Proceedings of the 2nd International Conference on Modern Medicine and Global Health
DOI: 10.54254/2753-8818/45/20240511

Expression of gene CG13427 in drosophila embryo and its regulation by the transcription factor Zelda

Zixuan Wang

Shanghai WLSA Academy, Shanghai, 200433, China

audrey0808@163.com

Abstract. This study focuses on the moiety influencing gene expression in Drosophila embryonic development, including RNA polymerase II, transcription factor (TF) Zelda, and interactive neighboring genes. The CG13427 gene in Drosophila was used as an experimental model. Chromatin Immunoprecipitation (ChIP) was conducted to examine the DNA sequence targeted by Polymerase II and Zelda at different time points. Since previous evidence indicates that Zelda may play a key role in activating the zygotic genome, a well-controlled experiment here was designed to investigate whether the expression of adjacent genes would be interrupted if Zld-binding sites in one gene are abolished hence determining the correlation. Based on the research works done by Y. Ma and Rushlow (unpublished data), the exact binding site of Zelda in the CG13427 was found at the CAGGTAG motif. Therefore, a series of experiments were designed to investigate how Zelda affects the expression of CG13427 in Drosophila embryos. Validation experiments were subsequently designed using gene editing tools to abolish CAGGTAG when the mutation was introduced to further verify the previous findings.

Keywords: GeneCG13427, transcription factor (TF) Zelda, CAGGTAG motif, Drosophila embryos

1. Introduction

During the early development of the organism, the expression level of genes in the embryo varies in different stages and locations. As a key factor of gene transcription, and this could be concluded from the basic description of transcription. RNA polymerases can be used in research to determine the expression level of a gene in the transcription region through its binding site. Such works normally could be carried out using chromatin immunoprecipitation (ChIP) techniques, which ChIP is an invaluable method for studying interactions between specific proteins or modified forms of proteins and a genomic DNA region [1]. With antibodies specifically targeting RNA Polymerase II (anti-Pol II). Pol II is responsible for transcribing genes into RNA, and a Pol II ChIP experiment can identify the genomic regions where Pol II is actively transcribing genes. Previously, a study done by Nien et. al has shown one single transcription factor Zelda may suffice to activate the zygotic genome. Zelda binds to CAGGTAG and related motifs which were referred to as TAGteam sites from *in vitro* experiments. By a ChIP experiment using a monoclonal antibody targeting Zelda, the DNA sequence that Zelda stuck to was identified deciphering the gene expression in this region.

The focus of this article is on the gene expression of CG13427. According to the database from FlyBase, this gene, also known as Dmel\CG134270034514) in Fly Base, is a protein-coding gene in

DOI: 10.54254/2753-8818/45/20240511

Drosophila melanogaster. It has one annotated transcript and one polypeptide. The gene sequence is located at 2R: 20538877..20539316. Its temporal expression profile shows a range from extremely high to nearly zero expression detected, with peak expression observed within the 00–06-hour embryonic stages.

One peculiar aspect of this gene apart from traditional understanding is that it does not contain a necessary TATA box and is closely adjacent to a down-regulated gene, CG13428. The article hereby further elucidated on the full aspects of this gene, including its genetic features, regulation by Zelda plus its binding sites, and the interactive impact to its neighboring gene.

2. Methodology

2.1. In Situ Hybridization (ISH)

To observe and analyze the expression patterns of the specific gene CG13427 in embryos from Y. Ma's experiment and Rushlow's unpublished result, researchers employed the experimental methods of *in situ* hybridization and immunofluorescent staining. According to the study of Nien et al, In the experiment, embryos were fixed and subjected to *in situ* hybridization using digoxigenin-UTP labeled RNA probes synthesized from subcloned cDNA sequences or genomic intron DNA sequences. Additionally, immunofluorescent staining was performed using a primary mouse monoclonal anti-Zelda antibody and a secondary Alexa Fluor 488-labeled donkey polyclonal anti-mouse IgG. The embryos were then observed using fluorescence microscopy and Nomarski differential interference contrast microscopy. Furthermore, fluorescence *in situ* hybridization (FISH) was conducted using intron probes for sog and rho, a sheep anti-DIG antibody, and an Alexa Fluor 555-labeled donkey anti-sheep IgG secondary antibody. Finally, images were acquired using the Leica TCS SP5 confocal microscope[2].

