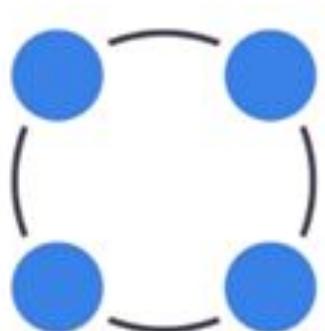


Изучение спектра вариантов генов, кодирующих структурные белки саркомера (TTN)



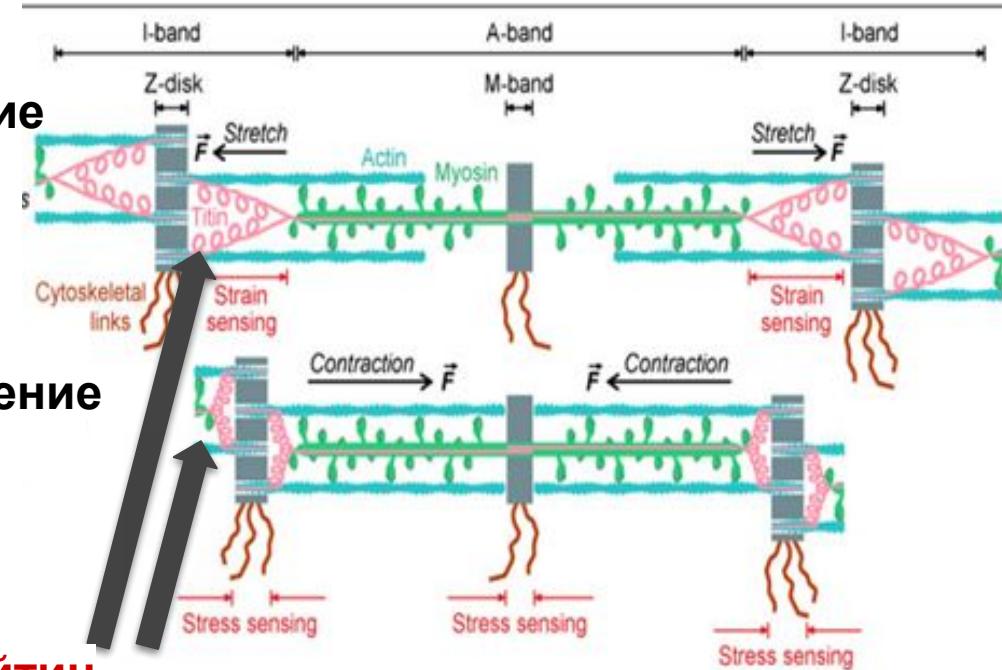
Лебеденко Ольга

Руководитель: Киселёв Артём
ФГБУ «НМИЦ им. В. А. Алмазова» Минздрава России

Тайтин структурный белок саркомера

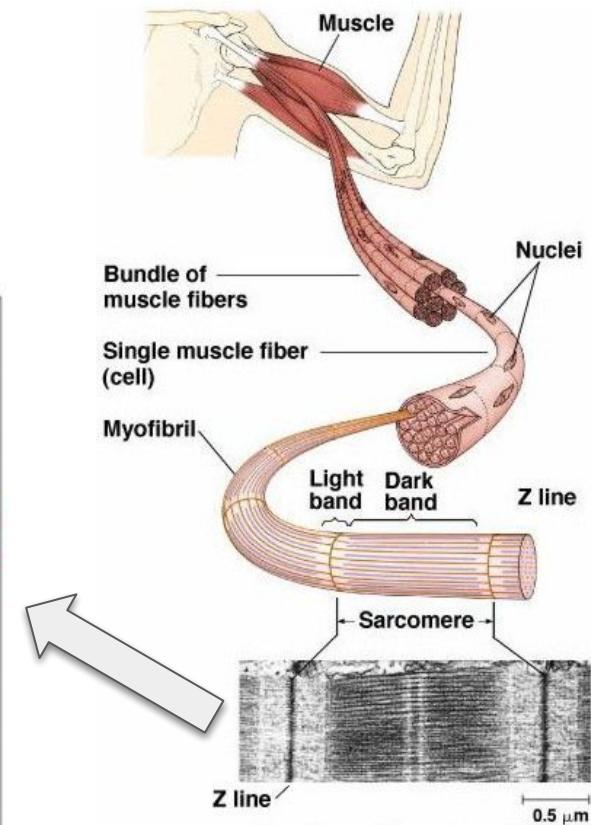
Отвечает за эластические свойства мускулатуры, имеет сигнальные функции, участвует в формировании саркомера

