

****

**PROJECT REPORT**

**SUBMITTED BY:**

**K.HARINI**

**Day 2: Gene Annotation**

Protein Name: **ACE2 isoform 3 precursor**

Protein ID – NP\_001373189.1

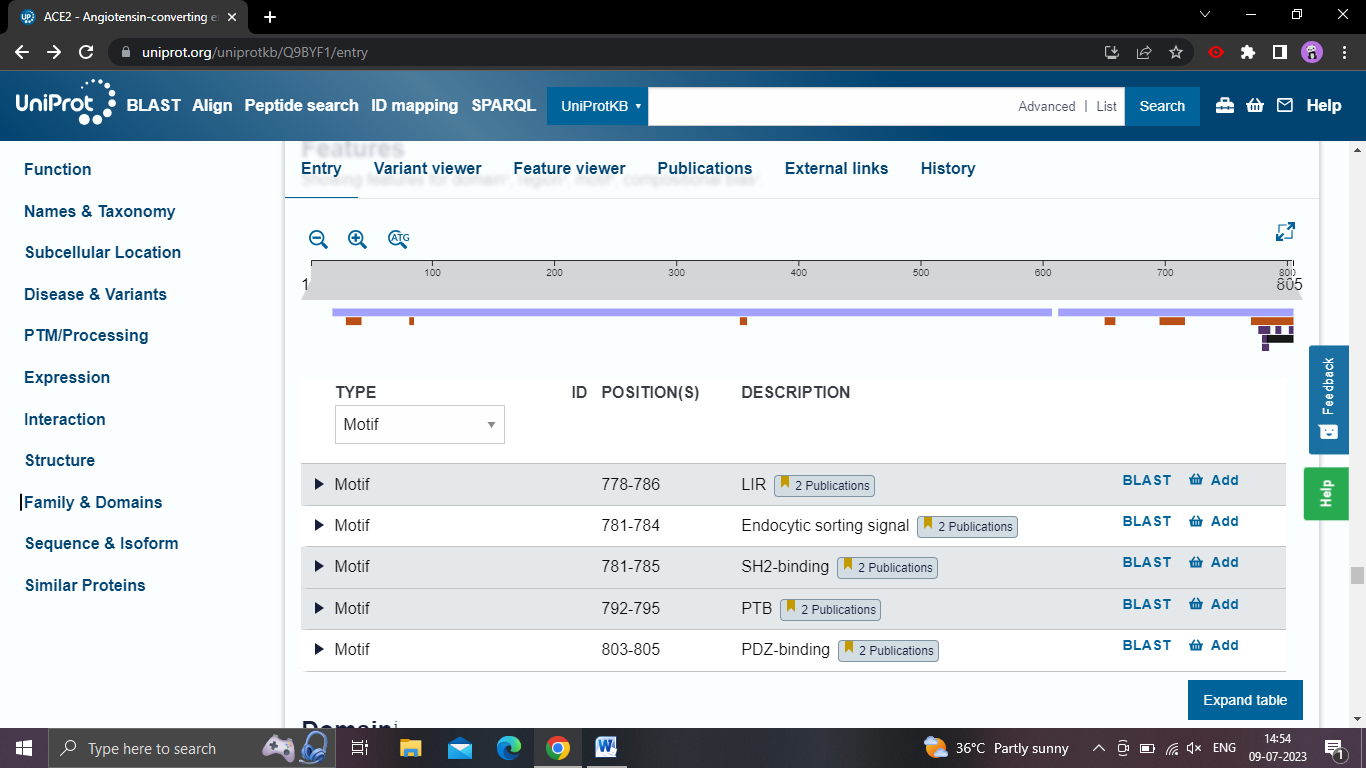
**Find the following Gene function details**Location- Start: 1 and End: 575