2.2. CRISPR/Cas9 gene editing

The hypothetical experiment to mutate the CAGGTAG site to CTCATAG is performed using CRISPR/Cas9 gene editing technology with the assistance of designing tools such as SnapGene and Chop-chop. CRISPR/Cas9 (clustered regularly interspaced short palindromic repeat)-Cas9 is a multipurpose technology for genetic engineering that relies on the complementarity of the guideRNA (gRNA) to a specific sequence and the Cas9 endonuclease activity [3].

The first step is to design a guide RNA. Start by extracting 60 base pairs containing the CAGGTAG site from the gene sequence using SnapGene. This sequence is then copied and pasted into the "target" field in the Chop-chop. In Chop-chop, the species is set as Drosophila melanogaster (dm6), the tool as CRISPR/Cas9, and the purpose as knock out. After submitting the sequence, Chop-chop provides results that include the position, sequence, and score of potential guide RNAs. To select a suitable guide RNA, it should be considered that CRISPR/Cas9 requires a PAM sequence, which is NGG, to be adjacent to the target site for cleavage. And the chosen guide RNA should not contain the PAM sequence. By analyzing the results from Chop-chop, a guide RNA can be selected that meets this requirement and has a high score.

The next step is to design the donor plasmid. To accomplish this, the initial step is to design two primers specifically for in vitro mutagenesis. Firstly, it is crucial to determine the location of the target mutation site within the donor plasmid. Once the mutation site has been determined, the next step is to design two primer sequences. For illustration purposes, the following primer sequences have been designed based on the guide RNA:

5' ATCCTGTGACAGGTAGAGAAGAAAA 3' changed into

5' ATCCTGTGACTCATAGAGAAGAAAA 3' (primer1)

5' TTTTCTTCTCTATGAGTCACAGGAT 3' (primer2)

In the HDR (Homology Directed Repair) step, after Cas9 has completed the cleavage guided by the sgRNA designed in the first step, CAGGTAG can undergo precise mutation through naturally occurring

DOI: 10.54254/2753-8818/45/20240511

crossover. To be more specific, HDR is a DNA repair mechanism, a journal about CRISPR gene editing described this process in detail. “Homology-directed repair (HDR) — enzymes patch the break using donor template DNA. Researchers design the donor template DNA, which may include a desired sequence flanked on both sides by “homology arms” that match the sequence upstream and downstream of the cut. A complementary DNA strand is created during repair” [4]. To ensure that the mutation is homologous and as much as possible that crossover happens on both sides of the CAGGTAG, 2000 kb long homology arms are selected in SnapGene.

2.3. Applications for gene visualization

Integrated Genome Browser (IGB) is used in the study as a powerful tool for visualizing and analyzing genomic data, including ChIP-seq data for proteins such as Zelda and Polymerase binding. By aligning the peaks of Zelda binding with the gene sequence of CG13427, the Zelda-binding site can be localized. Furthermore, the expression pattern of genes in Drosophila embryos can be determined by examining the binding of Polymerase II on the corresponding image. The start and end of the coding strand, as well as +1 of CG13427, could also be identified on IGB.

FlyBase is used as a comprehensive database that provides information on the genomics of the gene CG13427, including the whole genomic sequence of CG13427. It also offers a wide range of functions and features of CG13427.

SnapGene as a molecular biology software is used in designing the Guide RNA and donor plasmid in the study. In addition, based on the information provided on IGB, a more visual gene map of CG13427 was constructed by annotating the Zld-binding site, +1 position, 5’ UTR, coding strand, and Guide RNA from the gene sequence in SnapGene.

Chop-chop is used for designing guide RNA for CRISPR/Cas9 gene editing in the study. In the designed experiment, after inputting 60 base pairs containing CAGGTAG in CG13427, Chop-chop identifies potential Guide RNA sequences in a given DNA sequence and scores each sequence for reference.

