A Precursor to Genomic Engineering: Molecular Cloning Process of *Drosophila* Gene CG3165 into CRISPR Cas-9 Vector

Kiley Clark, Kaylen Karnes, Angelina Sherba, Arleigh Wood

BIOL324 Genetics

April 26, 2021

Abstract

The goal of the project was to create a tool using CRISPR Cas-9 that could knock out a gene in *Drosophila* cells. A guide RNA was designed using E-CRISP. The *Drosophila* plasmids were digested with the BspQ1 enzyme and the gRNA sequence was cloned into a vector. The vector had to undergo a series of purifications and annealing of the oligonucleotide gRNA sequences to prepare for ligation. The oligo sequences were ligated in the vector and the reaction was transformed with *E. coli*. Two of the four variants were successfully ligated into separate vectors at the intended genomic location, which was confirmed by DNA sequencing. The CRISPR-Cas9 tool with CG3165_19(F) and CG3165_15(R) was successfully made, but the ability for each specific CRISPR-Cas9 vector to perform a knockout is still unknown.

Introduction to Trex1

Drosophila melanogaster serves as the model organism for this study because of the similar, yet simplified genome comparable to that of humans. Using *Drosophila melanogaster* permits relatively easy genetic modifications to be made to genes with similar functions to that of humans, allowing conclusions to be made about human genes also. In humans, Trex1 is an exonuclease that repairs damaged DNA and indirectly affects immune responses. Previous studies support that Trex1 has an important role in repairing damaged DNA and regulating Type-1 interferon responses in the immune system. Mutations or deficiencies in this gene can trigger a variety of autoimmune disorders due to lack of checkpoint activity in the G2/M checkpoint (Yang, et al. 2007). The G2/M checkpoint requires the degradation of the ATM and ATR enzymes to continue through the cell cycle. A genomic analysis of human DNA showed that a loss of the Trex1 gene was correlated to the loss of other Type-1 Interferon inducing genes,

IFNα and IFNβ, are overexpressed of to compensate for the loss of Trex1 (Ivashkiv and Donlin, 2015).

The genome of *Drosophila melanogaster* expresses an equivalent gene with similar function to Trex1, CG1365, which is used as a model in this experiment. CG3165 and Trex1 have similar molecular functions as repair exonucleases in both humans and *Drosophila* and share similar protein alignments (FlyBase, 2021). Both CG3165 and Trex1 genes remove excess DNA that is not functional or has been damaged, which is vital to the survival of an organism and prevents errors in the genome from being expressed. Therefore, CG3165 is a suitable comparison for use in gene editing for the human equivalent Trex1. One of the major biological responsibilities of Trex1 is immune regulation and as a result, mutations of Trex1 can cause autoimmune diseases such as lupus (Malynn and Ma, 2010).

The gene also serves as a checkpoint for immune sensing of DNA that encodes for viral immunity and cancer immune resistance in humans (Marincola et al, 2017). Trex1 is involved with embryogenesis and organogenesis, so Trex1 mutations can induce random apoptosis or DNA degradation that can affect multiple organs (Arneth, 2019). Trex1 also has a role in viral immunity, specifically the degradation of viral RNA. If a patient were deficient in Trex1 or expressed a Trex1 mutation, many processes would be compromised and could have deadly consequences, such as HIV transmission and cerebral leukodystrophy (Wheeler et al, 2016). By studying the function of CG3165, the function of Trex1 and the mutations at this gene that cause imperative diseases can be better understood. To better comprehend the function of genes including CG3165, the gene editing tool CRISPR Cas-9 can be used by completing a knockout. CRISPR-Cas9 is a DNA sequence geared with a nuclease to cut desired target DNA. Using

CRISPR-Cas9, a specific sequence can be knocked out using the Cas9 protein, protospacer adjacent motif (PAM), and a specific gRNA sequence designed for the target sequence.

Through a knockout, a gene is made inoperative and the role of the gene in the organism can be understood. In the case of a CG3165 knockout in *Drosophila*, the organism would be compared to a wildtype organism with positive CG3165 expression. To determine the success of a knockout using CRISPR, the location of the gene must be known before the knockout. Then, a DNA sequence analysis of the genome after a CRISPR-Cas9 cut would indicate whether the gene was still in the same location or missing from the genome after the cut.

