План семинара:

- Знакомство с базой dbSNP и основными понятиями
    - SNP
    - аллель
    - MAF
    - Reference
- Разбор примеров влияния SNP на структуру белка из публикации
    - практика в pymol

Базы данных SNP и структурных вариантов человека. Влияние SNP на структуру белка.

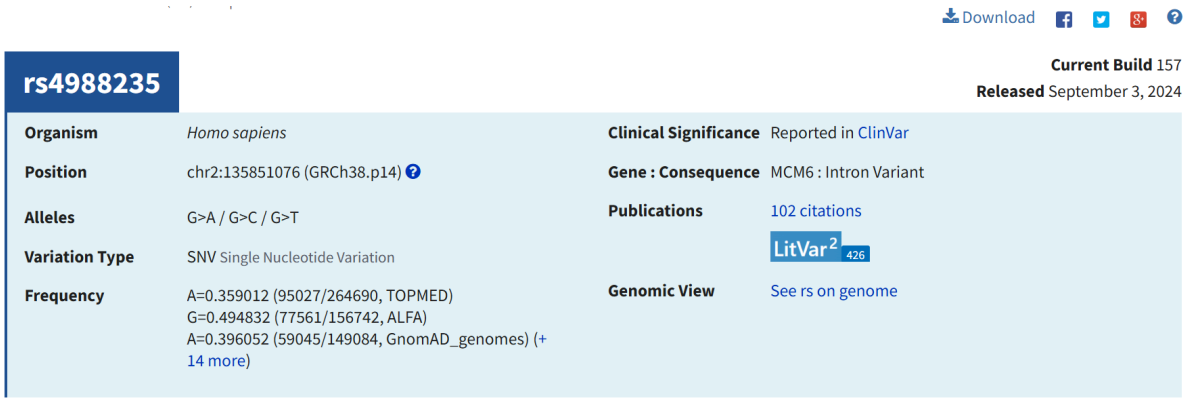
Databases of SNPs and human structural variants. Effect of SNP on protein structure.

Базы SNP:

- **dbSNP** https://www.ncbi.nlm.nih.gov/snp/

База данных однонуклеотидных полиморфизмов (dbSNP) — это публичная база данных, содержащая каталог генетических вариаций, включая однонуклеотидные полиморфизмы (SNP), короткие вставки и делеции (инделы) и другие незначительные геномные вариации. Созданная в 1998 году Национальным центром биотехнологической информации (NCBI), база данных dbSNP сыграла решающую роль в геномных исследованиях, исследованиях ассоциации заболеваний, фармакогеномике и популяционной генетике.
Ознакомьтесь с FAQ (после занятия**)** https://www.ncbi.nlm.nih.gov/snp/docs/faq/

Карточка записи rs4988235:



- в каких популяция референсный аллель встречается наиболее часто?
    - а наиболее редко?
- Попробуйте найти его в геномном браузере UCSC
    - На какой хромосоме ищем?
    - На какой позиции?
    - Где он находится в интроне, экзоне или межгенном пространстве?

MAF - частота второго аллеля по частоте. Другими словами, если есть три аллеля с частотами 0,50, 0,49 и 0,01, то MAF будет указан как 0,49.
MNV https://www.ncbi.nlm.nih.gov/snp/rs1800995

SNPedia https://www.snpedia.com/ почему-то не открыается

Другие данные:

Таблица 1. Шесть популярных наборов данных для инструментов, прогнозирующих эффект точечных мутаций..

Dataset	Compiled from	Size	Reference
MutPred	SwissProt and HGMD	65,657	[37]
SNPs&GO	SwissProt	38,460	[38]
PON-P	dbSNP, PhenCode, IDbases and 16 individual locus-specific databases	39,670	[39]
HumVar	SwissProt and dbSNP	41,918	[40]
Humsavar	SwissProt/UniProt	36,994	[41]
PredictSNP	SwissProt/UniProt	43,883	[42]

https://doi.org/10.1371/journal.pone.0171355.t001

Связь с заболеваниями/фенотипом

- https://www.ncbi.nlm.nih.gov/clinvar/
- https://www.ebi.ac.uk/gwas/home
- https://gnomad.broadinstitute.org/