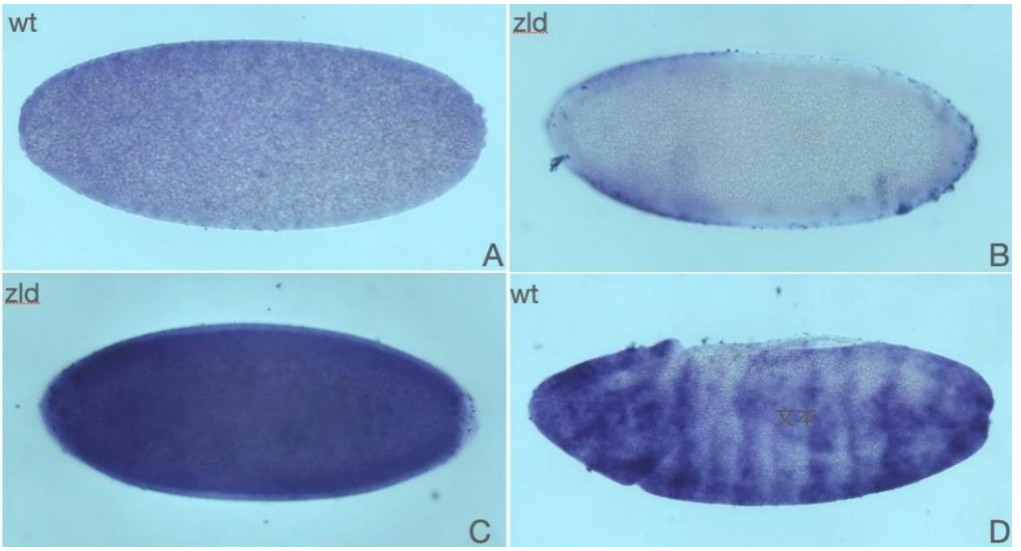
3. Results

3.1. Illustration of CG13427 expression in Drosophila embryo

Developmental geneticists turned to ‘network biology’ decades ago, q good example is the early patterning of the Drosophila embryo, a relatively well understood process involving the concerted action of about ten or so transcription factors (TFs)[5].

In the experiment conducted by Y. Ma, the Zld ChIP and Polymerase ChIP reflect the binding of Zelda and polymerase II to gene CG13427, indirectly indicating the gene expression levels. As shown in Figure (E), the Integrated Genome Browser (IGB) visualizes the experimental data, showing that in NC13, the binding level of polymerase II to CG13427 is comparatively the highest and the peak duration is the longest of all binding patterns. In NC12, the peak length is significantly shorter, and the binding level is lower (shown in pink). In zkd13, Polymerase II binding barely appears (shown in blue). The binding pattern of Zelda appears as a peak-like structure, indicating that the Zld-binding site is likely located at the peak position (shown in green).

DOI: 10.54254/2753-8818/45/20240511



E

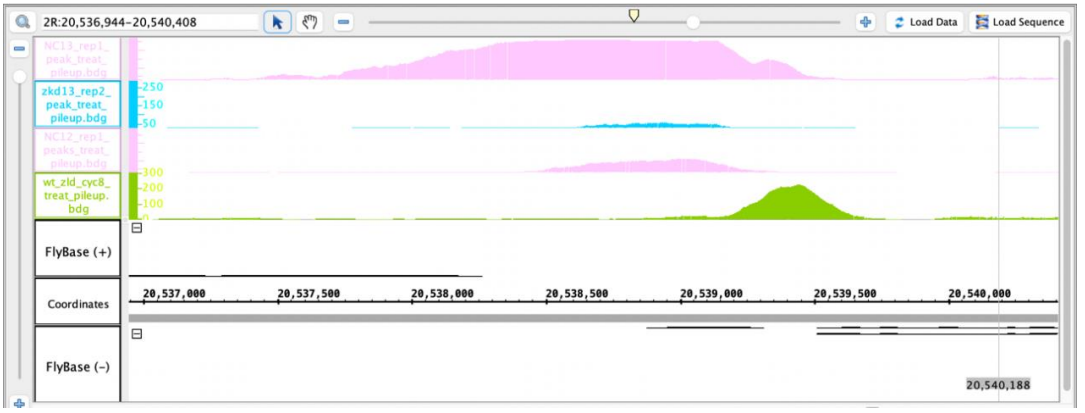


Figure 1. Zelda and polymerase II protein expression in Drosophila embryo. Wild-type (wt; A, B) and zld- (C, D) embryos were hybridized as indicated. (A) The early wild-type Drosophila embryo exhibits a pale purple hue. (B) The early zld- Drosophila embryo exhibits a very faint purple color, indicating low polymerase binding as shown in blue in Figure (E). (C) The older zld- Drosophila embryo exhibits a darker purple. (D) The older wild-type Drosophila embryo exhibits darker purple and distinct regions corresponding to the head and dorsal side.

