Spring 2024

Exploring the Potential Pathogenicity of a Type 2 Diabetes Mellitus Associated INSR Missense Variant of Uncertain Significance Through daf-2 in the Caenorhabditis elegans Model

Brittany White
bwhite11@stu.jsu.edu

Exploring the Potential Pathogeneicty of a Type 2 Diabetes Mellitus Associated *Insr* Missense Variant of Uncertain Significance Through *Daf*-2 in the *Caenorhabditis Elegans* Model

A Thesis Submitted to the
Graduate Faculty of Jacksonville State University
in Partial Fulfillment of the
Requirements for the Degree of
Master of Science
with a Major in Biology

By

Brittany Nicole White

Jacksonville, Alabama

May 3, 2024

Brittany Nicole White May 3, 2024

Abstract

Type 2 diabetes mellitus (T2DM) is hallmarked by insulin resistance, with the *INSR* gene identified as a key player in this condition in humans. This gene is known to harbor genetic variants with a wide range of clinical significance from pathogenic to variants of uncertain significance (VUS) to benign. This project investigates a VUS associated with T2DM identified through ClinVar. A gene mutational analysis, predictive amino acid substitution analysis, and protein modeling predict *INSR* c.1628C>T (p. Thr543Met) to be likely pathogenic or damaging. PolyPhen-2 predicts this variant to be probably damaging (HumDiv score of 1.000). Evolutionary conservation of the VUS loci across multiple species was confirmed through multiple sequence alignment carried out through Benchling. The VUS results in an amino acid class change from the polar, hydrophilic amino acid threonine to the nonpolar, hydrophobic amino acid methionine. Utilizing SWISS-MODEL, homology-based protein models were generated for both human native and VUS mutant INSR proteins. Subsequent analysis in PyMOL revealed an RMSD value of 0.001. The affected amino acid resides in the extracellular alpha subunit, which may contribute to its potential impact to protein structure and function. To understand the functional implications of this structural alteration, we plan to observe potential phenotypic differences between wildtype N2 *C. elegan*s nematodes and a CRISPR-Cas9 engineered *daf-2* VUS nematodes. This research aims to elucidate the functional impact of the *INSR* variant and provide valuable insight into its role in the context of insulin resistance and T2DM.

Acknowledgements

I would like to thank my peers in the Worm Hole Lab, especially my fellow graduates Himani and Vanessa, for being a constant source of motivation, and my family and my "Well fam" for being supportive of me throughout this journey, especially Blake, Stacey, Isaac, Hannah, Leigh Anne, and my band mates. I also want to thank my mother and Granny for staying up with me on long nights, making me laugh when things got hard, and pushing me to never give up. I would like to give a big thanks to Dr. Ashley Turner for her insights and guidance as a mentor, along with my other committee members, Dr. Chris Murdock, and Dr. Jenna Ridlen. Without them, this project would not have happened. I would like to thank Trinity Elston for beginning this project and trusting me to take it on, as well as the Jacksonville State University Department of Biology, Dr. Hensley, and Dr. Lindblom for allowing me to continue this research and providing the resources to do so.

I also would like to thank the man who taught me to put God first, to pursue my dreams, and to never take myself too seriously. You taught me how to be innovative to find a way to accomplish the impossible, all while having fun. You modeled how to be a successful adult while still being a kid at heart. I wish you were here, Daddy, and I know you would be so proud.

Brittany White

Table of Contents

List of Tables

Page

List of Figures

Page

Introduction

Type 2 diabetes mellitus (T2DM) is a metabolic disorder hallmarked by insulin resistance (Wei et al., 2020). T2DM accounts for approximately 90% of all cases of diabetes (Goyal et al., 2023). T2DM affects most vital organs and is found to be associated with diseases such as Alzheimer's disease, certain cancers, as well as hypertension and other forms of cardiovascular disease (Haeusler et al., 2018, Baghaie et al., 2023). T2DM may also lead to the development of neuropathy, retinopathy, and nephropathy. This can lead to end stage renal disease, blindness, and loss of lower limbs, respectively (Goyal et al., 2023). Homeostasis of plasma glucose levels is a complex process, with glucose uptake into cells being limited in patient with T2DM (Wei et al., 2020). In T2DM, the cells do not respond properly to insulin. Insulin increases the permeability of the cells to allow for glucose intake. While not the only contributing factor to insulin's inability to maintain glucose homeostasis, this study will focus on the inability of insulin to bind to the insulin receptor. When this binding is impaired, glucose intake into the cells does not occur. As such, the glucose remains in the blood plasma and levels begin to rise. This results in the loss of insulin's ability to properly maintain glucose homeostasis (Figure 1, Panel I). Insulin production by pancreatic beta cells is increased in response to increasing serum glucose levels, which in turn causes the beta cells of the pancreas to change and even become dysfunctional (Goyal et al., 2023). As insulin levels continue to increase, cell receptor sensitivity decreases, which leads to peripheral insulin resistance (Baghaie et al., 2023). While environmental factors such as obesity play a role in the development of T2DM, there could also be a genetic

contribution to disease development. One study found that defects in both insulin-dependent and insulin-independent mechanisms for glucose uptake could be detected more than a decade before the onset of the clinical disease. This same study found a high incidence of diabetes in first degree relatives of individuals with T2DM, including among identical twins. Researchers also found that factors such as insulin resistance, insulin deficiency, beta-cell failure, obesity, and excess caloric intake can act synergistically with genetic components, leading to the development of T2DM (Martin et al., 1992). Though not the only contributing factor, obesity has been found to be a major contributor to an increased risk for the development of T2DM. Obesity, particularly in patients with excess abdominal adipose tissue, can increase the risk for the development of T2DM as adipose tissue can promote insulin resistance through various inflammatory processes (Goyal et al., 2023). Obesity is inversely related to the concentration of insulin receptors on the surface of cells as well as plasma insulin levels (Payankaulam et al., 2019). Insulin resistance also plays a key role in the development of T2DM, with it being a main factor to separate prediabetic and nondiabetic patients (Martin et al., 1992). This separation between prediabetic, nondiabetic, and diabetic is defined by the American Diabetes Association. According to the ADA, an A1C level of 6.5% or higher is considered diabetes, while 5.7% to 6.4% is considered prediabetic levels. Levels under 5.7% are considered nondiabetic. Similarly, fasting plasma glucose (FPG) levels can be tested. The ADA states that a FPG of 126 mg/dL or higher is considered diabetic. 100mg/dL to 125 mg/dL is considered prediabetic levels, and below 100mg/dL is considered nondiabetic, or normal (Table 1) (Malkani and Mordes, 2011). Insulin resistance is also synonymous

with hyperinsulinemia. It is notable that insulin resistance is shown to dramatically decrease after periods of fasting. This could point to the role of nutrition in the development of T2DM (Haeusler et al., 2018).

