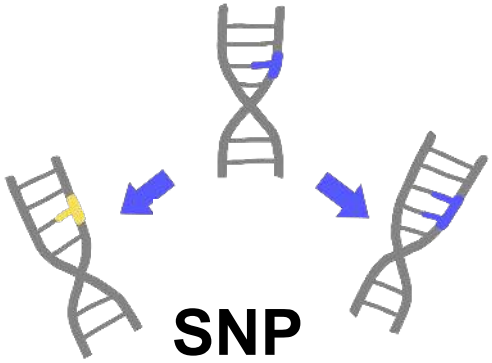
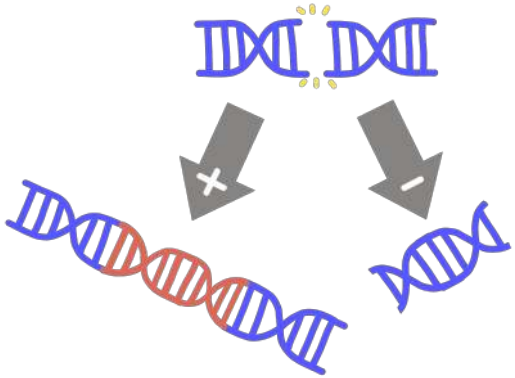
Поиск патогенных вариантов (SNPs, indels) в экзомах пациентов с различными вариантами идиопатических кардиомиопатий

Килина Д.А.
Руководитель: к.б.н., Киселёв А.М.
ФГБУ «НМИЦ им. В. А. Алмазова» Минздрава России



SNP

INDEL



Поиск ранее неизвестных мутаций

Полный экзом:
20 000 генов

Клинический
экзом: ~4000 генов

1 ген

5 – 500 генов

Анализ отдельных мутаций / генов

Генные панели

Экзом

Цель проекта

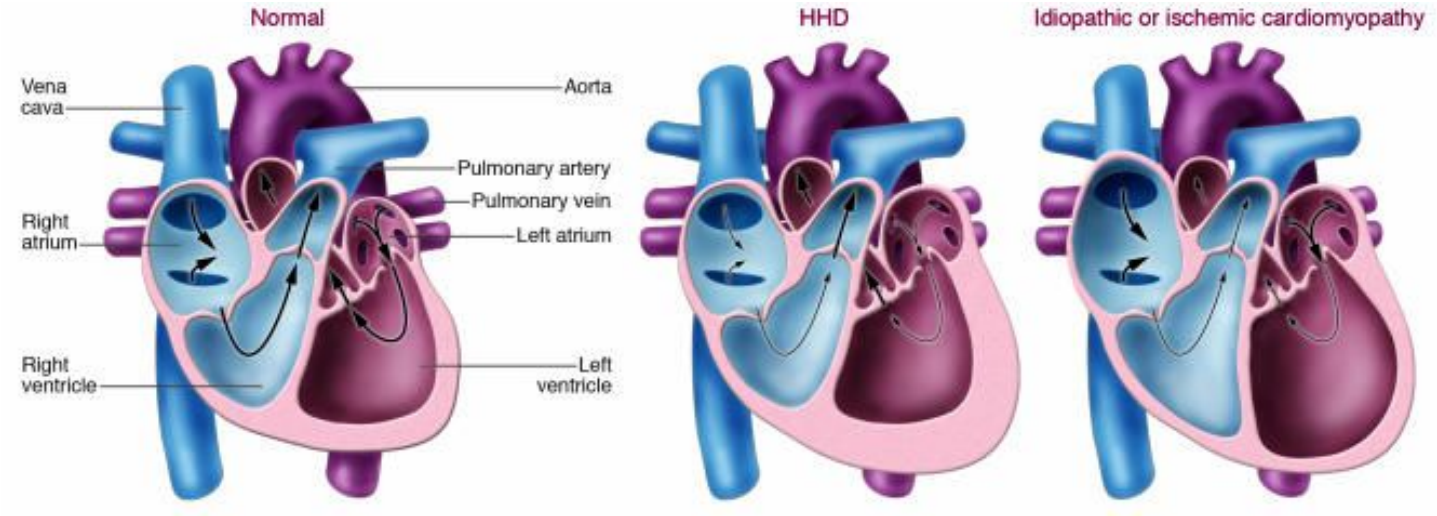
Выявление ранее неизвестных вариантов SNPs и indels у пациентов с различными вариантами идиопатических кардиомиопатий.