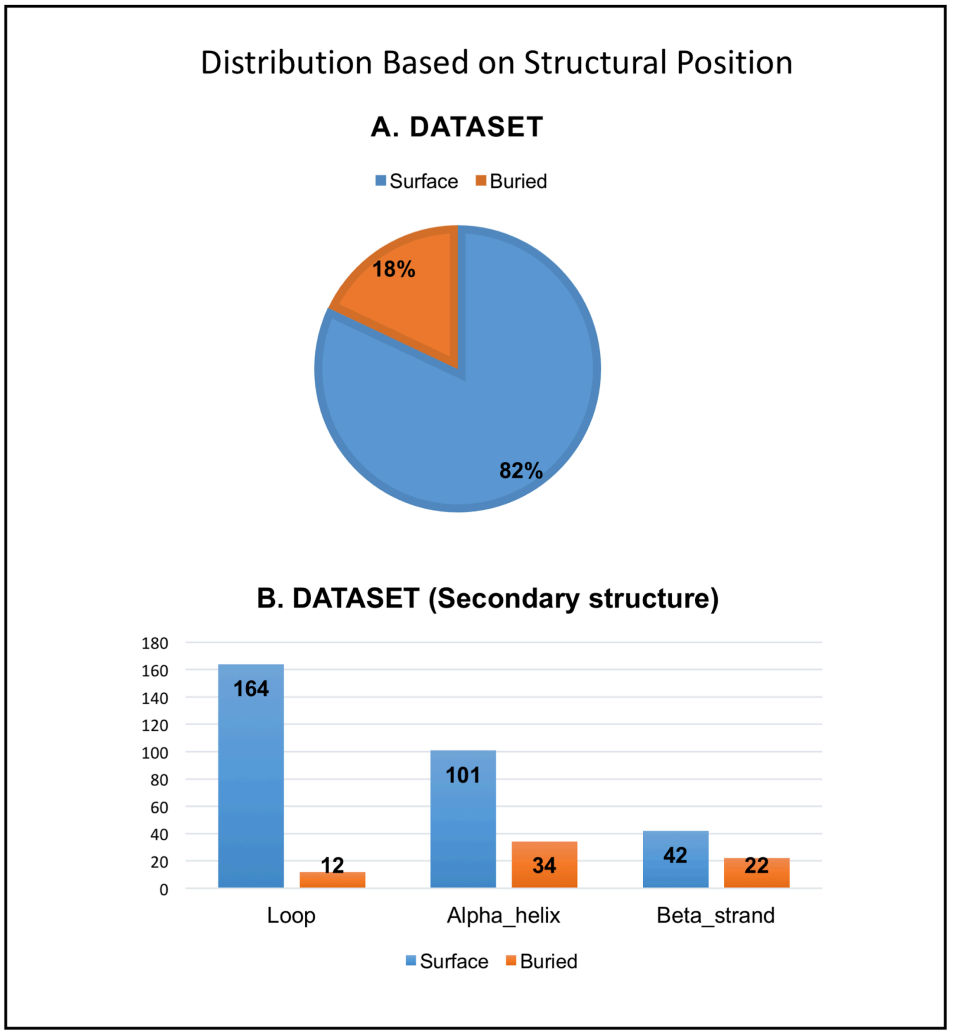
Как SNP влияют на структуру и функцию белка?

Impact of genetic variation on three dimensional structure and function of proteins

Roshni Bhattacharya,Peter W. Rose,Stephen K. Burley,Andreas Prlić
RCSB Protein Data Bank, San Diego Supercomputer Center, University of California San Diego, La Jolla, California, United States of America

Location of SNPs within 3D structures

In the *Surface* category, it was observed that **52%** (155 out of 297) of the SNVs map to *Loop* regions compared to ~34% for *Alpha_helix* and ~14% for *Beta_strand*.