According to the pattern of embryo developmental stages on FlyBase (Figure 2), it can be observed that CG13427 is actively expressed in 0-24h after egg laying, but hardly expressed in L1, L2, and L3 larva stages. At the same time, the expression is restored in adult eclosion, but the expression is not as active as 0-24. From this, it can be inferred that CG13427 mainly functions during the early growth of Drosophila, specifically during the embryonic stage. Additionally, it is observed that CG13428, located next to CG13427, shows a similar expression pattern to CG13427, except that CG13428 appears to be more active during the adult eclosion stage.

DOI: 10.54254/2753-8818/45/20240511

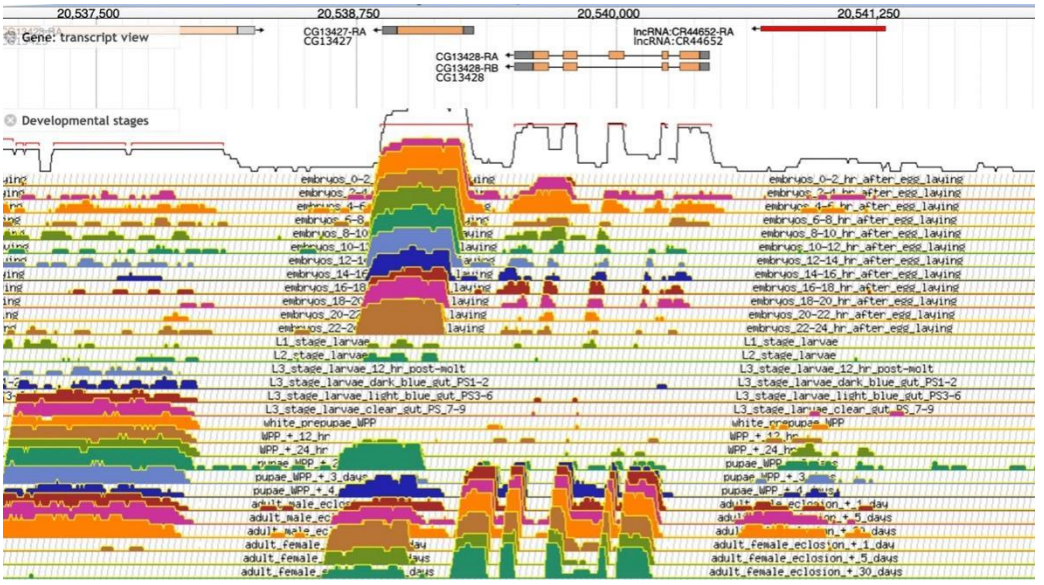
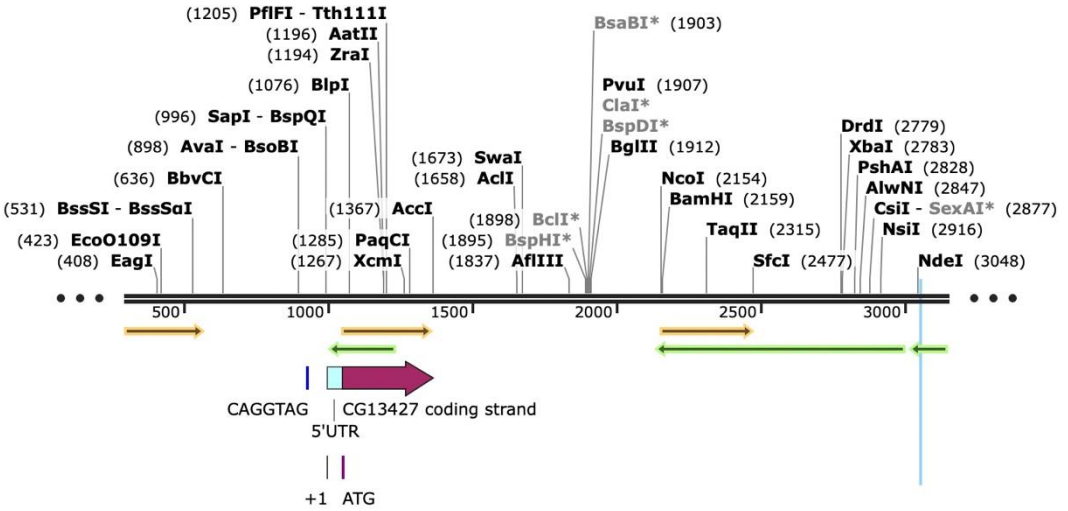


Figure 2. Expression of CG13427 and CG13428 in embryo developmental stages

3.2. Feature of CG13427 gene region

By observing the SnapGene map created based on the features of gene CG13427, it can be observed that the CAGGTAG motif, which is the Zelda binding site, is located at the +1 upstream of the gene's transcription start site (highlighted in dark blue). The 50kb long 5' UTR is located between the +1 site and the ATG start codon (highlighted in light blue), followed by a coding strand that is 318kb long. See Figure(G).