растяжение



сокращение

тайтин



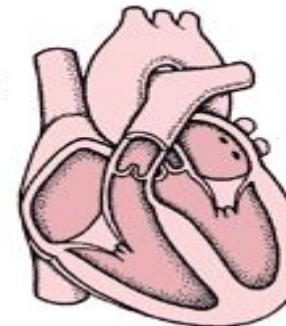
саркомер - единица
строения мышечного
волокна

Структурные изменения в сердечной мышце могут приводить к кардиомиопатии

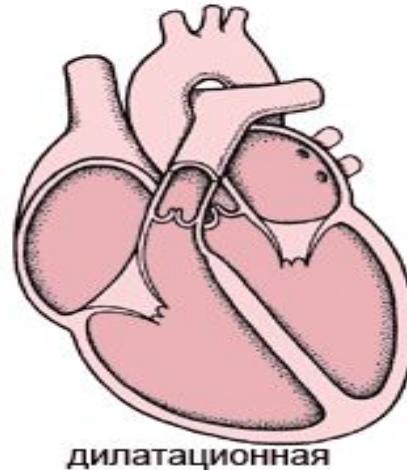
Cardio
heart

myopathy
weak muscle

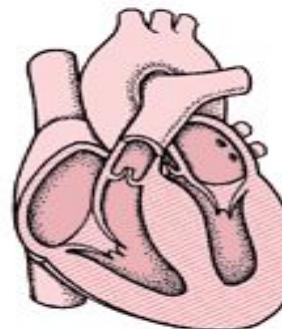
Нормальное сердце



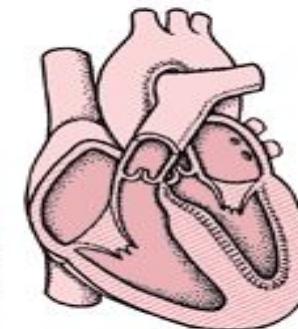
КАРДИОМИОПАТИИ:



дилатационная



гипертрофическая



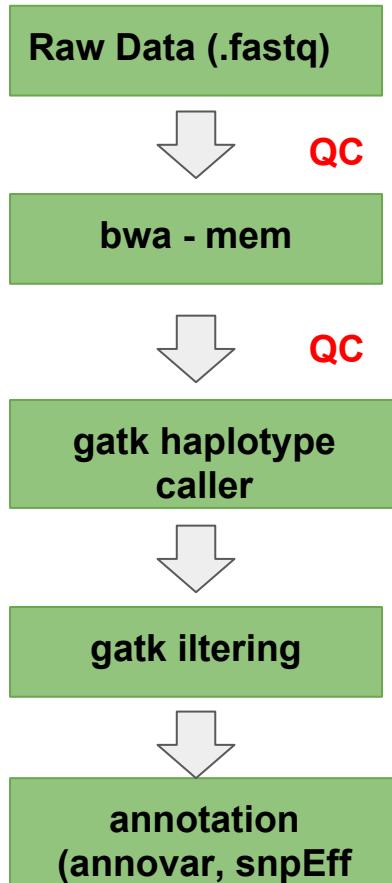
рестриктивная

Цель проекта: выделение и описание вариантов гена тайтина, характерных для больных с различными вариантами наследственных кардиомиопатий

Задачи: анализ сырых данных при помощи GATK, популяционный анализ, статистическая обработка



SNP Calling pipeline for Haploplex cardio panel (titin)



Snakemake pipeline implementation

Nsamples = 151

1) Digest genomic DNA.



2) Hybridize the HaloPlex probe library in presence of the Indexing Primer Cassette. Hybridization results in gDNA fragment circularization and incorporation of indexes and Illumina sequencing motifs.

