

Analysis of Class I Genes and Patterning Genes in *D. Melanogaster*

Evelyn Yilin Chen¹, Yichen Liu², Xinyi Tian³

¹International School of Beijing, Beijing 200022, China;

²The Perse School, Cambridge 01223, The United Kingdom;

³Western Reserve Academy, Hudson, Ohio 44236, US.

Abstract

In *D. melanogaster*, the DNA-binding protein Zelda is a master transcriptional activator that is required for the maternal-to-zygotic transition and promotes timely and robust transcriptional activation of embryogenesis. The aim of this paper is to analyse the differences and similarities between patterning genes and class I genes in terms of their *zld* regulated expression on genetic and physiological levels in *Drosophila*.

Keywords

Zld, *Drosophila*, Maternal-to-zygotic transition, Patterning genes, Class I genes.

1. Introduction

In *Drosophila*, the DNA-binding protein Zelda (also known as Vielfaltig) is a master transcriptional activator that is required for the maternal-to-zygotic transition and promotes timely and robust transcriptional activation of embryogenesis, allowing the early developmental events to be coordinated and cell differentiation properly regulated [1].

zld is first detected in 1-2 hour of embryogenesis (mitotic cleavage cycle 8-13), followed by another burst of activity in the third hour (cycle 14), which is characterised by the cellular blastoderm formation and cellularisation [2]. Physiologically, *zld* transcripts are first ubiquitously expressed in germline cells and later on restricted to the nervous system.

The temporal dynamic of *zld* regulates a wide range of genes: activating class 1 genes and potentiating the transcription of patterning genes. The Rushlow lab found that “Zld binds to genes involved in early developmental processes such as cellularization, sex determination, neurogenesis, and pattern formation” [2]. In *zld* absent mutants, many target genes remain deactivated, while others, particularly the patterning genes, exhibit delayed transcriptional activation or weak and sporadic expression.

Furthermore, *zld* binds to a cis-regulatory heptamer motif CAGGTAG and related sequences, referred to as TAGteam sites. It was recognised that *zld* has a stronger tendency to bind close, usually with 2kb, to the transcription start site (TSS), as higher the level of gene expression, closer to the TSS is *zld*'s binding sites located [2].

In this report, class I genes—*esg*, *sc*, and *Tsg*—are analysed in comparison with patterning genes—*ths*, *gt*, and *sog*—for defining differences and similarities between them, in terms of *zld* regulated expression and transcription factors' relationship with expression patterns.

2. Methods

Overall, the results of this paper is derived based on secondary information and genetic data analysis by utilising a variety of genomic browsers.

Integrated Genome Browser (IGB)

Six gene tracks and TAGteam sites tags are loaded into the browser to create a lucid, holistic view on the zld protein’s interaction with TAGteam sites, the relationship between zld and RNA polymerase’s binding sites, cycle 12-14 genetic activation’s differences, and the comparison between wild type and zld knocked down (zkd) in terms of their genetic expression.

JASPAR

Used to obtain transcription factor (TF) binding profiles of six taxonomic groups’ species stored as position frequency matrices (PFM) and transcription factor flexible models (TFFM).

Flybase

An bioinformatics database that contains the primary repository of genetic and molecular data for D. melanogaster. This browser is used to determine genes’ expression pattern to deduce each of their function and interaction.

Flyexpress

Image-matching search engine of spatiotemporal expression patterns data of D. melanogaster in embryogenesis, containing standardized images submitted from BDGP, Fly-FISH and other researchers’ publications. It is used in this paper to research expression patterns to hypothesis the relationship between a gene and its transcription factors.

3. Data analysis

3.1 Patterning Genes, ths, gt, sog, Zld Regulated Expression

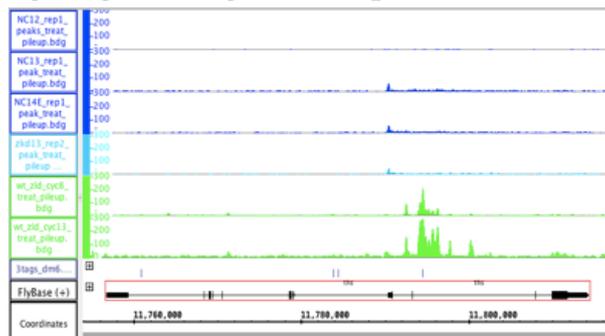


Figure 1: Thisbe (ths)