Задачи

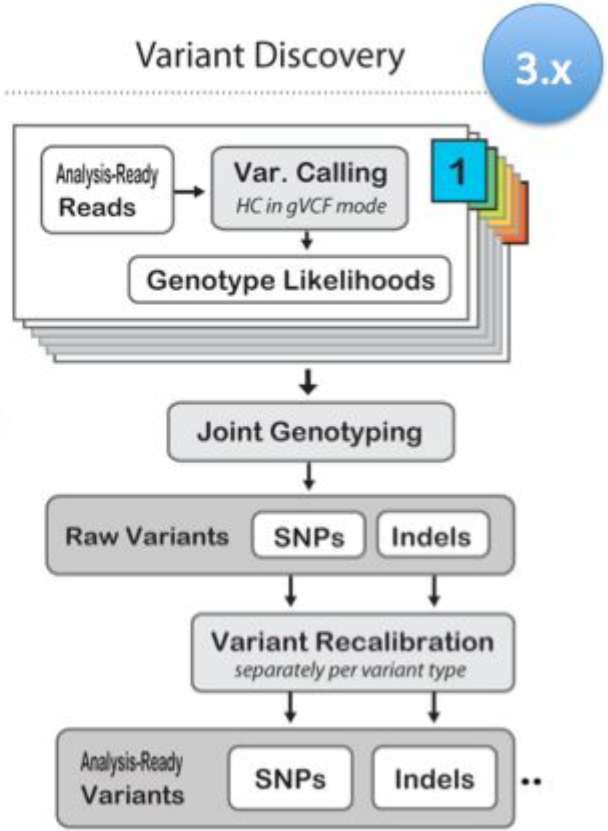
- Изучить принцип работы Snakemake
- Написать пайплайн для Variant Calling и Variant Filtering
- Провести анализ данных с помощью HaplotypeCaller
- Аннотация с помощью Annovar, SnpEff

Исходные данные

Данные полноэкзомного секвенирования 10 пациентов с диагнозом идиопатическая кардиомиопатия (bam файлы)