By creating guide RNA with the *Drosophila* gene, CG1365, and ligating it into a vector with promoter for Cas9 expression, CRISPR can successfully knockout the target sequence. Successfully knocking out the CG3615 gene and observing the effects of the model organism would display valuable information about the function of Trex1 in humans that lead to many harmful diseases. To do so, the target gRNA must first be created and molecularly cloned into the CRISPR-Cas9 tool and at the correct genomic location. This experiment describes the molecular cloning process of CG3165 guide RNA into the CRISPR-Cas9 vector on the 2L chromosome in preparation for a knockout.

Methods

Designing Guide RNA

After the Trex1 gene was identified as the human gene of interest, the homologous gene in *Drosophila* was identified using FlyBase (FlyBase.org) as CG3165 for the CRISPR-Cas9 knockout. A guide RNA sequence was created using the CRISPR design tool E-CRISP (e-

crisp.org) to ensure on-target activity during gene editing by CRISPR-Cas9. The gene name, CG3165, was entered into the E-CRISP design tool and the sgRNA sequence with the highest specificity and efficiency score was selected and used to determine where gRNA was binding the DNA sequence. To design the forward oligonucleotide, TTC was added to the 5' end of the sequence located 23 nucleotides upstream of the NGG codon. Then NGG was removed, and a single G nucleotide was added to the 5' end to allow transcription from the U6 promoter. The reverse primer was made by adding AAC to the 5' end of the reverse complement of the target sequence. The chromosome, genomic start, and genomic end were recorded for further use as a checkpoint throughout the experiment.

Molecular Cloning and Restriction Digest

The circular plasmid had to be linearized and digested with the BspQ1 enzyme in order to molecularly clone the gRNA sequence into the vector at the sticky ends of the restriction enzyme. To complete the restriction enzyme digest, 10 µL of 2 uL pAC-sgRNA-Cas9 (from Addgene, 49330) was added to a 0.5 mL centrifuge tube with 10 µL of BspQ1 Enzyme mix (from NE Biosciences, R0712), water, and buffer. The mixture was gently blended using a pipette and incubated for 2 hours at 50℃. After the incubation period, the sample was stored at -20℃ after 4 hours to preserve the enzyme mix. After the incubation period, the newly linearized plasmid was observed for the presence of restriction enzymes EcoR1 and Not1.

Purification of Linear Vector

The vector had to be purified to rid the sample of unnecessary enzymes and salts from the enzyme mix in order to continue with ligation and annealing of the forward and reverse oligonucleotide sequences. To purify the digested pAc-sgRNA-Cas9 vector, 44 µL of Binding

Buffer was added and mixed with the enzyme mix. The mixture was then added to a spin filter column with a charged silica matrix to separate DNA from the BspQ1 enzyme and spun for 1 minute at 13,000 rpm. The flow through was discarded and 200 µL of DNA Wash Buffer was added to the column and spun again for 1 minute at 13,000 rpm. Two more separate washes were completed with the wash buffer and binding buffer. The column was then transferred to a new 1.5 mL tube and 20 µL of sterile water was added. The Nanodrop was used to determine the success of the purification by measuring the concentration of DNA (ng/µL) and absorption at 260/280.

Annealing of gRNA Oligonucleotides

The forward and reverse oligonucleotide gRNA sequences were phosphorylated and annealed to create the double stranded DNA with the sequences that code for the gRNA sequence. The forward sequence for CG3165-19 (5'- TTCGGTCTCTAGGTCCAGAA-3') and reverse (5'-AACTTCTGGACCTAGAGACCACC-3') and forward sequence of CG3165-15 (5'-TTCGCCATTAAGAGCGCAACTGG-3') and reverse (5'-AACCCAGTTGCGCTCTTAATGGC-3') were annealed in the reaction. The enzyme T4 Polynucleotide Kinase was needed to manually add phosphates to the 5' end of the sequence because this DNA is not in a living organism. To create the kinase mix, 4 µL of Oligo 1 (100 µM), 4 µL of Oligo 2 (100 µM), 32 µL of T4 Polynucleotide kinase mix were added to the PCR tube containing the DNA mixture. Then the mixture was annealed in a thermocycler at 37℃ for 30 minutes and then heated to 95°C for 5 minutes. Lastly the mixture was cooled to 25°C at 5℃/min to allow the double stranded DNA to recombine in preparation for ligation into the Cas9 vector.