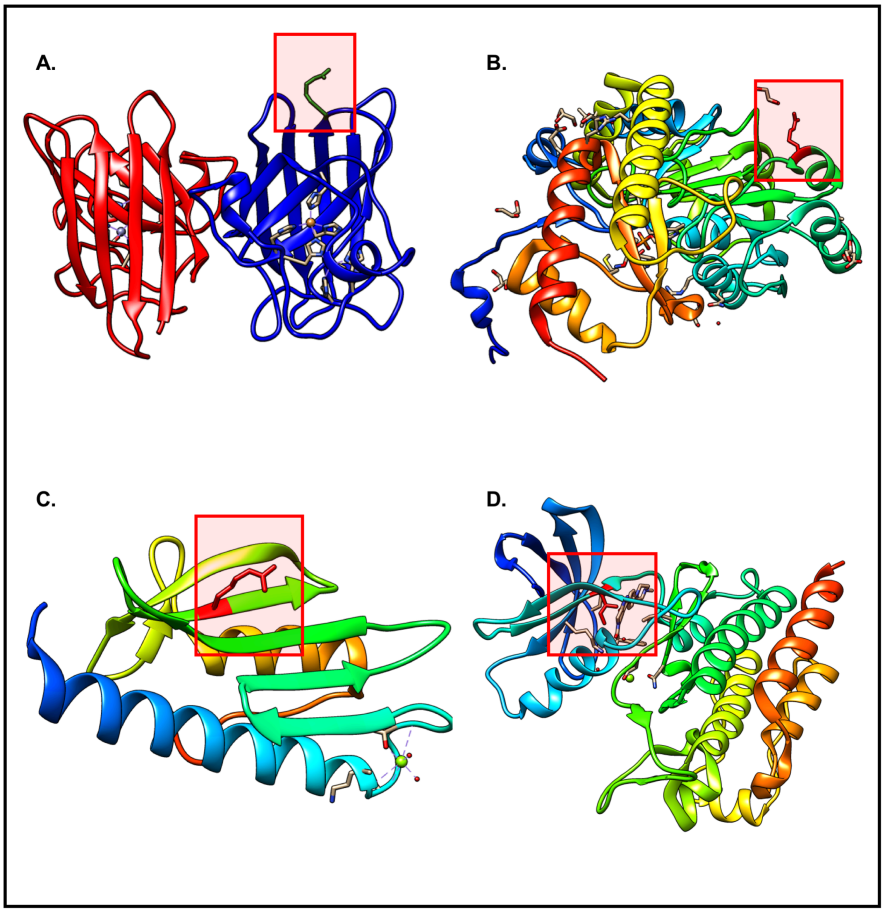


Fig 2. SNV consequences map to various locations within protein structures.

A) PDB: **1AZV**, SNV: rs121912431 (G37R) is present on the surface of the protein in the highlighted Loop segment, where it causes the neurological disease Lou Gehrig's disease. B) PDB: **1J04**, SNV: rs121908529 (G170R) is present on the surface of the protein in the highlighted Alpha_helix, where it causes hereditary kidney stone disease primary hyperoxaluria type 1. C) PDB: **3S5E**, SNV: rs138471431 (W155R) is present on the surface of the protein in the highlighted Beta_sheet, where it causes the neurodegenerative disease Friedreich's ataxia. D) PDB: **2V7A**, SNV rs121913459 (T315I) is present in the ATP-binding domain and causes resistance to the drug imatinib in patients with chronic myelogenous leukemia.

Table 2. Consequence of SNVs on protein structure and function for a dataset of 374 SNVs for which experimentally obtained atomic level data for the variation is available in the Protein Data Bank.

Each SNV can be scored for multiple categories.

Activity	52
Aggregation	28
Stability	58
Binding	44
Assembly	19
Rearrangement	25

https://doi.org/10.1371/journal.pone.0171355.t002

1. **Activity**—The SNV causes increase, decrease, or complete loss of protein activity.
2. **Aggregation**—The SNV renders the protein aggregation prone.
3. **Stability**—The SNV causes a change in protein stability. It may make the protein susceptible to proteolytic cleavage, or cause a change in thermal inactivation temperature, or cause a change in the energy of stabilization of the protein. It can also lead to destabilization of a protein oligomer, loss of packing or hydrophobic interactions, or change a mode(s) of protein-protein interaction.
4. **Binding/Dissociation**–The SNV leads to changes in affinity for a known binding partner, or alterations in association or dissociation kinetics. It can also cause structural changes in the binding site or affect specificity for a binding partner(s).
5. **Assembly**—The SNV affects the oligomeric assembly properties of the protein.
6. **Rearrangement**—The SNV causes local structural rearrangements (conformational changes) in the neighborhood of the amino acid change arising from the SNV.

Table 3. Examples for each SNV related effect category.

