Day 2:

Gene Annotation

Protein Name: epidermal growth factor receptor isoform a precursor [Homo sapiens]

Protein ID - 29725609

The following Gene function details

Location - 55019017bp-55211628bp

Family - PK_Tyr_Ser-Thr

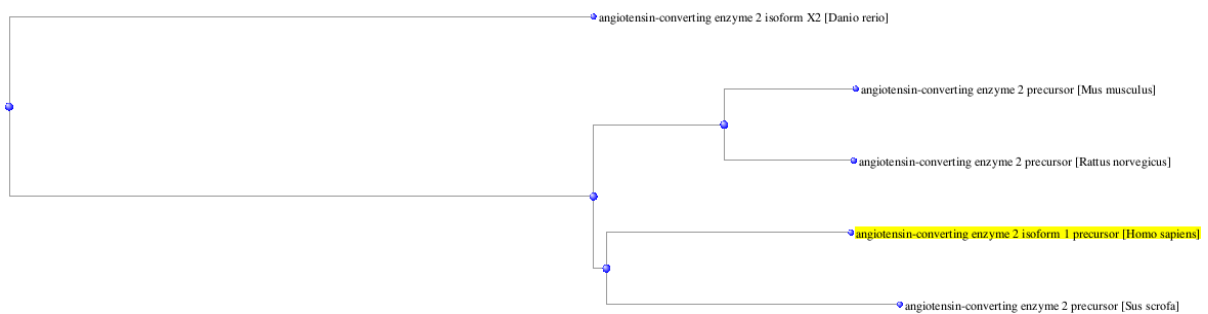Clan - CL0016

Domain - tyrosine kinase

Motif- 417 - 666

E value - 4.3e-95   1.1e-98

Description of function - Protein kinases catalyse the transfer of the gamma phosphate from nucleotide triphosphates (often ATP) to one or more amino acid residues in a protein substrate side chain, resulting in a conformational change affecting protein function

Day 3: Phylogenetics

Construct a Phylogenetic tree for components of Corona virus.

Phylogenetic tree of the protein associated with Corona virus for the species: Homo sapiens, Mus musculus, Rattus norvegicus, Danio rerio and Sus scrofa.

Day 4:

Genome name - Severe acute respiratory syndrome-related coronavirus

From RAST results:

1) Nucleotide sequence
   atggggatagcactactaaaattaattttacacattagggctcttccatataggcagctctccctagcattattcactgtaccctcgatcgtactccgcgtggcctcgatgaaaatgtggtggctctttcaagtcctccctaatgttacacattga
2) Location on the genome NC_004718.3_29697_29542
3) Start and end nucleotide 29697 to 29542
4) Function - hypothetical protein

Perform a BLAST on the nucleotide sequence and paste a screenshot of the obtained BLAST results
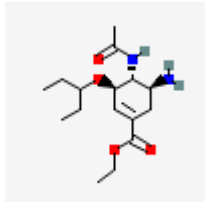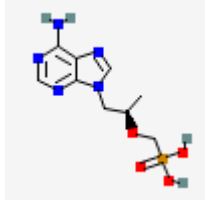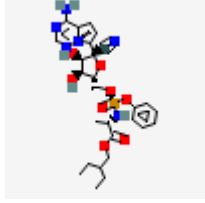
Day 5 & 6:

Molecular Docking

Protein Name: Angiotensin-converting enzyme 2

Protein ID - 7DQA

Ligand Name	Ligand ID	Follows Lipinski Rule?	Energy value	Dock Image
Oseltamivir	65028	Yes	-5.8	
Tenofovir	464205	Yes	-6.6	
Remdesivir	121304016	No	-7.9	