Haploplex sample preparation

3) Capture target DNA-probe hybrids. Biotinylation of probe DNA allows capture using streptavidin-coated magnetic beads.



4) PCR amplify targeted fragments to produce a sequencing-ready, target-enriched sample.

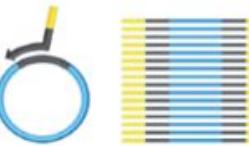


Figure 1 Overall HaloPlex target-enriched sequencing sample preparation workflow.

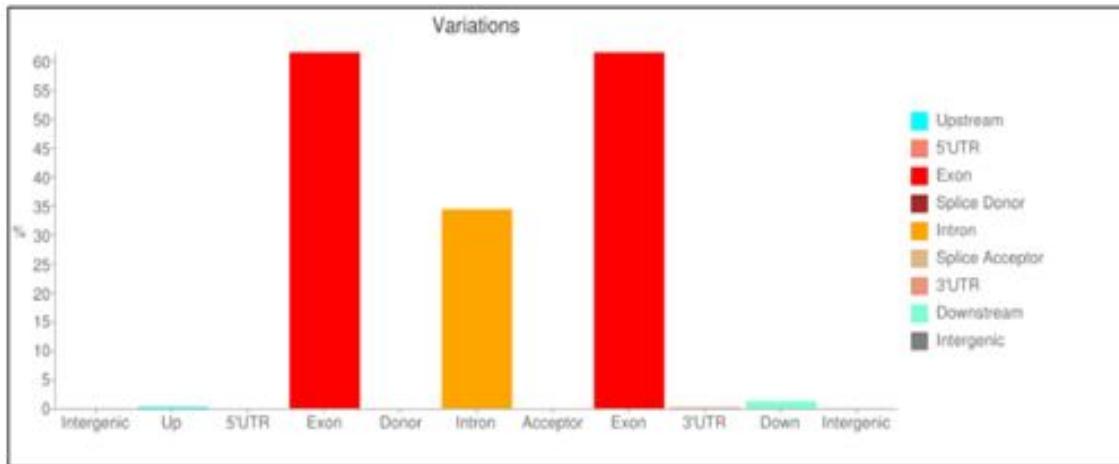
Annotation results 418 SNP by.snpEff

Number of effects by type and region

Type	Region				
Type (alphabetical order)	Count	Percent	Type (alphabetical order)	Count	Percent
DOWNSTREAM	42	1.201%	DOWNSTREAM	42	1.234%
EXON	17	0.486%	EXON	2,089	61.369%
INTERGENIC	1	0.029%	INTERGENIC	1	0.029%
INTRON	1,220	34.887%	INTRON	1,169	34.342%
NEXT_PROT	26	0.743%	SPLICE_SITE_REGION	63	1.851%
NON_SYNONYMOUS_CODING	1,285	36.745%	TRANSCRIPT	26	0.764%
PROTEIN_STRUCTURAL_INTERACTION_LOCUS	90	2.574%	UPSTREAM	8	0.235%
SPLICE_SITE_REGION	93	2.659%	UTR_3_PRIME	6	0.176%
SYNONYMOUS_CODING	709	20.275%			
UPSTREAM	8	0.229%			
UTR_3_PRIME	8	0.172%			

Variants rate details

Chromosome	Length	Variants	Variants rate
2	243,199,373	418	581,816
Total	243,199,373	418	581,816