G

DOI: 10.54254/2753-8818/45/20240511

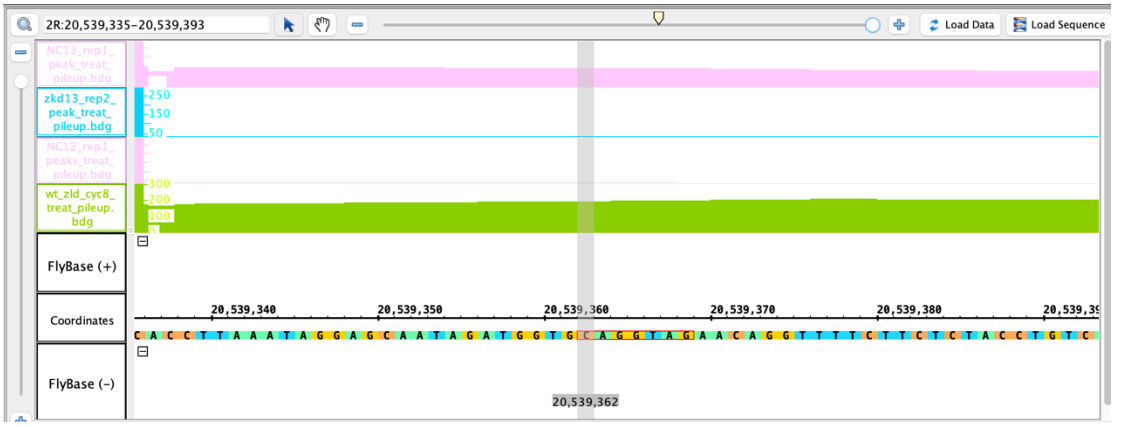
H 

Figure 3. The genomic feature of CG13427 and Zelda-binding peak (H) By corresponding the Zelda-binding peak with CG13427's DNA sequence, which could be inferred that the CAGGTAG site is the Zelda binding site on gene CG13427.

4. Discussion

4.1. Correlation of Gene Expression: from a Controlled Variable Experiment Targeting Zelda Binding Sites

Based on the observations from Figure 1, gene CG13427 exhibits a peak in expression levels during the embryo 0-24 hours after the egg-laying stage, while CG13428 shows minimal expression levels throughout all stages. However, it is interesting to note that the expression peak of CG13428 corresponds almost perfectly with CG13427. This raises the discussion topic of whether neighboring genes can influence each other's expression levels. To investigate this, several controlled experiments are designed using Zld binding sites as variables.

In the hypothetical experiments, the experimental targets are gene1 and gene2, and the control group consists of gene1 and gene2 with their Zld-binding sites knocked out. In the first group, the Zld-binding site of gene1 is knocked out while gene2's is preserved, and the expression of gene1 is observed. In the second group, the Zld-binding site of gene2 is knocked out while gene1's is preserved, and the expression of gene2 is observed. The real results should be analyzed after the experiment, but only the hypothetical results are analyzed here.

If in the first group, the expression of gene1 is not significantly down-regulated and there is a significant down-regulation in Zld-binding, it is likely that the expression of genes can mutually influence each other, meaning that an actively expressing gene may drive the expression of a neighboring gene with weaker expression. On the other hand, if the expression of gene1 is significantly down-regulated and there is a significant down-regulation in the Zld-binding site, it suggests that the expression of genes does not mutually influence each other. Another possibility is that the expression of gene1 is not significantly down-regulated and there is no significant down-regulation in the Zld-binding site, which may indicate that the Zld-binding site of gene2 affects the Zld-binding of gene1, but the expression of genes does not mutually influence each other. These three scenarios roughly cover all possible results and analyses that may arise in a real experiment, and the results of the second group would be similar to the first group, and so on.

