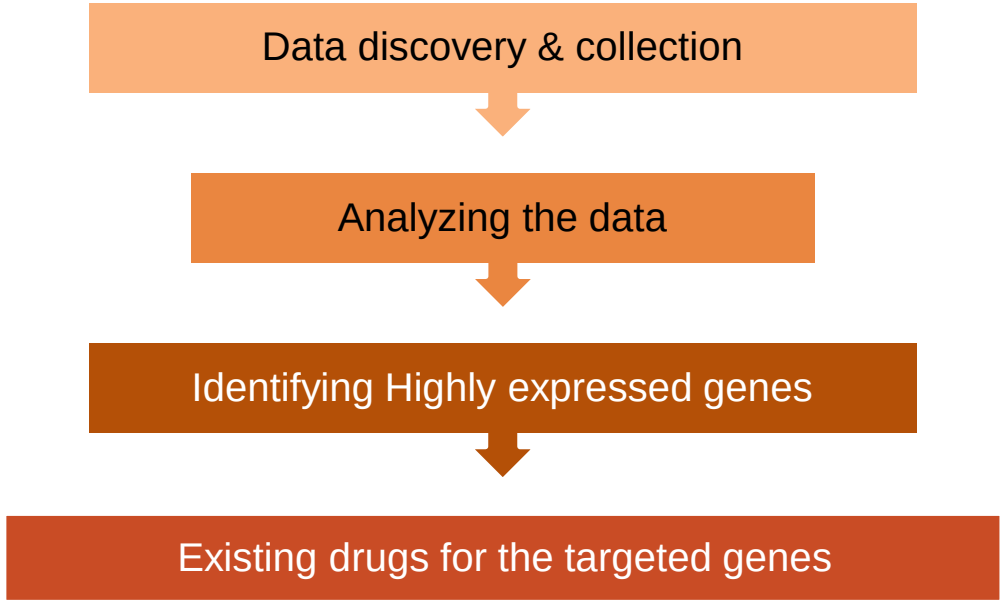
Covid'19 Project
Progress or Status

Repurposing of drugs and reducing the prognostic symptoms of the disease

STATUS **SUMMARY**



DATA **COLLECTION**

Looked for the whole genome data in various sources and latest published papers related to COVID and found 2 datasets in GSE data bank.

GSE – 152075

- It consists of RNA sequencing profiles of nasopharyngeal swabs of COVID positive and negative patients.
- Study consists of data for 430 positive and 54 negative patients.

GSE – 151161

- This data is of whole RNA sequencing blood samples of patients treated with abatacept.
- Data comprises of week 0 & week 12 transcriptional profile of 38 positive patients.

GSE - 152075

INITIAL ANALYSIS

- Calculated the mean, median and standard deviation of gene counts for positive as well as negative patients separately.
- We got mean & median as 0 for some genes, which we converted to 1 for calculating the log odd score.
- Then calculated the log odd score from mean counts as well as median counts.

SELECTING HIGHLY EXPRESSED GENES

- We then selected the genes having median log score >= 1, indicating the change of 2 in no. of genes from negative to positive patient.
- We got 537 genes having median log score >= 1, and 193 genes having mean and median score both >= 1 and 3863 genes were found having mean and median log score <= -1.
- Then calculated the 1^{st} and 3^{rd} quartile for the 537 genes and mean of the counts between this range to cross check our results. We also plotted the gene count distribution graph for each 193 genes having mean and median >= 1.

EXISTING **DRUGS**

After selecting the highly expressed genes, we converted Gene Id to Uniprot Id and Chembl id, which we got 519 Uniprot Ids for 513 unique genes and 131 Chembl Ids for the Gene Ids.

DRUG-BANK DATABASE

- Downloaded the whole drug bank database in xml format. Processed it to get existing drugs for 519 Uniprot Ids and their details.
- We got 784 drugs from the database for 99 unique Uniprot Ids along with their mechanism and the type of drug.

CHEMBL DATABASE

- Collected the relevant data from it's python api and searched the compounds that are active on those 131 Chembl Ids.
- Due to huge database we got 2641 drugs whose names were known for 55 unique Chembl Ids and in total 37686 compounds have activity for 88 unique Chembl Ids.

GSE - 151161

INITIAL ANALYSIS

- Some gene counts were 0 so converted them to 1 for calculating the log odd score.
- Calculated the log score for each patient taking week 0 value in denominator, indicating the expression of genes during the progress of infection as well as effect of Abatcept.
- At last, we calculated mean, median and standard deviation of log odd score for each gene.

CONCLUSION

Identified the target genes for the COVID-19 infection by various statistical and computational tools. Identified the existing drugs for these genes from the Drug Bank database as well as compounds that have an action on these genes from the Chembl database.