In Silico Analysis of Single Nucleotide Polymorphisms (SNPs) in the Human Titin (TTN) Gene

Lei Jo Presaldo
Ignite Minds, Bioinformatics4all
b2023ljppresaldo@pshs.edu.ph
Philippine Science High School

Monica Sing
Ignite Minds, Bioinformatics4all
monica.sing.2023@gmail.com
International School of Phnom Penh

Toma Ogawa
Ignite Minds, Bioinformatics4all
cap10toma@gmail.com
Pathways School Gurgaon

Aksh Garg (Research Mentor)
Ignite Minds, Bioinformatics4all
akshgarg@gmail.com
Palos Verdes Peninsula High School

Jason Apostol (Research Mentor)
Bioinformatics4all
Jason1712445@gmail.com
Palos Verdes Peninsula High School

Abstract

The Titin (TTN) gene provides instruction for the titin protein that is responsible for various aspects of muscle function. This study aims to identify the most deleterious single nucleotide polymorphisms (SNPs) in the TTN gene that hinder proper protein function. Using SNPedia, 598 SNPs were retrieved and tested for its deleteriousness using bioinformatics tools. Five bioinformatics tools were used to predict the most deleterious and damaging SNPs that might disrupt the function and structure of the human TTN gene through means of in-silico analysis. They include SIFT, SNPs&GO, PROVEAN, PhD-SNP, and Polyphen-2. GeneMANIA was used to find the most relevant gene-to-gene interactions of the TTN gene while Swiss Model and Swiss PDB Viewer were used for 3-D modelling. This study suggested that among 598 SNPs, 11 SNPs (Rs138060032, rs202094100, rs2555818, rs267607155, rs267607156, rs281864931, rs368277535, rs374615369, rs375159973, and rs72648247) are predicted to be most deleterious by the five bioinformatics tools. The study also proposes the 3D structure of the native TTN gene protein and the structure of the aforementioned SNPs. The TTN gene was found to have gene-to-gene interactions with other genes related to diseases that are caused by SNPs in the TTN gene. It was also concluded a majority (45%) of the deleterious SNPs were located in the A-band region of the TTN gene. This study will be considered useful when conducting large scale research on these proteins and in the process of precise drug development related to these mutations.

Abbreviations

1. SNP: Single Nucleotide Polymorphism
2. nsSNPs: Non-synonymous single nucleotide polymorphisms
3. NCBI: National Center of Bioinformatics
4. TTN: Titin
5. rsID: Rapid Stain Identification Series
6. PDB: Protein Data Bank
7. RI: Reliability Index
8. SVM: Support Vector Machine

Introduction

In the current field of genomics, understanding genomic variation is a major challenge as the organism's genome has an innumerable number of genetic variations. The most common form of genetic variation are the SNPs which are single base pair variations at specific locations in the genome.[1] NsSNPs, comprised of missense and nonsense SNPs, are important to study as the amino acid change occurs in the coding regions of a gene that can result in structural and functional changes of the proteins they code for.[2] Damaging nsSNPs can cause genetic diseases and affect an individual's response to some drugs.[3]

Titin (TTN), a gene located on the long arm (q) of chromosome 2 and localized at 2q31, provides instruction for the largest protein in the human body, Titin.[4] The Titin protein is encoded

[1] C. Børsting and N. Morling, "Single-Nucleotide Polymorphisms," in *Encyclopedia of Forensic Sciences: Second Edition*, 2013, https://doi.org/10.1016/B978-0-12-382165- 2.00042-8.
[2] Tikam Chand Dakal et al., "Predicting the Functional Consequences of Non-Synonymous Single Nucleotide Polymorphisms in IL8 Gene," *Scientific Reports*, 2017, https://doi.org/10.1038/s41598-017-06575-4.
[3] K. M. Giacomini et al., "The Pharmacogenetics Research Network: From SNP Discovery to Clinical Drug Response," *Clinical Pharmacology and Therapeutics*, 2007, https://doi.org/10.1038/sj.clpt.6100087.
[4] Müller-Seitz, M., K. Kaupmann, S. Labeit, and Harald Jockusch. "Chromosomal localization of the mouse titin gene and its relationto "muscular dystrophy with myositis" and nebulin genes on chromosome 2." *Genomics* 18,

by 363 exons of the TTN gene that produces an amino acid residue of 38, 138 (42,000 kDa).[5] Titin plays a role in the assembly and functioning of vertebrates' striated muscles through its essential role in the sarcomere, connecting the Z-disc to the M-band.[6]

Further research has shown that titin is important in hypertrophic signaling and in assembling new sarcomeres.[7] Another function of the TTN protein is its role in the elasticity of relaxed striated muscles and Titin acts as the molecular scaffold for thick filament formation.[8] In non-muscle cells, titin plays a role in mitosis through facilitating chromosome condensation and chromosome segregation.[9]

Titin plays an important role in proper muscle function, being the third most abundant protein in the human muscular system. [10] Despite its abundance, there is a lack of interpretation for the TTN gene due to its large size.[11]

Mutations of the TTN gene are linked to a range of medical conditions, often affecting muscles. Such mutations are associated with injurious diseases that involve myopathy (skeletal muscle disease), cardiomyopathy, and muscle disorders. Titinopathies caused by variants in the TTN gene comprises of Centronuclear myopathy, Early-onset myopathy, Familial dilated cardiomyopathy, Hereditary myopathy with early respiratory failure, Tibial muscular dystrophy, and Arrhythmogenic right ventricular cardiomyopathy.[12] According to studies, titin mutations are the most common genetic cause for dilated cardiomyopathy, heart failure, and premature death [13]

This study investigated deleterious SNPs of the TTN gene through the use of computer predictions, modeling, and simulations to predict and model their structural and functional consequences. The results of this study can aid future in-depth research on the human TTN gene and it can be useful in future studies on the treatment of diseases caused by these variations.

Literature Review

Though there are studies that discuss the role of TTN mutations in muscle disorders, few papers were analysing the human TTN gene using bioinformatics.[14] Currently, there are no studies focusing on the effects of SNPs on the TTN gene. The large size of the TTN gene is a barrier in studying and analysing the gene and protein in its entirety.[15]

no. 3 (1993): 559-561. https://doi.org/10.1016/S0888-7543(05)80356-8; Rossi, Elena, Antonio Faiella, Massimo Zeviani, Siegfried Labeit, Giovanna Floridia, Silvia Brunelli, Marina Cammarata, Edoardo Boncinelli, and Orsetta Zuffardi. "Order of six loci at 2q24-q31 and orientation of the HOXD locus." *Genomics* 24, no. 1 (1994): 34-40. https://doi.org/10.1006/geno.1994.1579.

[5] Marie Louise Bang et al., "The Complete Gene Sequence of Titin, Expression of an Unusual ≈700-KDa Titin Isoform, and Its Interaction with Obscurin Identify a Novel Z-Line to I-Band Linking System," *Circulation Research*, 2001, https://doi.org/10.1161/hh2301.100981.

[6] Bang et al.; Martina Krüger and Sebastian Kötter, "Titin, a Central Mediator for Hypertrophic Signaling, Exercise-Induced Mechanosignaling and Skeletal Muscle Remodeling," *Frontiers in Physiology*, 2016, https://doi.org/10.3389/fphys.2016.00076.

[7] Müller-Seitz, M., K. Kaupmann, S. Labeit, and Harald Jockusch. "Chromosomal localization of the mouse titin gene and its relationto "muscular dystrophy with myositis" and nebulin genes on chromosome 2." *Genomics* 18, no. 3 (1993): 559-561. https://doi.org/10.1016/S0888-7543(05)80356-8.