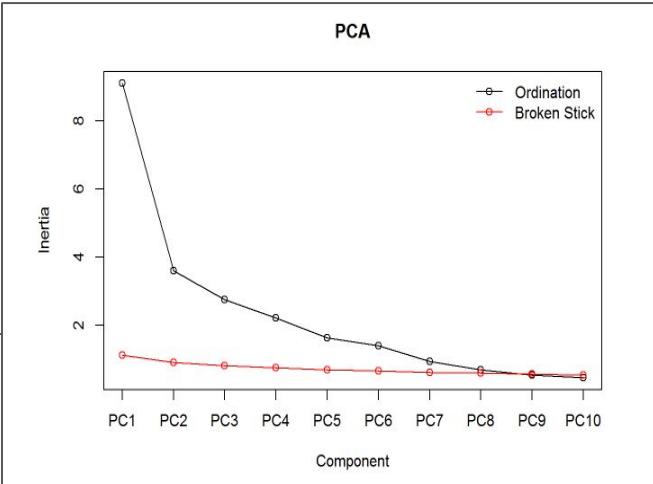
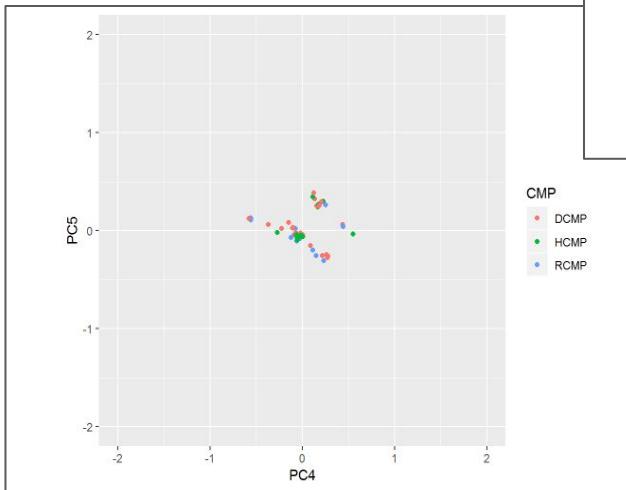
Missense / Silent ratio: 1.8124

Ts/Tv (transitions / transversions)

Transitions	6,893
Transversions	1,247
Ts/Tv ratio	5.5277

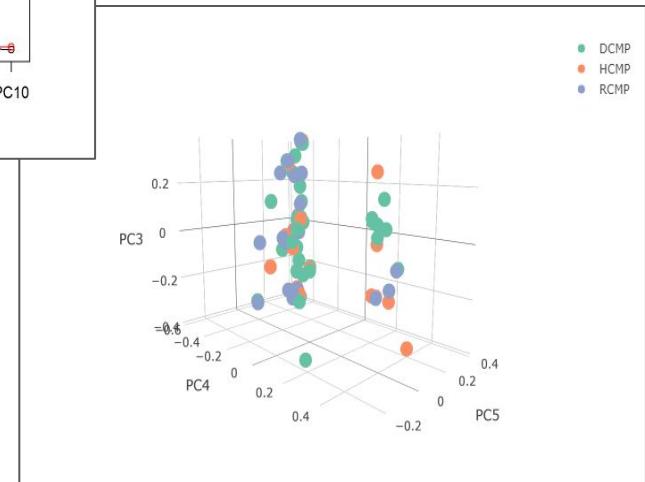
Результаты PCA по всем SNP - отсутствие кластеризации по типу кардиомиопатии