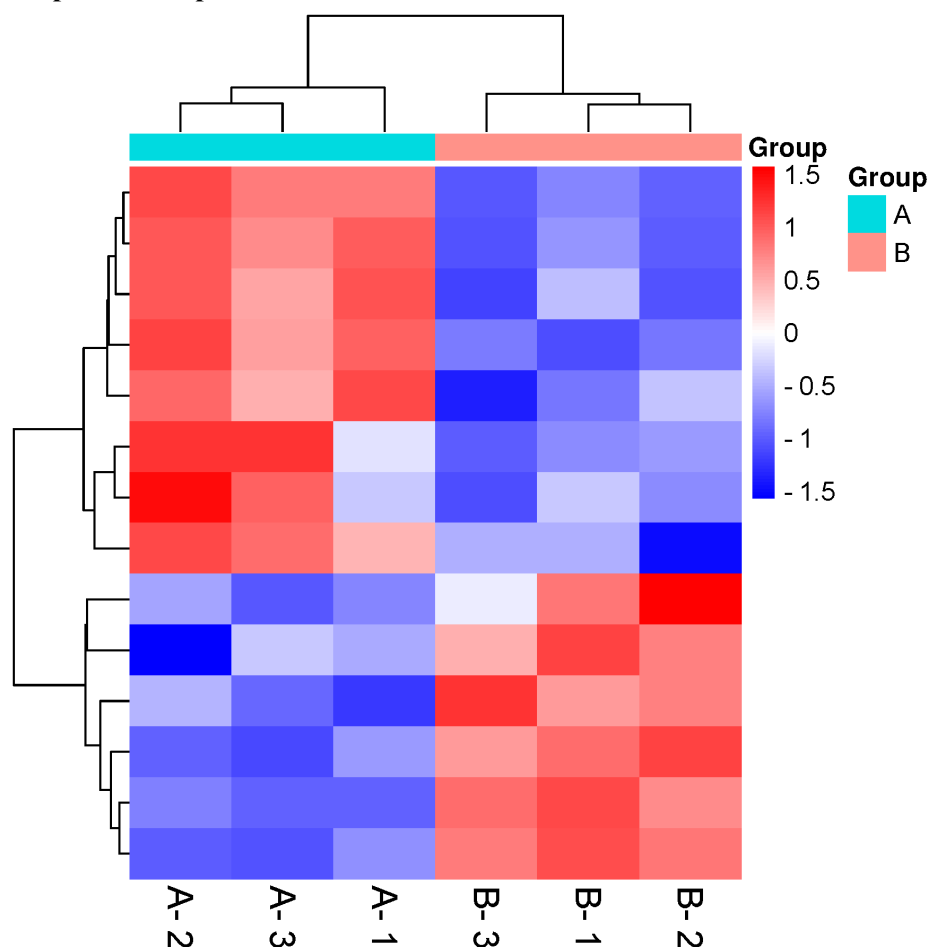**Day 7: Heatmap analysis**

Objective : To plot a heat map and understand the differential expression based on numbered data.

Input data Table

gene	Control-A	Control-A	A	Control-B	Control-B	B
sample	A-1	A-2	A-3	B-1	B-2	B-3
ICA1	5.7	11.7	9.9	5.7	4.5	3.2
DBNDD1	9.43	10.67	9.39	3.4	2.5	2.3
ALS2	10.59	9.89	8.5	4.2	5.75	2.5
CASP10	10.38	10.2	8.5	5.1	2.8	2.4
CFLAR	5.75	10.85	10.9	3.9	4.2	2.8
TFPI	9.82	10.45	8.5	2.5	3.4	3.5
NDUFAF7	8.9	11.02	10.33	5.75	2.4	5.75
RBM5	10.59	10.67	9.94	6.55	5.75	5.55
MTMR7	2.5	5.2	3.5	9.21	9.76	11.47
SLC7A2	3.5	4.2	2.5	9.21	11.78	5.75
ARF5	4.5	3.2	2.7	9.89	10.72	8.81
SARM1	3.5	2.2	1.9	10.85	9.76	9.73
POLDIP2	3.5	4.2	3.5	11.33	9.76	10.49
PLXND1	5.1	1.2	5.75	11.22	9.76	8.81

Output heatmap:

Five interpretation points understood:

1.Heat maps can reveal clustering patterns, where genes with similar expression patterns are grouped together. This can suggest functional relationships or shared regulatory mechanisms.

2. By comparing heat maps between different conditions or experimental groups, researchers can identify genes that are differentially expressed.

3. Heat maps can reveal co-expression patterns, where certain genes show consistent up- or down-regulation across multiple samples suggesting co-regulation or functional interactions between genes.

4. Heat maps can be used to identify gene sets or pathways that are coordinately regulated. Genes involved in the same biological pathway or functional process may exhibit similar expression patterns.

5. Heat maps can help identify genes or samples that deviate from the general expression patterns representing phenomena that require further investigation.
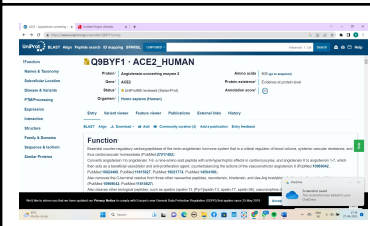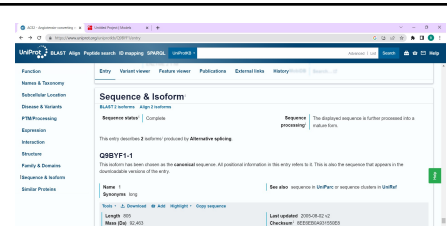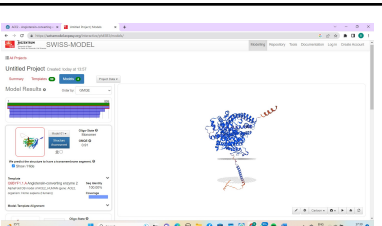
Day 8 & 9: Homology Modelling:

You can choose any protein which is involved in SARS CoV-2 Pathogenesis (Eg: ACE2 receptor, Any envelope protein) and can take at least 2 homologous sequences with sequence similarity >30%. Try to develop an hypothesis around it (Like Why you want to use Homology modelling for your protein of interest, Purpose and outcome of it) and more importantly how it is going to add value to your hypothesis.

Protein: Angiotensin-converting Enzyme 2
Gene: ACE2
PDB: Q9BYF1 (First Isoform)

Target	Sequence	Result

Day 10:

Please paste your GitHub account link: https://github.com/apoorvade/bversity23