



UIT

UNIVERSITY OF INFORMATION
TECHNOLOGY AND SCIENCES



Machine Learning and Network-Based Bioinformatics Models to Identify the Genetic Effect of SARS-COV-2 Infections to Idiopathic Pulmonary Fibrosis(IPF) Patients

Department of Information Technology
University of Information Technology & Sciences (UIT)
Holding 190, Road 5, Block J, Baridhara, Maddha, Naya Nagar Rd,
Dhaka 1212

Introduce Ourselves

Presented By

Imran Hossain (1734255010)
Musfikur Rahaman (16155035)
Tareq Mahmud Tonmoy (1734255015)
Rakibul Alam Limon (1734255012)

Supervised By

Sk. Tanzir Mehedi
Lecturer
Department of IT, UIT

Outlines

- ☐ Introduction
- ☐ Objectives
- ☐ Related Works
- ☐ Methodology
- ☐ Data Analysis
- ☐ Results Analysis
 - ☐ Identification of DEGs and Common Gene Between Covid-19 & IPF
 - ☐ Gene Ontology Analysis
 - ☐ Pathways Finding
 - ☐ PPI Network Analysis
 - ☐ Hub Genes Analysis
 - ☐ TF-Gene Interactions
 - ☐ TF-miRNA Coregulatory Network
 - ☐ Candidate Drug Identification
- ☐ Conclusion
- ☐ Reference

Introduction

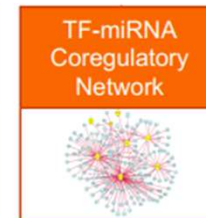
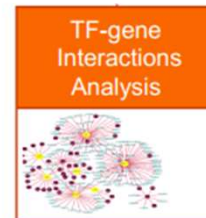
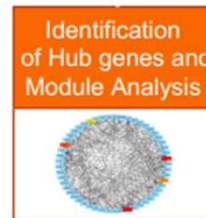
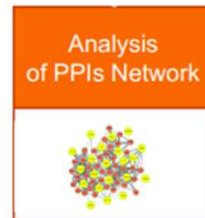
Introduction

- ❑ The ongoing coronavirus (COVID-19) pandemic has driven universal health and healthcare exigency, aside from the remarkable socio-economic effects.
- ❑ Virus genome sequencing is a vital and rapidly-developing tool in the diagnosis of COVID-19 and in understanding the spread and control of the new coronavirus.
- ❑ Idiopathic pulmonary fibrosis (IPF) is a risk factor for COVID-19 as patients with long-term lung injuries are more likely to suffer in the brutishness of the pestilence.



Introduction

- Machine learning helps positively by discovering an effective identification for Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) with IPF.



Objectives

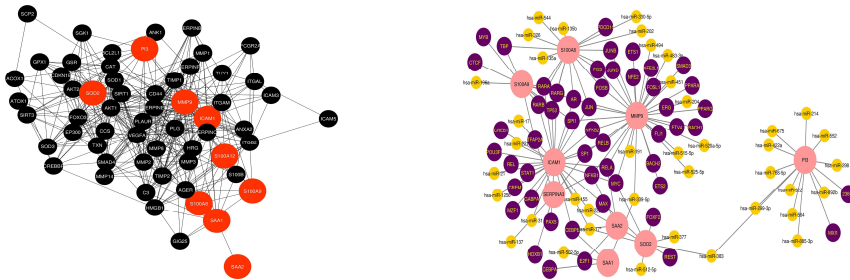
Objectives

- We will use **machine learning** methods which will provide a strong identification genetic effect of **SARS-CoV-2** infections with IPF diseases.
- Also, try to **find out the discrete effects of SARS-CoV-2 infections** with IPF diseases by network based analysis.
- Find Effect of SARS-CoV-2 virus for IPF patients.
- Find the biological pathways and the other relationship between COVID-19 and IPF patients.
- Finding common genes and possible targets drug.
- To present more accurate results by comparing network based analysis with others.

Related Works

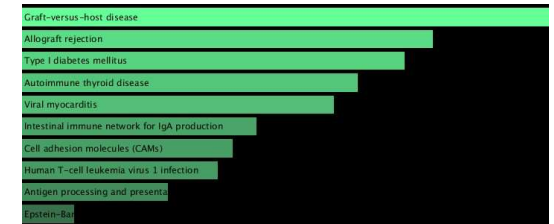
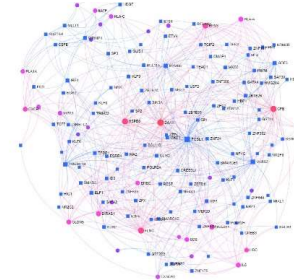
Related Works

Network Based Identification



- Accomplished analysis between two database
- Eleven Common Gene between Covid-19 and IFP
- Find gene ontology and pathway by enrichment analysis
- Identification performance is good