Family- Peptidase\_M2

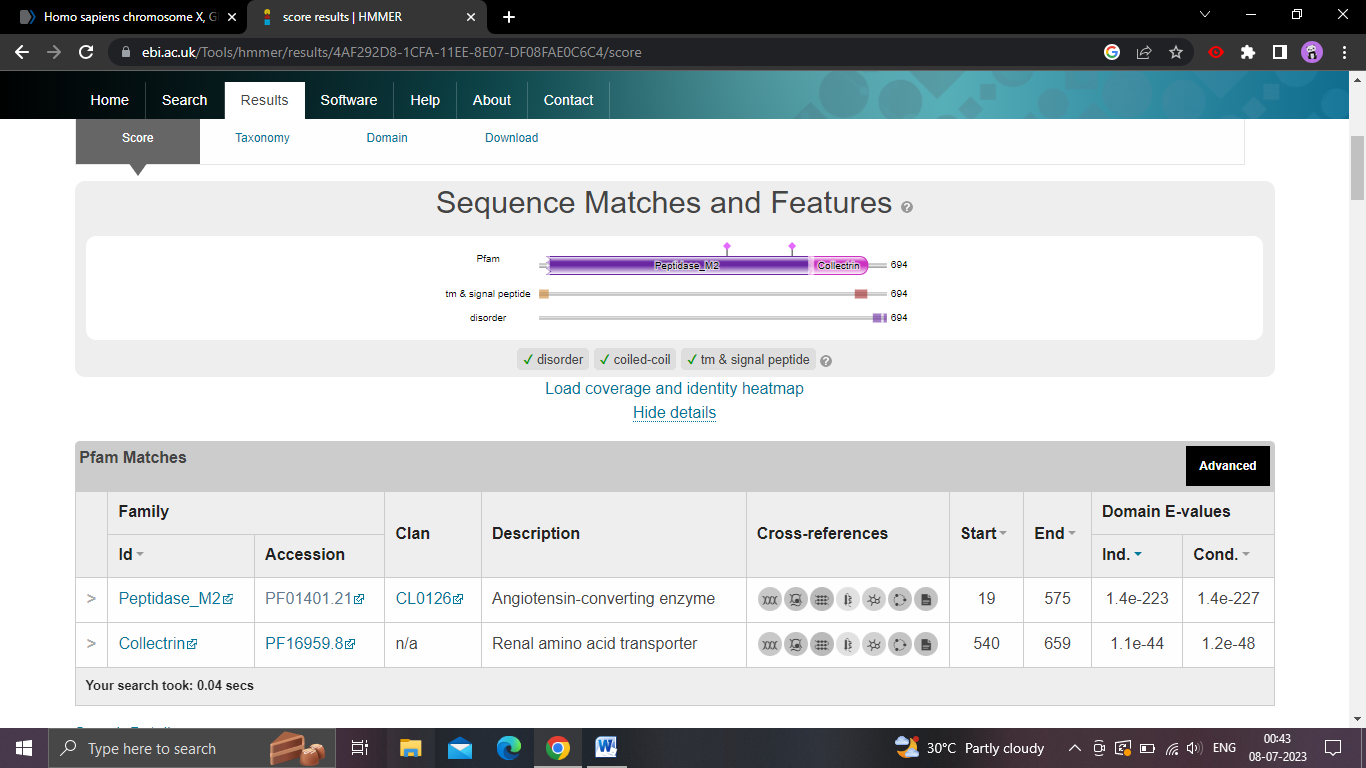
Clan – CL0126

Domain - The extracellular region of the ACE2 enzyme is composed of two domains. The first is a zinc metallopeptidase domain (residues 19-611). The second domain is located at the C-terminus (residues 612-740) and is 48% identical to human collectrin. The cytoplasmic tail contains several linear motifs such as LIR, PDZ-binding, PTB and endocytic sorting signal motifs that would allow interaction with proteins that mediate endocytic trafficking and autophagy.

Motif –



E value – Ind. 1.4e-223, Cond. 1.4e-227



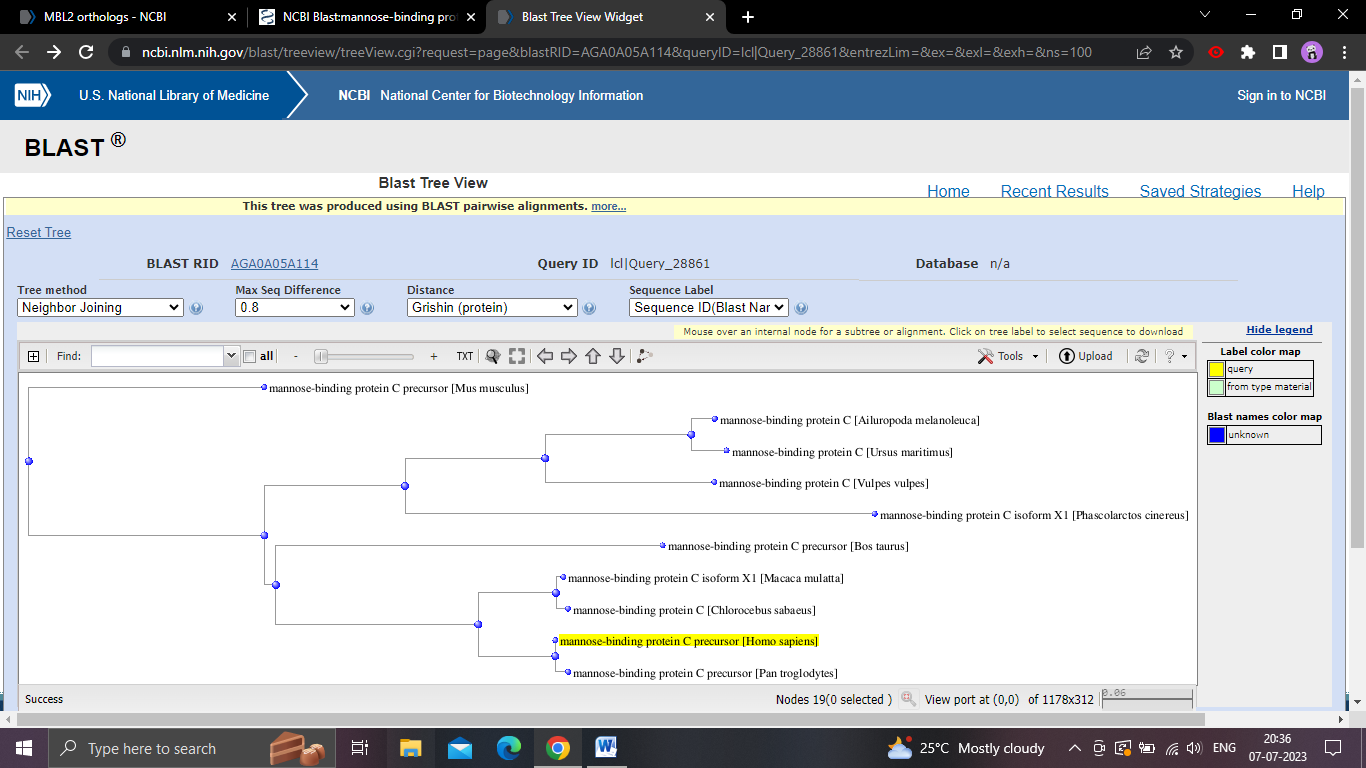
Description of function - The ACE2 protein encoded by this gene belongs to the angiotensin-converting enzyme family of dipeptidyl carboxydipeptidases and has considerable homology to human angiotensin 1 converting enzyme. This secreted protein catalyzes the cleavage of angiotensin I into angiotensin 1-9, and angiotensin II into the vasodilator angiotensin 1-7. ACE2 is known to be expressed in various human organs, and its organ- and cell-specific expression suggests that it may play a role in the regulation of cardiovascular and renal function, as well as fertility. In addition, the encoded protein is a functional receptor for the spike glycoprotein of the human coronavirus HCoV-NL63 and the human severe acute respiratory syndrome coronaviruses, SARS-CoV and SARS-CoV-2, the latter is the causative agent of coronavirus disease-2019 (COVID-19). Multiple splice variants have been found for this gene and the dACE2 (or MIRb-ACE2) splice variant has been found to be interferon inducible.