As previously stated, one role of insulin is to increase the permeability of the cell membrane to allow more glucose to move intracellularly from the plasma. This occurs in tissues such as muscular tissue, adipose tissue, liver tissues, and neuronal tissues. Insulin also acts as a negative regulator of insulin signaling; as insulin concentration in plasma increases, insulin receptors on peripheral cell membranes downregulate causing a decrease in signaling for insulin release and termination of the insulin bond to the receptor (Haeusler et al., 2018). The human insulin receptor gene (*INSR*) is the first critical node in this insulin signaling pathway (Taniguchi et al., 2006). *INSR* also plays a key role in metabolic regulation, control of growth, and neuronal function. At one point, the *INSR* gene was believed to be a house-keeping gene, but recent evidence shows that *INSR* is heavily regulated with dynamic transcriptional regulation. These regulatory elements are found both within the promoter and in the intronic region of the DNA sequence. Large DNA sequences are dedicated to encoding feedback for this sophisticated signaling system (Payankaulam et al., 2019). Furthermore, the. structure and placement of the insulin receptor could indicate its function in insulin signaling. The receptor protein is a single pass transmembrane receptor, crossing across the lipid bilayer of the cell membrane. (Payankaulam et al., 2019). Evidence of the insulin receptor's role in the insulin signaling metabolic pathway can be seen upon knockout of the receptor gene. This leads to metabolic dysfunction and impaired muscle growth (Haeusler et al.,

2018). The mammalian insulin receptor is a heterotrimer formed from two alpha subunits and two beta subunits. The alpha subunits are linked to each other via extracellular ligand binding domains and are the site of insulin binding (Murphy and Hu, 2013, Payankaulam et al., 2019). The alpha subunits also exhibit affinities for both insulin and insulin-like growth factor- II (IGF-II), making it stronger in glycogen synthesis promotion and in metabolic effects than the beta subunits (Haeusler et al., 2018). The beta subunits contain transmembrane and intracellular tyrosine-kinase domains. Cysteine on each alpha subunits link together and to beta subunits through disulfide bonds. Binding of insulin to receptors on the alpha subunit creates a conformational change on the alpha subunits and, sequentially, the beta subunits (Payankaulam et al., 2019). In this, insulin binding induces autophosphorylation of tyrosine residues. This action activates the tyrosine kinase activity intrinsic to the beta subunit of the insulin receptor (Murphy and Hu, 2013). Tyrosine residues in the protein are trans-phosphorylated and serve by acting as binding sites for insulin receptor substrates. These substrates can act as adaptor proteins (Payankaulam et al., 2019). These metabolic actions of insulin, such as glucose uptake by the cells and glycogen synthesis, require serine and threonine residues to be phosphorylated (Haeusler et al., 2018). Phosphorylation of insulin at the receptor leads to the production of insulin substrate proteins which activate the P13K-AKT/ protein kinase ß (PKB) pathway and the Ras-mitogen activated protein kinase (MAPK) pathway (Taniguchi et al., 2006). The P13K enzyme activate the serine and threonine kinase activity of AKT which is responsible for modulation of glucose uptake by glucose transporter type 4 (GLUT4), glycogen synthesis by GSK3, protein and fat synthesis by mTOR, and gene expression by

the forkhead family box genes (FOXO) (Haeusler et al., 2018). Particularly the exocytosis of GLUT4 allows for the increased permeability of glucose through the cell membrane. Protein tyrosine phosphatases (PTP) then dephosphorylate the tyrosine residues to decrease the activity of insulin (Taniguchi et al., 2006). For a simplified schematic of insulin signaling (Figure 1, Panel II).

The *daf-2* gene family is one part of the major metabolic pathways that help regulate lipid, protein, and energy metabolism in the nematode *C. elegans*. This gene family can also aid in response to stresses and in developing cell structure, such as heat stress and starvation (Halaschek-Wiener et al., 2005). *daf-2* encodes for an insulin receptor family member, hinting to its role as a regulator of metabolism (Kimura et al., 1997). Similar to insulin signaling in humans, *daf-2* kinase will autophosphorylate its own insulin receptor substrate (IRS-1) domain (Thomas and Inoue, 1998). Agonist insulin-like peptides bind to the *daf-2* receptors, recruiting insulin receptor substrates and activating the age-1/ p13k pathways, thereby increasing the level of phosphodylinositol (3,4,5) triphosphate (PIP_3). PIP_3 will become phosphodylinositol (4,5) biphosphate (PIP_2), activating the kinase cascade. This cascade consists of several key kinases, such as 3-phosphoinositide dependent protein kinase-1 (PDK-1), protein kinase B (akt-1/-2), and serum and glucocorticoid inducible kinase 1 (SGK-1). This cascade then phosphorylates and inactivates downstream *daf*-16 and FOXO transcription factors, thus promoting translocation of daf-16 from the nucleus to the cytosol of peripheral cells (Altintas et al., 2016). Nothing is known about the functions of the two isoform daf-2 proteins formed from alternative splicing. One study did find that *daf-2* is highly

expressed in neuroendocrine cells and in the nervous system of the nematode (Murphy and Hu, 2013). In another study, knock-in fluorescent receptors were used to identify the intestine as the main site for *daf-2* to *daf-16* signaling. This study also determined this signaling to have a major role in lifespan regulation in *Caenorhabditis elegans*, with decreased signaling leading to a 94% extended lifespan (Zhang et al., 2022).

The *daf-2* pathway is exclusively responsible for the regulation of life span in the adult nematode, and it plays a key role in larval development (Altintas et al., 2016). Downregulation of the *daf-2* pathway leads to a doubled lifespan. This pathway also upregulates the stress response to antimicrobials as well as upregulating genes associated with metabolism while simultaneously downregulating genes known to shorten the lifespan (Halaschek-Wiener et al., 2005). Loss of function found with *daf-2* mutants yields resistance to a variety of bacterial pathogens. Part of this could be due to serine-threonine kinases activity to regulate lifespan and resistance to pathogens, though the mechanism of this remains unknown (Evans et al., 2008). This pathway also regulates reproduction and lipid metabolism. It also controls the pathway to enter developmental diapause, forming a metabolically arrested dauer larva, making *daf-2* the central regulator of dauer formation (Halaschek-Wiener et al., 2005). This trigger to enter the dauer larval state is made by a decrease in the *daf-2* signaling pathway (Gottlieb and Ruvkun, 1994). These dauer larva are in a third stage of larval development, which could also be thought of as a diapause stage of arrested development of the larva until conditions are better fitting for development to continue occurring. As such, the dauer stage may be induced by stressful environment, such as food limitation, high population size, and high

Temperatures (Thomas and Inoue, 1998, Halaschek-Wiener et al., 2005). In the dauer larva stage, *C. elegans* can survive up to several months in a harsh environment (Thomas and Inoue, 1998). *daf-2* mutants have also exhibited an increase in thermotolerance and may express lower levels of transcripts associated with metabolism (Murphy and Hu, 2013, Halaschek-Wiener et al., 2005).

There are similarities between the sequences and phenotypes of *age-1* and *daf-2*, suggesting these genes participate in similar pathways as that of insulin signaling (Figure I, Panel III). *daf-2* remains the only homolog of its mammalian insulin receptor family (Gottlieb and Ruvkun, 1994). In one study, *daf-2* mutants were found to store fat and glucose in a form of glycogen. This same study found *daf-2* insulin signaling and mammalian insulin signaling to have opposing effects on metabolism, though both can cause a shift towards fat deposition (Thomas and Inoue, 1998). A decrease in *daf-2* signaling may induce metabolic and developmental changes analogous to metabolic control by the insulin receptor found in mammals (Kimura et al., 1997). The *insulin-like growth factor-1* homolog *daf-2* has been found to be a principal component on lifespan, with deceased signaling of *daf-2* being associated with an increase in lifespan (Halaschek-Wiener et al., 2005). In one study, it was found that glucose drastically shortens the lifespan of *C. elegans daf*-2 mutants by inhibiting FOXO family members, specifically *daf*-16 and the heat shock factor *HSF-1*. Insulin downregulates similar glycerol channels in mammals by inhibiting the expression of aquaporin glycerol channels. This would suggest glucose response and the pathways involved are evolutionarily conserved (Lee et al., 2009). These similarities show the *C. elegans*

species serves as an effective model for insulin signaling and could inform investigations into mammalian insulin signaling and other diseases associated with mutations and dysfunctions of the *INSR* gene (Murphy and Hu, 2013).

