

---

*Bioinformatics@Data Science A.Y. 2019-2020***Network Biology**

Hifsa Tahreem and Poompavai Chandraprakash

<sup>1</sup>Group no. 12**Abstract**

Parkinson's disease is a progressive nervous system disorder that affects movement. We try to study the possible neuronal proteins involved in the mutation for this disease by documenting their possible interactions and relationships. We have two interactions sources namely Integrated Interactions Database (IID) and BioGrid. Later, we perform a detailed analysis which involves the enrichment analysis associated to these proteins.

---

**PART – I (DATA COLLECTION)****1. Basic introduction about the disease/process**

Parkinson disease (PD) is a progressive neurodegenerative movement disorder that results primarily from the death of dopaminergic (DA) neurons in the substantia nigra pars compacta (SNc). Mutations in alpha-synuclein, UCHL1 (a ubiquitin carboxy-terminal hydrolase L1), parkin, DJ1 (a parkin-associated protein involved with oxidative stress), and PINK1 (a putative serine threonine kinase) are known to cause early-onset PD. Mutations or altered expression of these proteins contributes to the damage and subsequent loss of DA neurons through common mechanisms that result in proteasome dysfunction, mitochondrial impairment, and oxidative stress. The demise of DA neurons located in the SNc leads to a drop in the dopaminergic input to the striatum. This results in a reduced activation of the direct pathway and in a disinhibition of the indirect pathway, which is associated with the elevation of adenosine A2A receptor transmission. Such unbalanced activity of the striatal output pathway is at the basis of the motor impairment observed in PD.

**1.1. Seed genes****a. DisGeNet dataset**

We get 109 Seed Genes for the disease id **C0030567** that is Parkinson Disease from **DisGeNet dataset**

**b. Uniprot**

We collect the basic information from the **Uniprot**. Possible genes that could be involved in this disease are represented below in the table. For each of these genes, we have the information about the Gene Symbol, Uniprot AC, and Uniprot ID, Main protein, Entrez Gene ID and brief description of its functionality.

We got information about 90 out of 109 genes from Uniprot some of them are enlisted in following table:

Gene symbol	Uniprot AC	Protein name	Entrez Gene ID	Function
AIF1	P55008	Allograft inflammatory factor 1	199	Enhances membrane ruffling and RAC activation.
ATP1A3	P13637	Sodium/potassium-transporting ATPase subunit alpha-3	478	catalyzes the hydrolysis of ATP coupled with the exchange of sodium and potassium ions
BDNF	P23560	Brain-derived neurotrophic factor	627	Important signaling molecule that activates signaling
BST1	Q10588	ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase 2	683	involved in pre-B-cell growth
LYST	Q99698	Lysosomal-trafficking regulator	1130	required for sorting endosomal resident proteins
CP	P00450	Ceruloplasmin	1356	copper-binding glycoprotein. It has ferroxidase activity
CSF1R	P07333	Macrophage colony-stimulating factor 1 receptor	1436	acts as cell-surface receptor for CSF1 and IL34
CYP2D6	P10635	Cytochrome P450 2D6	1565	involved in the metabolism of fatty acids, steroids and retinoids
CYP2E1	P05181	Cytochrome P450 2E1	1571	involved in the metabolism of fatty acids
DCTN1	Q14203	Dynactin subunit 1	1639	Plays a key role in dynein-mediated retrograde transport of vesicles and organelles

**Table 1: Uniprot Dataset**

## 1.2. Summary on interaction data

We got the Interaction data for all our seed Genes from two different sources as discussed above. We also accessed the two sources programmatically. Also, we extract data related to “type of Interaction”. We saved data sets in three different files according to the data sources.

### 1. IID Integrated Interactions Database

IID is an on-line database of detected and predicted protein-protein interactions (PPIs) in 18 species including human, 5 model organisms, and 12 domesticated species. In each of these species, it can provide networks that are specific to tissues, sub-cellular localizations, diseases, and druggable proteins.

Following table represent IID interaction data.

UniProt1	UniProt2	symbol1	symbol2	evidence type
O14656	Q96EU7	TOR1A	C1GALT1C1	exp
O14656	P10909	TOR1A	CLU	exp
O14656	Q5JTV8	TOR1A	TOR1AIP1	exp
O14656	P27824	TOR1A	CANX	exp
O14656	Q9BT78	TOR1A	COPS4	exp;ortho
O14656	Q8IWL3	TOR1A	HSCB	exp
O14656	O75596	TOR1A	CLEC3A	exp
O14656	Q14999	TOR1A	CUL7	exp
O14656	Q14204	TOR1A	DYNC1H1	exp
O14656	Q8WXE9	TOR1A	STON2	exp;ortho

**Table 2 IID Dataset**

## 2. BioGrid

The Biological General Repository for Interaction Datasets (BioGrid) is a curated biological database of protein-protein interactions, genetic interactions, chemical interactions, and post-translational modifications. Following table represent BioGrid interaction data format.