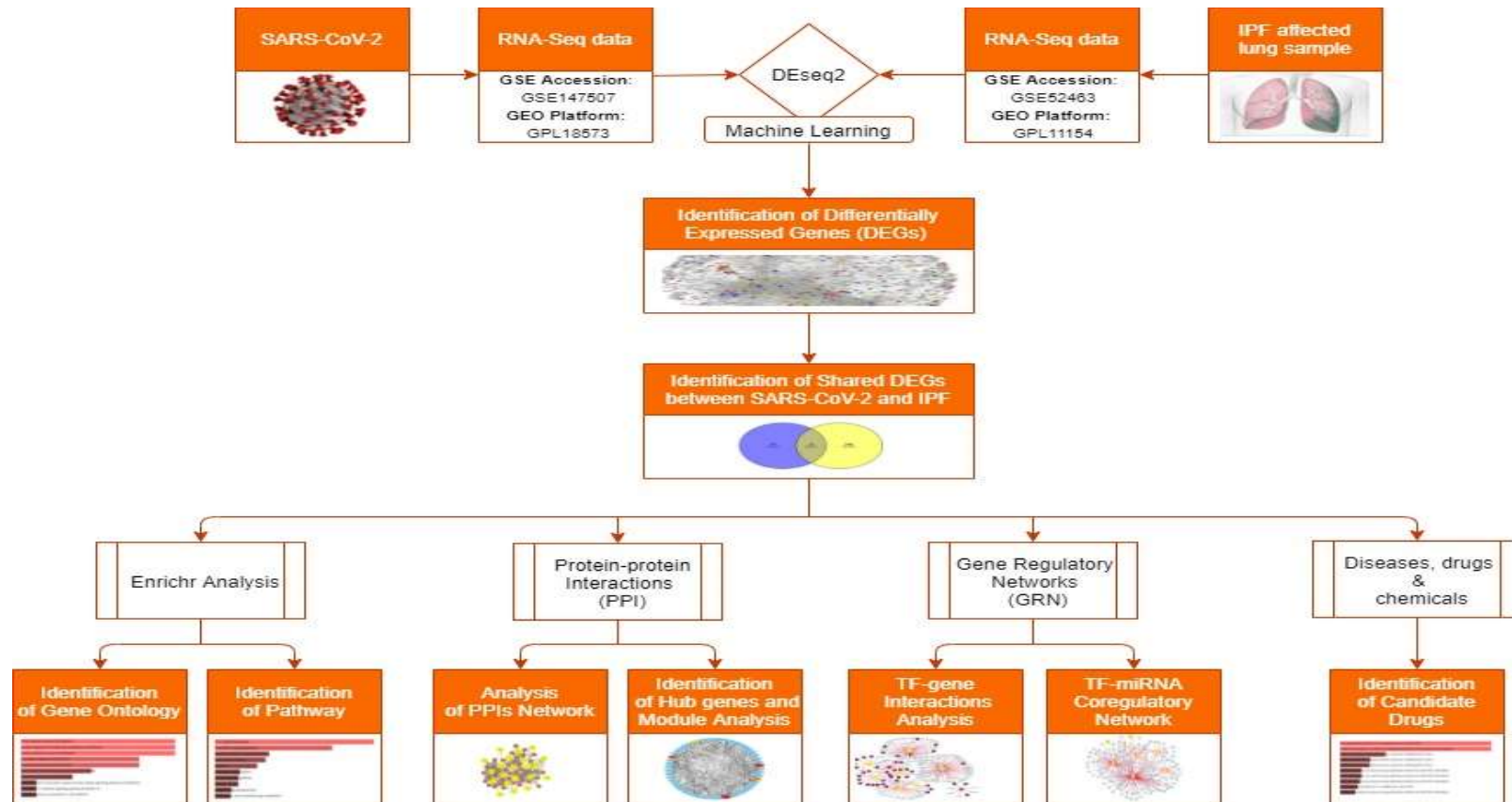
System Biology Approach Based



- Three dataset was used COVID-19, IFP & COPD
- Total 65 Common Gene are identified
- Find pathway and others analysis by enrichment analysis
- Drug-target interactions are suggested from Hub genes

Methodology

Methodology



Data Analysis

Data Analysis

❑ GEO Database

➤ GSE 147507

1. MetaData (Filtered)
2. CountData (Normalized)

➤ GSE 52463

1. MetaData (Filtered)
2. CountData (Normalized)

	GSM12672	GSM12672	GSM12672	GSM12672	GSM12672	GSM12672	GSM12672	GSM12672	GSM12672
A1BG	220	109	104	158	138	181	211	125	157
A1CF	24	22	16	30	35	39	56	18	40
A2M	55168	41228	29820	52334	93331	17203	48773	36323	17747
A2ML1	48	24	22	54	70	87	82	36	65
A2MP1	56	24	39	59	73	46	51	45	31
A3GALT2	1	2	3	4	4	2	4	0	1
A4GALT	186	95	94	348	92	97	212	156	42
A4GNT	15	1	2	4	7	10	3	6	7
AAAS	293	200	232	178	152	151	239	338	104
AACS	674	342	411	412	336	376	548	549	328
AACSP1	26	21	17	23	52	44	44	23	23
AADAC	32	29	173	10	61	44	70	163	29
AADACL2	30	18	14	46	75	16	77	95	11
AADACL3	14	4	7	7	11	30	18	3	14
AADACL4	5	1	1	4	4	5	5	1	7
AADACP1	3	8	4	18	18	6	17	10	5
AADAT	129	89	82	47	67	39	74	117	33
AAED1	173	309	240	267	224	197	275	345	189
AAGAB	331	366	323	295	338	205	399	465	213
AAK1	1751	903	948	1391	1164	1016	1457	1155	836
AAMDC	152	130	118	130	110	131	121	143	119
AAMP	574	616	478	472	310	359	396	689	275