C. elegans remains an effective choice for modeling human diseases by transgenesis, while also having the ability to generate null animals. While the complex organs found in mammals may be lacking in the nematode, these complex phenotypes associated with various mammalian diseases can be reduced and studied by their component parts and molecular events. This species also has a completely sequenced genome that shares a vast homology with mammals. They are also relatively easy to maintain, cost effective, and have a short lifespan. These reasons along with the ability to process large numbers of animals simultaneously further validate using *C. elegans* as a choice model (Silverman et al., 2009). This serves as the premise of this project. We aim to explore the pathogenicity of a variant of unknown significance in the *INSR* gene. A mutation was identified and located at c.1628C>T (p.Thr543Met) in the *INSR* gene in humans. A conserved genetic loci was also observed in the *daf-2* gene in *C. elegans* at c.2084C>T (p.Thr698Met) (Table 1). Upon confirmation of conservation, pathogenicity prediction was conducted in PolyPhen-2. Informed by these results, we predicted a change in the structure of the protein encoded by *INSR*, resulting in a change in the protein's function. To explore this hypothesis further, bioinformatic analyses were conducted to investigate the extent of structural changes. Furthermore, primers were designed and tested to target the VUS region in the *daf-2* genes. RNA guides were then designed to bring this study into the beginning stages of producing a CRISPR-Cas9

engineered *C. elegans daf-2* mutant to test the functionality of the protein encoded by the variant *daf-2* gene.

Materials and Methods

All experiments detailed in this study were conducted in accordance with established standard operating protocols and procedures as set forth by the research laboratory of Dr. Ashley Turner.

Confirmation of Conservation between *INSR* and *daf-2*

The identification of a missense VUS associated with T2DM within the human gene *INSR* was documented in ClinVar at position c.1628C>T (p.Thr543Met) by an undergraduate student (Landrum et al., 2018). Specifically, a mutation at the orthologous mutation in the *C. elegans* gene *daf-2* was identified at position c.2084C>T (p.Thr698Met). Leveraging the respective gene IDs for *INSR* (ENSG00000171105) and *daf-2* (WBGene00000898), the corresponding cDNA files were successfully imported into Benchling (Benchling, (RRID:SCR_013955)). To verify conservation between *INSR* and *daf-2*, a multiple sequence alignment (MSA) was executed. This alignment was facilitated by employing the Auto-MAFFT alignment algorithm within Benchling, enabling the creation of a consensus alignment (Katoh & Standley, 2013). Such an approach permitted a comprehensive evaluation of conservation across both DNA and amino acid sequences at the VUS loci.

PolyPhen-2 Analysis

To assess the potential pathogenic impact of the amino acid substitution on the human INSR protein, a prediction was conducted using PolyPhen-2. The protein identifier (P06213), specifying the human INSR protein, along with the substitution

position (543) and the specific amino acid alteration (threonine to methionine), were entered into the corresponding fields on the online bioinformatics server (Adzhubei et al., 2013). The HumDiv score was chosen over the HumVar score to account for the multifactorial nature of T2DM (Adzhubei et al., 2010). This predictive analysis aimed to elucidate the potential functional consequences of the identified amino acid change, providing valuable insights into the protein's stability and activity.

Protein Modeling and PyMOL Analysis

SWISS-MODEL, a platform for protein structure prediction, was employed to generate three-dimensional structural predictions for both the wildtype and mutant forms of the human *INSR* genes (Benkert et al., 2011; Bertoni et al., 2017; Bienert et al., 2017; Guex et al., 2009; Mariani et al., 2013; Studer et al., 2014, 2020, 2021; Waterhouse et al., 2018). Initially, the protein identifier for INSR (P06213) was submitted to the online bioinformatics server to generate the model of the wildtype protein. Subsequently, to create the mutant protein model, a FASTA sequence file with the specific amino acid substitution was provided to the online server. Template-based modeling and threading for both wildtype and mutant protein models were performed using the Local MetaThreading-Server (LOMETS). The template selected for modeling was the smallest unit forming the INSR tetramer with the highest coverage available, which in this case was a heterodimer (7bwa.1.A). PDB files obtained from SWISS-MODEL for both the wildtype and mutant proteins were then exported into PyMOL (Version 2.5.1, Schrödinger).

PyMOL, a robust molecular visualization tool, was utilized to visualize the three-dimensional protein models of both the wildtype and mutant forms. These models were superimposed to facilitate the visualization and analysis of structural differences, as illustrated in Figure 3. Additionally, PyMOL was employed to determine the root mean square deviation (RMSD) value, providing insights into the significance of structural alterations. This comprehensive approach enabled both visualization and bioinformatics analysis of structural changes, enriching the understanding of the molecular implications of the identified mutations.

DNA Lysis

N2 *C. elegans* were maintained within a controlled biological incubator set at a temperature of 20°C. These organisms were maintained on nematode growth media plates supplemented with OP51 *E. coli*. For experimental purposes, five to ten adult nematodes were delicately extracted and transferred into individual PCR tubes, each preloaded with 30 µL of lysis buffer fortified with proteinase K (Cat. # V302B, Promega). This assembly underwent an incubation period of two hours at 65°C, followed by a subsequent heat treatment at 95°C for ten minutes.

Polymerase Chain Reactions

Benchling's design tool was harnessed to design primers aimed at amplifying the target VUS locus within the *daf-2* gene. Two primer sets were selected based on the Primer 3 specificity score provided by Benchling (Table 3). Following primer selection, polymerase chain reaction (PCR) analyses were conducted. A master mix tailored for primer set 1 was formulated, comprising 25µL of 2x Taq Master Mix (Cat. # M0270S,

New England Biolabs), 1µL each of forward and reverse primers at a concentration of 10µM, and 50µL of nuclease-free water. Subsequently, 11.5µL of the master mix was dispensed into three PCR reaction tubes. Utilizing lysed DNA extracted from the N2 *C. elegans* strain as a template, 1µL was added to the first two PCR reaction tubes, totaling 12.5µL. A third tube received 1µL of nuclease-free water, serving as a crucial non-template control. This procedure was replicated for the second primer set.

The initial PCR reactions were executed with an annealing temperature of 55°C for 30 cycles, employing the Biorad thermal cycler. Subsequently, both primer sets underwent electrophoresis on a 6% polyacrylamide gel at 150V for 45 minutes. Post-electrophoresis, the gel was stained with GelRed (Cat. # 41003, Biotium) and imaged utilizing the Biorad imaging system.

To optimize the annealing temperature for each primer set, gradient PCR analyses were conducted. Primer set 1 underwent annealing temperature variation between 60°C to 70°C, while primer set 2 explored a range from 50°C to 60°C. These analyses ensured the fine-tuning of experimental conditions to maximize PCR efficiency and specificity, crucial for downstream analyses and interpretation.

CRISPR Guide Design

Within Benchling, precise CRISPR targets flanking the VUS locus within the *daf-2* gene were identified. RNA guides tailored to the target locus were designed. The selection process was rigorous, with a meticulous comparison of on-target scores, specifically optimized for 20-base-pair guides utilizing NGG protospacer adjacent motif (PAM), as outlined by Doench et al. (2016). Moreover, the specificity score, or off-target

score, as elucidated by Hsu et al. (2013) was evaluated. Two optimal guides were chosen based on their superior on-target and off-target scores, as illustrated in Table 4. This sets the stage for precise and effective CRISPR-mediated gene editing experiments.

Results

In our investigation, a VUS associated with T2DM was previously identified within the *INSR* gene, as documented in ClinVar. The ortholog of this gene was discovered in the nematode *C. elegans* model, specifically in the *daf-2* gene. By constructing a MSA, we confirmed conservation of the VUS location at the amino acid level (Table 2).

We employed PolyPhen-2 to predict the potential impact of the amino acid substitution on the functionality of the human INSR protein. The HumDiv score, chosen to account for the multifactorial nature of T2DM, indicated a probable damaging effect, with a score of 1.000 (Figure 2).

It was noted that there was a class change from the hydrophilic, polar amino acid threonine to the hydrophobic, nonpolar methionine. Furthermore, we analyzed the potential structural alterations in the human INSR protein by generating 3D models for both the wildtype and variant proteins. Subsequently, we compared these structures atom by atom, yielding an RMSD value of 0.001. This structural alteration results in the substitution of the polar amino acid threonine with the nonpolar amino acid methionine (Figure 3).