Ligation of Oligos in pAc-sgRNA-Cas9 vector

The oligo sequences were ligated into the pAc-sgRNA-Cas9 vector using the enzyme T4 DNA ligase to continue the molecular cloning process. The oligos were diluted tenfold with sterile water by mixing 45 µL of sterile water with 5 µL of annealed oligonucleotides from the PCR tube. A LIG+INS tube with gRNA oligos and a tube with no no gRNA oligos LIG-Neg were used to determine the success of vector digestion and phosphorylation. In the LIG+INS tube, 5 µL of digested PAC vector, 5 µL of 1:10 oligos, and 10 µL of ligase mix were added. In the LIG-Neg tube, 5 µL of digested PAC vector, 10 µL of ligase mix, and 5 µL of nuclease-free water were added. Both tubes were mixed with a pipette and incubated at room temperature for 15 minutes. The two tubes were used in comparison to determine success of ligation and restriction enzyme digestion in *E.coli*.

Transformation of E. coli

Approximately 15 minutes after the ligation into the vector, chemically competent E. coli was transformed with the ligation reaction onto agar plates to determine if the circular plasmid with the gRNA oligonucleotides was successfully taken up by E. coli. One tube with 25 µL of NEB 5-alpha Competent E. coli cells was combined with 5 µL of LIG+INS, then mixed and placed one ice for 10 minutes. In a separate tube, 25 µL of NEB 5-alpha Competent E. coli cells was combined with 5 µL of LIG-NEG, mixed, and placed on ice for 10 minutes. The samples were heat shocked for 30 seconds at 42 ℃ and then placed on ice again for 2 minutes. Then 450 µL of room temperature LB was mixed into each sample, placed at 37 ℃ for 45 minutes, then shaken at 250 rpm. Agar selection plates with LB containing ampicillin were heated to 37℃, the optimal bacterial growth temperature. The cells were mixed thoroughly once more and 250 µL of

each tube was placed on the selection plate. The plates were stored at 37℃ overnight and stored at 4℃ until the colonies could be counted. The presence of colonies on each plate indicated whether the bacteria took the plasmid, and therefore the success of the restriction digest and ligation were successful.

Plasmid Purification

After the transformation, the molecular cloning process continued with a purification of the plasmid from an E. coli colony grown on the LIG+INS plates. First, a frozen pellet of DH5alpha E. coli was collected from the LIG+INS plate. The pellet was resuspended in 250 µL of resuspension buffer and vortexed until the pellet was completely resuspended. To dissolve the cell and bacterial DNA, 250 µL of lysis buffer was added and inverted 5 times. Then to neutralize the lysis buffer, 350 µL of neutralization buffer was added and the sample was centrifuged for 5 minutes to separate the DNA. The clear supernatant was transferred to a spin column and centrifuged for 1 minute at maximum speed. The flow through was thrown out and the previous step with the wash buffer was repeated. The column was transferred into a 1.5 mL centrifuge tube and 50 µL of sterile water was added, then centrifuged again for 1 minute. The nanodrop was used to quantify the purity of the sample by recording the concentration (ng/µL) and absorbance at 260/280. The volume needed to deliver 800 ng of DNA to a sequencing tube was calculated based on the DNA concentration found on the nanodrop.

