IN SILICO ANALYSIS OF SINGLE NUCLEOTIDE POLYMORPHISMS (SNPs) IN THE HUMAN TITIN (TTN) GENE

OGAWA, PRESALDO, SING

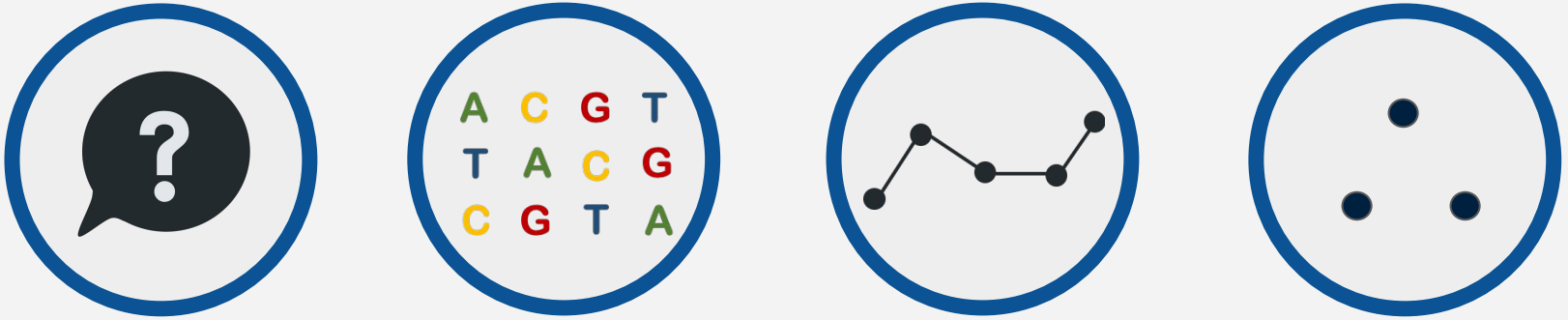
ACKNOWLEDGEMENTS

We would like to express our greatest gratitude to Aksh Garg and Jason Apostol for guiding us through the course of our research and for equipping us with skills that helped the completion of this project.

To IgniteMinds and Bioinformatics4All, thank you for this platform to present our results with fellow researchers and other students all over the world.

This research would not be possible without your assistance. Thank you.

FLOW OF THE PRESENTATION

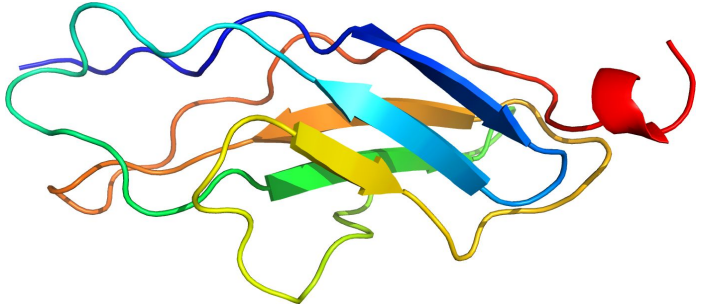


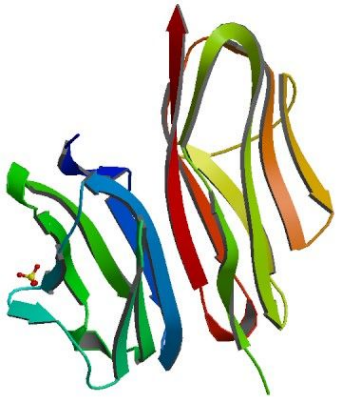
Introduction | Methodology | Results | Conclusions

TITIN (TTN)

Also known as connectin

- Largest protein in the human body
- 60,000 + character amino acid sequence
- Encoded by the TTN gene
- Functions as a ‘spring’
    - Responsible for the elasticity of our muscles

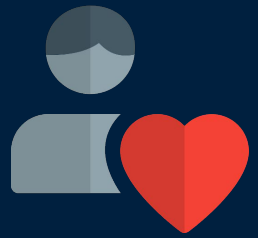




WHAT DO STATISTICS TELL US

Cardiomyopathy often goes undiagnosed. According to the Centers for Disease Control and Prevention, as many as 1 of 500 adults may have cardiomyopathy.

Dilated cardiomyopathy, the most common form, affects five in 100,000 adults and 0.57 in 100,000 children.