Activity	rs137852646	Glycyl-tRNA synthetase	2PMF	2ZT5	G526R	Loss of activity	Charcot-Marie-Tooth disease	[50]
Aggregation	rs121912442	Cu, Zn superoxide dismutase [HSOD	1N19	4FF9	A4V	Destabilization of protein and formation of aggregates.	Lou Gehrig's disease	[51]
Stability	rs74315351	DJ-1	2RK4	1P5F	M26I	Leads to decrease thermal stability and inactivation.	Rare forms of familial Parkinsonism	[52,54]
Binding	rs104894227	HRAS	2QUZ	2CE2	K117R	Increases the rate of nucleotide dissociation and results in constitutive activation of HRAS.	Costello Syndrome	[55]
Assembly	rs1141718	Manganese superoxide dismutase	1VAR	1MSD	I58T	The packing defects due to the mutation disrupt the dimer-tetramer equilibrium and favor the dimer over tetramer in solution.	Amylotrophic Lateral Sclerosis	[56]
Rearrangement	rs61749389	von Willebrand factor	1IJK	1OAK	I546V	The mutation causes a "Gain of Function" effect and produces a phenotype in which regulation is lost	von Willebrand disease	[57]

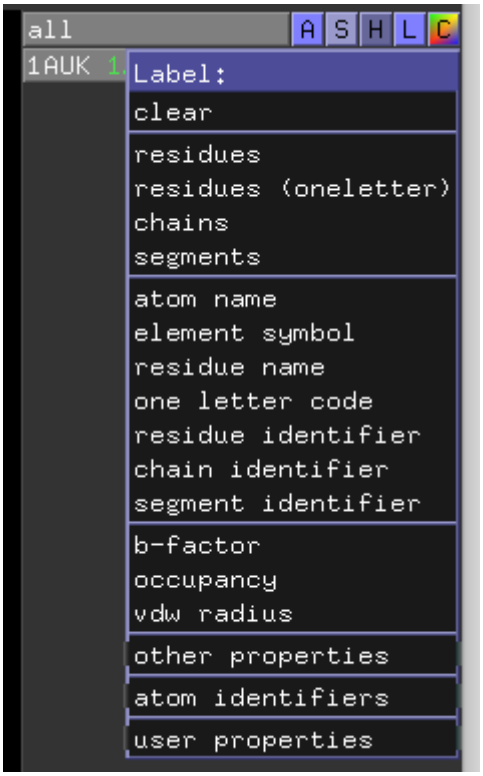
https://doi.org/10.1371/journal.pone.0171355.t003

Case Studies

Скачайте и установите Pymol если ещё не сделали это
https://pymol.org/2/
Selection algebra cheatsheet:
https://pymol.sourceforge.net/newman/user/S0220commands.html

Example 1. SNVs that affect both protein structure and function.

1. Load structures
   File->Get PDB… 1E33
   File->Get PDB… 1AUK
2. Align the structures
3. Press gray L button on the right part of the screen (L-labeling)



- Try different residue visualization modes. Choose one you find the best.
- Find Pro→Leu mutation (P428L) at 3D structure in Pymol.
- Find rs28940893 in dbSNP and UCSC genome browser.

> select resi ### in ####

Recreate the mutation via Wizard->Mutagenesis

Arylsulfatase A (gene: ARSA) breaks down sulfatides. The Pro→Leu mutation (P428L) (rs28940893) mapping to amino acid 426 in the PDB structure yields an oligomerization defect (preferred mutant assembly is dimer instead of octamer as for wild-type (Wildtype PDB: 1AUK)) that increases the susceptibility of the protein to degradation by lysosomal cysteine proteinases, leading to severe reduction in half-life and metachromatic leukodystrophy. Therefore, this SNV related change affects both Stability and the protein Assembly.

Example 2. **Activity.**

52 of 374 SNV related changes in our dataset (~14%) either increase or decrease protein activity.

human glycyl-tRNA synthetase (mutant PDB: **2PMF**) loses detectable enzymatic activity due to a G526R (rs137852646) mutation, which is causative of Charcot-Marie-Tooth disease. G526 is an evolutionarily conserved residue located in the midst of motif 3 that connects *Beta_strand* β19 with *Alpha_helix* α13. With the exception of the mutation site, the overall structure of the G526R mutant protein is almost identical to that of the wild type (Wildtype PDB: **2ZT5**) enzyme (alpha-Carbon atomic position root-mean-square deviation = 0.8Å). Although the G526R change does not disturb the positions of residues comprising the active site, the sidechain of the mutated residue (R526) interdicts access to the active site, thereby inactivating the enzyme.