Diagnostic Digestion, Sequencing, Transfection

A diagnostic digest was completed to determine if the gRNA target sequence was molecularly cloned into the vector successfully. The restriction digest gave DNA fragments from both the original vector and new pAc-sgRNA-Cas9 vector with the gRNA insert. In a centrifuge

tube, 4 µL of recombinant vector, 10 µL enzyme mix (HindIII, Xho1, BspQI), and 6 µL of sterile water were added and incubated at 37 ℃ for 20 minutes. The DNA mixture was heated to 50℃ for another 20 minutes. Then 4 µL of 6X DNA loading dye was added and mixed in preparation for gel electrophoresis. The gel chamber was loaded with about 275 µL 1X TAE. Each well was loaded with 5 µL of Quick Load Purple 100 bp DNA Ladder and 15 µL of each sample was loaded into separate wells. The gel was ran at 120 V for 30 minutes and the band widths were observed to determine presence of the gRNA insert.

A DNA sequence analysis was completed as a second line of evidence of the success of the molecular cloning process. DNA sequencing required a DNA template, primer, DNA polymerase, dNTPs, and a dideoxynucleotide. The vector sequences were aligned with the recombinant vector using the online gene editing tool Benchling. The results of the sequencing reaction between the template and DNA sequence were recorded.

Results

Preparation of Linear Vector

Flybase, an online repository for genetics of *Drosophila*, was used in order to find the gene in *Drosophila* that is most closely related to TREX1 in humans based on similar functions between the two genes. CG3165 was found to have the highest number of matches with the TREX1 sequencing. A match indicates similar protein sequences between the human gene, CG3165, and the resulting *Drosophila* gene, CG3165. The gene CG3165 exhibited the highest similarity in arrangement to the sequence of the desired human gene, Trex1 (**Figure 1**). The green lettering indicates a match between the human and fly gene. This alignment had a 35% similarity with a positive alignment score of 11/15.

Figure 1. Protein alignment of CG3165 (*Drosophila melanogaster*) and Trex1 (*Homo sapiens*) genes based on nucleotide sequence.

On the database E-CRISP, CG3165_15 and CG3165_19 nucleotide sequences had the highest specificity and efficiency score for the target gene, CG3165. While there are other CG3165 gRNA sequences, CG3165_19 and Cg3165_15 had the highest specificity and efficiency score in *Drosophila*. Exons in the CG3165-RB and Cg3165-RA coding regions were targeted due to their continuous expression. The genomic positions of CG3165_19, 15 are shown on the 2L chromosome in relation to the other sequences and coding regions on the chromosome

(**Figure 2**). The forward and reverse gRNA sequences from the oligos CG3165_19 and

Table 1. gRNA sequences created. The four oligonucleotide sequences were created in the 5' to 3' direction in order to be ligated into the vector.

CG3165_15 were created to ligate into the vector as the target sequence (**Table 1)**.

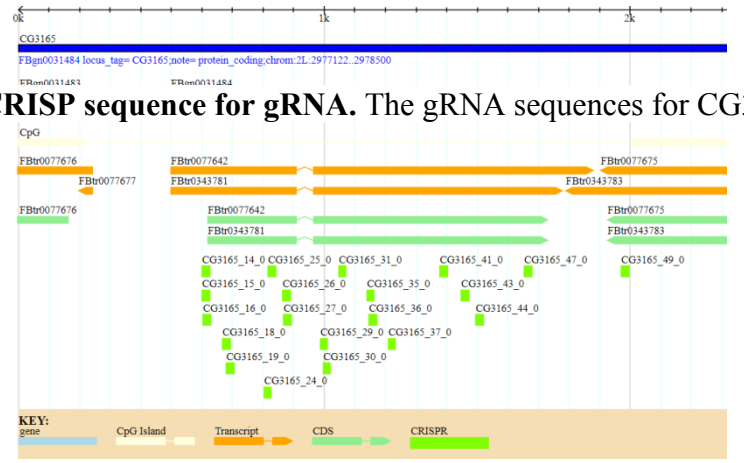


Figure 2. E-CRISP sequence for gRNA. The gRNA sequences for CG3165 on the 2L chromosome.

gRNA Variants	Nucleotide Sequences
CG3165_19gRNA_For	TTC GGT GGT CTC TAG GTC CAG AA
CG3165_19gRNA_Rev	AAC TTC TGG ACC TAG AGA CCA CC
CG3165_15gRNA_For	TTC GCC ATT AAG AGC GCA ACT GG
CG3165_15gRNA_Rev	AAC CCA GTT GCG CTC TTA ATG GC