Gene ID	Seed_Gene	GeneA	GeneB
0	AIF1	855811	852324
0	AIF1	855811	855784
0	AIF1	853222	855811
0	AIF1	855661	855811
0	AIF1	855811	841
0	AIF1	855811	843
0	AIF1	819749	843791
0	AIF1	819749	843791
0	AIF1	852169	855811

### Table 3 BioGrid Dataset

### c. Summary of main results

Here is the summary on interaction data from two sources.

Database	Seed Genes	Interacting Proteins including Seed Genes	Total Number of Interactions
BioGrid	86	6836	14789
IID	89	7686	21018

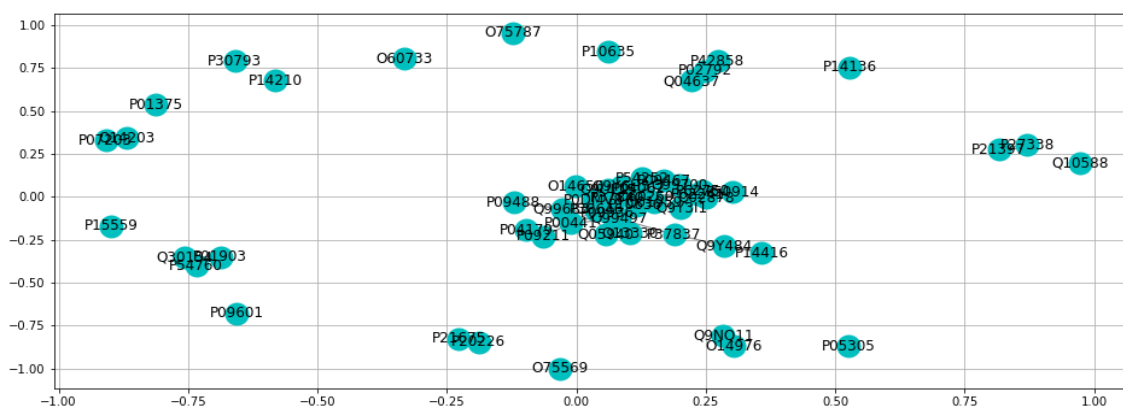
### Table 4: Interaction Summary

### 1.3. Interactome data

### a. Seed genes Interactome

Seed gene interactome data is in following format.

GeneA	GeneB	UniprotA	UniprotB	DataSource
BST1	BST1	Q10588	Q10588	BioGrid_SGI
SNCAIP	PRKN	Q9Y6H5	O60260	BioGrid_SGI
CYP2D6	CYP2D6	P10635	P10635	BioGrid_SGI
CYP2D6	CYP2D6	P10635	P10635	BioGrid_SGI
NQO1	NQO1	P15559	P15559	BioGrid_SGI
DRD2	WDR45	P14416	Q9Y484	BioGrid_SGI
TOR1A	TOR1A	O14656	O14656	BioGrid_SGI
TOR1A	TOR1A	O14656	O14656	BioGrid_SGI
TOR1A	SNCA	O14656	P37840	BioGrid_SGI