[8] Sebastian Kötter, Christian Andresen, and Martina Krüger, "Titin: Central Player of Hypertrophic Signaling and Sarcomeric Protein Quality Control," *Biological Chemistry*, 2014, https://doi.org/10.1515/hsz-2014-0178.

[9] Prilusky, Jaime, Eran Hodis, David Canner, Wayne A. Decatur, Karl Oberholser, Eric Martz, Alexander Berchanski, Michal Harel, and Joel L. Sussman. "Proteopedia: a status report on the collaborative, 3D web-encyclopedia of proteins and other biomolecules." *Journal of structural biology* 175, no. 2 (2011): 244-252. https://doi.org/10.1016/j.jsb.2011.04.011.

[10] UniProt Consortium. "UniProt: a worldwide hub of protein knowledge." *Nucleic acids research* 47, no. D1 (2019): D506-D515. https://doi.org/10.1093/nar/gky1049.

[11] Siegfried Labeit, Bernhard Kolmerer, and Wolfgang A. Linke, "The Giant Protein Titin: Emerging Roles in Physiology and Pathophysiology," *Circulation Research*, 1997, https://doi.org/10.1161/01.RES.80.2.290.

[12] Savarese, Marco, Lorenzo Maggi, Anna Vihola, Per Harald Jonson, Giorgio Tasca, Lucia Ruggiero, Luca Bello et al. "Interpreting genetic variants in titin in patients with muscle disorders." *JAMA neurology* 75, no. 5 (2018): 557-565. https://doi.org/10.1001/jamaneurol.2017.4899.

[13] Virginie Carmignac et al., "C-Terminal Titin Deletions Cause a Novel Early-Onset Myopathy with Fatal Cardiomyopathy," *Annals of Neurology*, 2007, https://doi.org/10.1002/ana.21089; Ozge Ceyhan-Birsoy et al., "Recessive Truncating Titin Gene, TTN, Mutations Presenting as Centronuclear Myopathy," *Neurology*, 2013, https://doi.org/10.1212/WNL.0b013e3182a6ca62; Peter Hackman et al., "Tibial Muscular Dystrophy Is a Titinopathy Caused by Mutations in TTN, the Gene Encoding the Giant Skeletal-Muscle Protein Titin," *American Journal of Human Genetics*, 2002, https://doi.org/10.1086/342380; Daniel S. Herman et al., "Truncations of Titin Causing Dilated Cardiomyopathy," *New England Journal of Medicine* 366, no. 7 (2012): 619–28, https://doi.org/10.1056/NEJMoa1110186; Ohlsson, Monica, Carola Hedberg, Björn Brådvik, Christopher Lindberg, Homa Tajsharghi, Olof Danielsson, Atle Melberg, Bjarne Udd, Tommy Martinsson, and Anders Oldfors. "Hereditary myopathy with early respiratory failure associated with a mutation in A-band titin." *Brain* 135, no. 6 (2012): 1682-1694. https://doi.org/10.1093/brain/aws103; Taylor, Matthew, Sharon Graw, Gianfranco Sinagra, Carl Barnes, Dobromir Slavov, Francesca Brun, Bruno Pinamonti et al. "Genetic variation in titin in arrhythmogenic right ventricular cardiomyopathy–overlap syndromes." *Circulation* 124, no. 8 (2011): 876-885. https://doi.org/10.1161/CIRCULATIONAHA.110.005405.

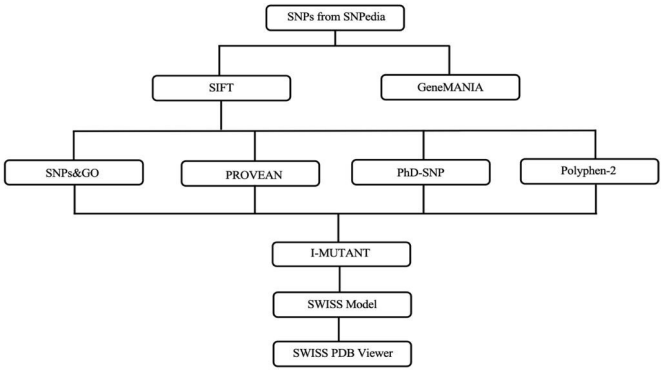
[14] John T. Hinson et al., "Titin Mutations in IPS Cells Define Sarcomere Insufficiency as a Cause of Dilated Cardiomyopathy," *Science* 349, no. 6251 (2015): 982–86, https://doi.org/10.1126/science.aaa5458.

[15] Savarese, Marco, Lorenzo Maggi, Anna Vihola, Per Harald Jonson, Giorgio Tasca, Lucia Ruggiero, Luca Bello et al. "Interpreting genetic variants in titin in patients with muscle disorders." *JAMA neurology* 75, no. 5 (2018): 557-565. https://doi.org/10.1001/jamaneurol.2017.4899.

Methodology

The methodology of this study is divided into three parts, namely the following: (1) selection of SNPs in the human TTN gene, (2) prediction and identification of deleterious SNPs, and (3) modelling functional and structural effects imposed on the TTN gene by various deleterious SNPs. The entirety of the process is summarized in Figure 1.

Figure 1. Bioinformatics software and servers used in TTN gene SNPs analysis



Selection of SNPs in human TTN gene

All SNPs were extracted from SNPedia in July 2020, with a total of 598 SNPs related to the TTN gene, available here. SNPedia is a wiki resource that serves as a database for SNPs, converting large scale genomic data from studies into comprehensible data.[16] Since the file came in a CSV format and included excess data not needed for the study, the data was altered through Python in Google Colab, ultimately deriving a batch of 598 rsIDs. Meanwhile, the FASTA file of the protein sequence of TTN was acquired from UniProt which is a central database of protein sequences with accurate sequences and functional annotation (Apweiler et al., 2004).[17]

Prediction and Identification of Deleterious SNPs

Five bioinformatics tools were used to identify and predict the effects of SNPs and amino acid substitutions on the human TTN gene. These tools are SIFT, SNPs&GO, PROVEAN, PhD-SNP, and PolyPhen-2.

SIFT (Sorting Intolerant from Tolerant)

SIFT (sift.bii.a-star.edu.sg/) is an online web-server used to predict if an amino acid substitution has an effect towards a protein's function.[18] SIFT's foundations stand on the theory that protein evolution and protein function are correlated.[19] When fed with a query sequence, it takes an amino acid's sequence homology and physical properties to predict and classify SNPs as either tolerated and deleterious substitutions (Herman et al., 2012).[20] In this paper, the server was used to differentiate tolerant from intolerant SNPs.

Using the SIFT's dbSNP rsIDs (SIFT4G predictions) tool, it will provide information about the submitted SNPs that will support the validity of this paper's results. Five-hundred ninety-eight SNPs were run through the SIFT server by submitting the previously mentioned FASTA file, eventually obtaining the needed information such as the amino acid changes and SIFT predictions. For each rsIDs submitted, SIFT will calculate the probabilities of all 20 possible outcomes when the residue is substituted with all 20 amino acids (Kumar et al., 2009).[21] These outcomes are then placed in a scaled probability matrix, normalizing the probabilities by the most frequent amino acid. SIFT provided other SNP-related information as well, namely: coordinates, reference alleles, alternate alleles, SIFT scores, SIFT medians, region, and allele frequencies. The SIFT score is essentially a scaled probability - based on the matrix previously mentioned, of the substitution being tolerated with a decision threshold of 0.05. A substitution that scores less than the decision threshold is classified as deleterious, while anything above is tolerated.