Furthermore, in the real situation, the lack of significant expression in CG13428 is due to down-regulation, but the reason for this down-regulation is unknown. By examining its gene sequence, it is found that it does not contain any Zld-binding sites. Therefore, in the hypothetical experiment, a Zld-binding site could be inserted into the CG13428 sequence to observe its expression. If the gene starts to express, it will not only confirm the importance of Zelda in regulating gene expression but also that the expression of genes does not mutually influence each other (based on the second and third scenarios in the experiment mentioned above).

DOI: 10.54254/2753-8818/45/20240511

4.2. CG13427 expression is down regulated in Zld-binding site CAGGTAG mutant embryos

To further justify the cruciality of CAGGTAG as a Zld-binding site and identify the influence of mutation on the Zld-binding process. In the designed experiment, by gene editing, CAGGTAG is mutated to CTCATAG by CRISPER-cas9, as CAGGTAG is known to be the most common Zld-binding site, according to the study of Nien et al. "Enrichment indices of all possible heptamers were calculated for Zld-bound regions using a 500 bp window centered around the middle of Zld-bound peaks. Using an enrichment score of 3.5 as the cutoff, 11 heptamers were recovered (Figure S1B). The highest-ranking heptamer was CAGGTAG"[2]. After mutating to CTCATAG, Zld-binding in CG13427 is expected to reduce since CG13427 expression is down-regulated in *zld-* mutant embryos (see fig1). After mutation, if a reduction of Zld-binding is observed, it could be concluded that Zelda normally binds the CAGGTAG site and activate the gene expression.

On the other hand, if there is hardly significant down-regulation in Zld-binding or CG13427 expression in the *zld-* mutant embryo is observed, this result could be explained by the following: 1. CRISPER-cas9 in gene editing process didn't function well, in this case, the mutated gene sequence should be retrieved and confirm the result of the CAGGTAG mutation. 2. CTCATAG could be predicted to replace CAGGTAG as the new Zld-binding site. While this situation is relatively difficult to verify, many subsequent comparison data between genes containing CTCATAG and Zld-binding peaks are required. 3. Zelda itself has opened entrances for other binding sites, "Zelda is responsible for "opening" the enhancer regions so that other regulatory factors can gain access and bind to them".

5. Conclusion

Through the utilization of In situ Hybridization (ISH), CRISPR/Cas9 gene editing, and gene visualization tools such as FlyBase, SnapGene, and Integrated Genome Browser (IGB), this study has centered on the investigation of gene CG13427 from three distinct perspectives. The first pertains to the examination of gene expression during drosophila embryonic development, revealing that CG13427 predominantly operates during the early stages of Drosophila growth, specifically within the embryonic phase. The second aspect involves the characterization of the gene CG13427 itself, uncovering the presence of the CAGGTAG motif located at the +1 position upstream of the gene's transcription start site, a 50kb long 5' UTR situated between the +1 site and the ATG start codon, followed by a 318kb long coding strand. The third dimension of this research entails the exploration of the impact of mutant Zelda binding site on gene expression, with the hypothesis positing that CG13427 expression is downregulated in Zld-binding site CAGGTAG mutant embryos. Furthermore, the correlation of gene expression was comprehensively investigated through a controlled experiment involving CG13427 and its neighboring gene CG13428. In this hypothetical experiment, a Zld-binding site could be incorporated into the CG13428 sequence to monitor its expression, revealing valuable insights regarding the regulatory role of Zelda in gene expression as well as the independent nature of gene expression.

References

[1] Carey, M. F., Peterson, C. L., & Smale, S. T. (2009). Chromatin immunoprecipitation (chip). Cold Spring Harbor Protocols, 2009(9). doi:10.1101/pdb.prot5279

[2] 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. doi:10.1371/journal.pgen.1002339

[3] Arora, L. (2017, November 8). Gene editing and crop improvement using CRISPR-Cas9 system. https://pubmed.ncbi.nlm.nih.gov/29167680/

[4] Ray Centi et al. "Using CRISPR Gene Editing to Modify the LacZ gene in. E.coli." https://www.researchgate.net/publication/351664466_Using_CRISPR_Gene_Editing_to_Modify_the_LacZ_gene_in_Ecoli

[5] Adryan, B., & Teichmann, S. A. (2010). The developmental expression dynamics of drosophila melanogaster transcription factors. Genome biology. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2884543/