Given the outcomes of our *in silico* studies, we deemed it necessary to design reagents to assess the functional implications of this structural change in the INSR protein through *in vivo* experimentation. To this end, we designed and executed various experiments aimed at creating a variant strain of *C. elegans* harboring the *daf-2* ortholog mutation for further analysis. Two primer sets were designed for amplifying the targeted

daf-2 VUS region in the nematode model (Table 3). Primer set 1, with an expected size of 575 bp, displayed robust bands at approximately 600 bp with minimal residual bands. Conversely, primer set 2, expected to yield a product of 791 bp, exhibited strong bands around 800 bp, along with more pronounced residual bands (Figure 4). Subsequent PCR amplification across a gradient of annealing temperatures allowed us to optimize the annealing temperature for each primer set, determining 64°C for primer set 1 and 54°C for primer set 2 (Figures 5 and 6, respectively).

To enable precise targeting of the *daf-2* VUS region in the nematode model, we designed RNA guides for CRISPR-Cas9 engineering. These guides, designed to flank the *daf-2* VUS locus within 30 base pairs, were tailored to complement the corresponding binding sites within the protospacer adjacent motif (PAM), where Cas9 would bind and induce DNA cleavage (Table 4, Figure 7). The guides chosen demonstrated the best potential on-target and off-target scores.

Conclusion

T2DM is a complex, multifactorial disease. The investigation outlined represents a crucial step towards elucidating the functional implications of the identified variant of the *INSR* gene associated with T2DM. By delineating the structural consequences of the threonine to methionine substitution within the INSR protein, we have laid the groundwork for further experimental exploration.

Primers were designed and annealing temperatures for each primer set were optimized. RNA guides were designed. The proposed experimental approach moving forward would include an in vitro assay to ensure the guide cuts efficiently at the targeted site. Previous literature utilized a sgRNA expressed from a U6 small nuclear RNA promoter to allow Cas9 to target the desired genomic sequences. It was also found that the optimal expression from the pol III promoter occurs when the first nucleotide base is a purine. A SV40 nuclear localization signal (NLS) was also inserted into the pUC57 vector at the 3' open reading frame. Microinjection techniques would then be performed to insert this into the gonads of the nematode, allowing for reproduction of future generations to contain this mutation (Friedland et al., 2013). This would allow the engineering of a CRISPR-Cas9 generated *daf-2* mutant strain. The N2 strain *C. elegans* will act as a control, along with a second engineered strain that lacks the VUS with a silent induced mutation. There is also a known *daf-2* knockout strain in the lab. A *rol-6* guide will also be microinjected into the C. elegans to act as a positive control to give a visual phenotype marker. Those with a positive phenotype will be selected for breeding

and downstream analysis of the VUS *daf-2*. Using phenotype assays, we aim to determine a correlation between the identified mutation and downstream insulin signaling alterations. A lifespan assay would be conducted to determine the presence of a positive phenotype. Animals with a positive phenotype would present with arrested development in the larval stage with the formation of dauer larva. Those with a positive phenotype would also display a dramatically extended lifespan compared to their wildtype counterpart. The integration of dauer and lifespan assays will provide valuable insights into the physiological consequences of this genetic variation, potentially validating its pathogenicity. It has been found that a decrease in *daf-2* signaling also leads to increased resistance to stresses such as oxidative stress, heat stress, hypoxic stress, osmotic stress, ultraviolet stress, and heavy metal toxicity (Altintas et al., 2016). Other phenotype assays could be conducted evaluating the stress response of *daf-2* mutants to these stressors. Along with these phenotype assays, RT-PCR will be performed to detect the expression of the mutation in the transcripts of the nematode (Friedland et al., 2013). Positive phenotypes would support the hypothesis that this mutation impacts the protein's functionality. Further additional experiments would include functional challenges with insulin and/ or glucose in the nematode *daf-2* mutant compared to the controls.

Furthermore, a proposed bioinformatics analysis utilizing Schrodinger's Molecular Operating Environment (MOE) offers a comprehensive understanding of the structural dynamics and functional implications of the mutated INSR protein. By visualizing protein interactions within both extracellular and intramembranous domains, we seek to uncover novel insights into the molecular mechanisms underlying insulin

resistance in T2DM. A secondary bioinformatics analysis option would be mCSM membrane, which is a free open-source webserver-based software. It allows for investigations of single-point mutations on membrane proteins, including the stability of the proteins and the likelihood of them being pathogenic. This is accomplished by using graph-based signatures to model the protein geometry along with its physiochemical properties (Pires et al., 2020).

The potential reclassification of the variant from unknown to likely pathogenic status underscores the clinical significance of our findings. Moreover, the prospect of extrapolating our results to other disease contexts harboring similar variants underscores the broader implications of our research. By unraveling the intricacies of INSR protein functionality, we pave the way for the development of targeted gene therapies tailored to individuals carrying this mutation, thereby advancing personalized medicine approaches in the management of T2DM.

Further reaching implications would include other conditions associated with *INSR* and *daf-2*, including aging. This is based off a study showing that there does exist a role of insulin signaling in mammalian aging (Altintas et al., 2016). There is a possibility that metabolism and aging are correlated (Thomas and Inoue, 1998). Hyperinsulinemia and the increase of bioavailable IGF-1 may have a role in the tumor formation and progression, particularly in patients who are known to have insulin resistance. This could be because insulin and IGF-1 inhibit the hepatic synthesis of sex hormone binding globulin (SHBG), which stimulates the synthesis of sex steroids. Excess insulin and IGF-1 also increase the overproduction of reactive oxygen species. This leads to damage to

DNA, which in turn may lead to mutagenesis and carcinogenesis. This excess can also lead to an overabundance of inflammatory cells in adipose tissue, creating a protumorigenic environment (Arcidiacono et al., 2012). This is further implicated in breast cancer, with the expression of estrogen receptors and IGF being positively correlated. The increased activation of the estrogen receptors may increase the stimulation of mammary carcinogenesis by enhancing proliferative activity in mammary epithelial cells. Abdominal obesity in women is also associated with increased concentrations of free IGF-1 and free estradiol, leading to further increased activation of estrogen receptors (Stoll, 2002).

This multiprong approach holds promise not only for advancing our understanding of T2DM pathogenesis but also for delineating novel therapeutic avenues with far-reaching implications for the broader spectrum of diseases associated with *INSR* variants. Through rigorous experimental validation and computational analyses, we endeavor to catalyze transformative advancements in the field of molecular medicine and personalized medicine, ultimately improving clinical outcomes for patients affected by T2DM and related disorders. This opens the door for potential targeted therapies for this *INSR* variant, specifically gene therapy or small molecule drugs that target the specific variant and mechanism.

References

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S., & Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, 7(4), 248–249. https://doi.org/10.1038/nmeth0410-248

Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. Current Protocols in Human Genetics / Editorial Board, Jonathan L. Haines ... [et Al.], 0 7, Unit7.20. https://doi.org/10.1002/0471142905.hg0720s76

Altintas, O., Park, S., & Lee, S.-J. V. (2016). The role of insulin/IGF-1 signaling in the longevity of model invertebrates, C. elegans and D. melanogaster. *BMB Reports*, *49*(2), 81–92. https://doi.org/10.5483/bmbrep.2016.49.2.261

Arcidiacono, B., Iiritano, S., Nocera, A., Possidente, K., Nevolo, M. T., Ventura, V., Foti, D., Chiefari, E., & Brunetti, A. (2012). Insulin resistance and cancer risk: An overview of the pathogenetic mechanisms. *Experimental Diabetes Research*, 2012, 789174. https://doi.org/10.1155/2012/789174

Baghaie, L., Bunsick, D. A., & Szewczuk, M. R. (2023). Insulin Receptor Signaling in Health and Disease. *Biomolecules*, 13(5), 807. https://doi.org/10.3390/biom13050807

Benchling (RRID:SCR_013955)