SNPS&GO

SNPS&GO (snps.biofold.org/snps-and-go/snps-and-go.html) is an online bioinformatics database that gathers data of the

[16] Michael Cariaso and Greg Lennon, "SNPedia: A Wiki Supporting Personal Genome Annotation, Interpretation and Analysis," *Nucleic Acids Research*, 2012, https://doi.org/10.1093/nar/gkr798.

[17] R. Apweiler, "UniProt: The Universal Protein Knowledgebase," *Nucleic Acids Research*, 2004, https://doi.org/10.1093/nar/gkh131.

[18] Ng, Pauline C., and Steven Henikoff. "SIFT: Predicting amino acid changes that affect protein function." *Nucleic acids research* 31, no. 13 (2003): 3812-3814. https://doi.org/10.1093/nar/gkg509

[19] Nimir, Mohammed, Mohanad Abdelrahim, Mohamed Abdelrahim, Mahil Abdalla, Wala eldin Ahmed, Muhanned Abdullah, and Muzamil Mahdi Abdel Hamid. "In silico analysis of single nucleotide polymorphisms (SNPs) in human FOXC2 gene." *F1000Research* 6 (2017). https://doi.org/10.12688/f1000research.10937.2.

[20] Herman et al., "Truncations of Titin Causing Dilated Cardiomyopathy."

[21] Prateek Kumar, Steven Henikoff, and Pauline C. Ng, "Predicting the Effects of Coding Non-Synonymous Variants on Protein Function Using the SIFT Algorithm," *Nature Protocols*, 2009, https://doi.org/10.1038/nprot.2009.86.

protein's sequence profile, 3D structure, functional and structural information and implements several other prediction methods, including PhD-SNP and PANTHER to calculate the output. It relies on an SVM (support vector machine) classifier to give predictions on whether an amino acid change is neutral or disease-related at 79% tested accuracy, as well as a RI and a probability.[22] This support vector machine is an implemented machine learning classifier trained specifically for identifying disease-related (desired output 0) and neutral (desired output 1) SNPs. The threshold value is 0.5, where anything higher is classified as disease-related. 52 values act as the input vectors, including 40 for mutations and local sequence information, 5 for information derived from the previous 40, 5 for PANTHER parameters, and 2 for gene oncology terms. The SVM output (O(D)) is the probability in which a polymorphism is disease-related, it is used to calculate the RI for each prediction through this equation,

$$RI = 20*|O(D)-0.5|,$$

where the lower the RIs are, the more accurate the predictions are.[23] This web-server was used to identify whether a mutation is associated with a disease or not.

All 228 amino acid substitutions acquired from the SIFT software were inputted into **SNPS&GO**. However, some of them were said to be incorrect when it was submitted to the SNPS&GO server. Any amino acid changes that appeared as "Incorrect mutation" were eliminated. As a result, with the results of 31 amino acid changes over 29 SNPs were removed. The rest were considered the most relevant and will be used in the proceeding in silico procedures.

It's important to note that SNPS&GO provided multiple results for each amino acid change from the three methods utilized and all have varying degrees of probability and RIs. However, since PANTHER predictions were extremely lacking, with only 4 out of 31 classified results, a decision was made to eliminate them, using only the SNPS&GO and PhD-SNP predictions.

PROVEAN (Protein Variation Effect Analyzer in Variants)

PROVEAN is an online software that can predict the effects of mutations on a protein's biological function and can be found here, provean.jcvi.org/index.php. It has a binary classification system that states whether an amino acid substitution is deleterious or neutral through scoring. The software is capable of giving predictions to amino acid substitutions, deletions, or insertions whether single or multiple. According to PROVEAN's procedure, which also utilized an SVM-based classifier similar to SNPS&GO, PROVEAN predictions are based on predefined thresholds that serve as a basis for computing the PROVEAN score. Anything that received a score below a predefined threshold is classified as deleterious, while anything above the threshold is predicted as neutral. For binary classification where predictions can be "deleterious" or "neutral", a default threshold of -2.5 is currently set and is used.[24] For this paper, the PROVEAN Protein Batch tool was utilized, specifically for the Homo Sapiens species. The protein query sequences from UniProt as well as the 31 amino acid substitutions obtained after running SNPS&GO were submitted.

PhD-SNP

PhD-SNP predicts the functional and structural consequences of an amino acid substitution on a protein using a similar process to SNPS&GO and with a tested 76% accuracy, accessible here: snps.biofold.org/phd-snp/phd-snp.html. It outputs the classification - whether the amino acid change is disease-related or neutral, in addition to the RI, again, through a very similar process to SNPS&GO and PROVEAN.[25] Likewise, the decision threshold is 0.5, with the same calculation method for RI. "SVM-based method using sequence and profile information" out of the 3 methods available was chosen to conduct on the data as it had more comprehensive methods.

Polyphen-2 (Polymorphisms Phenotyping v2)

Polyphen-2 is a web-server which predicts the possible impact of an amino acid substitution on the structure and function of a human protein found at genetics.bwh.harvard.edu/pph2/. It performs predictions and calculates the probability value based on several data points, including the protein sequence itself and its structural information and functional analysis of SNPs amongst others. The decision threshold of each substitution is based on the false positive rates (FPR), where substitutions with FPR≥10% is classified as "probably damaging", 10%>FPR≥20% is "possibly damaging", and 20%>FPR is

[22] Emidio Capriotti and Russ B. Altman, "Improving the Prediction of Disease-Related Variants Using Protein Three-Dimensional Structure," *BMC Bioinformatics*, 2011, https://doi.org/10.1186/1471-2105-12-S4-S3.

[23] Remo Calabrese et al., "Functional Annotations Improve the Predictive Score of Human Disease-Related Mutations in Proteins," *Human Mutation*, 2009, https://doi.org/10.1002/humu.21047.

[24] Yongwook Choi and Agnes P. Chan, "PROVEAN Web Server: A Tool to Predict the Functional Effect of Amino Acid Substitutions and Indels," *Bioinformatics*, 2015, https://doi.org/10.1093/bioinformatics/btv195.

[25] Calabrese et al., "Functional Annotations Improve the Predictive Score of Human Disease-Related Mutations in Proteins."

"benign". Lastly, it provides the probability of the substitution being damaging. [26]

I-Mutant Server

I-Mutant (gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) is an online server that predicts protein stability after a mutation. This web-server serves a purpose of predicting the stability of each single point protein mutation inputted. Its results are classified as either increase or decrease, signifying increase and decrease in stability respectively. It provides a RI ranging from 0 to 10 with 0 as the lowest.[27] All predictions are simulated at a temperature of 25°C and a pH level of 7.0 so as to simulate the real life environment.

GeneMANIA

GeneMANIA (http://www.genemania.org/) is a web-server that hypothesizes gene function and associates genes with similar ones using genomics data.[28] The software focuses on giving predictions on gene-to-gene interactions. In this paper, GeneMANIA is used to predict TTN's interactions with other genes.

Modelling Amino Acid Substitutions on a Structural and Functional Level

Selection of amino acid sequences to model

In order to generate a fair system to select deleterious amino acid substitutions for modelling and visualization, it was decided that amino acid substitutions marked as deleterious by at least 4 out of 6 bioinformatics softwares were considered as adverse to proper gene function. 4 was a reasonable threshold considering SIFT was not able to locate a large majority of SNPs and amino acid changes. For a lot of those labelled 'Not Found' by SIFT, the other softwares was able to give data on them. After this filtering process, the list of over 30 substitutions filtered out to 11 that passed the threshold, 1 of which predicted either deleterious, probably damaging, or disease-related for all 6 softwares, another 3 substitutions predicted as such for 5 out of 6 softwares, and the other 7 with such classification for 4 out of 6 softwares.