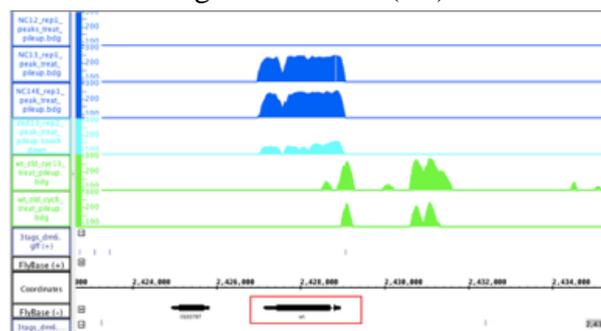


Figure 2: Giant (gt)

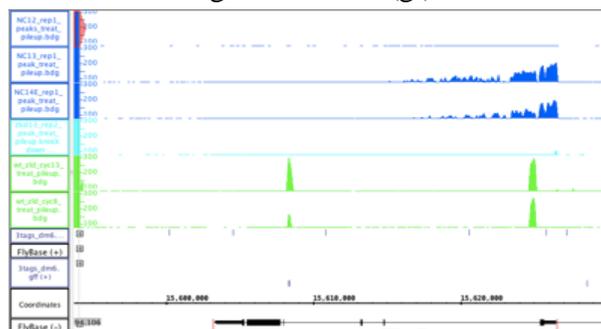


Figure 3: Short gastrulation(sog)

Ths is a patterning gene localised on chr2R: 11,790,366-11,812,015, and, according to RNA Pol II chip, *ths-RA* is the one out of two isoform that is expressed. The majority of the gene is composed of introns, so only a small portion of the 21,650 bp gene is expressed to synthesise functional protein. Furthermore, the *zld* peak and its enhancer are both situated in the introns region instead of 2kb upstream of TTS. There are 4 TAGteam sites within *ths*, but only 1, CAGGTAG, attracts *zld* protein binding and is located 4135 bp downstream of TSS. *zld*'s enrichment is the lowest at the beginning of MZT (cycle 8-9). However, as illustrated in JASPAP's RNA polymerase chip-seq tracks (colour indigo), transcription presents in none of the cycles, suggesting that *ths* is not expressed. The *zkd* track (blue) is not affected, for it is equally not transcribed.

Gt is a relatively small gene situated on the negative strand of chrX:2,427,113-2,428,967. *Gt* is mainly composed of exons, with one CAGGTAG site upstream. *Gt*'s expression is first initiated in NC 13 and more strongly expressed in the later stage of MZT. It is apparently relying on *zld*'s regulation, as indicated by the subdued expression on *zkd*, but the moderate level of expression denotes the presence of other transcription factors.

Sog is located on the negative strand of chrX:15,604,576-15,626,540. Similar to *ths*, *sog* is a relatively large gene mainly composed of introns, where the enhancer and *zld* peak are. It has one CAGGTAG site located 100,545 bp downstream of the TSS, within an intron section. The *zld* peaks from mitotic cycle 8 to 14 are of similar heights and shapes, indicating that *sog* is enriched with *zld* throughout the MZT. However, its transcription does not begin until NC 13, which implies that *sog*'s expression is not mandated until NC 13, in addition to its requirement of other TFs' cooperation to be transcribed.

The reason why that patterning genes' TAGteam sites are usually localised within the gene' intron instead of following the *zld* binding sites' 2 kb upstream rule, is likely because *zld* is not directly involved in the activation of class 2 genes' transcription, so it is unnecessary for *zld* to bind in the proximity of RNA polymerase's promotor. However, since it potentiates the process, it binds to an enhancer, which can be anywhere around or within the gene.