**Day 3: Phylogenetics**

**Construct a Phylogenetic tree for components of Corona virus.**

**You can choose any gene/protein/component associated with Corona virus for atleast 5 different species/variants. Add the screenshot of the tree here.**

Phylogenetic tree construction: MBL2 - Mannose Binding Lectin 2



**Day 4: Genome Annotation**

**Genome name (the one of your interest): Human papillomavirus type 60, complete genome**

**NCBI Taxonomy ID: 333930**

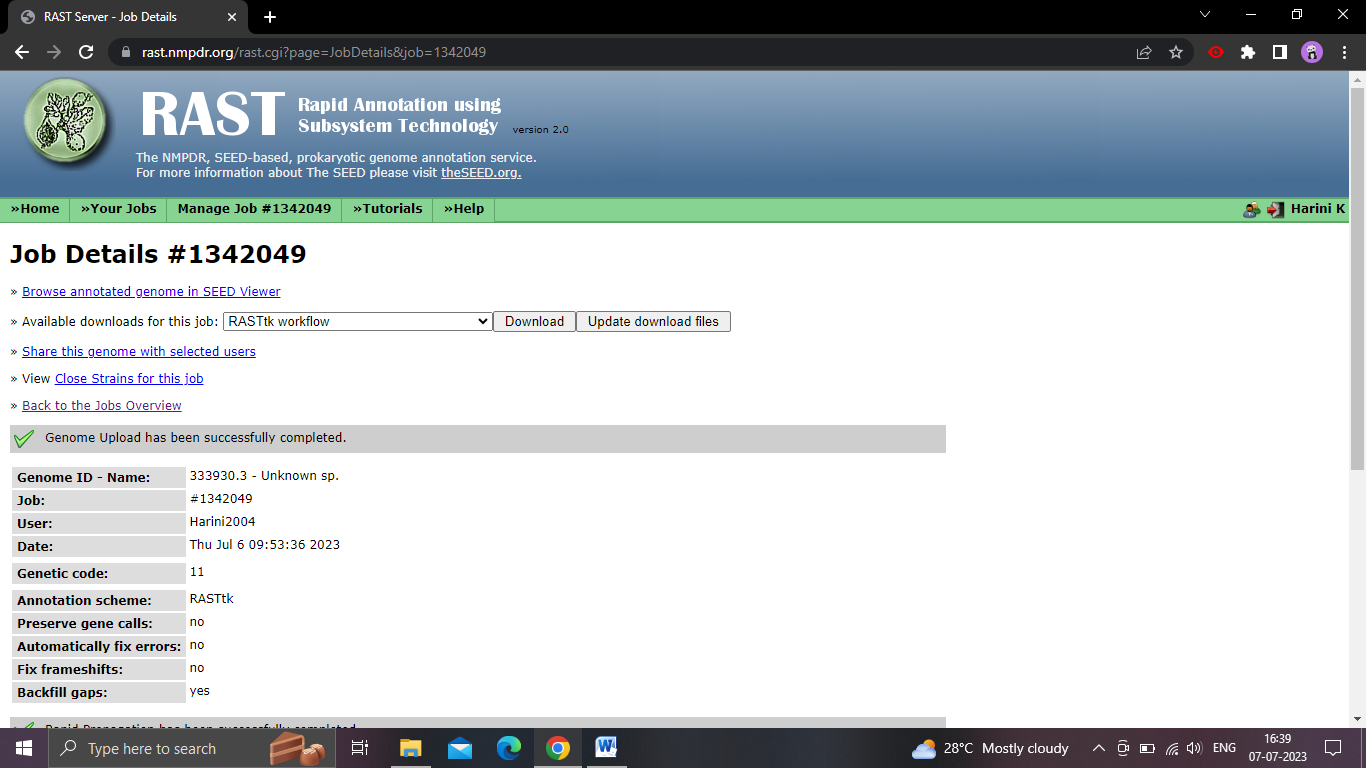
**From RAST results:**

*Mention the desired nucleotide sequence that you choose to perform BLAST on and fill in the following:*