Swiss Model

Swiss Model (swissmodel.expasy.org/) is an online database for protein modelling and visualization. IThe initial plan was to extract pre-existing program database (PDB) files of TTN sequences acquired from Swiss Model. However, the issue of the inability to figure out specific positions the substitutions hold in the models that were already made arose. Fortunately, Swiss Model has a program that allows users to generate models of their own from a submitted amino acid sequence.

It works by accessing pre-existing structurally characterized templates and models from Protein Database (PDB) and uses comparative modelling to generate the models of the input protein sequence. However, there are over 850,000 known protein sequence entries and only 20,000 of these pre-existing templates from PDB, which makes it extremely difficult to access a majority of models and sequences.[29] Therefore, comparative modelling is only as effective and substantial as the PDB templates allow it. As a result, only certain sections of the protein sequence that have prior templates can be modelled. In addition, since the modelling process is entirely computer generated, there could be inaccuracies in the model.

Since the sequence limit is at 5,000, it became a necessity to submit the 34,350 residue long TTN sequence in groups of 5,000 residues, which resulted in 7 batches in total. Calculations were made in order to find the new positions of each substitution in their new respective batches as seen in Table 4.

As expected, upon the return of the Swiss Model query, only some of the amino acid substitutions were resided in areas that had coverage from the Swiss Model visualizations. 7 of the 11 substitutions were included in varying numbers of models, the final generated model to mutate was chosen by selecting the highest QMEAN4 value for each substitution. The QMEAN4 is an absolute quality measurement method that computes the 'degree of nativeness' of a generated model against its experimental counterparts. In Swiss Model, a QMEAN4 value

[26] Ivan A. Adzhubei et al., "A Method and Server for Predicting Damaging Missense Mutations," *Nature Methods*, 2010, https://doi.org/10.1038/nmeth0410-248.

[27] Mehran Akhtar et al., "Identification of Most Damaging NsSNPs in Human CCR6 Gene: In Silico Analyses," *International Journal of Immunogenetics* 46, no. 6 (December 1, 2019): 459–71, https://doi.org/10.1111/iji.12449.

[28] Warde-Farley, David, Sylva L. Donaldson, Ovi Comes, Khalid Zuberi, Rashad Badrawi, Pauline Chao, Max Franz et al. "The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function." *Nucleic acids research* 38, no. suppl_2 (2010): W214-W220. https://doi.org/10.1093/nar/gkq537.

[29] Schwede, Torsten, Jurgen Kopp, Nicolas Guex, and Manuel C. Peitsch. "SWISS-MODEL: an automated protein homology-modeling server." *Nucleic acids research* 31, no. 13 (2003): 3381-3385. https://doi.org/10.1093/nar/gkg520.

higher than -4 is considered high quality, while a score lower than -4 is considered low quality (Benkert et al., 2011).[30] All necessary PDB files acquired from Swiss Model were downloaded and imported into Swiss PDB viewer.

Swiss PDB Viewer

Swiss PDB Viewer (www.expasy.org/spdbv/) is a free visualization application that allows for a variety of different processes concerning protein modelling. It allows mutations of protein sequences, and has the ability to simulate what the structural consequences of these alterations are.

For each of the 7 amino acid changes, the corresponding PDB files acquired from Swiss Model were inputted into Swiss PDB viewer and searched for its position based on the previous calculations in Table 4. Afterwards, the amino acid changes were performed on the 3D model of the respective area of the gene.

Results and Discussion

Selection and Predictions of SNPs effects on Human TTN gene

Table 1. Quantitative Data Analysis of SIFT Results

SIFT Results	Frequency	Percentage*
Number of SNPs that were found within SIFT database	282	47.16%
Number of SNPs that were not found within SIFT database	316	52.84%
Number of Unique SNPs with a SIFT Prediction	9	1.51%
Number of Unique SNPs with no SIFT Prediction	589	98.49%

** percentage of total SNPs submitted*

All SNPs recruited from SNPedia were subjected to six bioinformatics softwares and servers. As apparent in the data (Table 1), more than half of the SNPs were not found within the SIFT database. These SNPs were labelled as "Not Found" by the server. Only 9 unique SNPs had prediction information, with 37 other SNPs having very limited data and figures. 228 unique amino acid substitutions were provided amongst all 46 SNPs. Out of the 10 SNPs with predictions from SIFT, 7 were classified as deleterious and 3 as tolerant. The detailed outputs can be found here. Additionally, the SIFT results identified that rs2244492 SNP was the most common out of those identified on SIFT, with an average allele frequency of 0.413. It is most prevalent amongst African alleles with a frequency of 0.607. However the most common SNP with a SIFT prediction of 'Deleterious' is rs139517732 with an average allele frequency of 0.001, being most prevalent in the East Asian alleles at a frequency of 0.003. Unfortunately, SIFT did not provide allele frequency data on a large number of SNPs, especially those that were given SIFT predictions, where 6 deleterious and 1 tolerated SNPs had no data.

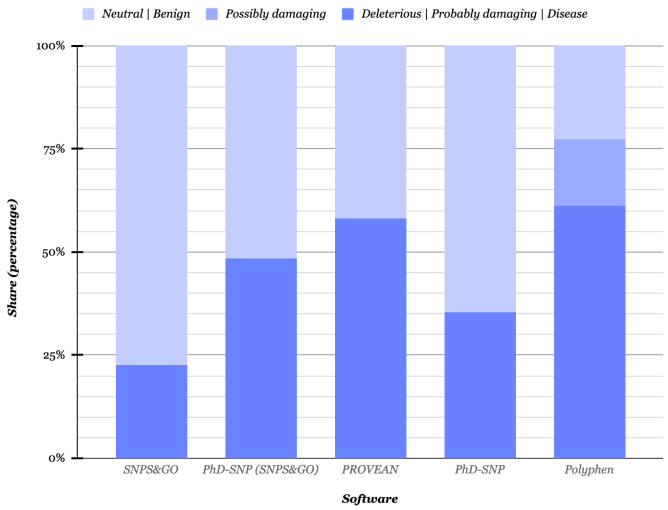
Table 2. SNPs with population allele information

SNP	*Prediction*	*Ave allele freq*	*Highest allele prevalence*	*Freq of highest prevalence*
rs2244492	*N/A*	*0.413*	*African*	*0.607*
rs4893853 K3154R	*Tolerated*	*0.086*	*American*	*0.154*
rs12476289 R1572Q	*Tolerated*	*0.075*	*American*	*0.172*
rs16866412	*N/A*	*0.073*	*East Asian*	*0.156*
rs17355446	*N/A*	*0.052*	*American*	*0.144*
rs184412722	*N/A*	*0.006*	*South asian*	*0.027*
rs72648247	*N/A*	*0.002*	*South asian*	*0.007*
rs72646867	*N/A*	*0.002*	*South Asian*	*0.004*
rs139517732 V54M	*Deleterious*	*0.001*	*East Asian*	*0.003*
rs56391938	*N/A*	*0*	*Europe*	*0.001*
rs371678190	*N/A*	*0*	*Europe*	*0.001*
rs147879266		*0*	*East Asian*	*0.001*

For a comprehensive data table of all results, click here. For the table with the 11 substitutions for modelling and visualisation, click here or view Appendix 2.