Dilated cardiomyopathy third leading cause of heart failure in the United States behind coronary artery disease (CAD) and hypertension

SNPs

- A variation in a single base pair in a DNA sequence
- A type of mutation
- A single nucleotide substitution that is present in a sufficiently large fraction of the population
- Occurs in more than 1% of a population

VS

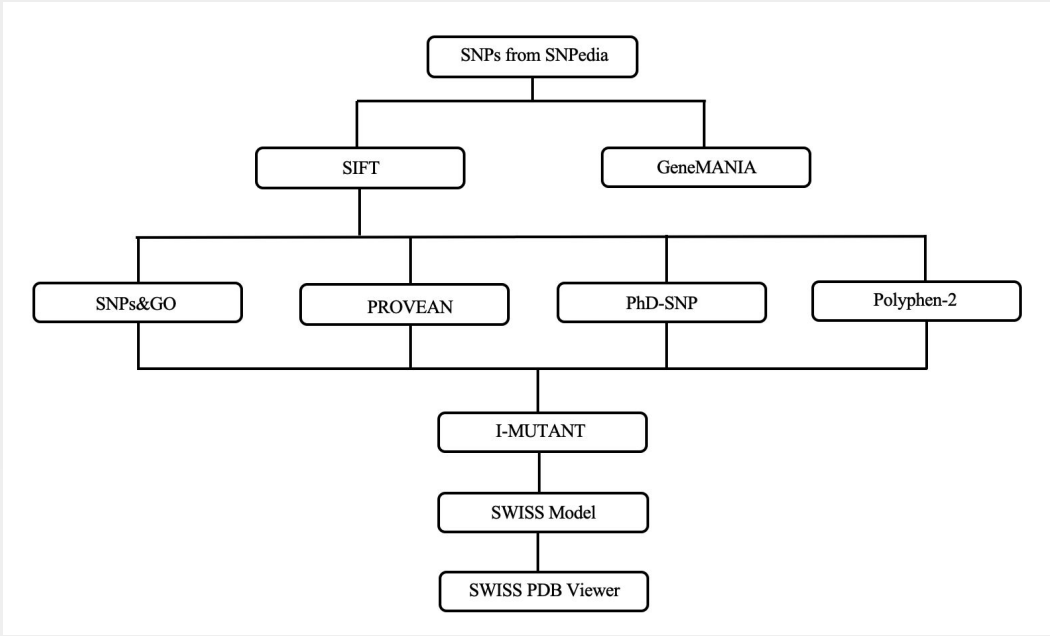
MUTATIONS

- Variation in a DNA's sequence leading to deletion, insertion, duplication, or substitution of one or more nucleotides
- Occurs in less than 1% of the population

What happens when something goes wrong in the TTN protein?

- Centronuclear Myopathy
- Tibular Muscular Dystrophy
- Early-onset Myopathy
- Arrythmogenic Right Ventricular Dystrophy
- Familial Dilated Cardiomyopathy
- Hereditary Myopathy

METHODOLOGY



METHODOLOGY

A. Selection of SNPs in the human TTN gene
   - SNPedia database

B. Prediction and Identification of Deleterious SNPs
   - SIFT
   - SNPs&GO
   - PROVEAN
   - PhD-SNP
   - PolyPhen-2
   - I-MUTANT
   - GeneMANIA

C. Modelling Amino Acid Substitutions on a Structural and Functional Level
   - Swiss Model
   - Swiss PDB Viewer

DATA COLLECTION RESULTS

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

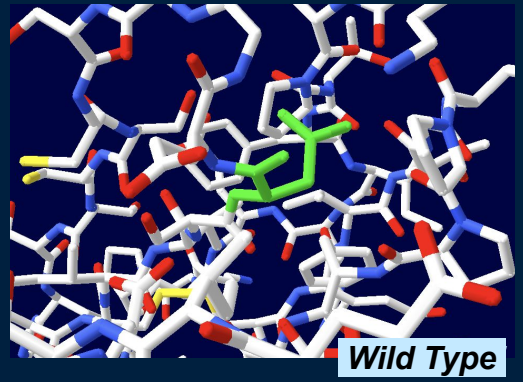
PROCESS OF SELECTING SNPS

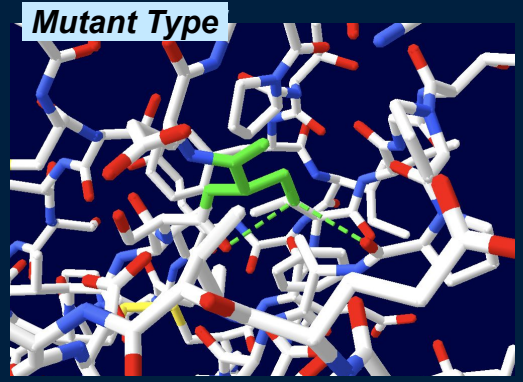
MODELLING

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

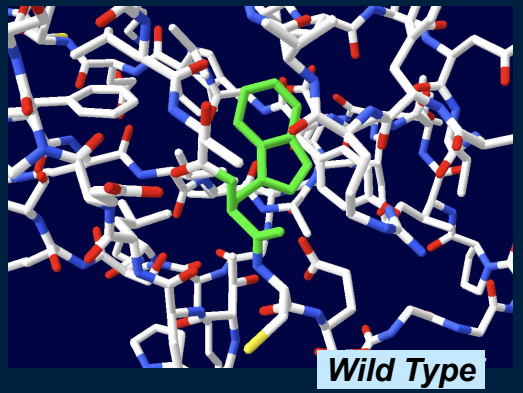
- Submitted batches of 5000 nucleotides in Swiss Model
- For each respective SNP position, found a model with the highest QMEAN4 value
- Input into swiss PDB viewer