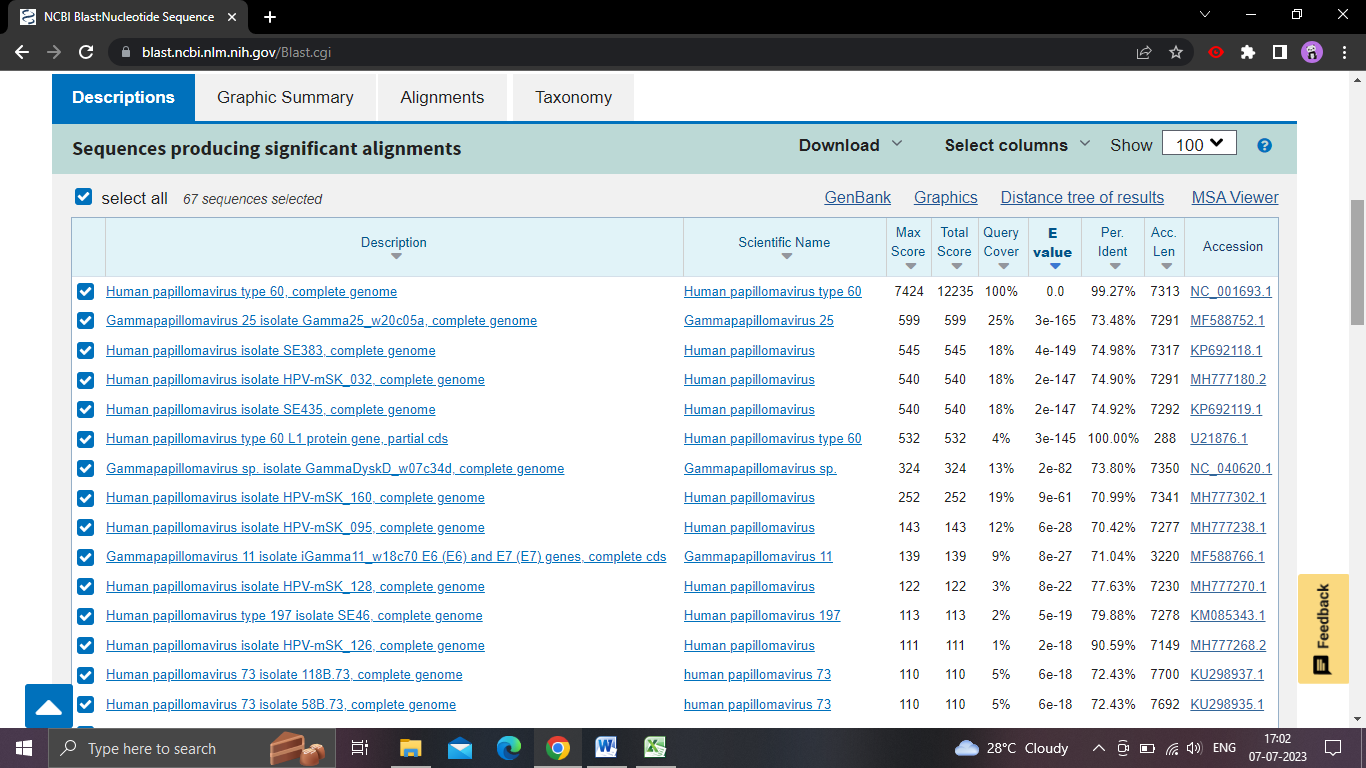
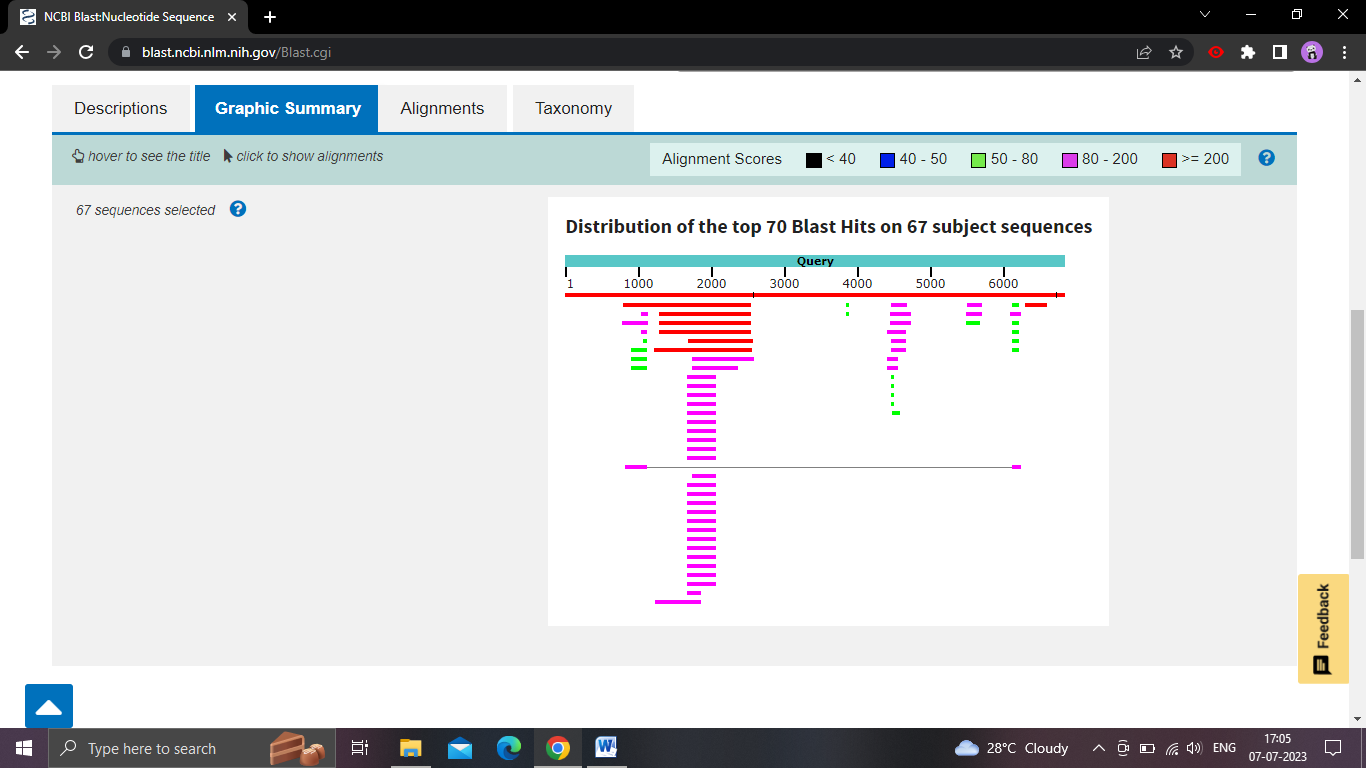
1. Nucleotide sequence