Expression Pattern Explained in Relation to TFs

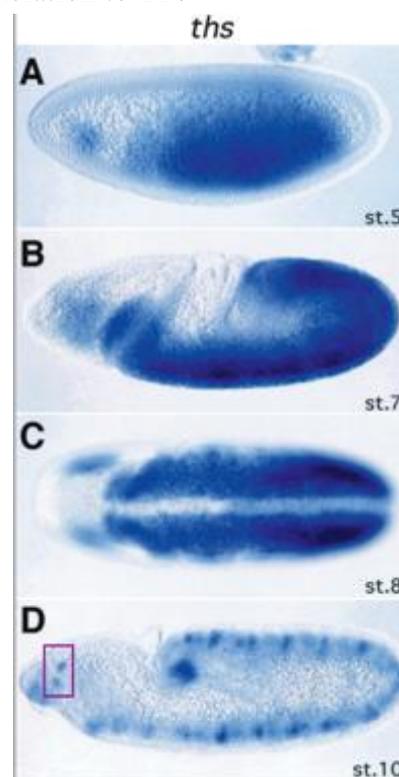


Figure 4. *Ths*' expression (a) stage 1-3, (b) stage 4-6, (c) dorsal view, (d) later embryonic development.

Source: Stathopoulos, 2004 [3]



Figure 5. Gt expression stage 4-6

Source: flyexpress



Figure 6 Sog expression (a) stage 1-3. (b) stage 4-6

Source: flyexpress

Throughout the course of development, *ths'* expression is mainly concentrated in the mesodermal region of the embryo, which includes the cardioblast, neuroblast, and other appendages' primordium. It's in situ hybridisation stain is visible in the germband and absent in the lateral-ectoderm region. From a dorsal view, *ths'* expression is absent from the midline. Transcriptional repressors *Sna* and *twi* inhibit the expression of genes promoting dorsal cell fates. Thus the combination of *dl*, *Twi*, and *Sna* in the ventral-most nuclei not only promotes mesoderm formation but also prevents cells from adopting the mesectoderm and lateral ectoderm fates. The darkened blue dots are attributable to the involvement of gap genes even-skipped (*eve*) and hunchback (*hb*), as they counteract the expression pattern by repressing each other at antagonistic regions [4]. Furthermore, the expression pattern of the elongated germ band is the result of multiple homeotic genes: Sex combs reduced (*Scr*) is expressed in prothoracic segments of the ectoderm, and *pb* expressed in the mouthpart anlagen. Lastly, *ths'* gene codes for observable patterns but also functional organ systems cardioblast and neuroblast, the formation of which is the combined result of TFs such as *tin*, *bap*, *Gsc*, and *onecut*.

Gt first becomes up-regulated throughout the anterior half of the embryo and in a cap of posterior cells. By stage 4-6, *gt* is lost from the anterior-dorsal-lateral regions of the embryo, where the future brain is patterning. These three stripes on the embryonic head correspond to the future ventral furrow lip, epipharynx and hypopharynx. Specifically, *gt* is expressed in antennal anlage, central brain anlage, dorsal head epidermis anlage, visual anlage, ventral ectoderm, head mesoderm. From previous works, it had been speculated that *gt*'s expression in the anterior and posterior are separately regulated. In this case, we see that high *hb* transcription stimulated by *Bicoid* in the anterior, forming a *hb* protein gradient, which likely is responsible for the activation of *gt* in the anterior region. Since *hb* does not reach its peak expression until NC12, explaining why *gt* has a low level of transcription at this stage as shown in the IGB.

Short gastrulation (*sog*) is a key dorsal-ventral patterning gene that promotes neural development by preventing other *dl* regulated proteins from diffusing into the neuroectoderm and suppressing expression of neural genes. It is involved in the development of neurogenic structure, and then migrate to be sparsely distributed in the elongated germband and mainly localised in the brain, visual, and wing discs.

The *dl* binding motif has the highest score and frequency, which precisely is expressed in regions other than dorsal-ventral gradient, where *sog* is localised. *Hkb* is transcribed in two domains in the cellular blastoderm, forming an anterior and a posterior cap, regions where *sog*'s lateral strip is absent. It's possible that *hkb* represses *sog*'s expression on a physiological level, but, functionally, it acts to regulate the development of neurogenic structures and brain epidermis. Lastly, *hb* and *bcd* are highly expressed in the anterior region and are activated relatively late, explaining *sog*'s stage 7-8's frontal focalised expression.