Data Analysis

❑ Analysis with Machine Learning

❑ Cut-Off :

❑ P-Value < 0.05

❑ $|\log_2FC| > 1.0$

```
library(DESeq2)
library(ggplot2)
library(dplyr)
library(readr)
library(RColorBrewer)
```

Read gene counts data

```
countData <- read.csv('GSE147507_filtered_countdata.csv', header = TRUE, sep = ",")
head(countData)
```

	X	GSE4432378	GSM4432379	GSM4432380	GSM4432381	GSM4432382	GSM4432383	GSM4432384	GSM4432385	GSM4432386
	<chr>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>
1	DDX11L1	0	0	0	0	0	0	0	0	0
2	WASH7P	29	24	23	34	19	44	68	43	33
3	FAM138A	0	0	0	0	0	0	0	0	0
4	FAM138F	0	0	0	0	0	0	0	0	0
5	OR4F5	0	0	0	0	0	0	0	0	0
6	LOC729737	112	119	113	127	84	270	11	3	6

Check row and columns of gene counts data

Data Analysis

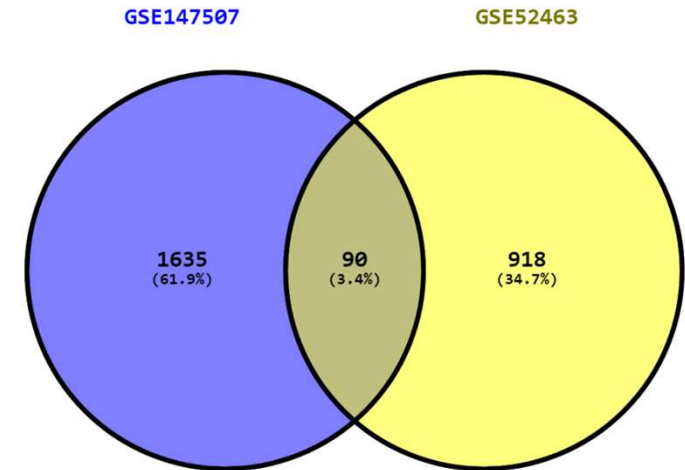
❑ GSE-147507

- ❑ Total - 1725
- ❑ Up Regulated – 926
- ❑ Down Regulated – 799

❑ GSE-52463

- ❑ Total - 1008
- ❑ Up Regulated – 669
- ❑ Down Regulated - 339

❑ Common Gene: 90



Result Analysis

Gene Pathways Finding

☐ Pathways

- ☐ KEGG
- ☐ Wiki Pathways
- ☐ Reactome
- ☐ BioCarta

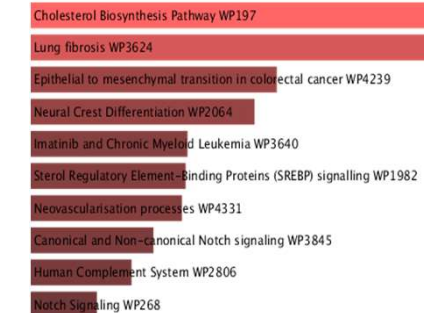
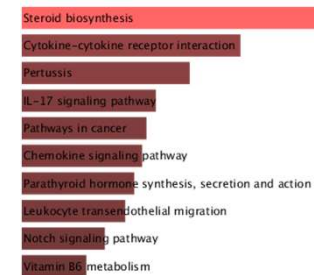
Gene Pathways Finding

- Pathway analysis refers to a method of initial characterization and interpretation of pathological condition that was studied with omics tools.
- This studies might identify long lists of altered genes.
- Interaction among molecules in a cell that leads to a certain change in a cell.
- This methods of pathway analysis might be used to identify key genes/ proteins .