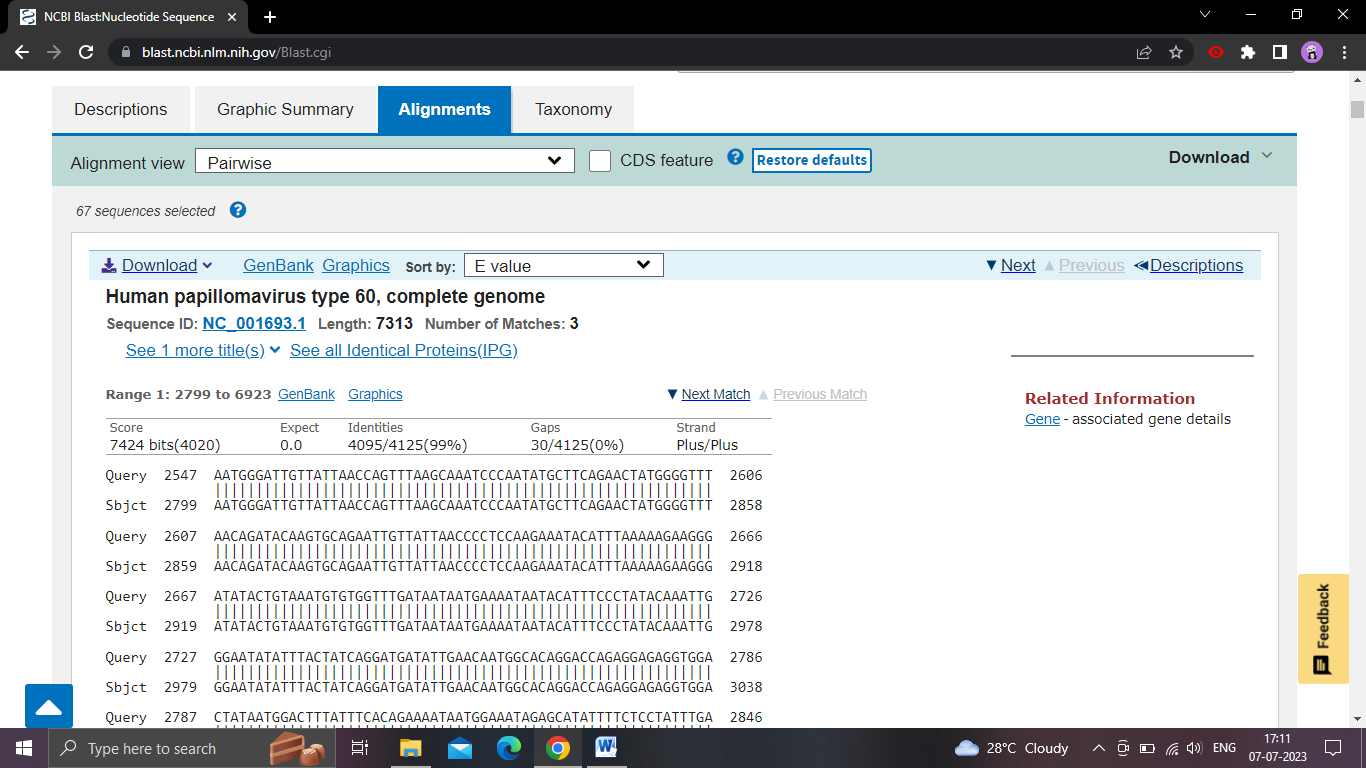
atggaagaagacaggtttccaacaacagtggctgattattgctctgaatttgatattcctttaaaagatcttaagttaaaatgtgtattttgtagattttacttaactgaacagcagttggctgcattttatattaagaatttaaagttagtttggaaaaaccgttattgctttgcttgttgtactccgtgtcttagacttactgctaaatttgaagctgaaaactattttcagtgtatgtgtaaaggagaagtattagaagttctaactcgtattcctttgagttctctttctgtacgctgttttgattgccttacattgttatcttttgcagaaaaaatagattgtataattagtggtcaaaacttttatcttgtaaggggtcgctggagatcatattgtagaaattgcattgagaaatgaatgattggtaatcagcctaatgttaataaccttgatgtaaatttggaggagttggttttacctgtcagtcttttagctgatgaggagttgtcacccgatggtgatcctgaggaggaggagcactatccctatacaatagacacctgttgcaaaccttgtggagcaggtgttagatttactattattgctactccttctgctgtaataactcttcgtcaactattgcttcaagaagtgttcctgacctgtttgaggtgctccagatccctttttcgacatggcagatcctaaatggcagatcctaataaaggtattaattctcttgaattaaatgaggggcatagcgaatggtatgttgtgacagaagctgagtgtattaatagtttggatacaatggaagagctatttgaagaaagtacagatggatctattgtgtccaatctgatagacgattccgaggaactggaggagggaaattccctggcactctacaatgaacagttaacagaggattgcaatagagctattttagcgctaaaacgaaagttaactaaaacacctttgaaaagccaggacagaacggttgctgacctaagtccaaggttggaagctgtcacaatttcacctcagagacaaagcaaaaggagattatttgaggacagtggattaggagaagatgaagctacaaattctattgaaaaaaaggtagtttcgaactctttagagagtaatgaaagtgggaccttggttgtggaaacggacagtatatttcgcagtactaatagaaaagccactttattggctaaatttaaagaatactttggtgtagcatatggagatttaacgcgaccgtttaaaagtgatagatcatgttgtgaaaactgggttataagtgtgtgtgccgctgcagaagaagtaatagaagcttcaaaaacagttatgcagcaacattgtgattttttacaggttatttcatatgggttttatgctttatatttggtaaaatttaaaacagcaaaaagtagagatactataatgaaattattttctttaacattgaatgtacaagaacaacaattaatgtgtgatccaccaaaatctagaagtaccccaacagctctttatttttatagaaggtcatttggaaatgcatcgtttatttatggaccgtttccagattggttagcaaaattaactatgttagaccatgaatccgccgcgagctcagaacagtttgaacttgctcaaatgattcaatttgcttatgacaataacttgacaactgaatctgaaatagcttataaatatgcattgttagctgattctgatgctaatgcagctgcatttttaaaaagtaatcaacaggtaaaatatgtacgagactgctatgcaatgttaagatattataaaagacaagaaatgaaagatatgtcaatttctgagtggatatggaaatgttgcgatgattgtaatcaagagggcaactggaaattaatagcacaatttttacgttatcaagaagttaatttcatttcatttttatgtgcattaaagacattgtttaaaggtattccaaaaagaaattgtttagtgttttggggacctccagatacaggaaagtcttatatttgttcttcacttacacggtttatgcagggaaaagtggtatctttcatgaatagacacagtcaattttggctacaacctttacaagattgtaagcttggattcttagatgatgctacatttcaatgttggcaatacatggatgtaaatatgcgaaatgcattagatggtaatcatatttcattagatttaaaacataaagctcctttacaaataaaattacctcctttattaataactaccaatgttgatgttgaaaatgaagctagcttaatgtatttaaagagtagactagtattttttaagtttcctaacaaattacctttaaaagaaaatgatgaagtactttatgaaattactgatgcatcatggaaatgtttttttatcaaatttgcaagccatttagagctgacagccagaggagacgagcagcatgaatcaggccgatctgaccgagcgttccgatgcactgcaggaacaaatactgaatctatatgaatgggattgttattaaccagtttaagcaaatcccaatatgcttcagaactatggggtttaacagatacaagtgcagaattgttattaacccctccaagaaatacatttaaaaagaagggatatactgtaaatgtgtggtttgataataatgaaaataatacatttccctatacaaattgggaatatatttactatcaggatgatattgaacaatggcacaggaccagaggagaggtggactataatggactttatttcacagaaaataatggaaatagagcatattttctcctatttgatagtgatgcacaaacatattcacaaactgggacatggacagtgcattataaaaaccaaattatttctgctcctgttaccagctcctcaaaacaatcctccgacgactacacttccaaagccgggcagcaaccccacttcttcgcctcatcctcatcgccgactactaccgacggaggacagacctcacaagagggagtctctagctctaccacgtcgccgagtgctgttcgattacgacgccgaagatccaacgagcaacaaagagaactatcctccagagagtcgccccgtaccaaaagacgcagagtacccgacgaagtcgaccgacagagtgccgtggggtctgccccaactgctgaagaggtgggaagcagacatagatctcttcctagaagcggtatatcaagacttgcaagacttcaaggagaagctcgggatcctccaatcctgctaattaaaggtctagcaaactcccttaaatgttggagatatcgattgaaaaagtatacacgttattttaaatgtatgagtacagtatttagatgggtagacatagacgtaccagaatcttctagacataaattgttagttgtttttaatgataccacacaacgagatgtttttatgaaattagtaactttgccaagaggttgtacatatacatttggaacattaaactctttgtaaatgtatgctagagtaaagcgtgttaaaagagactctgttgaaaacttatataaacaatgtcaacttggtgctgactgccctcctgatgtacgaaataaagtagaaggaaccacacttgcagatcgattattacaaatatttggaagtatcttatacctggggaatttgggtatagggacaggtaaaggatcggggggtgctacaggctatacacctcttggtactgcaagagtgccagcgtctacacctggaacagttataaaacccacacgaccgttttcggttcctttagatccaataggttctggaattccatcgcaacctgtaggaggtcggttacctgtggacattatagatgccagtgcttcctctattatacccttgcaggaggtcctgccagagaccaccattatagtaggcggtgacagcggtcctggattgggtgcaagtgaaattgatatagtttcagaaccaagaccagatgtggtcggtgttgatactcaaccaacagtatatacctccattgataatacagttgctacattagacattacaccagctacaccaccagtgaagaaaataatattagaccctataagttcaggttctgaaggtgctgcagcaataacattttcagatataagtgctgcagatttaaatgtatttgttgatcctcagggagcaggagatagaataagttttggagaagaaatagagttaggccctattaatcagcctgctcaatttgaaatagaagaacctccacgtaccagtacacctggggagggatttcaacgtgttacaacacgggcaagggaattatataatagatttgtgcagcaacagccaacacaaaatattgacttcttaggacgtccttcccgcgctgtacaatttgaatttgaaaatcccgccttttttaatgatgaggtgactatgcagtttgagcaagatttgcaagaggtggctgcagcgccagaccaggattttgcagatgtcagggaattaggacgagctcggttttcagaaacatctgcaggaacaatacgagtcagcagattaggaacaaaaggaacaatgaaaacacgaagtggactaactattggtcagaaagttcacttttattttgatattagtgatattccagctgcagaaaccatacaactaagaacattaggcgagtcttcccacgatttttctgcagtagataatattactgaaagtacatatattaatttaactgaaactacaaatgaaggcctaatacctgacaatatactagaggatgaatttacagaaaactttaataatgcacagctgatatttgcaacaatagatgaaggagaatctatgatcatgcccacaattcctccaggtgtagctttaaagttatttattccagaaatagctgcaagcgtgttaaatgtggttcatccttcttctgagtggactattttgattcctaatgttccagatgaaattattcagcctgctatggcagtagatgtttatgatgacttttatttacatcctcatcttcttaggcgacgcaaacgaaaacggttggattttttttaaatggctctttggttgcagacagctggacaattgtatcttccaccaagcaagcctgttgctcgtgtactcagtacggatgaatatgtacagccaacaaatttagtttttcatacaggaactgacagaatgctaattgtaggacacccttattttgatattatagacagtggttcaaataatattactgtacctaaatgtagtggaaatcagtttagagttatgcgactactatttccagatccaaacaaatttgcaatgatagacagagctgtctttaacccagaaagagagagattagtctggagattggaaggccttgaaattggtagaggtgggcctttaggaattggaacttctggccatccattatttaataaatatggagatacagaaaatcctgccgcatatcctttaaagcagaataatggtgatgacaatagaatggatgtttcaatggatccaaaacaaatgcaattgtttattgtaggatgtaaaccagccactggtgaacactgggatattgctaaaccttgtgatcctgctccagcaaaaggatcctgtcccccaattaaattaacacagtctataattcaggacggtgaaatgtgcgatacaggctttggaaatgcaaattttattaccttgcaagaagataaatcaggtgttcctctagatattaccaatgaaatttgtaaatatcctgatttactaaaaatgacaaaagatatctatggagatgcagtatttttctttggtaaacgtgaacaaatttattccaggcattattttgtaagaggagggatagatggagatagtttgccagattctggatattacttagcaccacaaactgataaacctcaaaataaccttggtggttatagttacttccctacaccaagtggttctgtagcttctagtgataatcaactttttaatagaccttattggcttcacagagcccaaggtgcaaataatggtatttgttggggtaaccaattattcattactatagtagataatactcgaaatactaatttatctatttcagtttacaaacaagatgctgctattgataacagatataaatataaacaggaagattttagacagtatttacgccatactgaagaatatgaagtagagttaattctgagactctgtaaagtgcctttgaatccagatgttctagctcatttaaatgtaatggacaaaaatattttagaggattggcaactttcatttgtacctccaccaccacaaggtattgaagatgcttacagatatataatgtctcaggcaacaatgtgtccaacagatgttcctaatactgaaagggaagacccttataaacagtatacattttggacaatagatttgcaagaacgtttttctaacgaattatcccagttttctcttggaaaaagatatttatatcagtatggcttacttaacggacgaaaacggtctgctagtagttttgtaacgaagaaatcaaaaactgtgaaacgtaaaagaactaaataaatgcaatgccaagataattatgaaggcgcggtagacaaaaatgtagaccccgtgagtcagacacttgctgatgtgcaatttattaatccttcattttacaacagcagcattattaattaa