2 кластера в проекции
на плоскость
PCA4,PCA5

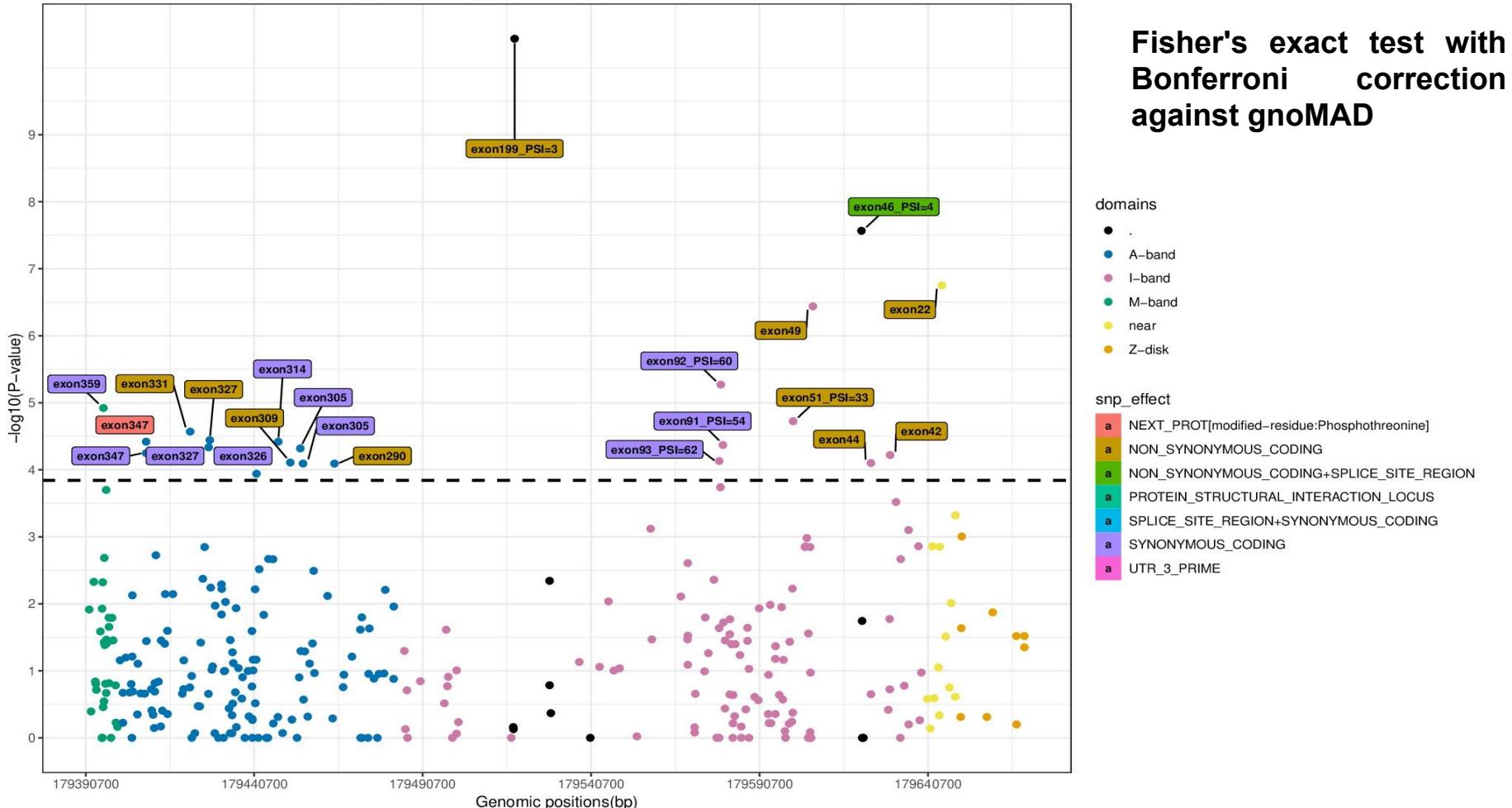


компоненты PCA(1-7)
практически
полностью
объясняют
изменчивость в
данных

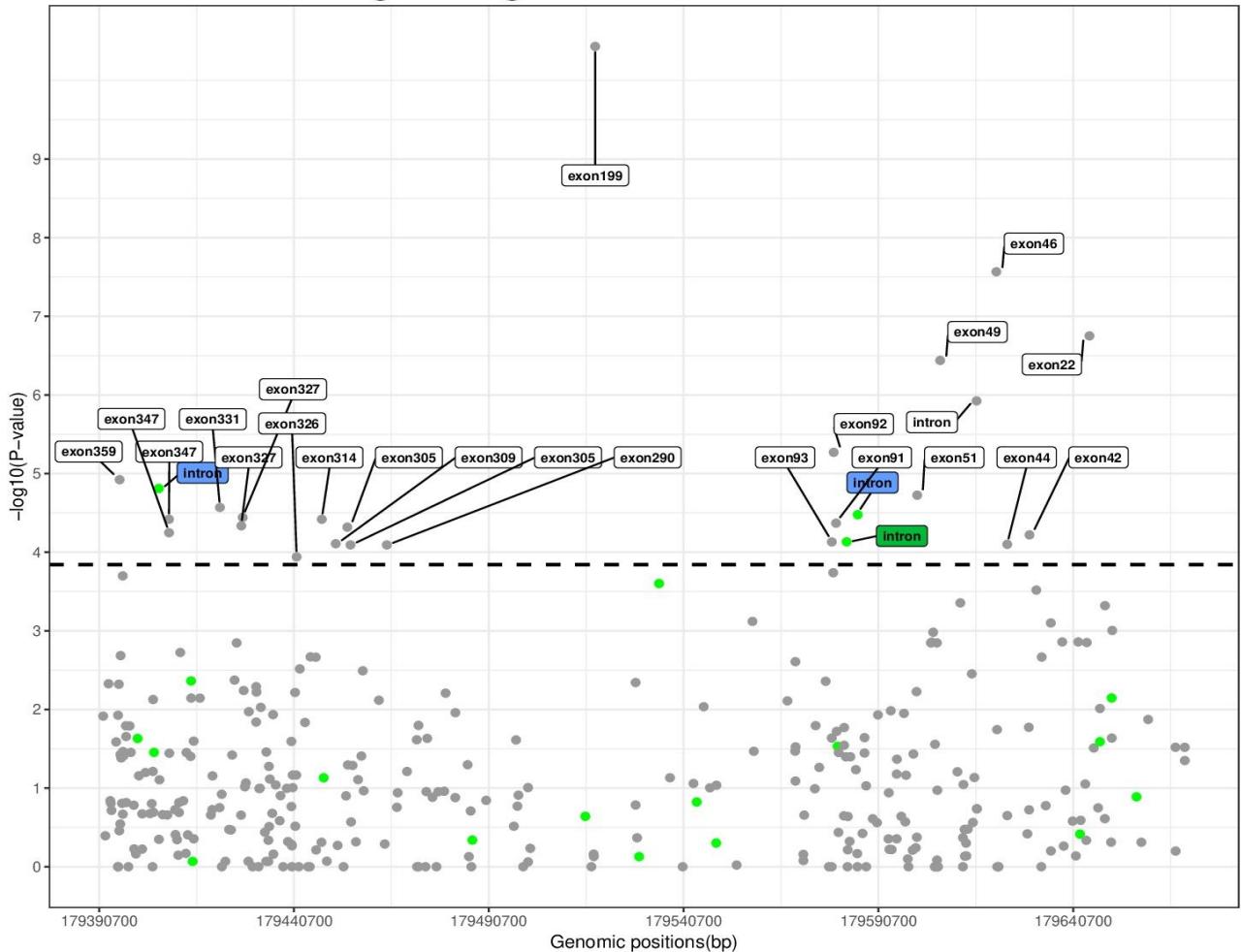
3D визуализация



Manhattan plot SNP in coding region TTN from cardimyopathy patient against gnomad data