Pathways Finding

Pathways (Top 10 Pathways for each of the subsections)

Database	Pathways	P-Value	Gene
KEGG	Steroid biosynthesis	1.41E-06	CYP24A1;CYP51A1;MSMO1;LSS
	Cytokine-cytokine receptor interaction	2.13E-03	ACVRL1;CXCL6;IL1RN;CCL11;TNF
	Pertussis	4.89E-03	SF8;CCL18
	IL-17 signaling pathway	0.008541751	CXCL6;CCL11;MMP1
	Pathways in cancer	0.009944168	DLL4;MMP1;EPAS1;GADD45A;N
	Chemokine signaling pathway	0.010692617	OTCH4;STAT4;PIM2
	Parathyroid hormone synthesis, secretion and action	0.01217993	CXCL6;CCL11;ARRB1;CCL18
	Leukocyte transendothelial migration	0.014115739	CYP24A1;PDE4B;ARRB1
	Notch signaling pathway	0.019753727	CLDN5;RHOH;CLDN2
	Vitamin B6 metabolism	0.026701191	DLL4;NOTCH4
WikiPathways	Cholesterol Biosynthesis Pathway WP197	5.04E-07	PSAT1
	Lung fibrosis WP3624	9.44E-06	FDPS;CYP51A1;MSMO1;LSS
	Epithelial to mesenchymal transition in colorectal cancer WP4239	7.61E-04	CCL11;SFTPA2;SFTPC;SFTPA1;S
	Neural Crest Differentiation WP2064	0.001122419	MAD7
	Imatinib and Chronic Myeloid Leukemia WP3640	0.003609434	DLL4;CLDN5;CDH2;NOTCH4;CLD
	Sterol Regulatory Element-Binding Proteins (SREBP) signalling WP1982	0.003725975	N2
	Neovascularisation processes WP4331	0.003977743	DLL4;TFAP2A;CDH2;NOTCH4
	Canonical and Non-canonical Notch signaling WP3845	0.006533184	GADD45A;PIM2
	Human Complement System WP2806	0.009581233	FDPS;CYP51A1;LSS
	Notch Signaling WP268	0.017488237	ACVRL1;DLL4



Gene Ontology Finding

☐ Gene Ontology

- ☐ GO Biological Process
- ☐ GO Molecular Function
- ☐ GO Cellular Component

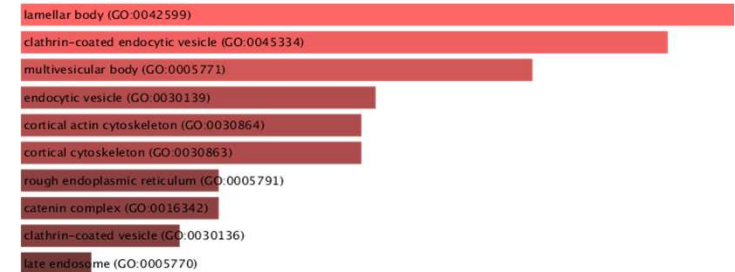
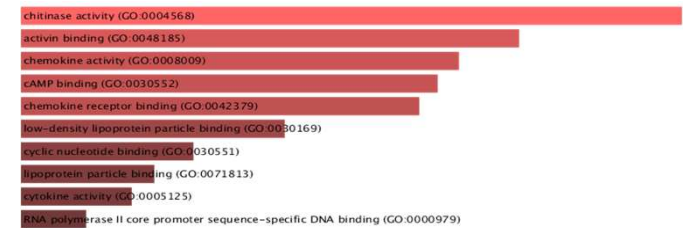
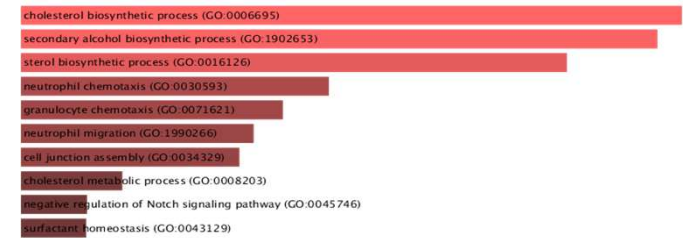
Gene Ontology Finding

- The Gene Ontology is a major bioinformatics initiative to unify the representation of gene and gene product attributes across all species.
- The gene ontology allows users to describe a gene to gene products in detail

Gene Ontology Finding

Gene Ontology (Top 10 GO terms for each of the subsections)