**Table 5: SGI Dataset**

**Fig 1: SGI Graph**



## 1.4. Enrichment Analysis

## a. Seed Genes

## Gene Ontologies

## GO Biological Process 2018

Term	Overlap	P-value	Adjusted P-value	Old P-value	Old Adjusted P-value	Odds Ratio	Combined Score	Genes	Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11	Column12	Column13	
negative regulation of neuron death (GO:1901215)	17/83	1.36E-22	6.94E-19	0	0	37.581519	1892.184077	PRKN,BDNF,LRRK2,GBA,HTRA2,PARK7,NGF,SLC30A10,SOD2,SOD1,PINK1,GDNF,VP35,ATP13A2,FBXO7,SNCA,EIF4G1														
negative regulation of apoptotic process (GO:0043066)	23/485	1.41E-15	3.59494E-12	0	0	8.7014092	297.5527827	HSPA9,PRKN,CSF1R,GPX1,BDNF,HGF,INSR,GSTP1,IGF2,PARK7,SLC30A10,NGF,SOD2,IGF1R,INS,SOD1,NRAA2,IL6,PINK1,GDNF,EPHB4,SNCA,HSPA1A														
negative regulation of oxidative stress-induced cell death (GO:1903202)	44/98	3.70E-15	6.29E-12	0	0	68.807339	2286.553715	PRKN,PINK1,BAG5,LRRK2,HTRA2,PARK7,SOD2,FBXO7,INS														
regulation of neuron death (GO:1901214)	26/634	4.31E-15	5.50E-12	0	0	30.58104	1011.542198	PRKN,LRRK2,DDIT4,GBA,HTRA2,VP35,ATP13A2,PARK7,MAPT,MAP3K5,SNCA,EIF4G1														
cellular response to oxidative stress (GO:0034599)	13/115	5.52E-14	5.63888E-11	0	0	20.741923	633.2049418	GPX1,LRRK2,GSTP1,HTRA2,PARK7,SOD2,SOD1,PINK1,HMOX1,ATP13A2,MAPT,SNCA,HSPA1A														
dopamine metabolic process (GO:0042417)	44/63	7.15E-14	6.08E-11	0	0	73.394495	2221.603331	PRKN,GC11,MAOB,MAOA,DRD1,SNCA,IP,DRD2,SNCA														
negative regulation of cell death (GO:0060548)	29/891	6.63E-13	4.83E-10	0	0	24.917884	698.7475804	PRKN,PINK1,LRRK2,GBA,HTRA2,VP35,ATP13A2,PARK7,SNCA,HSPA1A,EIF4G1														
negative regulation of oxidative stress-induced intrinsic apoptotic signaling pathway (GO:1902176)	18/101	3.56E-12	2.27E-09	0	0	71.355759	1881.087664	PRKN,PINK1,BAG5,HTRA2,PARK7,SOD2,INS														
regulation of autophagy (GO:0010506)	14/203	5.40E-12	3.06E-09	0	0	12.654223	328.309954	PRKN,HGF,LRRK2,GBA,HTRA2,RAB39B,C9ORF72,GFAP,UCHL1,PINK1,VP35,ATP13A2,MAPT,EIF4G1														
negative regulation of intrinsic apoptotic signaling pathway (GO:2001243)	1-Sep	4.83E-11	2.47E-08	0	0	26.635099	632.6596164	PRKN,PINK1,BAG5,LRRK2,HTRA2,PARK7,SOD2,INS,HSPA1A														

negative regulation of neuron death (GO:1901215)

negative regulation of apoptotic process (GO:0043066)

negative regulation of oxidative stress-induced cell death (GO:1903202)

regulation of neuron death (GO:1901214)

cellular response to oxidative stress (GO:0034599)

dopamine metabolic process (GO:0042417)

negative regulation of cell death (GO:0060548)

negative regulation of oxidative stress-induced intrinsic apoptotic signaling pathway (GO:1902176)

regulation of autophagy (GO:0010506)

negative regulation of intrinsic apoptotic signaling pathway (GO:2001243)

Bar Graph sorted by p-value ranking.

## GO Molecular Function 2018

Term	Overlap	P-value	Adjusted P-value	Old P-value	Old Adjusted P-value	Odds Ratio	Combined Score	Genes	Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9
growth factor activity (GO:0008083)	Jun-69	2.07E-06	0.002381189	0	0	15.95532509	208.8319391	GRN,IL6,BDNF,HGF,IGF2,NGF									
protein homodimerization activity (GO:0042803)	14/664	1.51E-05	0.008668816	0	0	3.868685752	42.95503263	CSF1R,GSTM1,GCH1,LRRK2,PARK7,THAP1,SOD1,NRAA2,GDNF,GSTA4,PRKRA,HMOX1,MAPT,MAP3K5									
protein kinase binding (GO:0019901)	12/495	1.65E-05	0.006315773	0	0	4.448151237	48.99406922	PRKN,CSF1R,PINK1,BAG5,GCH1,DCTN1,INSR,GSTP1,MAPT,FBXO7,EPHB4,IGF1R									
transmembrane receptor protein tyrosine kinase activity (GO:0004714)	May-61	2.05E-05	0.005886287	0	0	15.03985562	162.388583	CSF1R,INSR,EPHB4,IGF2R,IGF1R									
ubiquitin protein ligase binding (GO:0031625)	9/284	2.55E-05	0.005874922	0	0	5.814704742	61.4963814	HSPA9,PRKN,ATXN3,UCHL1,PINK1,BAG5,SNCAIP,FBXO7,HSPA1A									
ubiquitin-like protein ligase binding (GO:0044389)	9/297	3.62E-05	0.006946911	0	0	5.560189046	56.8589666	HSPA9,PRKN,ATXN3,UCHL1,PINK1,BAG5,SNCAIP,FBXO7,HSPA1A									
purine ribonucleoside triphosphate binding (GO:0035639)	10/396	6.17E-05	0.010152001	0	0	4.633490872	44.91039146	HSPA9,PINK1,GCH1,INSR,LRRK2,ATP1A3,TUBB4A,MAP3K5,EIF4G1,HSPA1A									
MAP kinase kinase activity (GO:0004709)	May-78	6.75E-05	0.009704615	0	0	11.76193837	112.9628157	CSF1R,INSR,EPHB4,IGF1R,MAP3K5									
insulin-like growth factor receptor binding (GO:0005159)	15-Mar	6.83E-05	0.00873574	0	0	36.69724711	351.9814482	INSR,IGF2,INS									
copper ion binding (GO:0005507)	Apr-41	7.24E-05	0.008329293	0	0	17.90109644	170.6650795	ATP13A2,PARK7,SNCA,SOD1									