The BspQ1 enzyme successfully cut and linearized the circular plasmid in order to insert the gRNA oligonucleotide sequences. After a purification of the linearized vector, a nanodrop was conducted to determine purity and concentration of the DNA. After purification, a 260/280 absorption measured the residual of contaminant enzymes that remained in the DNA at 260/280 nm. Purified DNA should be absorbed between 260-280 nm and have a measurement between 1.8 and 2.0. (Nanodrop reference). The 260/280 measurements from the purified vector samples were 1.94, 1.8, 1.89, and 2.02. The percent yield of purified DNA made was 34.4%, 15.8%, 32.8%, and 39% for forward and reverse sequences of CG3165-19 and CG3165-15. The preparation of the vector was a pivotal part of the molecular cloning process because this allowed for the insertion of the gRNA sequence at the target site.

Insertion of gRNA Sequences

The data from the transformation supported whether the restriction enzyme digest and ligation were successful based on the presence and absence of *E.coli* colonies (**Table 2**). The ligase positive plate contained the linear vector with the gRNA oligonucleotide sequence inserted, where cells should have been able to survive the ampicillin if the restriction enzyme digest and ligation were successful. The negative plate served as the control where no cells survive the ampicillin because they lack the ligase enzyme and gRNA DNA oligonucleotide. The reaction between the CG3165-19 Forward and Reverse vectors and *E.coli* created multiple

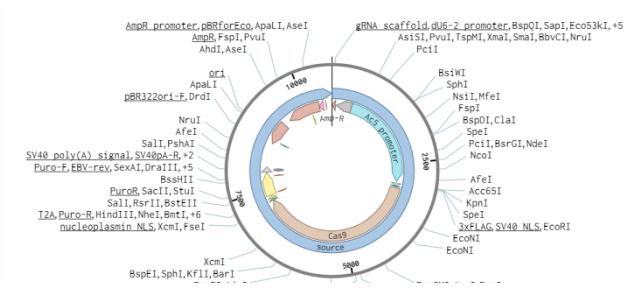
colonies on the positive plate and none on the negative ampicillin plate. The CG3165_15 Reverse ligation was partially unsuccessful, which is evident by the 2 colonies on the LIG-Neg plate. A purification was conducted to cleanse both purified vector samples of unwanted contaminants before being run on the Nanodrop. The concentrations of the CG3165_19 Forward, Reverse, and CG3165_15 gRNA DNA were above 300 ng/uL and the 260/280 for each sequence was above 1.90.

Table 2. DNA oligos plates with and without insert. The absence of any colonies on the negative plates is an indicator that the insertion was most likely successful. Each of the variants had no colonies of negative plates.

gRNA Sequence	LIG-Neg	LIG+INS
CG3165_19gRNA_For	0	5
CG3165_19gRNA_Rev	0	12
CG3165_15gRNA_Rev	2	18

A vector map was created to visualize the positions of the enzymes and coding regions within the vector using specific landmarks: the gene of interest CG3165_15 (**Figure 3A**) and CG3165_19 (**Figure 3B**) and restriction enzyme sites (BspQ1) as a guide throughout the molecular cloning process.

3A



3B

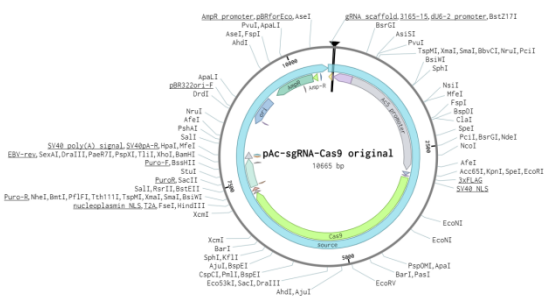


Figure 3. Vector map for CG3165. Vector maps show the markers for different enzymes and genes. 3A is an image of the vector with CG3165_15 and 3B is the vector with the CG3165_19 gene insert.