Manhattan plot SNP in noncoding region TTN from cardimyopathy patient against gnomad data



Fisher's exact test with Bonferroni correction against gnoMAD

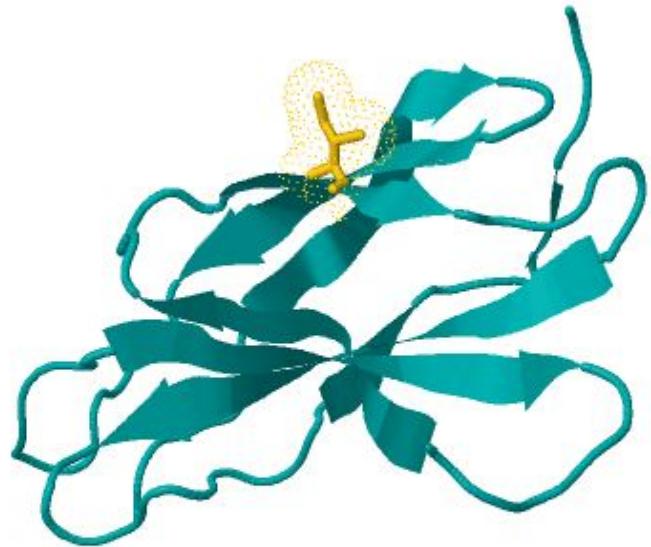
snp_effect
a
a NON_SYNONYMOUS_CODING+SPlice_SITE_REGION
- SPlice_SITE_REGION_INTRON