Figure 2. Share of Predictions across SNPs&GO, PROVEAN, PhD-SNP, and PolyPhen

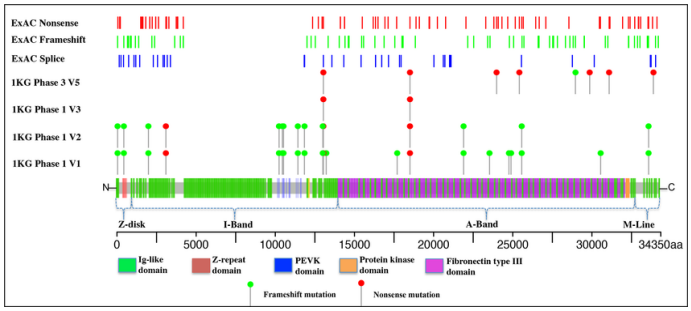
[30] Pascal Benkert, Marco Biasini, and Torsten Schwede, "Toward the Estimation of the Absolute Quality of Individual Protein Structure Models," *Bioinformatics*, 2011, https://doi.org/10.1093/bioinformatics/btq662.



Among the eleven SNPs that were predicted to be most deleterious, the replacement of Adenine to Guanine was the most prevalent as depicted in three mutations (Table 3). For the Guanine-Adenine and Cytosine-Thymine replacements, each had two mutations that depicted the change. Lastly, substitutions from Thymine-Cytosine, Cytosine-Adenine, Adenine-Thymine, and Thymine-Guanine only occur once.

Analysis of the Most Deleterious Substitutions

Figure 3. Spatial distribution of titin frameshift, nonsense and splice-site mutations in reference populations



Note. Reprinted from "Prevalence of Titin Truncating Variants in General Population," by Oyediran Akinrinade, Juha W. Koskenvuo, and Tero-Pekka Alastalo., 2015, PloS one

According to this figure from "Prevalence of Titin Truncating Variants in General Population"[31], the corresponding locations of each amino acid change within this study that passed the set threshold was derived, listed in Table 3.

Table 3. Location of Modeling Substitutions in the TTN gene

SNP	Amino Acid Change	Count	Location on TTN
rs138060032	R279W	4	Z-disk
rs184412722	N5958S	4	I-band
rs202094100	W14830C	5	A-band
rs2555818	C30037R	4	A-band
rs267607155	W976R	6	Z-disk
rs267607156	L34315P	5	M-line
rs281864931	H34305P	4	M-line
rs368277535	G34293R	4	M-line
rs375159973	W32431R	5	A-band
rs72648247	P29085S	4	A-band
rs72648263	A30244T	4	A-band

The data suggests the most prevalent region in which amino acid changes occur is the A-band, followed by the M-line, Z-disk, and lastly, the I-band. This consensus complements the findings of the study, "A Review of the Giant Protein Titin in Clinical Molecular Diagnostics of Cardiomyopathies", where truncating mutations in the A-band region were found to be responsible for up to 30% of all Dilated Cardiomyopathy cases out of 43% total contribution from truncating TTN mutations, making the A-band overrepresented.[32] The A-band region is responsible for stabilizing and regulating the thick filament within the protein, as well as providing binding sites for myosin and MyBP-C (Cardiac myosin-binding protein).[33] Studies also show that mutations in the A-band region of TTN is associated with hereditary myopathy with early respiratory failure.[34]

Protein Modelling and Visualization of amino acid substitutions

Seven of the following substitutions listed in the table had enough information to be modelled in Swiss PDB, the values

[31] Oyediran Akinrinade, Juha W. Koskenvuo, and Tero-Pekka Alastalo, "Prevalence of Titin Truncating Variants in General Population," *Plos One* 10, no. 12 (2015), https://doi.org/10.1371/journal.pone.0145284.

[32] Marta Gigli et al., "A Review of the Giant Protein Titin in Clinical Molecular Diagnostics of Cardiomyopathies," *Frontiers in Cardiovascular Medicine*, 2016, https://doi.org/10.3389/fcvm.2016.00021.

[33] LeWinter, Martin M., and Henk Granzier. "Cardiac titin: a multifunctional giant." *Circulation* 121, no. 19 (2010): 2137-2145. https://doi.org/10.1161/CIRCULATIONAHA.109.860171.

[34] Ohlsson, Monica, Carola Hedberg, Björn Brådvik, Christopher Lindberg, Homa Tajsharghi, Olof Danielsson, Atle Melberg, Bjarne Udd, Tommy Martinsson, and Anders Oldfors. "Hereditary myopathy with early respiratory failure associated with a mutation in A-band titin." *Brain* 135, no. 6 (2012): 1682-1694. https://doi.org/10.1093/brain/aws103.

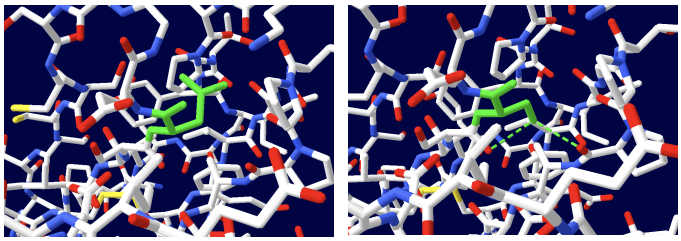
and data alongside the substitutions are to keep track of the computer generated models from Swiss Model (Figure 5). Figure 4 demonstrates the wild and mutant types for two of the substitutions with the highest QMEAN4 (model quality) value (-3.99 and -5.83 respectively). These figures show the spatial effects of the substitutions

Table 4. Swiss Model Inputs and Results

SNP	Amino Acid Change	Batch	New position	Model	QMEAN4 value
rs138060032	R279W	1	279	N/A	N/A
rs184412722	N5958S	2	958	2	-3.99
rs202094100	W14830C	3	4830	8	-5.83
rs2555818	C30037R	7	37	N/A	N/A
rs267607155	W976R	1	976	N/A	N/A
rs267607156	L34315P	7	4315	24	-6.92
rs281864931	H34305P	7	4305	24	-6.92
rs368277535	G34293R	7	4293	24	-6.92
rs375159973	W32431R	7	2431	6	-3.92
rs72648247	P29085S	6	4085	5	-6.26
rs72648263	A30244T	7	244	N/A	N/A

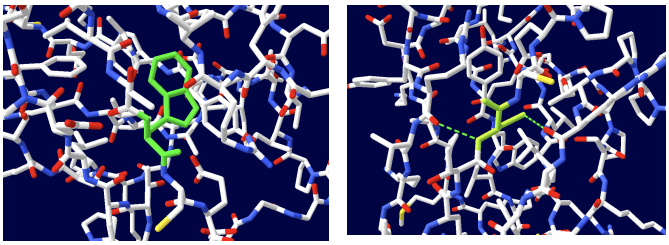
Figure 4. Swiss PDB Results

(N5958S, QMEAN4=-3.99)



Wild Type Mutant Type

(W1480C, QMEAN4=-5.83)

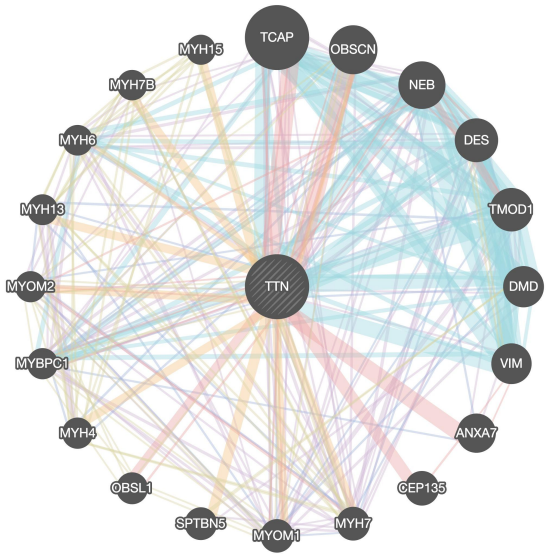


Wild Type Mutant Type

Interactions and relationship analysis between TTN and other genes

GeneMANIA detected that the TTN gene was associated with 20 other genes (Figure 5). Based on the output results, the human TCAP gene had the most interactions with the human TTN gene. This finding is relevant since the TCAP gene is associated with a rare form of recessive limb girdle muscular dystrophy.[35] The rest of the gene interactions are also found to be pertinent in this paper as they are also associated with diseases that correlate with the type of diseases caused by mutations in the TTN gene.