A sequential analysis of the DNA grown from *E. coli* colonies was completed to test if the target gRNA DNA sequence matched the current pAc-sgRNA-Cas9 vector DNA. A comparative representation of the sequencing between vector DNA and the original gRNA DNA is shown in **Figure 4**. The dark gray bars in the figure above show that the gRNA target sequence matches each nucleotide in the vector DNA. Light gray bars in place of the dark gray bars would indicate a discrepancy between the two sequences. The gRNA sequence in Figure 4A and 4B was inserted into the vector at the correct genomic location which is visible by the lack of the enzyme BspQ1 at the site of the gRNA sequence. Previously the BspQ1 enzyme was at the same location in the genome before being cut, then replaced with gRNA through ligation. The sequential analysis of the vector DNA and gRNA was one of the last conformational steps of the molecular cloning process.

4A

4B

Figure 4. Comparison of synthesized DNA and vector DNA. The DNA created in lab and the DNA from the vector were matched to see if the new DNA had the correct sequence. Each colored line shows a match nucleotide by nucleotide. 4A is the match for CG3165_19(F) and 4B is for CG3165_15(R). Only these two are mentioned because the other two were not successful.

A second confirmation of the success of the molecular cloning process was done through gel electrophoresis. Gel electrophoresis confirmed that the DNA matched the vector DNA. Three enzymes were used in the restriction enzyme digestion and all are present in Lane 1 (**Figure 5**). When certain base pair sequences are present, the enzymes will bind and cut the DNA. The two vector samples pAc5sg316519Cas9 (Lane 2) pAc6sg3165_19Cas9 (Lane 3) do not exhibit a band at 3,000 bp, showing the lack of enzyme BspQ1 from the original vector. This missing enzyme was cut and replaced by the gRNA sequence at the site where BspQ1 is expected (3000bp). The other two markers are expected at 6,000 bp (HindIII) and 1,000bp (Xhol), and both are visible at those bandwidths in the vectors. The BspQ1 enzyme was successfully cut

from the CG3165_19 vector sequences, which was a second confirmation of the success of two critical parts of the creation of this tool: the restriction enzyme digestion and ligation.

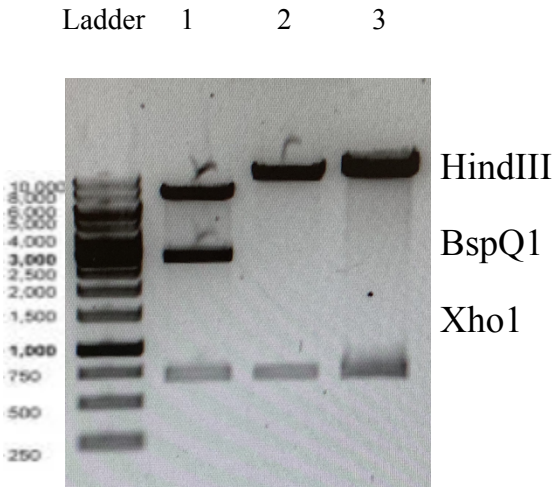


Figure 5. Gel electrophoresis results. The first well shows the three markers for the proteins that should be found in synthesized DNA. The second and third wells are the forward and reverse sequences of CG3165-19 that lack BspQ1.

Discussion

The purpose of the lab was to create a tool that can potentially knock out the CG3165 gene in *Drosophila*. The creation of the tool itself was successful via the molecular cloning process, however the ability of the tool to perform a knockout is not yet known. Purification after the transformation yielded 260/280 results that were between 1.8-2, meaning the DNA was relatively pure and the cloning process was perceived as successful up to this point. As seen in **Table 2**, the presence of colonies on the LIG+INS plate and lack of colonies on the LIG-Neg

plate confirmed the success of the restriction enzyme digest and ligation. No colonies on the LIG+ plate would show that the bacterium did not take the oligonucleotide sequence and no colonies would survive. For CG3165_19 (R,F) and CG3165_15 (R) there was more growth on the plates with the insert and none on the control plates, showing that the digestion of BspQ1 and insertion of the gRNA sequence was successful.