Category	GO ID	GO pathways	P-value	Genes
GO biological process	GO:0006695	Cholesterol biosynthetic process	1.80E-05	FDPS;CYP51A1;MSMO1;LSS
	GO:1902653	Secondary alcohol biosynthetic process	2.02E-05	FDPS;CYP51A1;MSMO1;LSS
	GO:0016126	Sterol biosynthetic process	3.10E-05	FDPS;CYP51A1;MSMO1;LSS
	GO:0030593	Neutrophil chemotaxis	9.49E-05	CXCL6;CCL11;PDE4B;CCL18
	GO:0071621	Granulocyte chemotaxis	1.18E-04	CXCL6;CCL11;PDE4B;CCL18
	GO:1990266	Neutrophil migration	1.35E-04	CXCL6;CCL11;PDE4B;CCL18
	GO:0034329	Cell junction assembly	1.44E-04	CLDN5;CDH2;FLNC;SMAD7
	GO:0008203	Cholesterol metabolic process	2.51E-04	FDPS;CYP51A1;MSMO1;LSS
	GO:0045746	Negative regulation of Notch signaling pathway	2.96E-04	DLL4;EGFL7;ARRB1
	GO:0043129	Surfactant homeostasis	2.97E-04	SFTPA2;SFTPA1
GO Molecular Function	GO:0004568	Chitinase activity	4.14E-04	CHIT1;CHI3L2
	GO:0048185	Activin binding	8.80E-04	ACVRL1;SMAD7
	GO:0008009	Chemokine activity	0.001162919	CXCL6;CCL11;CCL18
	GO:0030552	cAMP binding	0.001283478	POPDC3;PDE4B
	GO:0042379	Chemokine receptor binding	0.001397778	CXCL6;CCL11;CCL18
	GO:0030169	Low-density lipoprotein particle binding	0.00260634	TNFAIP6;PCSK9
	GO:0030551	Cyclic nucleotide binding	0.003977743	POPDC3;PDE4B
	GO:0071813	Lipoprotein particle binding	0.004764343	TNFAIP6;PCSK9
	GO:0005125	Cytokine activity	0.005292261	CXCL6;IL1RN;CCL11;CCL18
	GO:0000979	RNA polymerase II core promoter sequence-specific DNA binding	0.006533184	TFAP2A;POU2AF1
GO Cellular Component	GO:0042599	Lamellar body	3.86E-05	SFTPA2;SFTPC;SFTPA1
	GO:0045334	Clathrin-coated endocytic vesicle	7.54E-05	SFTPA2;SFTPC;CD207;SFTPA1
	GO:0005771	Multivesicular body	0.000295817	SFTPA2;SFTPC;SFTPA1
	GO:0030139	Endocytic vesicle	0.001438507	SFTPA2;SFTPC;CD207;SFTPA1
	GO:0030864	Cortical actin cytoskeleton	0.001660448	CLDN5;CDH2;MYZAP
	GO:0030863	Cortical cytoskeleton	0.001660448	CLDN5;CDH2;MYZAP
	GO:0005791	Rough endoplasmic reticulum	0.007015271	SFTPA2;SFTPA1
	GO:0016342	Catenin complex	0.007015271	CDH2;SMAD7
	GO:0030136	Clathrin-coated vesicle	0.010407203	SFTPA2;SFTPC;SFTPA1
	GO:0005770	Late endosome	0.025350026	SFTPA2;PCSK9;SFTPA1

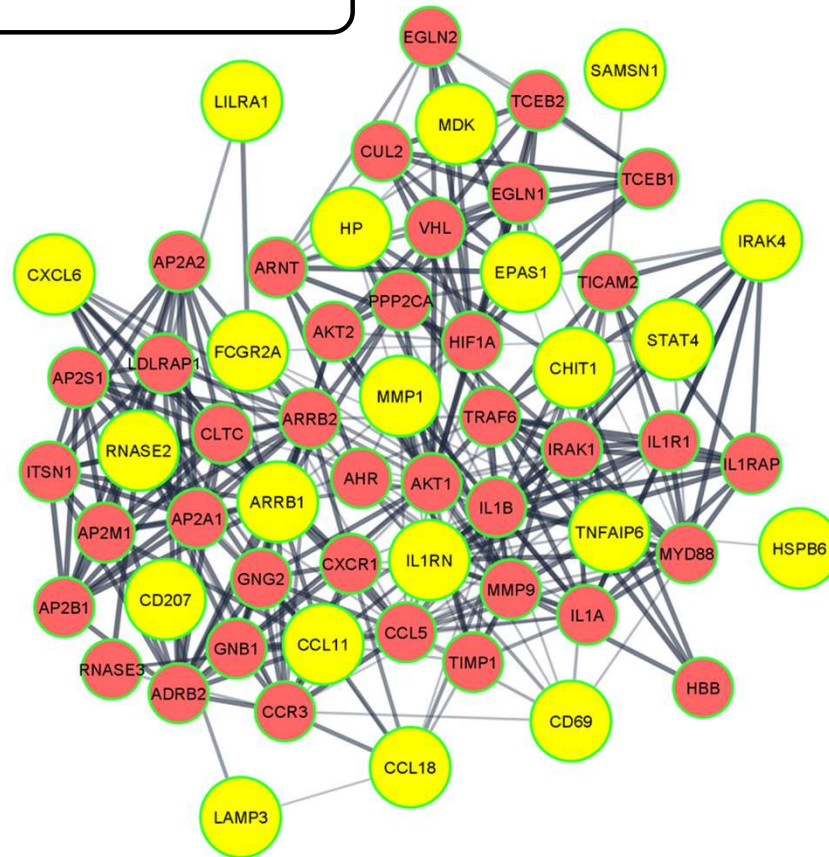


PPIs Network Analysis

- A protein-protein interaction (PPI) network represents a platform by which we have this chance to systematically identify disease-related genes from the relations between genes with similar functions.