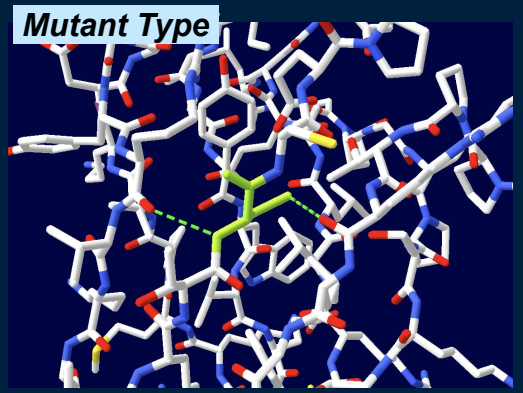
(N5958S, QMEAN4=-3.99)





(W1480C, QMEAN4=-5.83)



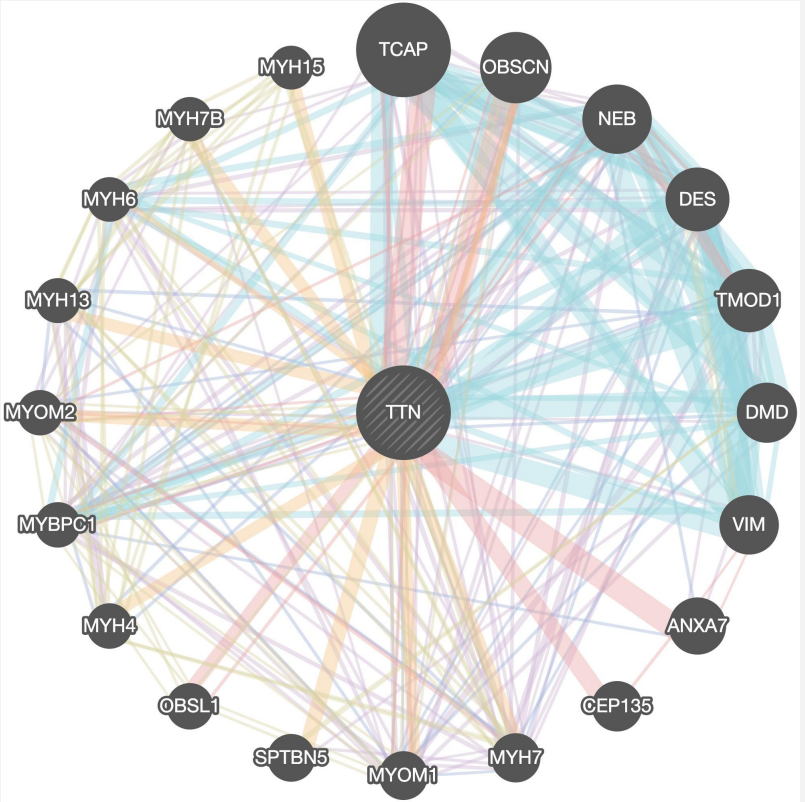


RESULTS

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

- SIFT
  - rs2244492 SNP was the most common out of those identified on SIFT, with an average allele frequency of 0.413
    - most prevalent amongst African alleles with a frequency of 0.607.
  - 11 SNPs (Rs138060032, rs202094100, rs2555818, rs267607155, rs267607156, rs281864931, rs368277535, rs374615369, rs375159973, and rs72648247) are predicted to be most deleterious
- I-Mutant
  - stability has nothing to do with deleteriousness
  - a majority (45%) of the deleterious SNPs were located in the A-band region of the TTN gene

GENE MANIA RESULTS



LIMITATIONS AND IMPROVEMENTS

- More specific types of SNPs to experiment with rather than analyzing them in general
- Couldn’t find a verified pdb model
  - Due to its large size
  - Indexing did not align with our frame of work
- Cutting down amino acid seq into smaller bits
  - Less accurate structural representation
  - Computer generated model is not verified by humans
  - Can not ensure same, high quality of models across different SNPs
  - Can not ensure all SNPs have a model to work with
- All analysis within this research is computer generated and theory based. Experimental studies can greatly aid in verifying these results

THANK YOU

THANK YOU

THANK YOU