Figure 5. Genemania result of functional interaction between TTN gene and its related genes



Improvements

Due to the protein's massive size, most bioinformatics tools were not able to analyze TTN as a whole. For Swiss Model,

[35] Ferreiro et al., "Telethonin-Deficiency Initially Presenting as a Congenital Muscular Dystrophy," *Neuromuscular Disorders*, 2011, https://doi.org/10.1016/j.nmd.2011.03.005.

there was an inability to cut off parts of TTN's 63206-character protein sequence to analyze specific parts of the protein. Only certain sections and domains have been investigated and modelled, leaving an incomprehensible knowledge of the structural and functional components of the TTN gene as a whole. Furthermore, QMEAN4 value did not remain consistent and retain the highest quality possible for the Swiss PDB-generated models. However, with human involvement to check if the model is coherent, higher quality and accuracy for the Swiss PDB-generated models could be ensured. Another improvement is to also specify certain types of SNPs (ie. missense, non-synonymous, etc.) to investigate for a more focused aim and experiment. The use of bioinformatics tools during the course of this research only provides predictions based on computational and mathematical algorithms. In vitro analysis would support the credibility of the in silico findings in this study, as there were many independent variables that aren't taken into account in the bioinformatics tools used. Additionally, other possible future research includes investigating the effects of individual SNPs based on the results of this study or investigating the effects of SNPs on miRNA target sites.

Conclusion

This analysis of SNPs in the TTN gene was conducted using bioinformatics softwares and servers that operate through mathematical algorithms. After multiple layers of in silico analyses on over 30 amino acid changes of the TTN gene, it's concluded that the most deleterious mutations are most prevalent in the A-band region of the TTN protein. As aforementioned, this is significant as the A-band region is responsible for stabilizing and regulating the filament within the protein, in addition to providing binding sites for myosin and MyBP-C.[36] Another study found that on a cellular level, deletions on TTN A-band/I-band junction in mice resulted in passive stiffness in the muscle. While on an overall functional level, it resulted in several symptoms and features of heart failure, such as exercise dysfunction and concentric cardiac hypertrophy.[37] This demonstrates that changes to the A-band region can result in significant consequences, however research and knowledge of the protein can aid in gene therapy methods and drug development for the several diseases related to TTN.

[36] LeWinter, Martin M., and Henk Granzier. "Cardiac titin: a multifunctional giant." *Circulation* 121, no. 19 (2010): 2137-2145. https://doi.org/10.1161/CIRCULATIONAHA.109.860171.

[37] Granzier, Henk L., Kirk R. Hutchinson, Paola Tonino, Mei Methawasin, Frank W. Li, Rebecca E. Slater, Mathew M. Bull, et al. "Deleting Titin's I-Band/A-Band Junction Reveals Critical Roles for Titin in Biomechanical Sensing and Cardiac Function." *Proceedings of the National Academy of Sciences* 111, no. 40 (2014): 14589–94. https://doi.org/10.1073/pnas.1411493111.

Acknowledgements

This research would not have been possible without the help and support from IgniteMinds and our mentors, Aksh Garg and Jason Apostol.

Reference List

Adzhubei, Ivan A., Steffen Schmidt, Leonid Peshkin, Vasily E. Ramensky, Anna Gerasimova, Peer Bork, Alexey S. Kondrashov, and Shamil R. Sunyaev. "A Method and Server for Predicting Damaging Missense Mutations." *Nature Methods*, 2010. https://doi.org/10.1038/nmeth0410-248.

Akhtar, Mehran, Tazkira Jamal, Hina Jamal, Jalal Ud Din, Muhsin Jamal, Muhammad Arif, Maria Arshad, and Fazal Jalil. "Identification of Most Damaging NsSNPs in Human CCR6 Gene: In Silico Analyses." *International Journal of Immunogenetics* 46, no. 6 (December 1, 2019): 459–71. https://doi.org/10.1111/iji.12449.

Akinrinade, Oyediran, Juha W. Koskenvuo, and Tero-Pekka Alastalo. "Prevalence of Titin Truncating Variants in General Population." *Plos One* 10, no. 12 (2015). https://doi.org/10.1371/journal.pone.0145284.

Apweiler, R. "UniProt: The Universal Protein Knowledgebase." *Nucleic Acids Research*, 2004. https://doi.org/10.1093/nar/gkh131.

Bang, Marie Louise, Thomas Centner, Friderike Fornoff, Adam J. Geach, Michael Gotthardt, Mark McNabb, Christian C. Witt, et al. "The Complete Gene Sequence of Titin, Expression of an Unusual ≈700-KDa Titin Isoform, and Its Interaction with Obscurin Identify a Novel Z-Line to I-Band Linking System." *Circulation Research*, 2001. https://doi.org/10.1161/hh2301.100981.

Benkert, Pascal, Marco Biasini, and Torsten Schwede. "Toward the Estimation of the Absolute Quality of Individual Protein Structure Models." *Bioinformatics*, 2011. https://doi.org/10.1093/bioinformatics/btq662.

Børsting, C., and N. Morling. "Single-Nucleotide Polymorphisms." In *Encyclopedia of Forensic Sciences: Second Edition*, 2013. https://doi.org/10.1016/B978-0-12-382165-2.00042-8.

Calabrese, Remo, Emidio Capriotti, Piero Fariselli, Pier Luigi Martelli, and Rita Casadio. "Functional Annotations Improve the Predictive Score of Human Disease-Related Mutations in Proteins." *Human Mutation*, 2009. https://doi.org/10.1002/humu.21047.

Capriotti, Emidio, and Russ B. Altman. "Improving the Prediction of Disease-Related Variants Using Protein Three-Dimensional Structure." *BMC Bioinformatics,* 2011. https://doi.org/10.1186/1471-2105-12-S4-S3. P

Cariaso, Michael, and Greg Lennon. "SNPedia: A Wiki Supporting Personal Genome Annotation, Interpretation and Analysis." *Nucleic Acids Research*, 2012. https://doi.org/10.1093/nar/gkr798.

Carmignac, Virginie, Mustafa A.M. Salih, Susana Quijano-Roy, Sylvie Marchand, Molham M. Al Rayess, Maowia M. Mukhtar, Jon A. Urtizberea, et al. “C-Terminal Titin Deletions Cause a Novel Early-Onset Myopathy with Fatal Cardiomyopathy.” *Annals of Neurology*, 2007. https://doi.org/10.1002/ana.21089.

Ceyhan-Birsoy, Ozge, Pankaj B. Agrawal, Carlos Hidalgo, Klaus Schmitz-Abe, Elizabeth T. Dechene, Lindsay C. Swanson, Rachel Soemedi, et al. “Recessive Truncating Titin Gene, TTN, Mutations Presenting as Centronuclear Myopathy.” *Neurology*, 2013. https://doi.org/10.1212/WNL.0b013e3182a6ca62.

Choi, Yongwook, and Agnes P. Chan. “PROVEAN Web Server: A Tool to Predict the Functional Effect of Amino Acid Substitutions and Indels.” *Bioinformatics*, 2015. https://doi.org/10.1093/bioinformatics/btv195.