There were two methods used to confirm the success of the molecular cloning process of CG3165 into the vector: DNA sequencing and gel electrophoresis. Sequencing of the vectors and gRNA sequences was done to see if the desired target sequence aligned with the ligated gRNA DNA in the vector. A perfect nucleotide alignment for CG3165_19 (F) and CG3165_15 (R) was observed in the DNA sequence analysis. From this, it can be concluded that the molecular cloning process was successful because the target sequence was observed at the intended genomic location. The gel electrophoresis results also indicate that the creation of the CRISPR-Cas9 tool was a success. The absence of the BspQ1 shows that the sequence was successfully cut out and replaced with the gRNA oligonucleotide sequence.

A limitation of this project was that the tool has not been tested to see if a knockout of CG3165 can be completed. Although the molecular cloning of two oligonucleotides gRNA sequences was successful, the ability for the Cas9 enzyme to cut and knockout the target gene cannot be confirmed and must be tested in the future. *Drosophila* do not have complex genomes or systems like humans, so it is unavoidable that there are low matches for similar genes in *Drosophila* and humans. Another limitation is that CG3165 only has a 35% similarity to TREX1 in humans so while it is the highest match, it is not necessarily the best comparison for the Trex1 function in humans.

There is not a lot known concerning the role of CG3165 in *Drosophila* other than the 3'-5'-exodeoxyribonuclease activity (FlyBase) which helps maintain stability of the genome (Shevelev 2002). There are still many unknowns with Trex1 that can be further explored in future studies. One is that researchers do not know what cell types contribute to the autoimmune disorders that are triggered by Trex1 deficiency. Studies have found that Trex1 knockout in dendritic cells alone is enough to cause an autoimmune disease but not in B cells, cardiomyocytes, neurons, or astrocytes alone. Through research, scientists can determine how to treat diseases that originate from Trex1 deficiencies and understanding how the different cell types are impacted can be useful as well (Yan 2017). There is also a lack of understanding into how Trex1 deficiency results in DNA polynucleotides that are "disease-driven". They trigger an autoimmune disorder by preventing the degradation of DNA. Studies have already been done proposing models for how this is potentially done but there have yet to be any concrete results (Grieves et al. 2015). Trex1 deficiencies can cause a wide range of problems for people but there is still a lot of uncertainty concerning the full effect it can have. Further research is needed in order to gain a better understanding for the affects the deficiency can cause in the body and how it impacts treatment.

References Cited

FlyBase, 2021. FlyBase Gene Report. NCBI. http://flybase.org/reports/FBgn0085170.html

FlyBase, 2021. Gene: Dmel\CG3165. http://flybase.org/reports/FBgn0031484

Grieves, J. Fye, J. Harvey, S. et al (2015). Exonuclease TREX1 degrades double-stranded DNA to prevent spontaneous lupus-like inflammatory disease. *Proceedings of the National Academy of Sciences of the United States of America.* 112(16). 5117-5122.

Hall, B., Limaye, A., & Kulkarni, A. B. (2009). Overview: generation of gene knockout mice. *Current protocols in cell biology*, *Chapter 19*, Unit–19.12.17.

Lino, C. A., Harper, J. C., Carney, J. P., & Timlin, J. A. (2018). Delivering CRISPR: a review of the challenges and approaches. *Drug delivery*, *25*(1), 1234–1257.

Ramlee, M., Yan, T., Cheung, A. *et al.* (2015). High-throughput genotyping of CRISPR/Cas9-mediated mutants using fluorescent PCR-capillary gel electrophoresis. *Sci Rep* 5, 15587

Shevelev, I. Hubscher, U. (2002). The 3'-5' exonucleases. *Nature Reviews*. 364-376.

Uddin, F., Rudin, C. M., & Sen, T. (2020). CRISPR Gene Therapy: Applications, Limitations, and Implications for the Future. *Frontiers in oncology*, *10*, 1387.

Yan, N. (2017). Immune Diseases Associated with TREX1 and STING Dysfunction. *Journal of Interferon & Cytokine Research.* 37(5). 198-206.