PPI Network

□ Node- 60
□ Edge- 308

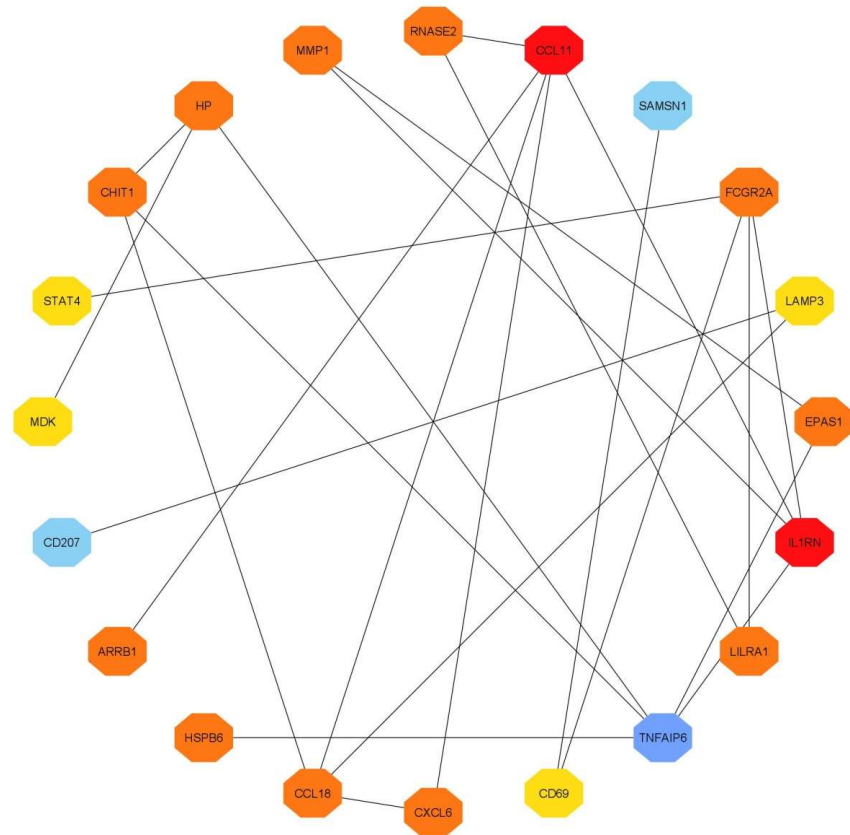


Identification of Hub Genes

- The Nodes that have most interactions in PPI considered Hub Gene.
- Hub genes are defined as genes with high correlation in candidate modules. High connectivity means that the connectivity ranked at top 10%.

Hub Genes

- ❑ Top-18
- ❑ Nodes-20
- ❑ Edges-24
- ❑ Effectful gene: 13

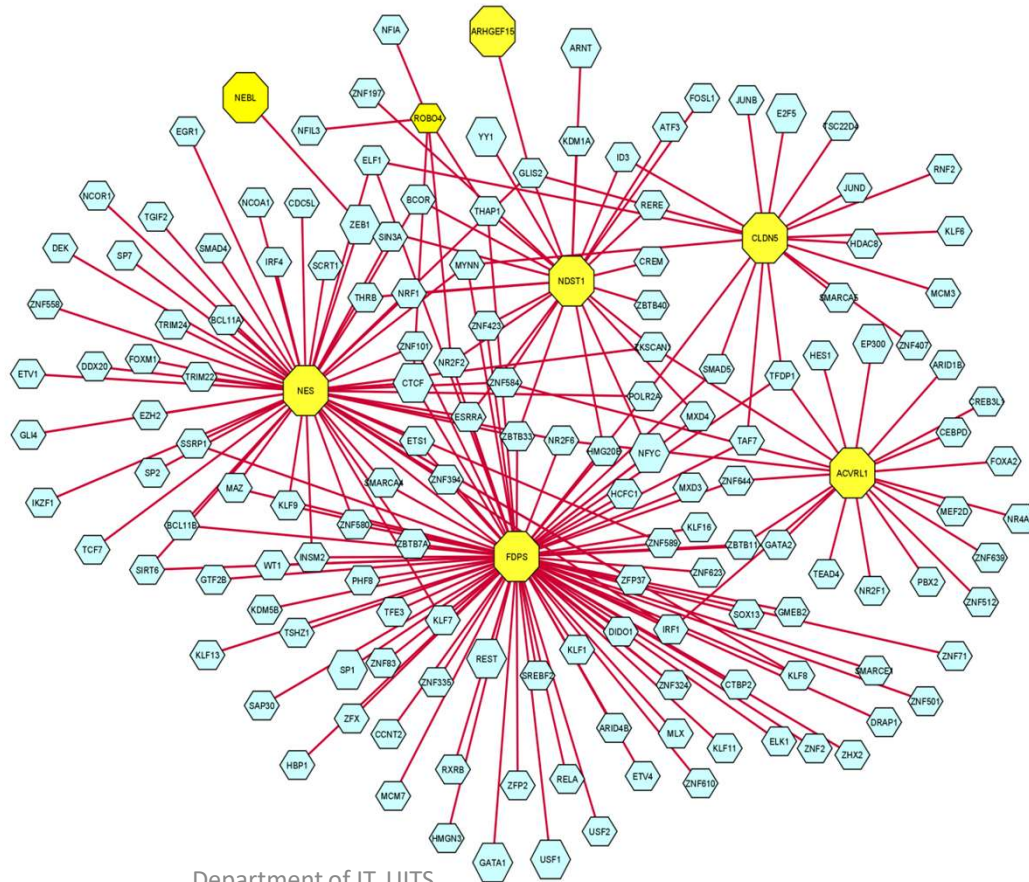


TF-Gene Interactions

- TF gene interaction with the identified common DEGs evaluates the outcome of TF on functional pathways and expression levels of the genes.