Benkert, P., Biasini, M., & Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics (Oxford, England)*, *27*(3), 343–350. https://doi.org/10.1093/bioinformatics/btq662

Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L., & Schwede, T. (2017). Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports*, *7*(1), 10480. https://doi.org/10.1038/s41598-017-09654-8

Bienert, S., Waterhouse, A., de Beer, T. A. P., Tauriello, G., Studer, G., Bordoli, L., & Schwede, T. (2017). The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Research*, *45*(D1), D313–D319. https://doi.org/10.1093/nar/gkw1132

Doench, J., Fusi, N., Sullender, M. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnology* **34**, 184–191 (2016). https://doi.org/10.1038/nbt.3437

Evans, E. A., Chen, W. C., & Tan, M.-W. (2008). The *DAF-2* insulin-like signaling pathway independently regulates aging and immunity in C. elegans. *Aging Cell*, 7(6), 879–893. https://doi.org/10.1111/j.1474-9726.2008.00435.x

Friedland, A. E., Tzur, Y. B., Esvelt, K. M., Colaiácovo, M. P., Church, G. M., & Calarco, J. A. (2013). Heritable genome editing in C. elegans via a CRISPR-Cas9 system. *Nature Methods*, 10(8), 741–743. https://doi.org/10.1038/nmeth.2532

Gottlieb S, Ruvkun G. *daf-2, daf-16,* and *daf-23*: genetically interacting genes controlling Dauer formation in Caenorhabditis elegans. *Genetics*. 1994 May;137(1):107-20. https://doi.org/10.1093/genetics/137.1.107. PMID: 8056303; PMCID: PMC1205929.

Goyal R, Singhal M, Jialal I. (2023). Type 2 Diabetes. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK513253/

Guex, N., Peitsch, M. C., & Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis*, *30 Suppl 1*, S162-173. https://doi.org/10.1002/elps.200900140

Haeusler, R. A., McGraw, T. E., & Accili, D. (2018). Biochemical and cellular properties of insulin receptor signaling. Nature Reviews. *Molecular Cell Biology*, 19(1), 31–44. https://doi.org/10.1038/nrm.2017.89

Halaschek-Wiener, J., Khattra, J. S., McKay, S., Pouzyrev, A., Stott, J. M., Yang, G. S., Holt, R. A., Jones, S. J. M., Marra, M. A., Brooks-Wilson, A. R., & Riddle, D. L. (2005). Analysis of long-lived C. elegans daf-2 mutants using serial analysis of gene expression. *Genome Research*, 15(5), 603–615. https://doi.org/10.1101/gr.3274805

Hsu, P., Scott, D., Weinstein, J. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases.*Nat Biotechnol* **31**, 827–832 (2013). https://doi.org/10.1038/nbt.2647

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kimura, K. D., Tissenbaum, H. A., Liu, Y., & Ruvkun, G. (1997). Daf-2, an insulin receptor-like gene that regulates longevity and diapause in Caenorhabditis elegans. *Science* (New York, N.Y.), 277(5328), 942–946. https://doi.org/10.1126/science.277.5328.942

Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W., Karapetyan, K., Katz, K., Liu, C., Maddipatla, Z., Malheiro, A., McDaniel, K., Ovetsky, M., Riley, G., Zhou, G., … Maglott, D. R. (2018). ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*, *46*(D1), D1062–D1067. https://doi.org/10.1093/nar/gkx1153

Lee, S.-J., Murphy, C. T., & Kenyon, C. (2009). Glucose shortens the life span of C. elegans by downregulating *DAF-16*/FOXO activity and aquaporin gene expression. *Cell Metabolism*, 10(5), 379–391. https://doi.org/10.1016/j.cmet.2009.10.003

Malkani S, Mordes JP. Implications of using hemoglobin A1C for diagnosing diabetes mellitus. Am J Med. 2011 May;124(5):395-401. https://doi.org/10.1016/j.amjmed.2010.11.025. PMID: 21531226; PMCID: PMC3086708.

Mariani, V., Biasini, M., Barbato, A., & Schwede, T. (2013). lDDT: A local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics (Oxford, England)*, *29*(21), 2722–2728. https://doi.org/10.1093/bioinformatics/btt473

Martin, B. C., Warram, J. H., Krolewski, A. S., Bergman, R. N., Soeldner, J. S., & Kahn, C. R. (1992). Role of glucose and insulin resistance in development of type 2 diabetes mellitus: Results of a 25-year follow-up study. Lancet (London, England), 340(8825), 925–929. https://doi.org/10.1016/0140-6736(92)92814-v

Murphy, C. T., & Hu, P. J. (2013). Insulin/insulin-like growth factor signaling in C. elegans. *WormBook: The Online Review of C. Elegans Biology*, 1–43. https://doi.org/10.1895/wormbook.1.164.1

Payankaulam, S., Raicu, A.-M., & Arnosti, D. N. (2019). Transcriptional Regulation of *INSR*, the Insulin Receptor Gene. *Genes*, 10(12), E984. https://doi.org/10.3390/genes10120984

Pires, D. E. V., Rodrigues, C. H. M., & Ascher, D. B. (2020). mCSM-membrane: Predicting the effects of mutations on transmembrane proteins. *Nucleic Acids Research*, 48(W1), W147–W153. https://doi.org/10.1093/nar/gkaa416

Silverman, G. A., Luke, C. J., Bhatixla, S. R., Long, O. S., Vetica, A. C., Perlmutter, D. H., & Pak, S. C. (2009). Modeling molecular and cellular aspects of human disease using the nematode Caenorhabditis elegans. *Pediatric Research*, 65(1), 10–18. https://doi.org/10.1203/PDR.0b013e31819009b0

Stoll, B. A. (2002). Oestrogen/insulin-like growth factor-I receptor interaction in early breast cancer: Clinical implications. *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, 13(2), 191–196. https://doi.org/10.1093/annonc/mdf059

Studer, G., Biasini, M., & Schwede, T. (2014). Assessing the local structural quality of transmembrane protein models using statistical potentials (QMEANBrane). *Bioinformatics (Oxford, England)*, *30*(17), i505-511. https://doi.org/10.1093/bioinformatics/btu457

Studer, G., Rempfer, C., Waterhouse, A. M., Gumienny, R., Haas, J., & Schwede, T. (2020). QMEANDisCo-distance constraints applied on model quality estimation. *Bioinformatics (Oxford, England)*, *36*(6), 1765–1771. https://doi.org/10.1093/bioinformatics/btz828

Studer, G., Tauriello, G., Bienert, S., Biasini, M., Johner, N., & Schwede, T. (2021). ProMod3-A versatile homology modelling toolbox. *PLoS Computational Biology*, *17*(1), e1008667. https://doi.org/10.1371/journal.pcbi.1008667

Taniguchi, C. M., Emanuelli, B., & Kahn, C. R. (2006). Critical nodes in signalling pathways: Insights into insulin action. Nature Reviews. *Molecular Cell Biology*, 7(2), 85–96. https://doi.org/10.1038/nrm1837

Thomas, J. H., & Inoue, T. (1998). Methuselah meets diabetes. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 20(2), 113–115. https://doi.org/10.1002/(SICI)1521-1878(199802)20:2<113::AID-BIES3>3.0.CO;2-U

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes.