Dakal, Tikam Chand, Deepak Kala, Gourav Dhiman, Vinod Yadav, Andrey Krokhotin, and Nikolay V. Dokholyan. “Predicting the Functional Consequences of Non-Synonymous Single Nucleotide Polymorphisms in IL8 Gene.” *Scientific Reports*, 2017. https://doi.org/10.1038/s41598-017-06575-4.

Ferreiro, Ana, Monica Mezmezian, Montse Olivé, Danielle Herlicoviez, Michel Fardeau, Pascale Richard, and Norma Beatriz Romero. “.” *Neuromuscular Disorders*, 2011. https://doi.org/10.1016/j.nmd.2011.03.005.

Giacomini, K. M., C. M. Brett, R. B. Altman, N. L. Benowitz, M. E. Dolan, D. A. Flockhart, J. A. Johnson, et al. “The Pharmacogenetics Research Network: From SNP Discovery to Clinical Drug Response.” *Clinical Pharmacology and Therapeutics*, 2007. https://doi.org/10.1038/sj.clpt.6100087.

Gigli, Marta, Rene L. Begay, Gaetano Morea, Sharon L. Graw, Gianfranco Sinagra, Matthew R.G. Taylor, Henk Granzier, and Luisa Mestroni. “A Review of the Giant Protein Titin in Clinical Molecular Diagnostics of Cardiomyopathies.” *Frontiers in Cardiovascular Medicine*, 2016. https://doi.org/10.3389/fcvm.2016.00021.

Granzier, Henk L., Kirk R. Hutchinson, Paola Tonino, Mei Methawasin, Frank W. Li, Rebecca E. Slater, Mathew M. Bull, et al. “Deleting Titin’s I-Band/A-Band Junction Reveals Critical Roles for Titin in Biomechanical Sensing and Cardiac Function.” *Proceedings of the National Academy of Sciences* 111, no. 40 (2014): 14589–94. https://doi.org/10.1073/pnas.1411493111.

Hackman, Peter, Anna Vihola, Henna Haravuori, Sylvie Marchand, Jaakko Sarparanta, Jerome De Seze, Siegfried Labeit, et al. “Tibial Muscular Dystrophy Is a Titinopathy Caused by Mutations in TTN, the Gene Encoding the Giant Skeletal-Muscle Protein Titin.” *American Journal of Human Genetics*, 2002. https://doi.org/10.1086/342380.

Herman, Daniel S., Lien Lam, Matthew R.G. Taylor, Libin Wang, Polakit Teekakirikul, Danos Christodoulou, Lauren Conner, et al. “Truncations of Titin Causing Dilated Cardiomyopathy.” *New England Journal of Medicine* 366, no. 7 (2012): 619–28. https://doi.org/10.1056/NEJMoa1110186.

Hinson, John T., Anant Chopra, Navid Nafissi, William J. Polacheck, Craig C. Benson, Sandra Swist, Joshua Gorham, et al. “Titin Mutations in IPS Cells Define Sarcomere Insufficiency as a Cause of Dilated Cardiomyopathy.” *Science* 349, no. 6251 (2015): 982–86. https://doi.org/10.1126/science.aaa5458.

Kötter, Sebastian, Christian Andresen, and Martina Krüger. “Titin: Central Player of Hypertrophic Signaling and Sarcomeric Protein Quality Control.” *Biological Chemistry*, 2014. https://doi.org/10.1515/hsz-2014-0178.

Krüger, Martina, and Sebastian Kötter. “Titin, a Central Mediator for Hypertrophic Signaling, Exercise-Induced Mechanosignaling and Skeletal Muscle Remodeling.” *Frontiers in Physiology*, 2016. https://doi.org/10.3389/fphys.2016.00076.

Kumar, Prateek, Steven Henikoff, and Pauline C. Ng. “Predicting the Effects of Coding Non-Synonymous Variants on Protein Function Using the SIFT Algorithm.” *Nature Protocols*, 2009. https://doi.org/10.1038/nprot.2009.86.

Labeit, Siegfried, Bernhard Kolmerer, and Wolfgang A. Linke. “The Giant Protein Titin: Emerging Roles in Physiology and Pathophysiology.” Circulation Research (1997). https://doi.org/10.1161/01.RES.80.2.290.

LeWinter, Martin M., and Henk Granzier. "Cardiac titin: a multifunctional giant." *Circulation* 121, no. 19 (2010): 2137-2145. https://doi.org/10.1161/CIRCULATIONAHA.109.860171.

Müller-Seitz, M., K. Kaupmann, S. Labeit, and Harald Jockusch. "Chromosomal localization of the mouse titin gene and its relationto “muscular dystrophy with myositis” and nebulin genes on chromosome 2." *Genomics* 18, no. 3 (1993): 559-561. https://doi.org/10.1016/S0888-7543(05)80356-8.

Ng, Pauline C., and Steven Henikoff. "SIFT: Predicting amino acid changes that affect protein function." *Nucleic acids research* 31, no. 13 (2003): 3812-3814. https://doi.org/10.1093/nar/gkg509

Nimir, Mohammed, Mohanad Abdelrahim, Mohamed Abdelrahim, Mahil Abdalla, Wala eldin Ahmed, Muhanned Abdullah, and Muzamil Mahdi Abdel Hamid. "In silico analysis of single nucleotide polymorphisms (SNPs) in human FOXC2 gene." *F1000Research* 6 (2017). https://doi.org/10.12688/f1000research.10937.2.

Ohlsson, Monica, Carola Hedberg, Björn Brådvik, Christopher Lindberg, Homa Tajsharghi, Olof Danielsson, Atle Melberg, Bjarne Udd, Tommy Martinsson, and Anders Oldfors. "Hereditary myopathy with early respiratory failure associated with a mutation in A-band titin." *Brain* 135, no. 6 (2012): 1682-1694. https://doi.org/10.1093/brain/aws103.

Prilusky, Jaime, Eran Hodis, David Canner, Wayne A. Decatur, Karl Oberholser, Eric Martz, Alexander Berchanski, Michal Harel, and Joel L. Sussman. "Proteopedia: a status report on the collaborative, 3D web-encyclopedia of proteins and other biomolecules." *Journal of structural biology* 175, no. 2 (2011): 244-252. https://doi.org/10.1016/j.jsb.2011.04.011.

Rossi, Elena, Antonio Faiella, Massimo Zeviani, Siegfried Labeit, Giovanna Floridia, Silvia Brunelli, Marina Cammarata, Edoardo Boncinelli, and Orsetta Zuffardi. "Order of six loci at 2q24-q31 and orientation of the HOXD locus." *Genomics* 24, no. 1 (1994): 34-40. https://doi.org/10.1006/geno.1994.1579.

Savarese, Marco, Lorenzo Maggi, Anna Vihola, Per Harald Jonson, Giorgio Tasca, Lucia Ruggiero, Luca Bello et al. "Interpreting genetic variants in titin in patients with muscle disorders." *JAMA neurology* 75, no. 5 (2018): 557-565. https://doi.org/10.1001/jamaneurol.2017.4899.

Schwede, Torsten, Jurgen Kopp, Nicolas Guex, and Manuel C. Peitsch. "SWISS-MODEL: an automated protein homology-modeling

server." *Nucleic acids research* 31, no. 13 (2003): 3381-3385. https://doi.org/10.1093/nar/gkg520.

Taylor, Matthew, Sharon Graw, Gianfranco Sinagra, Carl Barnes, Dobromir Slavov, Francesca Brun, Bruno Pinamonti et al. "Genetic variation in titin in arrhythmogenic right ventricular cardiomyopathy–overlap syndromes." *Circulation* 124, no. 8 (2011): 876-885. https://doi.org/10.1161/CIRCULATIONAHA.110.005405.

