

[Файл](#) пациента для семинара.

<https://drive.google.com/open?id=1Ldqv2A9wy8x3D2UOXBv7lwecZI75Zh7->

Познакомимся с новым форматом хранения генетических вариантов VCF
(https://en.wikipedia.org/wiki/Variant_Call_Format)

Пример строки:

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	ng1657
chr1	752721	rs3131972	A	G	.	PASS	.	GT	0/1

Chr1 -- хромосома SNP

752721 -- позиция на хромосоме

Rs3131972 -- идентификатор SNP в базе dbSNP

A -- референсный нуклеотид в данной позиции

G -- альтернативный нуклеотид

Ng1657 -- идентификатор исследуемого образца

0/1 -- генотип образца

0 соответствует референсному аллелю, 1 соответствует альтернативному аллелю.

В данном случае у человека генотип AG.

1/1 соответствует генотипу GG

0/0 соответствует генотипу AA

Является ли данный человек носителем болезни Тея-Сакса?

- Используем базу OMIM (<https://www.omim.org>), чтобы определить ген, ассоциированный с болезнью Тея-Сакса
- Используем NCBI (<https://www.ncbi.nlm.nih.gov/gene/>), чтобы найти координаты данного гена в геноме (используется сборка GRCh37.p13)
- Фильтруем из VCF человека позиции, которые попадают в данный ген
- Ищем все позиции, в которых у человека присутствует хотя бы 1 альтернативный аллель (0/1 или 1/1)
- Проверяем все найденные позиции на патогенность с помощью базы varsome.com или ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>)

Какой наиболее вероятный цвет глаз у этого человека?

- Будем предсказывать цвет глаз с помощью системы IrisPlex (<https://hirisplex.erasmusmc.nl>)
- Определим список SNP, необходимых для предсказания цвета глаз (6 штук)
- Извлечем необходимые SNP из VCF клиента и определим генотипы
- Введем генотипы в веб интерфейс и найдем вероятности для каждого цвета глаз.

Повышен или понижен риск тромбоза у данного человека?

- Определять риск будем на основании статьи <https://ashpublications.org/blood/article-lookup/doi/10.1182/blood-2011-12-397752>
- Предлагается использовать модель из 5 SNP:
These SNPs were rs6025 (F5, factor V Leiden), rs1799963 (F2, 20210 G > A), rs8176719 (ABO), rs2066865 (FGG, 10034 C > T), and rs2036914 (F11).
- Необходимо определить, какие аллели являются рисковыми для каждого из SNP. Это удобно сделать с помощью ресурса SNPedia
- Для клиента необходимо вычислить количество рисковых аллелей
- На основании графика 3 из статьи делаем вывод о риске тромбоза.