- coding region
- intron

Prediction snp effect on protein structure TTN

rs	gnomAD frequency	nucl change	protein pos	domen	amino change	Provean	Mutationassessor	SIFT
rs9808377	0.3507	A 179421694 G	29396	Fn3-102	I62T	Neutral	Neutral	NA
rs3829746	0.3522	T 179427536 C	27775	Fn3-90	I10V	Neutral	Neutral	NA
rs2042996	0.3511	G 179451420 A	21403	Fn3-43	T38I	Neutral	Neutral	NA
rs1001238	0.3553	T 179464527 C	18701	Ig-116	N14D	Neutral	Neutral	NA
rs2742347	0.0995	T 179600648 C	4842	Ig-31	R48K	Neutral	Neutral	NA
rs2291310	0.8854	C 179623758 T	3419	Ig-24	S75N	Neutral	Neutral	Neutral
rs2291311	0.8398	C 179629461 T	3261	Ig-23	V23M	Neutral	Neutral	Neutral

Fn3-102 rs9808377 I62T



FN3-90 rs3829746 I10V

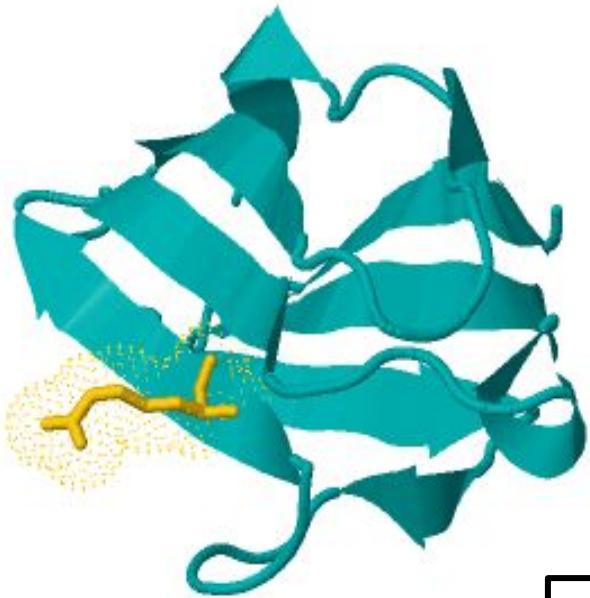


structure by TTN database
<http://fraternalilab.kcl.ac.uk/TITINdb/>
stability prediction by I/I-Mutant3.0

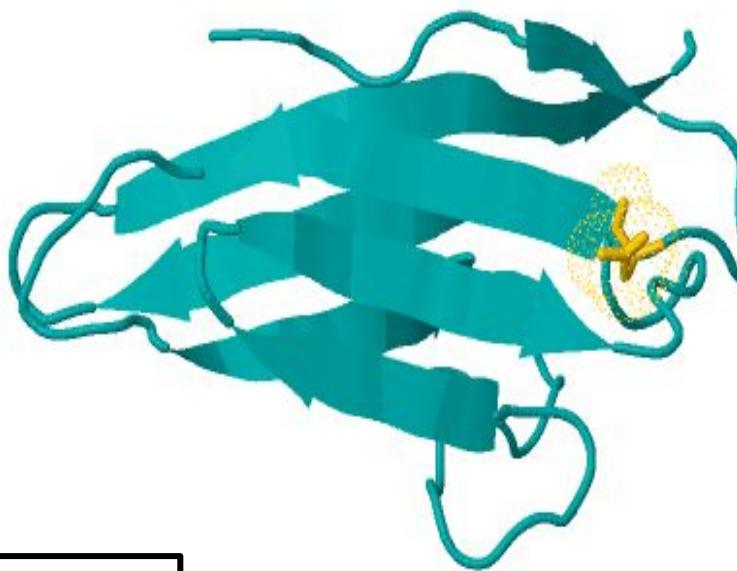
Stability Predictors ▲ G -1.29 Kcal/mol
RI = 8