TF-Gene Interactions

- Node- 155
- Edge- 208
- TF-Gene- 147
- DEGs Gene -8

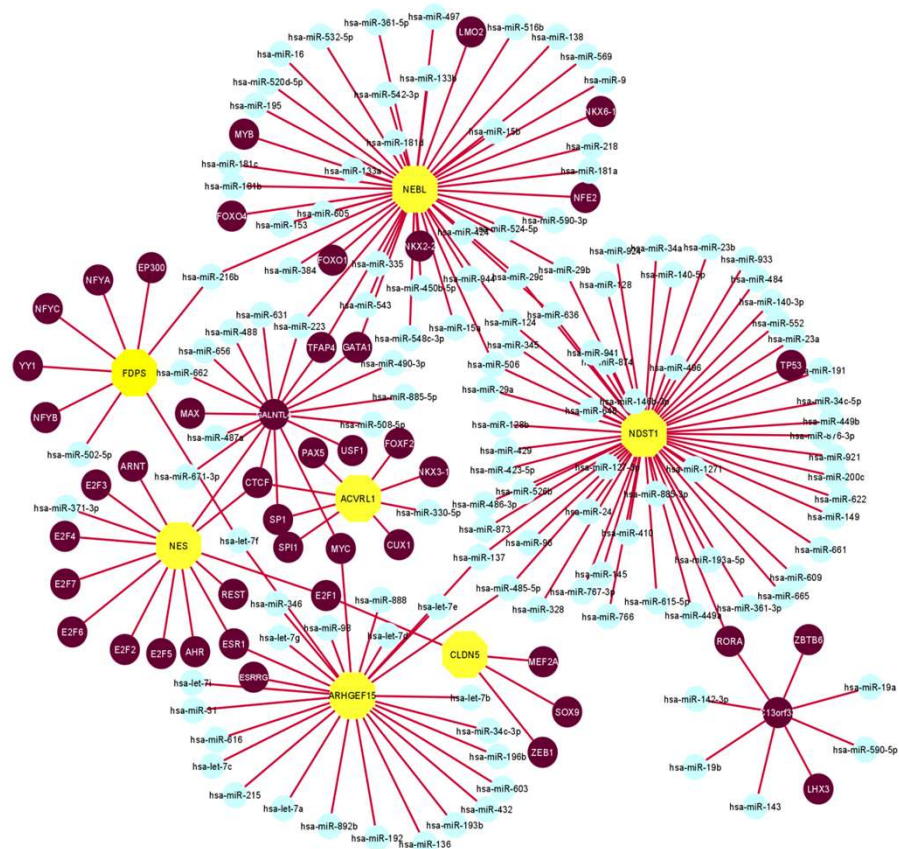


TF-miRNA Coregulatory Network

- The analysis of the TF-miRNA coregulatory network delivers miRNAs and TFs interaction with the common DEGs.

TF-miRNA Coregulatory Network

- Node- 179
- Edge- 189
- miRNA- 125
- DEGs – 7
- TF Gene - 47



Identification of Candidate Drug

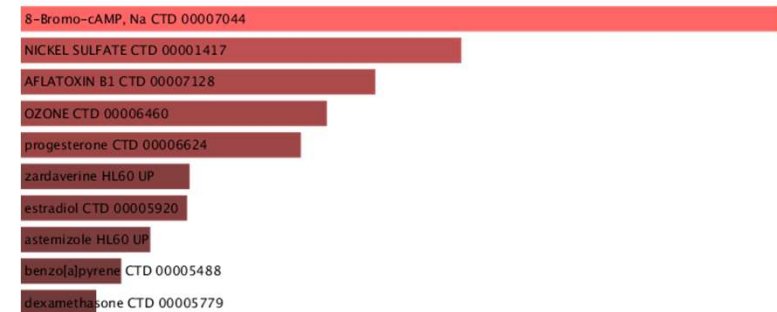
- According to P-value and adjusted P-value, the results from the candidate drugs were generated.
- The analysis depicts that two drug molecules are mostly interacted with most of the genes.
- As these signature drugs were detected for the common DEGs, these drugs represent common drugs for COVID-19 and IPF.