UniProt Consortium. "UniProt: a worldwide hub of protein knowledge." *Nucleic acids research* 47, no. D1 (2019): D506-D515. https://doi.org/10.1093/nar/gky1049.

Warde-Farley, David, Sylva L. Donaldson, Ovi Comes, Khalid Zuberi, Rashad Badrawi, Pauline Chao, Max Franz et al. "The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function." *Nucleic acids research* 38, no. suppl_2 (2010): W214-W220. https://doi.org/10.1093/nar/gkq537.

Biography

Lei Jo Presaldo

Lei Jo Presaldo, a high school student from the Philippines, is a Department of Science and Technology scholar and studies at Philippine Science High School - Main Campus. She is passionate in the sciences especially Biology, Chemistry, Research, and Computer Science. As a student with excellent academic standing, she is awarded by her school as part of the Director's List.

Monica Sing

Monica is a DP student from the International School of Phnom Penh, Cambodia. She is interested in Math, Particle Physics, Computer Science, and Economics. In the future, she hopes to pursue a career that involves the intersection of the aforementioned subjects. She has consistently demonstrated high academic achievement, having won two math subject awards and obtaining the Director's List all 5 years. She also enjoys partaking in extracurricular activities with a number of science and humanities related online student run organizations. Her deep interest in the sciences has inspired her to create STEM kits for underprivileged students as her MYP Personal Project, with the goal of fostering STEM education and awareness in Cambodia.

Toma Ogawa

Toma Ogawa, an IBDP student at Pathways School Gurgaon is highly passionate about Mathematics, Biology, and Chemistry. He is awarded by the school in academic excellence in the aforementioned subjects. He is the curator and the head of marketing for the stem magazine "Newston." His passion for biology motivated him to pursue biology as his extended essay in which he is currently conducting experiments regarding plant biology.

Aksh Garg - Mentor

Aksh Garg is extremely passionate about Math, Physics, Chemistry, and Machine Learning. He's worked on a multitude of machine learning projects ranging from computer vision–image recognition for autonomous driving vehicles, object detection systems, and COVID-19 classifiers with over 99.5% accuracy– to natural language processing–trigger word detectors, stock market predictors, and language translation systems. One of his most prominent projects is "Machine Learning Coupled Trajectory and Communication Design for UAV Facilitated Wireless Networks," which earned him 1st place at LA County Science and Engineering Fair and 1st place at Palos Verdes Peninsula Science and Engineering Fair, and bestowed upon him the honor of being an ISEF Finalist. His current project, "CoronaCare: An Integrated Mobile Platform for Rapid, Accurate, and Accessible COVID-19 Diagnosis," deals with using convolutional neural networks for diagnosing COVID-19.

Appendix

Appendix 1. GeneMANIA results of genes physical interactions, co-expression, and shared protein domain with the TTN gene

Gene	Description	Physical Interactions	Co-expression	Shared Protein Domains
TCAP	titin-cap	Yes	Yes	No
OBSCN	obscurin, cytoskeletal calmodulin and titin-interacting RhoGEF	Yes	Yes	Yes
NEB	nebulin	Yes	Yes	No
DES	desmin	Yes	Yes	Yes
TMOD1	tropomodulin	Yes	Yes	No
DMD	dystrophin	No	Yes	Yes
VIM	vimentin	Yes	Yes	Yes
ANXA7	annexin A7	Yes	Yes	No
CEP135	Centrosomal protein 135	Yes	No	No
MYH7	Myosin, heavy chain 7, cardiac muscle, beta	Yes	Yes	Yes
MYOM1	myomesin 1	Yes	Yes	Yes
SPTBN5	Spectrin beta, non-erythrocytic 5	No	Yes	Yes

OBSL1	Obscurin-like 1	Yes	No	Yes
MYH4	Myosin, heavy chain 4, skeletal muscle	No	Yes	Yes
MYBPC 1	Myosin binding protein C, slow type	Yes	Yes	Yes
MYOM2	Myomesin 2	Yes	Yes	Yes
MYH13	Myomesin, heavy chain 13, skeletal muscle	No	Yes	Yes

MYH6	Myomesin, heavy chain 6, cardiac muscle, alpha	No	Yes	Yes
MYH7B	Myomesin, heavy chain 7B, cardiac muscle, beta	No	Yes	Yes
MYH15	Myosin, heavy chain 15	No	Yes	Yes

Appendix 2. *Highly damaging SNPs as predicted by SIFT, SNPs&GO, PROVEAN, PhD-SNP, Polyphen-2, and I-MUTANT*

SNP	Amino Acid Change	COUNT (x/6)	SIFT		SNPs&GO			PhD-SNP (on SNPS&GO)			PROVEAN		PhD-SNP		PolyPhen-2		I-Mutant		
			Prediction	Score	Score	Probability	RI	Prediction	Probability	RI	Prediction	Score	Prediction	RI	Effect	Score	DDG	RI	Prediction
rs138060032	R279W	4	DELETERIOUS	0.000	NEUTRAL	0.429	1	DISEASE	0.717	4	DISEASE	-5.028	NEUTRAL	6	PROBABLY DAMAGING	1	-0.2	2	DISEASE
rs184412722	N5958S	4	N/A	N/A	DISEASE	0.644	3	DISEASE	0.692	4	DISEASE	-2.602	DISEASE	0	BENIGN	0.312	N/A	N/A	ERROR
rs202094100	W14830C	5	N/A	N/A	DISEASE	0.665	3	DISEASE	0.907	8	DISEASE	-10.456	DISEASE	7	PROBABLY DAMAGING	1	-1.34	6	DECREASE
rs2555818	C30037R	4	N/A	N/A	NEUTRAL	0.381	2	DISEASE	0.582	2	DISEASE	-8.853	DISEASE	7	PROBABLY DAMAGING	1	0.18	0	DECREASE
rs267607155	W976R	6	DELETERIOUS	0.001	DISEASE	0.554	1	DISEASE	0.791	6	DISEASE	-9.811	DISEASE	3	PROBABLY DAMAGING	1	-0.36	3	DECREASE
rs267607156	L34315P	5	N/A	N/A	DISEASE	0.775	6	DISEASE	0.856	7	DISEASE	-5.133	DISEASE	1	PROBABLY DAMAGING	1	-0.41	3	DECREASE
rs281864931	H34305P	4	N/A	N/A	NEUTRAL	0.281	4	DISEASE	0.626	3	DISEASE	-6.622	DISEASE	0	PROBABLY DAMAGING	0.998	0	2	DECREASE
rs368277535	G34293R	4	N/A	N/A	NEUTRAL	0.315	4	DISEASE	0.565	1	DISEASE	-5.6	DISEASE	4	PROBABLY DAMAGING	1	-0.7	6	DECREASE
rs375159973	W32431R	5	N/A	N/A	DISEASE	0.530	1	DISEASE	0.813	6	DISEASE	-10.727	DISEASE	7	PROBABLY DAMAGING	1	-1.48	7	DECREASE
rs72648247	P29085S	4	N/A	N/A	NEUTRAL	0.205	6	DISEASE	0.563	1	DISEASE	-5.98	DISEASE	4	PROBABLY DAMAGING	1	-1.37	8	DECREASE
rs72648263	A30244T	4	N/A	N/A	DISEASE	0.629	3	DISEASE	0.840	7	DISEASE	-3.093	NEUTRAL	1	PROBABLY DAMAGING	1	-1.76	9	DECREASE