Nucleic Acids Research, *46*(W1), W296–W303. https://doi.org/10.1093/nar/gky427

Wei, Q., Li, J., Zhan, Y., Zhong, Q., Xie, B., Chen, L., Chen, B., & Jiang, Y. (2020). Enhancement of glucose homeostasis through the PI3K/Akt signaling pathway by dietary with Agaricus blazei Murrill in STZ-induced diabetic rats. *Food Science & Nutrition*, 8(2), 1104–1114. https://doi.org/10.1002/fsn3.1397

Zhang, Y.-P., Zhang, W.-H., Zhang, P., Li, Q., Sun, Y., Wang, J.-W., Zhang, S. O., Cai, T., Zhan, C., & Dong, M.-Q. (2022). Intestine-specific removal of DAF-2 nearly doubles lifespan in Caenorhabditis elegans with little fitness cost. *Nature Communications*, 13(1), 6339. https://doi.org/10.1038/s41467-022-33850-4

APPENDIX A: TABLES

Table 1

Nondiabetic, prediabetic, and diabetic levels as defined by the ADA

	Nondiabetic	Prediabetic	Diabetic
A1C	< 5.7%	5.7% to 6.4%	≥ 6.5%
FPG	< 100mg/dL	100 to 125mg/dL	≥ 126mg/dL

Table 2

Genetic location of VUS within INSR orthologs of H. sapiens and C. elegans

Species	**Gene Ortholog**	**Variant Location**
Humans (*H. sapiens*)	*INSR*	c.1628C>T (p.Thr543Met)
Nematode (*C. elegans*)	*daf-2*	c.2084C>T (p.Thr698Met)

Table 3

Evolutionary Conservation of INSR VUS loci across multiple species

Human Disease Variant	AAT	GTG	ATG	GAG	TTC
	Asn	Val	Met	Glu	Phe
Human (*H. sapiens*)	AAT	GTG	ACG	GAG	TTC
	Asn	Val	Thr	Glu	Phe
Nematode (*C. elegans*)	AAC	ATG	ACG	ATC	GAA
	Asn	Met	Thr	Ile	Glu

Table 4

DNA primer design for targeting the INSR VUS locus within daf-2

	Primer Set 1	Primer Set 2
Forward Primer	5' ACGACGGATTTGACGCTTGATCGT 3'	5' TCGACTTGCGCTAGTGCCAACG 3'
Reverse Primer	5' CAGGATTCGGGAATGGCGTCCG 3'	5' GCGTCCGAATCGACTTGCGCTA 3'
Expected PCR Product Size	575 base pairs	791 base pairs

Table 5

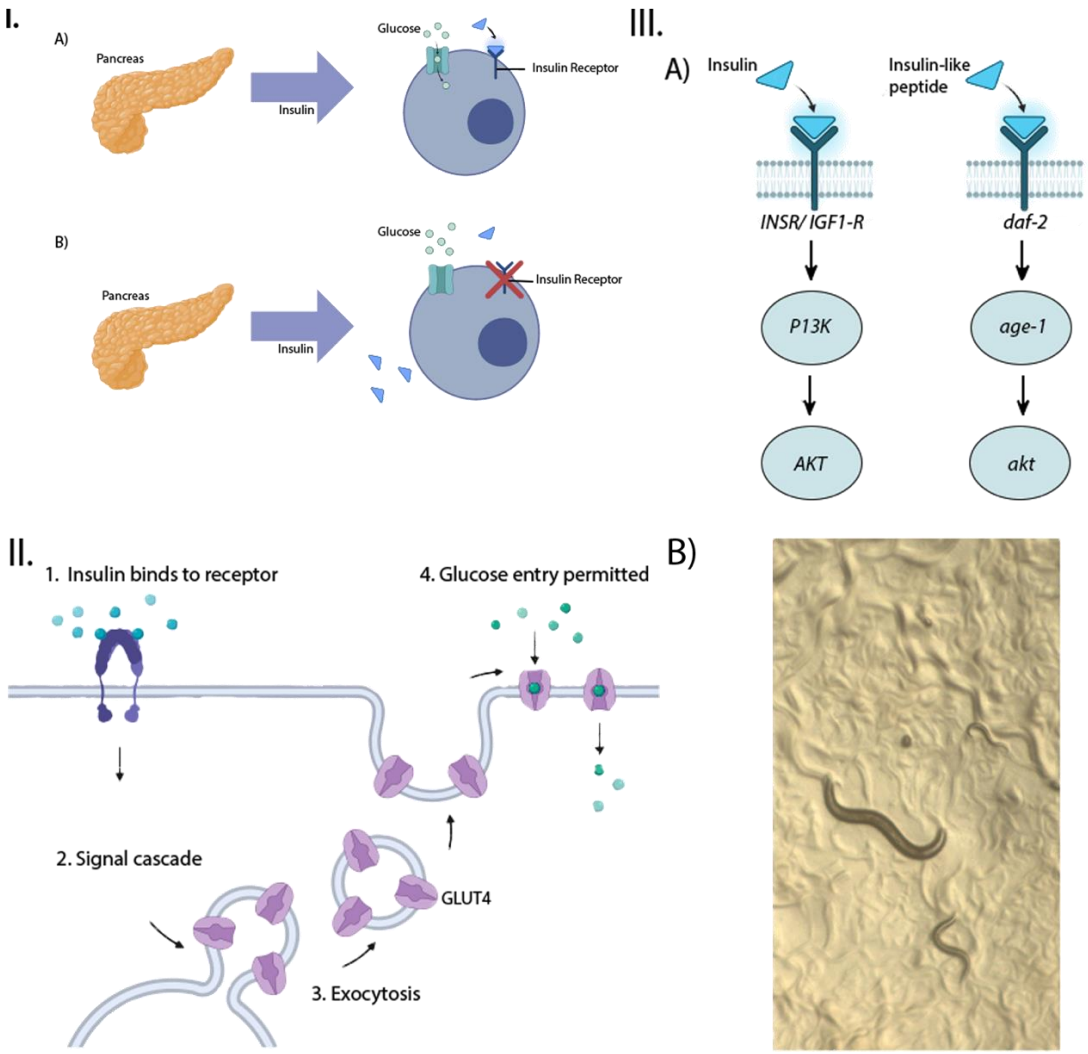
CRISPR RNA guide design for daf-2 targeting

Cut Position (DNA bp)	DNA strand	CRISPR RNA Guide Sequence	PAM
16316	Bottom	TCATGTTCTCATCGATTCGT	GGG
16339	Top	TGAGAACATGACGATCGAAG	AGG

APPENDIX B: FIGURES

Figure 1

Overview of T2DM, the insulin receptor in the insulin signaling pathway in humans, C. elegans, and the insulin-like growth factor 1 receptor in the insulin-like signaling pathway in C. elegans

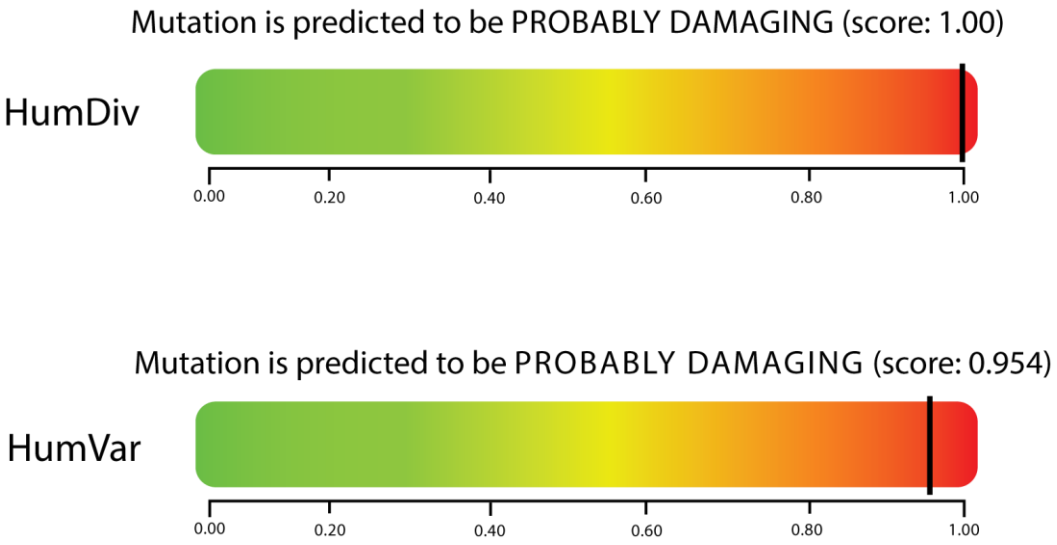


Panel I: A) Insulin intake into the cell allows for appropriate usage of glucose from the bloodstream. B) The ineffectiveness of the insulin receptor causes cells to not take in and utilize glucose, leading to increased blood glucose levels and T2DM. Panel II: Simplified schematic of the insulin signaling pathway. Panel III: A) Comparison of the insulin

signaling pathway to the ortholog insulin-like signaling pathway in *C. elegans*. B) Image of N2 *C. elegans* viewed using a dissecting microscope and taken with a Motic camera.