Candidate Drug

Name of drugs	P-value	Adjusted P-value	Genes
8-Bromo-cAMP, Na CTD 00007044	1.22358E-06	0.001936931	SFTPA2;TRPA1;GADD45A;SFTPC;CYP51A1;MSMO1;LSS;RASIP1;MEDAG;CAPN6;PSAT1;LAMP3;PAPPA;SLC39A8
NICKEL SULFATE CTD 00001417	1.02815E-05	0.008137833	CXCL6;IL1RN;CCL11;TNFAIP6;MMP1;EPAS1;LAMP3;CD207;PDE4B;STAT4;CD69;SAMS1
AFLATOXIN B1 CTD 00007128	1.82186E-05	0.00947667	POPDC3;IL1RN;RTKN2;HP;PCSK9;MSMO1;WFDC2;RGS4;SELENBP1;VSIG1;MDK;LAMP3;SLAMF7;PIM2;FLNC;TNS4;FDPS;TRPA1;GADD45A;LSS;CYP24A1;PSAT1;ID3;MNS1;COL9A2;STEAP1;NES;CFB;FREM2;KANK3
OZONE CTD 00006460	2.51744E-05	0.00947667	TFAP2A;SFTPA2;TRPA1;SFTPA1
progesterone CTD 00006624	2.99326E-05	0.00947667	FDPS;TNFAIP6;MMP1;NEBL;GADD45A;CYP51A1;HP;MSMO1;LRRC32;LSS;WFDC2;MEDAG;RGS4;CLDN5;CYP24A1;CAPN6;PSAT1;PAPPA;PDE4B;ID3;SLC39A8;NES
zardaverine HL60 UP	6.27591E-05	0.014433048	IL1RN;TNFAIP6;PDE4B;ID3;SAMS1
estradiol CTD 00005920	6.38227E-05	0.014433048	POPDC3;IL1RN;TNFAIP6;HHLA2;HP;PCSK9;MSMO1;LRRC32;PWP2;WFDC2;RGS4;SELENBP1;NDST1;CAPN6;MDK;LAMP3;PDK4;PDE4B;STAT4;SLC39A8;FLNC;MMP1;NEBL;GADD45A;ELOVL2;CYP51A1;LSS;CLDN5;CYP24A1;PSAT1;PAPPA;ID3;MNS1;STEAP1;CFB;FREM2
astemizole HL60 UP	8.15096E-05	0.016128719	TNFAIP6;CYP51A1;MSMO1;SLC39A8
benzo[a]pyrene CTD 00005488	9.88782E-05	0.017391584	ROBO4;IL1RN;TNFAIP6;EPAS1;GALNT18;RTKN2;HP;PCSK9;ARRB1;MSMO1;MEDAG;SELENBP1;CDH2;MDK;LAMP3;PDK4;PDE4B;PIM2;FLNC;TNS4;CDR1;FDPS;MMP1;GADD45A;CYP51A1;LSS;CLDN5;CYP24A1;PSAT1;PAPPA;POU2AF1;TNIP3;ID3;MNS1;CD69;CFB
dexamethasone CTD 00005779	0.000116758	0.018482771	SFTPA2;MMP1;LAMP3;SFTPC;HP;SLC39A8;SFTPA1;CD69;NES

Drug molecules that most genes are interacted with :

CTD 00005920
CTD 00005488



Conclusion

Conclusion

- A new dimension of machine learning with bioinformatics in the field of medical technology especially for **SARS-Cov2** pandemic.
- In the context of transcriptomic analysis, a few research has been done so far on SARS-CoV-2 and IPF.
- We have accomplished DEGs analysis between two datasets and filtered the materials through common gene identification and attempted to find infection responses between SARS-CoV-2 and IPF affected lung cells.
- Analyses regarding SARS-CoV-2 and IPF to predict the way of detecting infections for various diseases.
- Our analysis ensure **high identification** performances between SARS-CoV-2 and IPF.
- As SARS-CoV-2 is a recent discovery, there has been little research on its risk factors and infections.
- This research will become more and more important with the availability of exceeding datasets.

References

- 1) Coronaviridae Study Group of the International Committee on Taxonomy of Viruses.. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* 2020;**5**(4):536–44 .
- 2) Walls AC, Park YJ, Tortorici MA, *et al.* Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 2020;**181**(2):281–92 .
- 3) George PM, Wells AU, Jenkins RG. Pulmonary fibrosis and COVID-19: the potential role for antifibrotic therapy. *Lancet Respir Med* 2020;**8**(8):807–15 .
- 4) Sheng G, Chen P, Wei Y, *et al.* Viral infection increases the risk of idiopathic pulmonary fibrosis: a meta-analysis. *Chest* 2019;**157**(5):1175–87 .
- 5) Sun P, Lu X, Xu C, *et al.* Understanding of COVID-19 based on current evidence. *J Med Virol* 2020;**92**(6):548–51
- 6) Zou L, Ruan F, Huang M, *et al.* SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *N Engl J Med* 2020;**382**(12):1177–9.
- 7) Rodriguez-Morales AJ, Gallego V, Escalera-Antezana JP, *et al.* COVID-19 in Latin America: the implications of the first confirmed case in Brazil. *Travel Med Infect Dis* 2020;**35**:101613.

THANK YOU !