3.2 Class 1 Genes: esg, sc, tsg

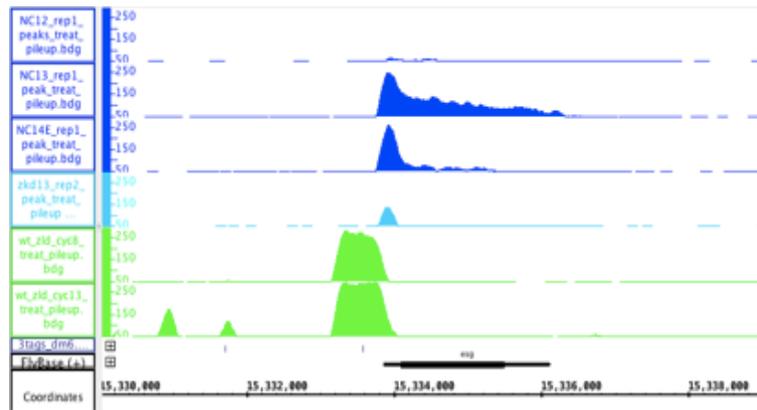


Figure 7: Escargot (esg)

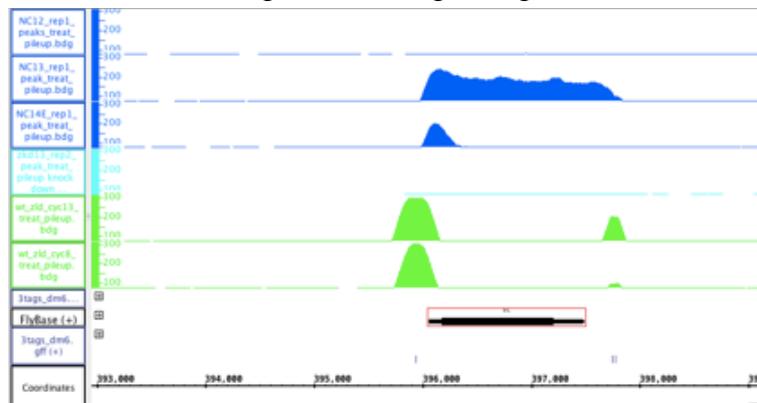


Figure 8: Scute (sc)

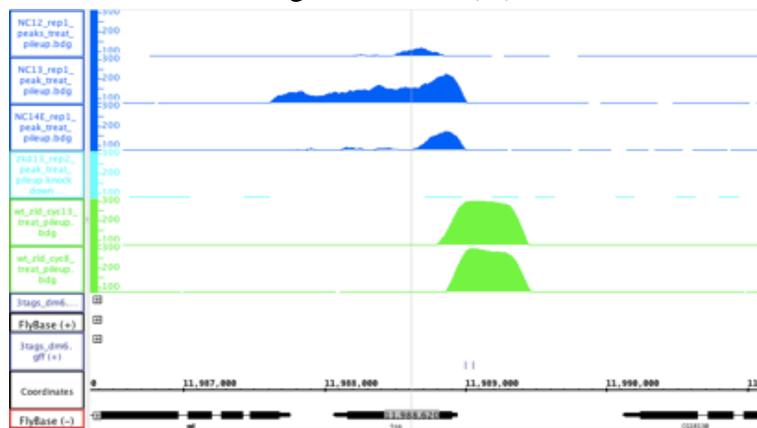


Figure 9: Twisted gastrulation(tsg)

Esg is localized on chr2L:15,333,866-15,336,138, with a single exon encoding an RNA polymerase II transcription factor that possesses five Zn²⁺-finger DNA binding domains. Sc is on the positive strand of chrX:396,060-397,497, with a CAGGTAG site 120bp upstream of its TSS. Tsg is located on the negative strand of chrX:11988062-11988948. There are 2 zld binding sites prior to Tsg, explaining the exceptionally broad peak, but both of them are on the positive strand, revealing that zld binding regulation can occur on the opposite strand of the gene.

Cycle 8 and 13 marks the beginning and end of MZT transition, and, as shown in IGB snapshots, zld transcription factor enrichment is high in both of these cycles, suggesting its activity throughout the MZT, which supports the previously discovered phenomenon that zld potentially regulate embryonic development as it's one of the earliest activated genes with long-term expression. Additionally, although zld is operative since mitotic cycle 8, all three class I genes' transcription does not start until

NC 13 then RNA polymerase’s activity diminishes immediately, shown by the absence of RNA polymerase peak in NC 12 and the relatively lowered peak in NC 14, indicating that their expression is relatively late in embryogenesis and each has only a short-expression duration.