1. Clean working area -
   File -> Reinitialize everything

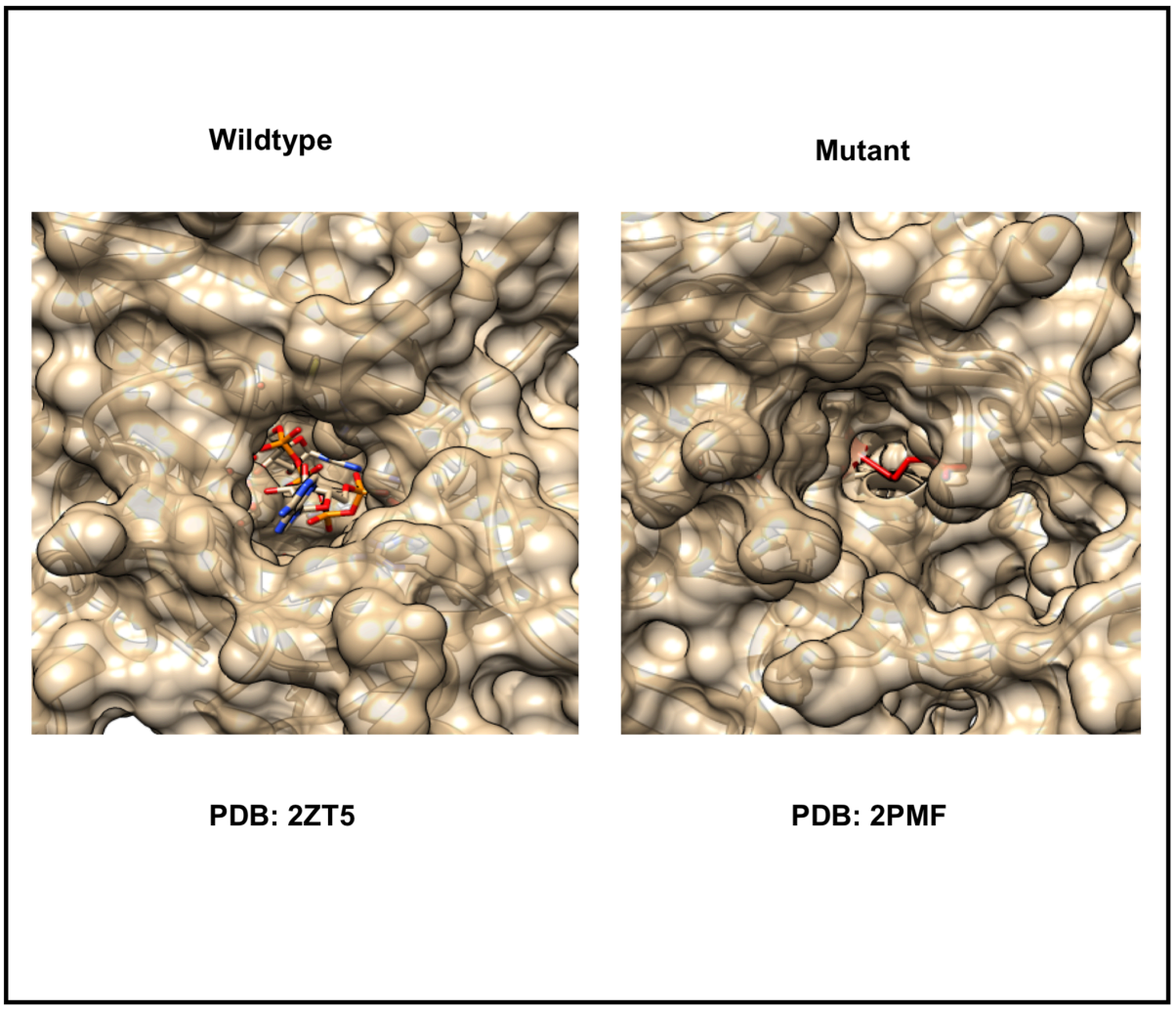
Mutant form

2. Load the structure into Pymol

File->Get PDB… 2PMF

3. Find position 526, mutation G526R

Сравнить с
File->Get PDB… 2ZT5
Воспользуйтесь
set transparency, 0.5, (selection)



