1	ap	10.3463	0.983604319	NT_033778.4:23044751-23045498	134	140	+	TTAATTA
MA0209.1	ap	8.95769	0.933573083	NT_033778.4:23044751-23045498	439	445	+	ATAATTA
MA0209.1	ap	8.95769	0.933573083	NT_033778.4:23044751-23045498	440	446	-	ATAATTA
MA0209.1	ap	7.96684	0.897873395	NT_033778.4:23044751-23045498	555	561	+	ATAATGA
MA0209.1	ap	7,45417	0.879401985	NT_033778.4:23044751-23045498	135	141	-	GTAATTA

Table 4 Relative scores, positions, and predicted sequences for ap binding sites upstream the gene twist.[4,5]

Reference

[1] Gene: Dmel\twi https://flybase.org/reports/FBgn0003900.

[2] Integrated Genome Browser.

[3] UCSC Genome Browser on D. melanogaster Aug. 2014 (BDGP Release 6 + ISO1 MT/dm6) Assembly http:// gander. wustl. edu/ cgi -bin/ hgTracks? Db =dm6&lastVirtModeType =default&lastVirtModeExtra State=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr2R%3A23046108%2D23 048325&hgsid=253861_YBa3N35ESrj9G8OA4xCMCA7NaScc.

[4] Search profile: Drosophila m http:// jaspar. genereg. net/ search? q=drosophila+m&collection=all&tax _group=all&tax_id=all&type=all&class=all&family=all&version=all.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202012_6(12).0036

[5] Drosophila melanogaster chromosome 2R NCBI Reference Sequence: NT_033778.4 https://www.ncbi.nlm.nih.gov/nuccore/NT_033778.4?report=fasta&from=22985374&to=23048325.

[6] Drosophila melanogaster chromosome 3L NCBI Reference Sequence: NT_037436.4 https://www.ncbi.nlm.nih.gov/nuccore/NT_037436.4?report=fasta&from=15509061&to=15509513.

[7] Uniprot. (2020) UniProtKB - P10627 (TWIST_DROME). https://www.uniprot.org/uniprot/P10627.

[8] UCSC Genome Browser on D. melanogaster Aug. 2014 (BDGP Release 6 + ISO1 MT/dm6) Assembly https://genome.ucsc.edu/cgi-bin/hgTracks? Db =dm6 &lastVirtModeType =default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr2R%3A22985374%2D23048325&hgsid=871992855_xiW6WbZx5MKJId2uvcW1j4QCw5Im.

[9] Gene: Dmel\Z600 https://flybase.org/reports/FBgn0004052.

[10] UCSC Genome Browser on D. melanogaster Aug. 2014 (BDGP Release 6 + ISO1 MT/dm6) Assembly https://genome.ucsc.edu/cgi-bin/hgTracks? Db =dm6 &lastVirtModeType =default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr3L%3A15509061%2D15509513&hgsid=871994109_OkYfBF4Ic1GQY8Sgy76Z37AAugXH.

[11] Scitable by nature education https://www.nature.com/scitable/definition/tata-box-313/.