SNPs and indels Calling pipeline



HaplotypeCaller in GVCF mode

Combine GVCF files

Genotype cogort GVCF file

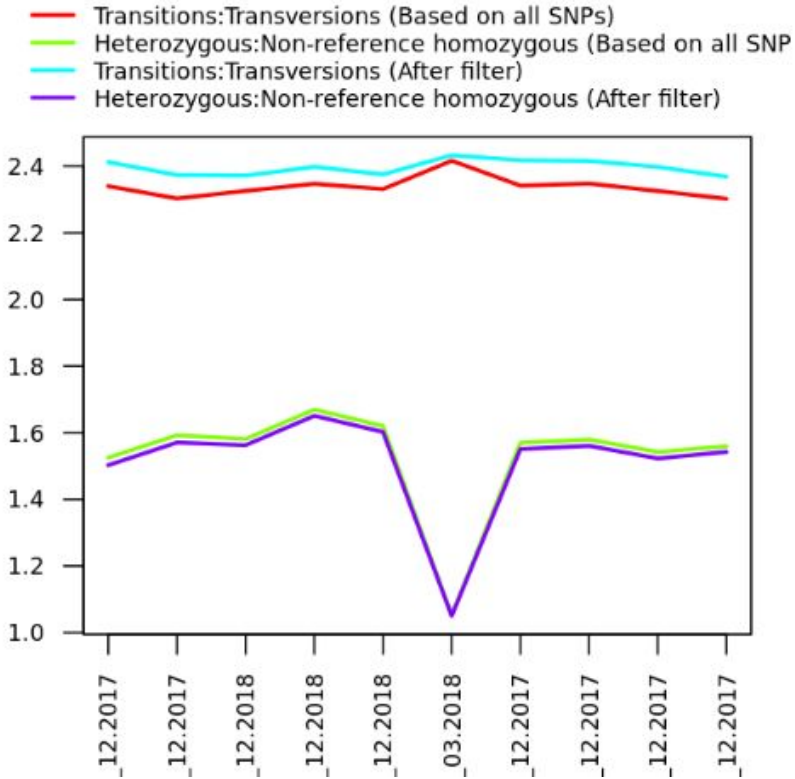
Select variants to SNP and INDELS

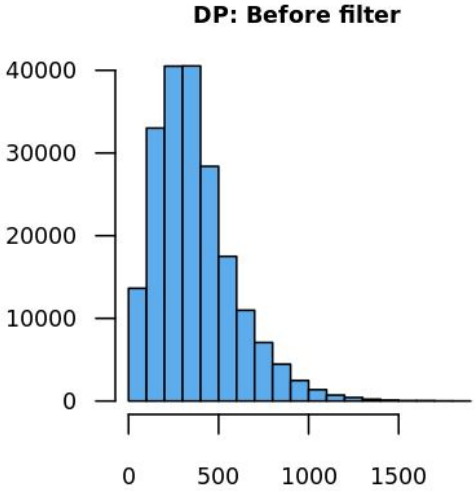
Filtering variants to SNP and INDELS

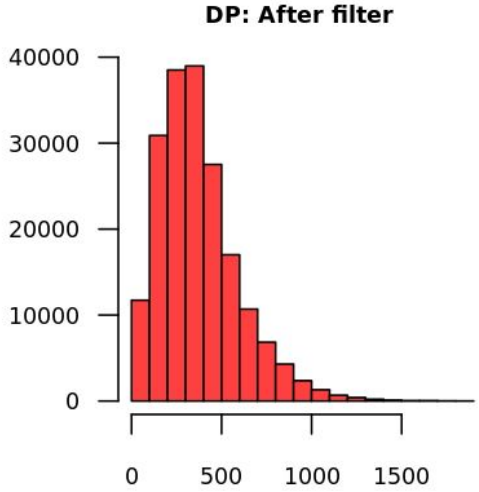
Annotation by Annovar and SnpEff

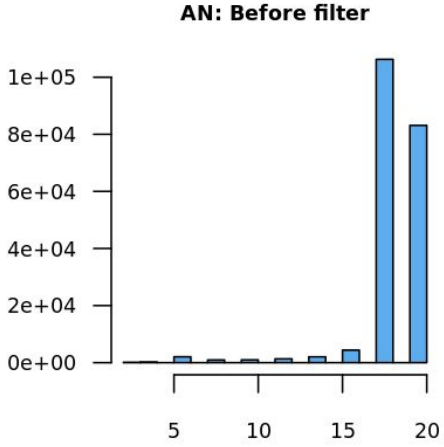
Select clinically relevant variant

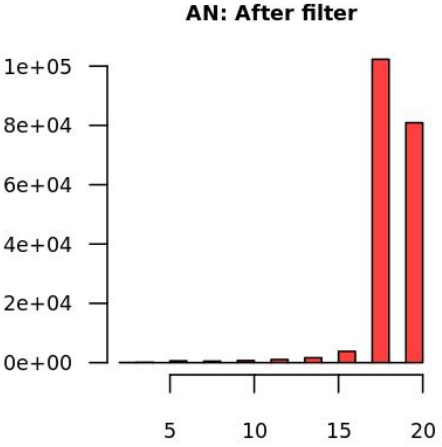
Result from QC3











Result from snpEff

Transitions	823,683
Transversions	358,768
Ts/Tv ratio	2.2959

Base changes (SNPs)