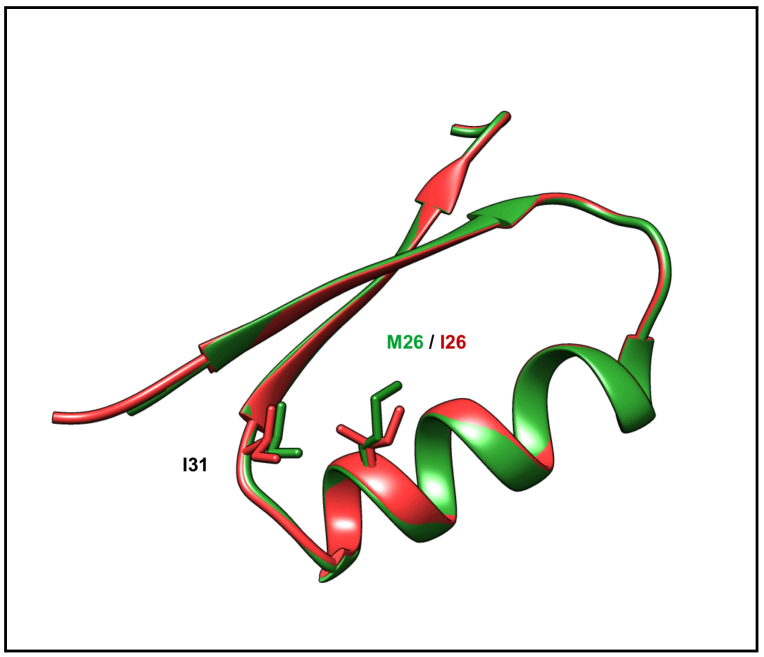
Example 3. Stability.

58 of 374 SNV related changes in our dataset (~16%) lead to reduced protein stability.

DJ-1 (mutant PDB: **2RK4**) is a small conserved protein (189 amino acids), whose absence or inactivation leads to rare forms of familial Parkinsonism in humans [52]. It is also a Ras-dependent oncogene and has been associated with several types of cancers [53]. The Met→Ile (**M26I**) mutation (rs74315351) decreases thermal stability and enhances formation of DJ-1 aggregates [54]. M26 (Wildtype PDB: **1P5F**) is a conserved residue, located in the hydrophobic core of the protein. Although M26 lies near the dimer interface, it does not directly participate in intermolecular protein-protein interactions across the dimer interface. The M26I mutation introduces a β-branched amino acid (isoleucine) into the tightly packed hydrophobic core of the DJ-1 monomer. The steric clash between I26 and the sidechain of I31 displaces the residues slightly and causes loss of optimal packing contacts in the interior of the protein resulting in lower stability.

- **Visualize the mutation** Met→Ile (M26I) in protein 1P5F
- Find SNP rs74315351 record in dbSNP database

- Try to recreate the image below using selection algebra and commands **cartoon** and **color**



Example 4. Binding.

44 of 374 SNV related changes in our dataset (~12%) affect ligand or macromolecule binding properties of the protein. A SNV can change the affinity of binding to partners, such as activators, repressors, or substrates. Such changes can also affect the kinetics of interactions with partners or alter binding specificity. Structurally, a SNV can alter the binding site of the protein, which can in turn affect interactions with partner proteins, ligands, etc. The Lys→Arg (K117R) (rs104894227) substitution in HRAS (mutant PDB: **2QUZ**) does not alter either intrinsic Ras GTPase activity or responsiveness to GTPase activating proteins, but instead causes constitutive activation of HRAS (and downstream targets) by markedly increasing the rate of GDP dissociation. This mutant HRAS protein activates the RAF/MEK/ERK signaling cascade, leading to growth factor independent cellular proliferation. Although lysine and arginine are both positively charged amino acids, even this conservative substitution results in constitutive activation of HRAS. Clinically, the K117R change in HRAS leads to constant and unchecked cell division causing Costello Syndrome, which is a rare genetic disorder affecting many parts of the body.

The Lys→Arg substitution at position 117 maps to the nucleotide-binding consensus sequence NKXD. In wild-type HRAS (Wildtype PDB: 2CE2), K117 stabilizes nucleotide binding when its aliphatic portion interacting with the base, while its terminal amino group interacts with ribose oxygen O4 of N85 and with a main chain segment (Gly13, CO) from the phosphate binding loop (P-loop). Destabilization of nucleotide binding is a consequence of subtle rearrangements due to introduction of a larger sidechain capable of making additional polar interactions.

- **Visualize the mutation** Lys→Arg (K117R) в белке 2CE2

- Find SNP rs104894227 record in dbSNP database

Если останется время, то можно изучить другие SNVs из статьи.

Дополнительно:

https://www.science.org/doi/10.1126/science.adg7492