Figure 2

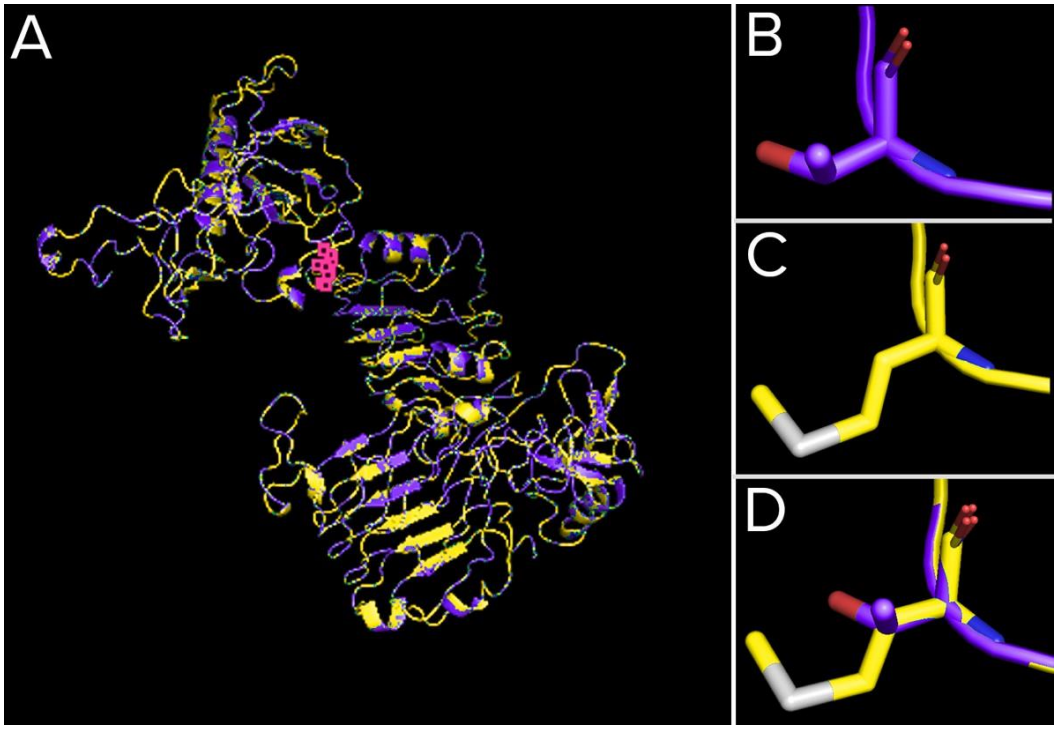
Results of the PolyPhen-2 analysis predicting the potential impact of the INSR VUS p.Thr543Met substitution on the human INSR protein



HumDiv and HumVar scores are reported from the analysis.

Figure 3

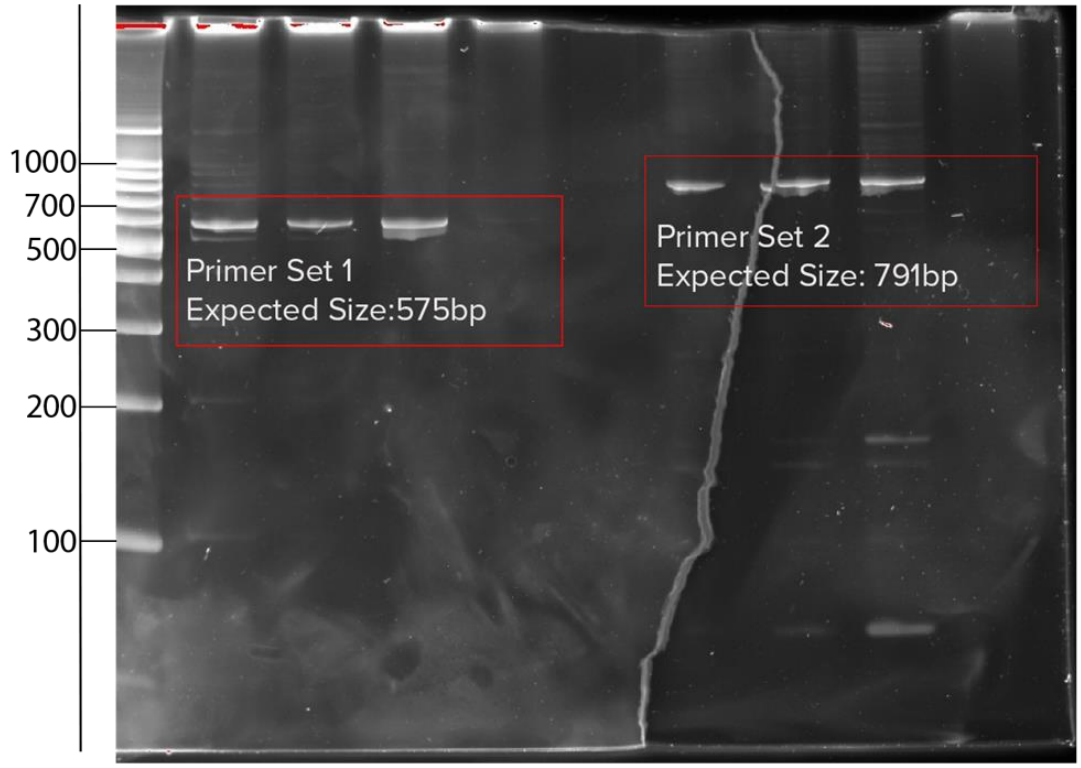
Predicted 3D structure of the human wildtype INSR protein and the mutant VUS protein



Protein structure generated with the INSR amino acid sequence for wildtype and the identified variant INSR VUS. A) Both the wildtype and the variant models were aligned atom for atom and superimposed (wildtype= green, variant= teal). The pink represents the targeted VUS region. B) Structure of threonine (a polar amino acid) in the targeted region of the wildtype INSR protein. C) Structure of methionine (a nonpolar amino acid) in the targeted region of the INSR variant protein. D) Superimposed structures of threonine and methionine in the targeted region (wildtype= green, variant= teal).

Figure 4

Initial PCR results of the two designed DNA primer sets amplifying the INSR VUS locus in daf-2 through gel electrophoresis



Initial PCR amplification for both primer sets annealed at 55°C and ran on a 6% polyacrylamide gel stained with GelRed. N2 *C. elegans* wildtype DNA was used across six reactions, three for each primer set. Lane 1 consisted of a 100bp ladder. Lanes 2-4 consisted of reactions for primer set 1 (1-1, 1-2, 1-3). Lane 5 consisted of a no template (water) control for primer set 1. Lanes 7-9 consisted of reactions for primer set 2 (2-1, 2-2, 2-3). Lane 10 consisted of a no template (water) control for primer set 2.