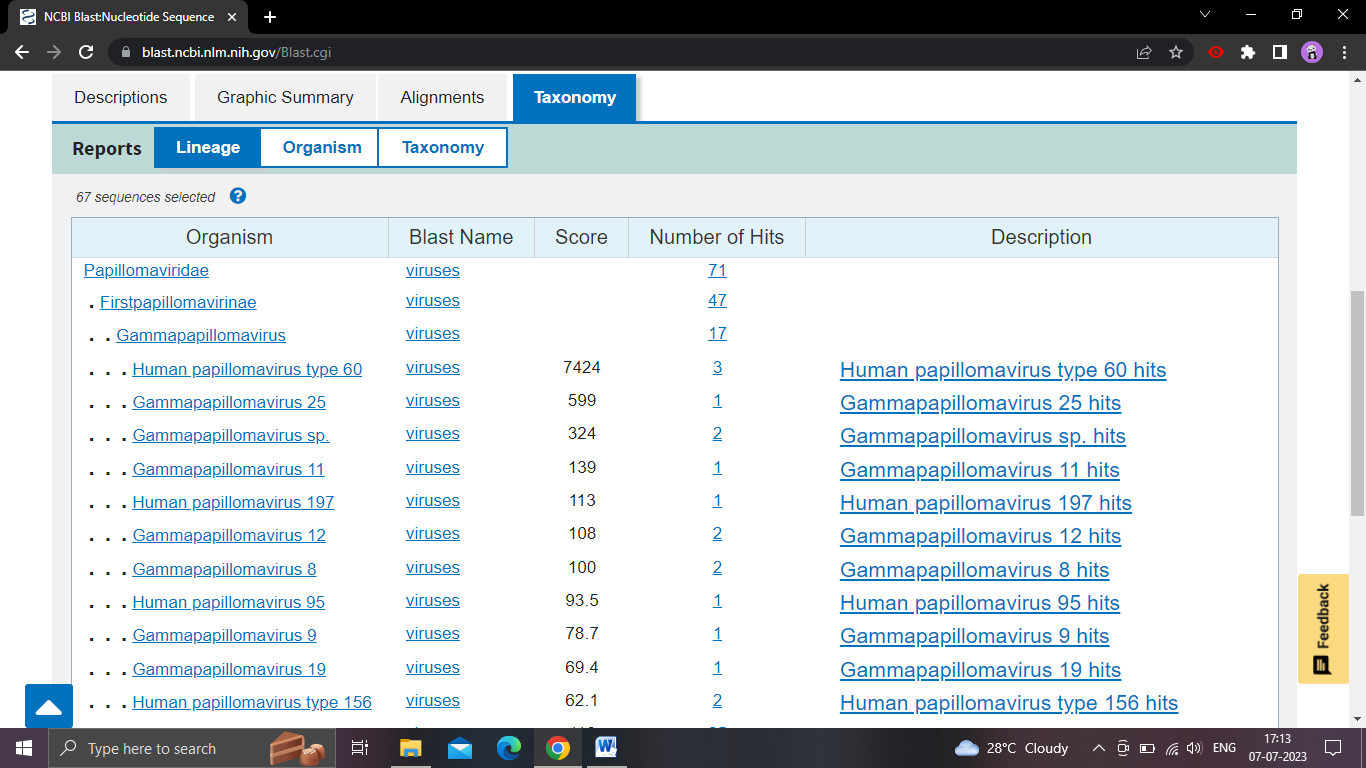
1. Location on the genome- U31792.1\_108\_530
2. Start and end nucleotide- START: 108, END: 530
3. Function- Hypothetical protein



