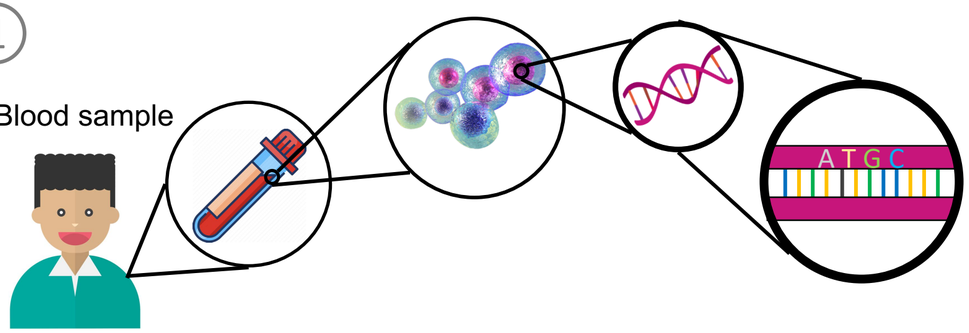
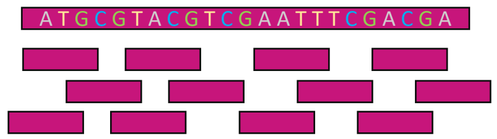
1

Blood sample

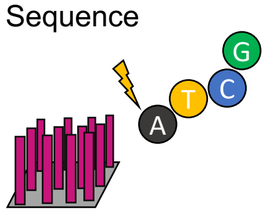


2

Prepare DNA



Sequence

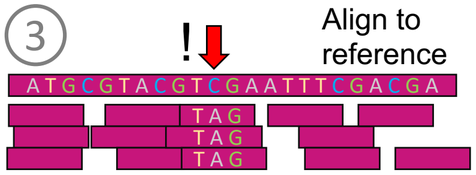


FASTQ data

```
>sample1
ATG CGC TAC
TGC GCT ACG
GCT ACG TAC
ATG CGC TAC
```

3

Align to reference



VCF data

Position	Ref	Alt
Chr 1: 1	A	-
Chr 1: 2	T	-
Chr 1: 3	G	-
Chr 1: 4	C	-
…		
Chr 11:	C	A

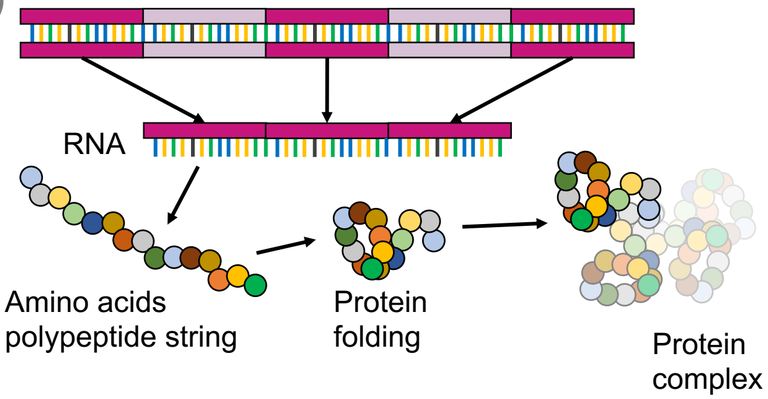
4

VCF data annotated

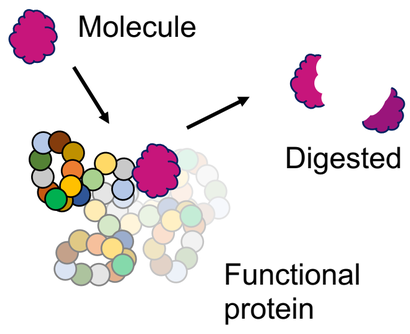
Position	Ref	Alt	Gene	Consequence
Chr 1: 1	A	-	RAG1	None
Chr 1: 2	T	-	RAG1	None
Chr 1: 3	G	-	RAG1	None
Chr 1: 4	C	-	RAG1	None
…				
Chr 1:11	C	A	RAG1	missense
…				
Chr 1: 21	T	G	RAG1	stop

5

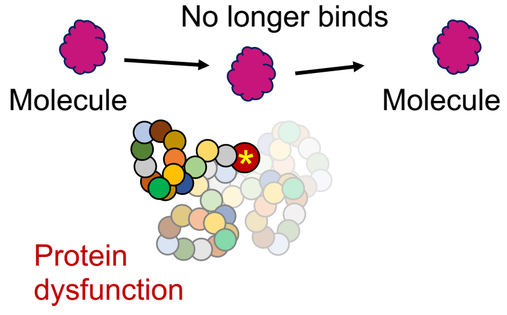
DNA in nucleus



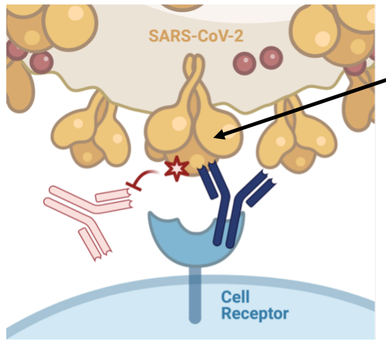
6



Example 1



7



Variant in gene
for virus spike
protein

New viral protein shape
Not bound by immune
receptor protein

Example 2

1. Get DNA.
2. Use sequencing technology to get the DNA sequences. Save as FASTQ format.
3. Align to reference genome to see what is “different”. Process until we have reliable interpretation of the sequence and variants. Save as VCF format.
4. We know the nucleotide change, how do we interpret it? Annotate with names of gene/protein that variants affect, and the consequence of change on that protein.
5. What effect does a damaging variant have? DNA is copied to RNA, then to protein. The damaged protein may not function correctly and have a biological effect because of misfolding.
6. Most biological mechanisms require protein interactions, so protein structure is critical. DNA variants affect everything.
7. SARS-CoV-2 spike protein is an example where variants change the protein structure.

Chr = Chromosome
Ref = Reference
Alt = Alternate
VCF = Variant call format
Missense = Amino acid change
Stop = Protein translation stops