	A	C	G	T
A	0	8,152	29,753	5,230
C	8,997	0	10,246	39,620
G	40,005	10,282	0	8,665
T	5,336	29,442	8,099	0

Number of effects by impact

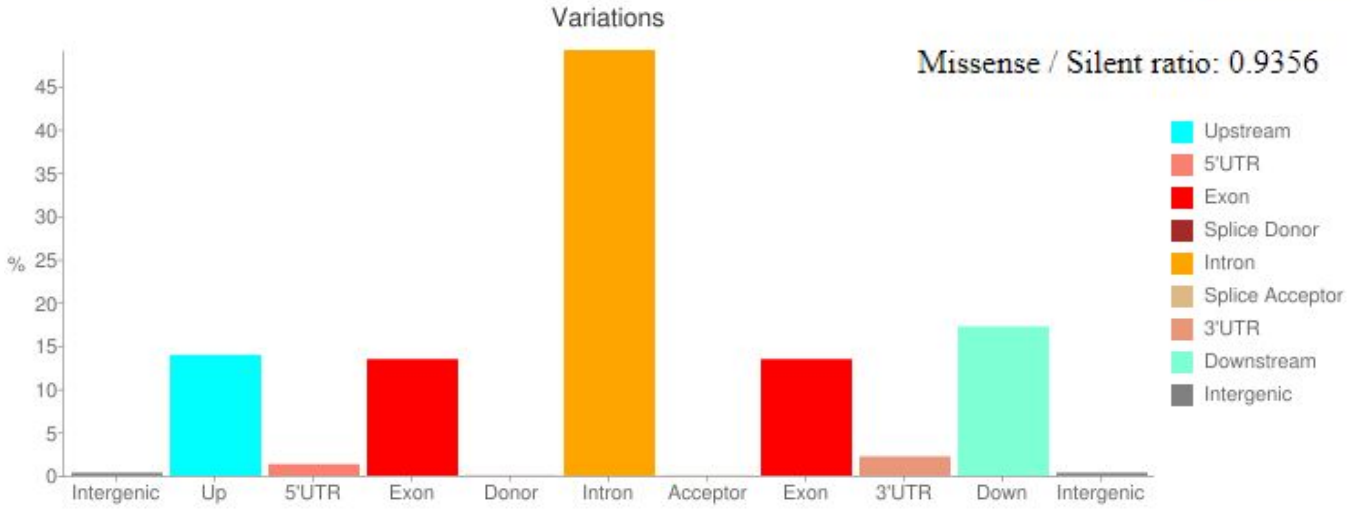
Type (alphabetical order)	Count	Percent
HIGH	15,434	0.879%
LOW	114,109	6.499%
MODERATE	76,237	4.342%
MODIFIER	1,550,097	88.281%

Number of effects by functional class

Type (alphabetical order)	Count	Percent
MISSENSE	72,159	48.071%
NONSENSE	823	0.548%
SILENT	77,128	51.381%

Number variants by type

Type	Total
SNP	203,827
MNP	0
INS	14,352
DEL	19,891
MIXED	0
INV	0
DUP	0
BND	0
INTERVAL	0
Total	238,070