Expression Pattern Explained in Relation to TFs

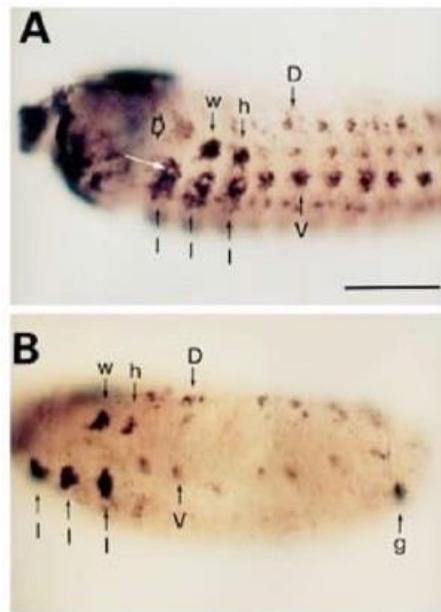


Figure 10 Esg expression (a) Stage 14 embryo. (b) Stage 16 embryo.

Source: Hayashi 2013 [5]

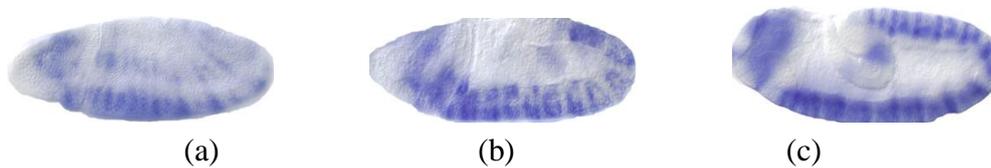


Figure 11 Scute expression (a) stage 1-3, (b) stage 4-6, (c) stage 9-10

Source: flyexpress

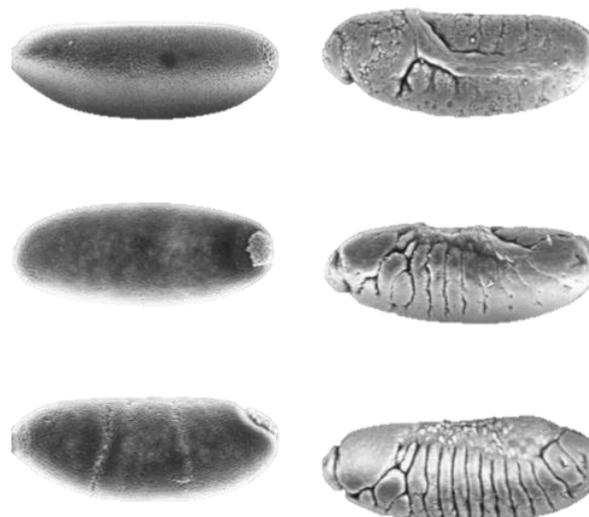


Figure 12 Tsg’s expression throughout the embryogenesis

Source: Hong 2015 [6]

Esg is a class I gene involved in early embryonic development, the expression of which is primarily ectodermal with a dynamic pattern, as it encodes for multiple anatomical features. esg transcript is detected in the presumptive wing, leg, genital, and haltere discs, as well as in the central and peripheral nervous system, cephalic and dorsal thoracic region, and digestive system cells (intestinal, gut, oesophagus). Specifically, nub is expressed in the wings and legs disc, and trx is involved in the formation of genital and haltere imaginal disc. The growth of central and peripheral nervous systems are regulated by slp1, bgb, fkh, and trx. Also, the dorsal thoracic region stained as shown in fig. 10 consists of drosophila's digestive and cardiovascular systems, explaining the presence of TFs tin, trx, fkh, and kni. In general, the pleiotropic gene esg's expression requires multiple TFs and cascade-like signalling functions that affect various targets.

Sc transcripts are responsible for the formation of embryonic nervous systems, wing discs, and a variety of sensory organs. Its expression is focused in the elongated germband region, which corresponds to a variety of central organ systems and later appendages' primordium. Tin, responsible for the development of cardiovascular tissues, is expressed chiefly in the germband. Trx is a central regulatory element, and it contributes to the formation of ventral nerve cord primordium and central brain primordium, but, more importantly, it also regulates the development of wing discs, haltere discs, and imaginal discs. Pdp1 regulate expression in muscles and circadian clock neurons, but its binding site is also highly enriched on sc. In summary, TFs with highly enriched binding sites on sc are all tightly related to sc's expression, yet each TF may only contribute to parts of the formation pattern.