**Perform a BLAST on the nucleotide sequence and paste a screenshot of the obtained BLAST results:**





**Day 5 & 6: Molecular Docking**

**Protein Name**: 7U0E Crystal Structure of C13C9 Fab in complex with SARS-CoV-2 S fusion peptide

**Protein ID** – 7UOE

**PDB DOI**: https://doi.org/10.2210/pdb7U0E/pdb

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Ligand Name** | **Ligand ID** | **Follows Lipinski Rule?** | **Energy value** | **Dock Image** |
| Flurbiprofen | 3394 | Yes | -7.0 |  |
| Amprenavir | 65016 | Yes | -7.1 |  |
| Dorzolamide | 5284549 | Yes | -5.9 |  |

**Day 7: Heatmap analysis**

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

**Problem statement or research question: To generate the heat map and understand the differential expressions of the gene given in the numbered dataset.**

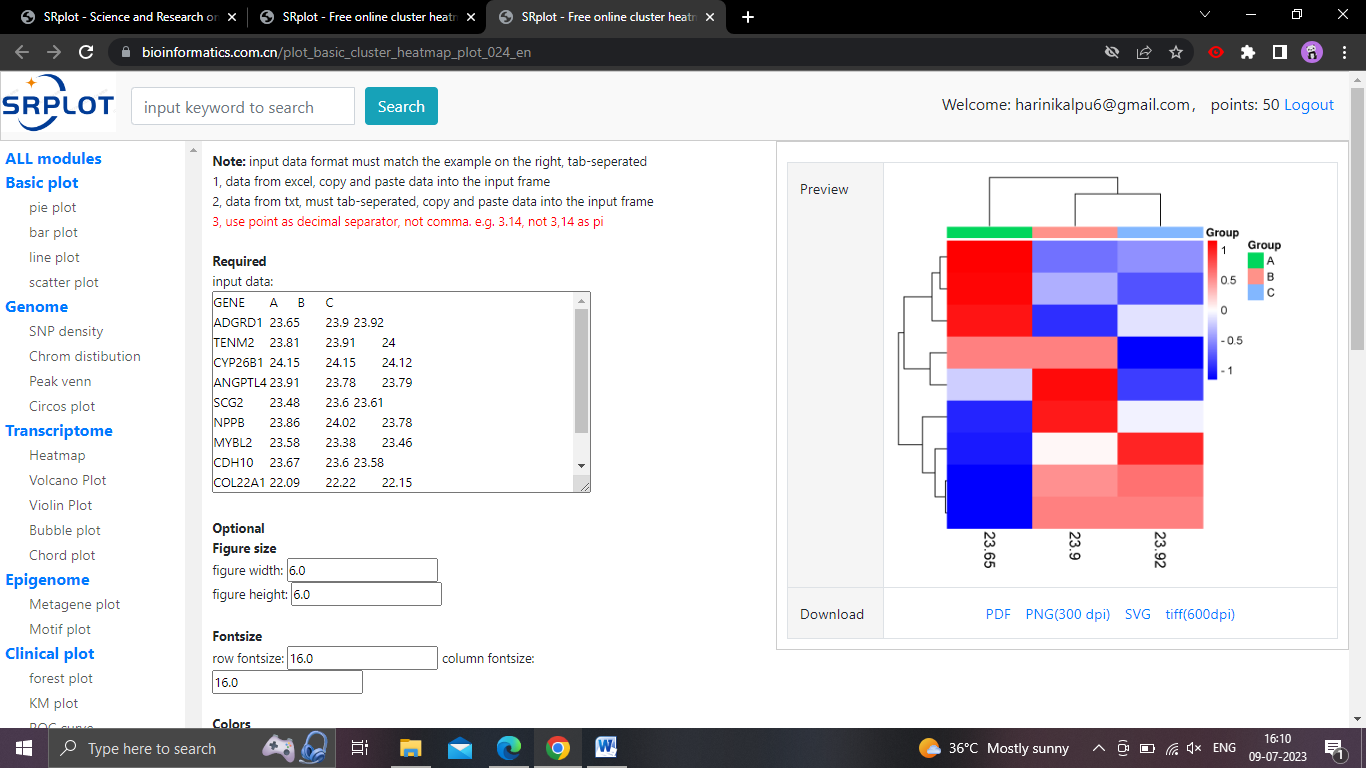
**Input details  
1. Gene(s) Name:** ADGRD1, TENM2, CYP26B1, ANGPTL4, SCG2, NPPB, MYBL2, CDH10, COL22A1, MK167

**2. Variants:** A, B, C

**Input data Table:**

|  |  |  |  |
| --- | --- | --- | --- |
| **GENE** | **A** | **B** | **C** |
| ADGRD1 | 23.65 | 23.9 | 23.92 |
| TENM2 | 23.81 | 23.91 | 24 |
| CYP26B1 | 24.15 | 24.15 | 24.12 |
| ANGPTL4 | 23.91 | 23.78 | 23.79 |
| SCG2 | 23.48 | 23.6 | 23.61 |
| NPPB | 23.86 | 24.02 | 23.78 |
| MYBL2 | 23.58 | 23.38 | 23.46 |
| CDH10 | 23.67 | 23.6 | 23.58 |
| COL22A1 | 22.09 | 22.22 | 22.15 |
| MK167 | 23.01 | 23.04 | 23.04 |

**Output heatmap:**



**Discussion points:** Heat map plot generated using SR PLOT can be an effective method to represent complex and huge amount

of data graphically. SR PLOT has many features like CLUSTER HEATMAP used to create such graphs as illustrated above. This heat map demonstrates heat map graph generation with the protein IL6 interleukin along with its sample dataset. With 10 different genes, the heat map provides the function by mapping the amount of energy at a particular region i.e. A, B, and C.

**Five interpretation points understood:**

* Relative Expression: In ‘A’, ADGRD1 gene has the highest virulence and MK167 gene has the lowest virulence in.
* In ‘B’, SCG2 gene has the highest virulence and CYP26B1 gene has the lowest virulence.
* In ‘C’, CDH10 gene has the highest virulence and ANGPTL4 gene has the lowest virulence.
* In General Expression, the ‘B’ has more virulence than the other variants and ‘C’ has the least virulence of all.
* There are two genes with zero virulence (MYBL2 in ‘B’ and NPPB in ‘C’).

.

**Day 8 & 9: Homology Modelling:**

You can choose any protein which is involved in SARS CoV-2 Pathogenesis pathway or cancer (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.

**Problem statement:** Develop a precise homology modelling approach to visualize and predict the 3D structure of HER2 receptor and to investigate the therapeutic actions of Drug X.

**Protein:** MYC proto-oncogene protein

**Gene:** MYC

**PDB:** P01106

|  |  |  |
| --- | --- | --- |
| **Target** | **Sequence** | **Result** |
|  |  |  |

**Day 10: GITHUB**

Please paste your GitHub account link:

https://github.com/haaarini