Selecting SNPs from annovar results

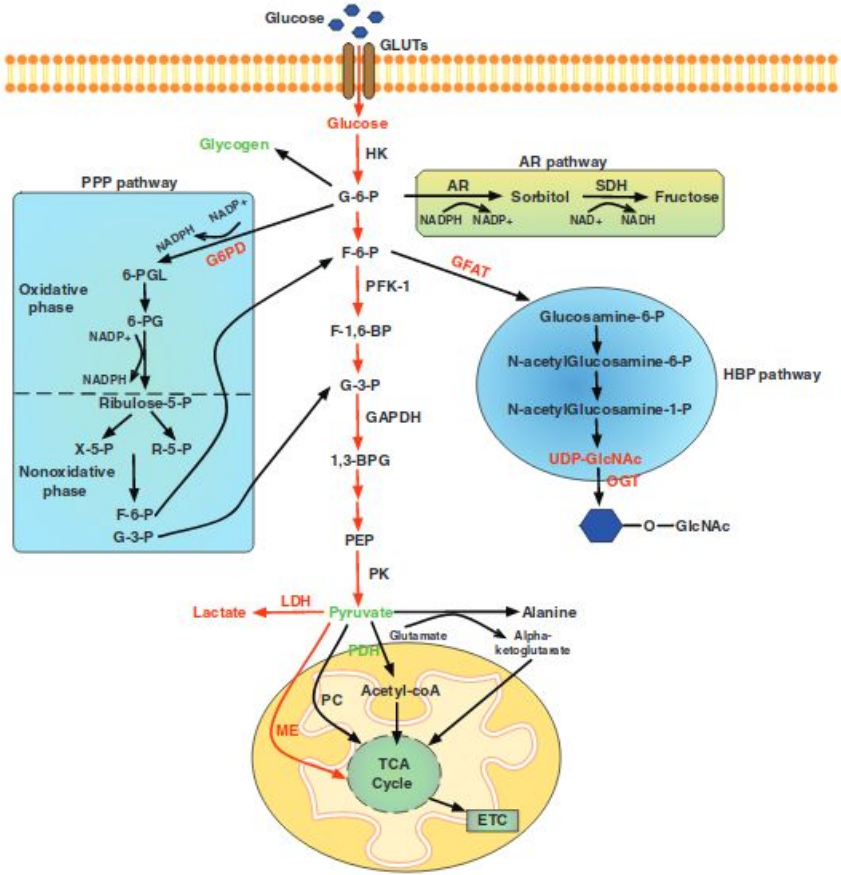
CHR	POS	REF	ALT	AF	ExAC	Func	Gene	Exonic Func	SIFT	Poly-phen2	LRT	PROVEAN	FATHMM	MetaSVM	CADD
2	74760062	C	T	0.05	8.237e-06	exonic	HTRA2	nonsynonymous	D	D	D	D	D	D	27.6
2	157370000	A	G	0.1	8.259e-06	exonic	GPD2	nonsynonymous	D	D	D	D	D	D	24.4
7	56087374	C	T	0.15	5.900e-02	exonic	PSPH	nonsynonymous	D	D	D	D	D	D	28.5
8	29046043	T	C	0.111	5.000e-04	exonic	KIF13B	nonsynonymous	D	D	D	D	D	D	27.3
9	**130162198**	**G**	**A**	**0.111**	**4.970e-05**	**exonic**	**SLC2A8**	**nonsynonymous**	**D**	**D**	**D**	**D**	**D**	**D**	**29.8**
12	9638097	G	A	0.05	8.254e-06	exonic	HAL	nonsynonymous	D	D	D	D	D	D	33.0

SLC2A8 (Solute Carrier Family 2 Member 8)

This gene belongs to the solute carrier 2A family, which includes intracellular glucose transporters and expression of glucose transporters.

The loss of metabolic flexibility associated with increased reliance on glucose utilization contribute to the development of cardiac dysfunction.

The changes in glucose metabolism in hypertrophied hearts include altered glucose transport and increased glycolysis*



*Shao, D., & Tian, R. (2015). Glucose Transporters in Cardiac Metabolism and Hypertrophy. Comprehensive Physiology, 331–351. doi:10.1002/cphy.c150016

Summary result from PROVEAN, dbSNP indels annotation

		Found in dbSNP?		PROVEAN Prediction		
	Total	Yes	No (novel)	Neutral	Deleterious	Not available
1. Protein-coding	**123**	**91**	**32**	**81**	**13**	**29**
(1) Single AA Change	0	0	0	0	0	0
(2) Synonymous	0	0	0	0	0	0
(3) Deletion	56	48	8	46	10	0
(4) Insertion	37	24	13	34	3	0
(5) Multiple AA Change	1	1	0	1	0	0
(6) Frameshift	28	17	11	0	0	28
(7) Nonsense	0	0	0	0	0	0
(8) Unknown	0	0	0	0	0	0
(9) Input error	1	1	0	0	0	1
2. Non protein-coding	**5**	**4**	**1**	**0**	**0**	**5**
3. Input format error	**0**	**0**	**0**	**0**	**0**	**0**

Результаты

1) Реализован GATK Variant Calling pipeline на Snakemake
https://bitbucket.org/DariaKilina/project-ib/src/master/

2) С помощью annovar и snpEff проаннотированы полученные варианты SNP и indels, среди них 48.07% missens и 51,38% silence, общее количество SNP - 203,827, INS - 14,352, DEL -19,891

3) В результате сравнения частот вариантов с частотами ExAC, учета аннотаций Sift, Polyphen2, LRT, PROVEAN, FATHMM, MetaSVM, CADD выявлен 6 ранее неизвестный SNPs в различных генах, из них наиболее интересной представляется мутация в гене SLC25A5

4) Новых indels вариантов выявлено не было

Спасибо за внимание!