growth factor activity (GO:0008083)

protein homodimerization activity (GO:0042803)

protein kinase binding (GO:0019901)

transmembrane receptor protein tyrosine kinase activity (GO:0004714)

ubiquitin protein ligase binding (GO:0031625)

ubiquitin-like protein ligase binding (GO:0044389)

purine ribonucleoside triphosphate binding (GO:0035639)

MAP kinase kinase activity (GO:0004709)

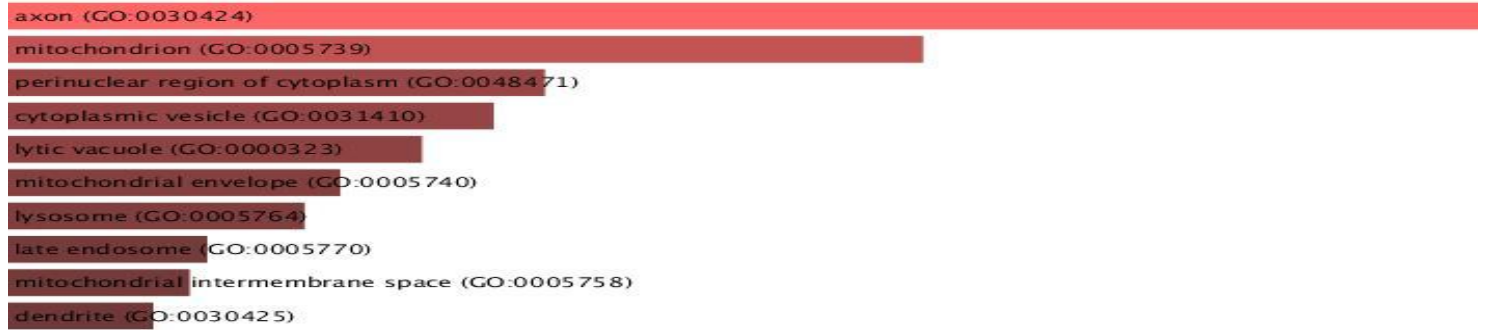
insulin-like growth factor receptor binding (GO:0005159)

copper ion binding (GO:0005507)

Bar Graph sorted by p-value ranking.