Stability Predictors ▲ G -0.18 Kcal/mol
RI = 1

Ig-31 rs2742347 R48K



Ig-116 rs1001238 N14D

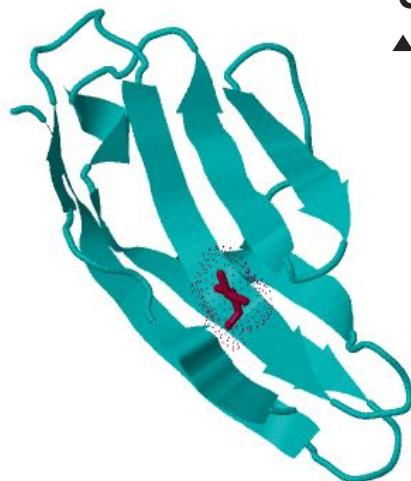


structure by TTN database
<http://fraternalilab.kcl.ac.uk/TITINdb/>
stability prediction by I/I-Mutant3.0

**Stability Predictors ▲ G -0.96 Kcal/mol
RI = 4**

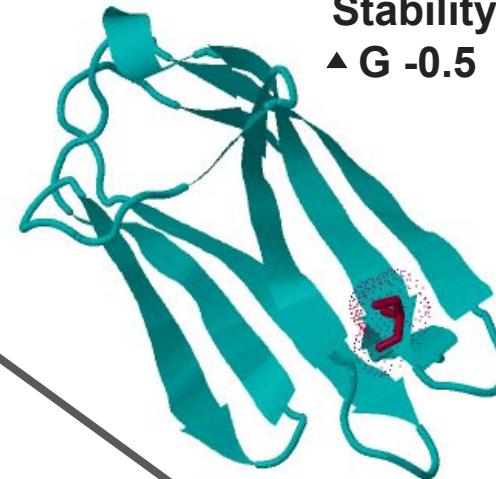
**Stability Predictors ▲ G -0.52 Kcal/mol
RI = 4**

Ig-23 rs2291311 V23M

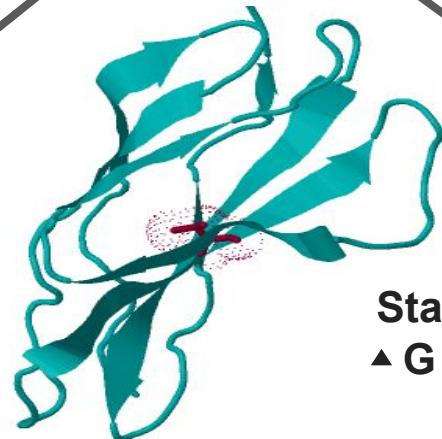


Stability Predictors
▲ G -0.24 Kcal/mol
RI = 4

Ig-24 rs2291310 S75N



Stability Predictors
▲ G -0.5 Kcal/mol
RI = 2



Stability Predictors
▲ G -0.83 Kcal/mol
RI = 5

structure by TTN database
<http://fraternalilab.kcl.ac.uk/TITINdb/>
stability prediction by I/I-Mutant3.0

Fn-43 rs2042996 T38I

Результаты

- 1) Реализован GATK Variant Calling pipeline для Haploplex cardio panel (titin) на Snakemake
- 2) С помощью snpEff проаннотированы полученные варианты - 418 SNP
- 3) Среди них 64,44% missens и 35,56% silence, genome effective length 243,199,373 chrom 2
- 4) PCA анализ по всем SNP показал отсутствие кластеризации по типу кардиомиопатии
- 5) В результате сравнения частот вариантов с частотами gnomAD при помощи точного критерия Фишера с поправкой Бонферрони статистически значимыми оказалось 26 SNP
- 6) Среди них 3 мутации в некодирующей области, в кодирующей области 12 missens и 11 silence
- 7) Наиболее интересной представляется мутация rs9808377 I62T, предположительно влияющая на стабильность субъединицы Fn3-102 ▲ G -1.29¹⁴

И ТАК МЫ ЗАГОНИМ МАМОНТА В ЯМУ.
ПЕРЕЙДЁМ К СЛЕДУЮЩЕМУ СЛАЙДУ



Спасибо за внимание



ЗА 12 000 ЛЕТ ДО Power Point'a