Tsg is required for specification of a narrow strip of the cells towards the dorsal midline, which give rise to the amnioserosa. tsg first appears in stage 4 embryos, expressed chiefly in two domains—the anterior cap and a broad mid-dorsal area with the expressions continuous across the dorsal midline. Then towards stage 5, the expression has reconstructed into 4 shadow stripes in the mid-dorsal region as well as a paired domain in the anterior region. During stages 7 and 8, the anterior expression dims and the mid-dorsal stripes are located between the anterior and posterior transverse furrow [7]. By sharing common characteristics and interacting with sog, tsg plays a role in Bone Morphogenetic Protein signalling, and create a complex of an antagonist of BMP activity [8].

4. Discussion

Zld regulation

While both patterning and Class I genes contain zld peaks, they are regulated by zld differently. Class I genes' transcription depends on zld activation, which is why zld sites are located in 300 upstream of TSS and is at the vicinity of a TATA box. On the other hand, zld only potentiates patterning genes' transcription and binds to their enhancer. However, it's certain that, after the evaluation of 2 types of genes, zld's involvement in embryonic gene activation is vital, for, regardless it's a patterning gene or a class I gene, the zld peaks are significantly lower than the zld present tracks, signifying subdued or loss of expression.

5. Conclusion

Unlike patterning genes' TFs' functions can be hypothesised via observable features as stained in the in situ hybridisation image, class I genes TFs' interaction are deduced from their anatomical functions, for class I genes' expression is relatively more dynamic and are not localised to a specific pattern. Also, patterning genes are expressed throughout NC 13 and 14, but Class I genes are frequently transcribed only in NC 13.

6. Further Researches

For ths, its expression is tightly related to pyramus (pyr), we would like to research more on their linkage. Additionally, cardioblast and neuroblast developments are regulated by both ths and esg, we wonder if they employ the same pathway or independent ones. Moreover, all of the regulatory

pathways are deductions made based on genome browsers and preview researches, which render them relatively unsubstantiated, so future lab-work is needed before any reliable conclusions can be made. Methods such as transgenesis/mutagenesis and Yeast one-hybrid assay can be used to verify the hypotheses.

References

- [1] Liang, Hsiao-Lan et al. "The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*." *Nature* vol. 456, 7220 (2008): 400-3. doi:10.1038/nature07388
- [2] Nien, Chung-Yi et al. "Temporal coordination of gene networks by Zelda in the early *Drosophila* embryo." *PLoS genetics* vol. 7,10 (2011): e1002339. doi: 10.1371/journal.pgen.1002339
- [3] Stathopoulos, Angelike et al. "pyramus and thisbe: FGF genes that pattern the mesoderm of *Drosophila* embryos." *Genes & development* vol. 18,6 (2004): 687-99. doi:10.1101/gad.1166404
- [4] Wilson, Megan J., et al. "Giant, Krüppel, and Caudal Act as Gap Genes with Extensive Roles in Patterning the Honeybee Embryo." *Developmental Biology*, Academic Press, 21 Dec. 2009
- [5] Hayashi, Shigeo et al. "Control of imaginal cell development by the escargot gene of *Drosophila*." *Development* 118 1 (1993): 105-15 .
- [6] Hong, D.-H. S.-W. (2015, Aug 6). Midline enhancer activity of the short gastrulation shadow enhancer is characterized by three unusual features for cis-regulatory DNA. Hsiao-Lan Liang, C.-Y. N.-Y. (2008, October 19).
- [7] Mason E.D., K. K. (1994). NUCLEOTIDE SEQUENCE [GENOMIC DNA], FUNCTION, DEVELOPMENTAL STAGE, TISSUE SPECIFICITY. 1489-1501.
- [8] P Vilmos, K. G. (2001, Oct). The Twisted gastrulation family of proteins, together with IGFBP and CCN families, comprise the TIC superfamily of cysteine rich secreted factors. *Journal of Clinical Pathology*, 54(5), 317-323.