Figure 5

PCR results and annealing temperature optimization of primer set 1 amplification of the INSR VUS locus in daf-2 through gel electrophoresis

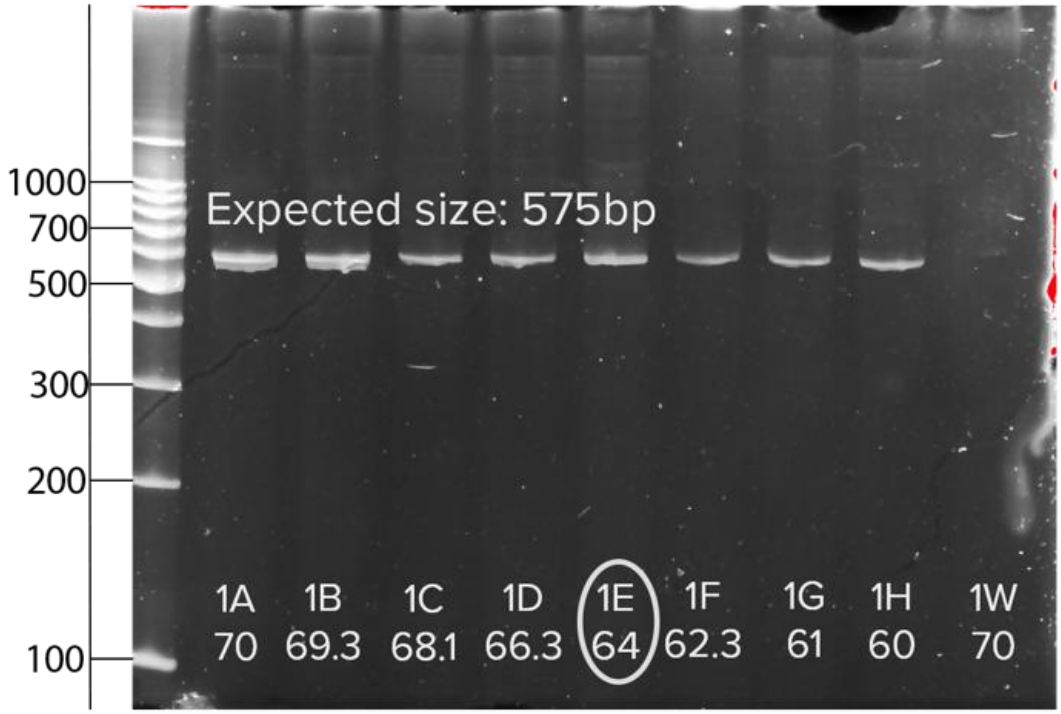


Image of GelRed stained PCR amplification of primer set 1 ran on 6% polyacrylamide gel across a thermal gradient from 60°C-70°C. N2 *C. elegans* DNA was used as a wildtype template across eight reactions, with the corresponding annealing temperatures labeled in each lane (1A= 70°C, 1B= 69.3°C, 1C= 68.1°C, 1D= 66.3°C, 1E= 64°C, 1F= 62.3°C, 1G= 61°C, 1H= 60°C). 1W was a no template (water) control ran at 70°C.

Figure 6

PCR results and annealing temperature optimization of primer set 2 amplification of the INSR VUS locus in daf-2 through gel electrophoresis

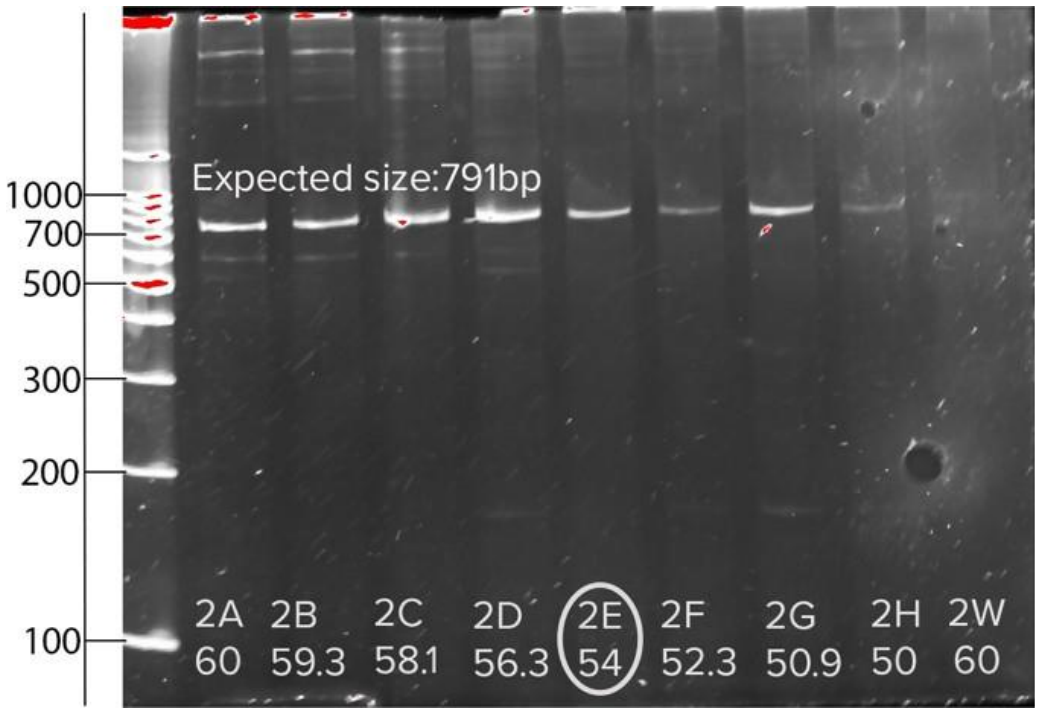
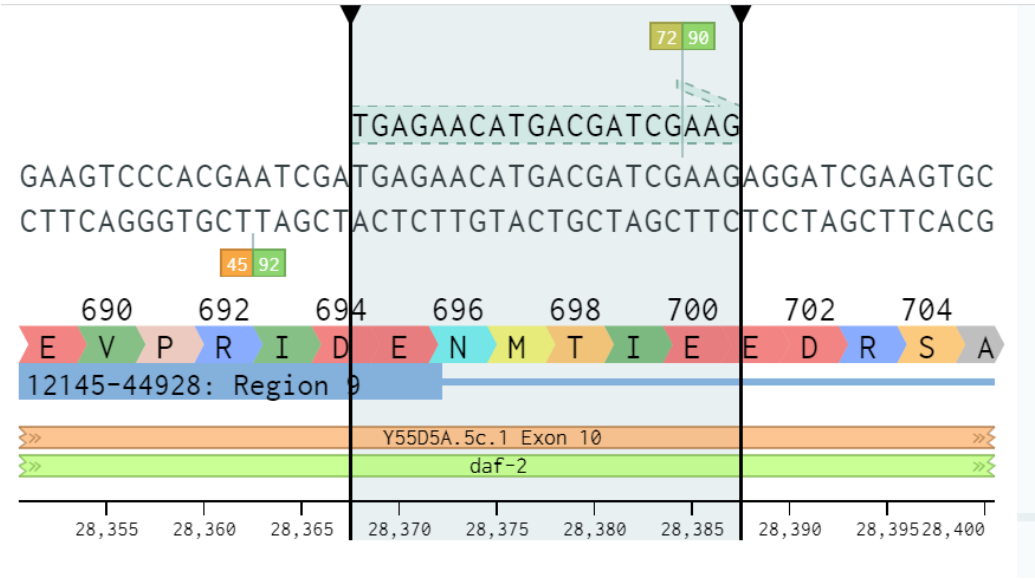


Image of GelRed stained PCR amplification of primer set 2 ran on 6% polyacrylamide gel across a thermal gradient from 60°C-70°C. N2 *C. elegans* DNA was used as a wildtype template across eight reactions, with the corresponding annealing temperatures labeled in each lane (2A= 60°C, 2B= 59.3°C, 2C= 58.1°C, 2D= 56.3°C, 2E=54°C, 2F= 52.3°C, 2G= 50.9°C, 2H= 50°C). 2W was a no template (water) control ran at 60°C.

Figure 7

CRISPR-Cas9 targeting in the daf-2 gene for generation of the daf-2 VUS C. elegans model



Schematic showing the VUS locus of the *daf-2* mutation. The amino acid substitution occurs at amino acid 698. The RNA guide has an on-target score of 72 and an off-target score of 90.