## GO Cellular Component 2018

Term	Overlap	P-value	Adjusted P-value	Old P-value	Old Adjusted P-value	Odds Ratio	Combined Score	Genes	Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10	Column11	Column12	Column13	Column14	Column15	Column16
axon (GO:0030424)	13/141	7.95E-13	3.55E-10	0	0	16.91717093	471.3211129	BDNF,INSR,LRRK2,ATP1A3,HTT,PARK7,NGF,SLC6A3,GF1R,PINK1,MAPT,DRD2,SNCA																
mitochondrion (GO:0005739)	24/1026	1.07E-09	2.39E-07	0	0	4.292075755	88.65405343	HSPA9,PRKN,PANK2,GPX1,MAOB,MAOA,BDNF,LRRK2,GSTP1,HTRA2,PARK7,SOD2,SOD1,CHCHD2,PINK1,BAG5,CYP2D6,OPA3,VP35,C19ORF12,MAPT,FBN1,SNCA,HSPA1A																
perinuclear region of cytoplasm (GO:0048471)	13/378	1.45E-07	2.16E-05	0	0	6.310373283	99.35018875	PRKN,BDNF,HFE,HTT,PARK7,AIF1,ATXN2,PINK1,BAG5,SYN1,HMOX1,SNCA,HSPA1A																
cytoplasmic vesicle (GO:0031410)	10/215	2.82E-07	3.15E-05	0	0	8.534243653	128.7028431	PRKN,GCH1,BDNF,HFE,LRRK2,VP35,NGF,SPG11,IGF1R,SOD1																
lytic vacuole (GO:0000323)	9/283	7.17E-07	6.40E-05	0	0	9.02391337	127.670416	GRN,LRRK2,GBA,HLA-DRA,VP35,ATP13A2,C9ORF72,GFAP,SNCA																
mitochondrial envelope (GO:0005740)	Jun-69	2.07E-06	1.54E-04	0	0	15.95532509	208.8319391	PANK2,CHCHD2,PINK1,MAOB,HTRA2,SOD1																
lysosome (GO:0005764)	12/422	3.30E-06	2.10E-04	0	0	5.217618157	65.8606128	HLA-DQB5,GRN,LRRK2,GBA,HLA-DRA,VP35,ATP13A2,C9ORF72,CP,SPG11,GFAP,SNCA																
late endosome (GO:0005770)	7/140	1.16E-05	6.49E-04	0	0	9.174311927	104.2249349	HLA-DQB5,HLA-DRA,HTT,VP35,ATP13A2,IGF1R,TMEM230																
mitochondrial intermembrane space (GO:0005758)	May-57	1.46E-05	7.26E-04	0	0	16.09528408	179.1592994	PANK2,CHCHD2,PINK1,HTRA2,SOD1																
dendrite (GO:0030425)	8/215	2.35E-05	0.001046276	0	0	6.827394922	72.78175134	SGCE,BDNF,INSR,LRRK2,HTT,MAPT,DRD2,NGF																



Bar Graph sorted by p-value ranking.

## Pathways

### KEGG 2019 Human

Term	Overlap	P-value	Adjusted P-value	Old P-value	Old Adjusted P-value	Odds Ratio	Combined Score	Genes	Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8
Parkinson disease	13/142	8.71E-13	2.68E-10	0	0	16.79803592	466.4643592	PRKN,LRRK2,HTRA2,PARK7,SNCA,IP,SLC6A3,UCHL1,PINK1,GNAL,DRD1,DRD2,SLC18A2,SNCA								
Cocaine addiction	Aug-49	2.25E-10	3.46E-08	0	0	29.9569369	665.5381975	DDC,MAOB,MAOA,BDNF,DRD1,DRD2,SLC6A3,SLC18A2								
MAPK signaling pathway	12/295	7.43E-08	7.63E-06	0	0	7.463846991	122.5153373	CSF1R,BDNF,HGF,INSR,IGF1,MAPT,NGF,TNF,IGF1R,MAP3K5,INS,HSPA1A								
Dopaminergic synapse	8/131	5.96E-07	4.59E-05	0	0	11.20526648	160.6040545	DDC,GNAL,MAOB,MAOA,DRD1,DRD2,SLC6A3,SLC18A2								
Fluid shear stress and atherosclerosis	8/139	9.36E-07	5.76E-05	0	0	10.56035905	146.597332	NQO1,EDN1,GSTM1,GSTA4,GSTP1,HMOX1,TNF,MAP3K5								
Amphetamine addiction	Jun-68	1.90E-06	9.74E-05	0	0	16.18996222	213.3044267	DDC,MAOB,MAOA,DRD1,SLC6A3,SLC18A2								
Drug metabolism	7/108	2.09E-06	9.18E-05	0	0	11.89262657	155.5621176	GSTM1,MAOB,GSTA4,CYP2D6,MAOA,GSTP1,CYP2E1								
Hepatocellular carcinoma	8/168	3.88E-06	1.49E-04	0	0	8.73743993	108.8701507	NQO1,GSTM1,GSTA4,HGF,GSTP1,IGF2,HMOX1,IGF1R								
Ras signaling pathway	9/232	5.08E-06	1.74E-04	0	0	7.11800633	86.77482407	CSF1R,BDNF,HGF,INSR,IGF2,PLA2G6,NGF,IGF1R,INS								
Alcoholism	8/180	6.45E-06	1.99E-04	0	0	8.154943935	97.4570171	DDC,MAOB,MAOA,BDNF,DRD1,DRD2,SLC6A3,SLC18A2								



KEGG 2019 Human bar graph sorted by P-Value Ranking



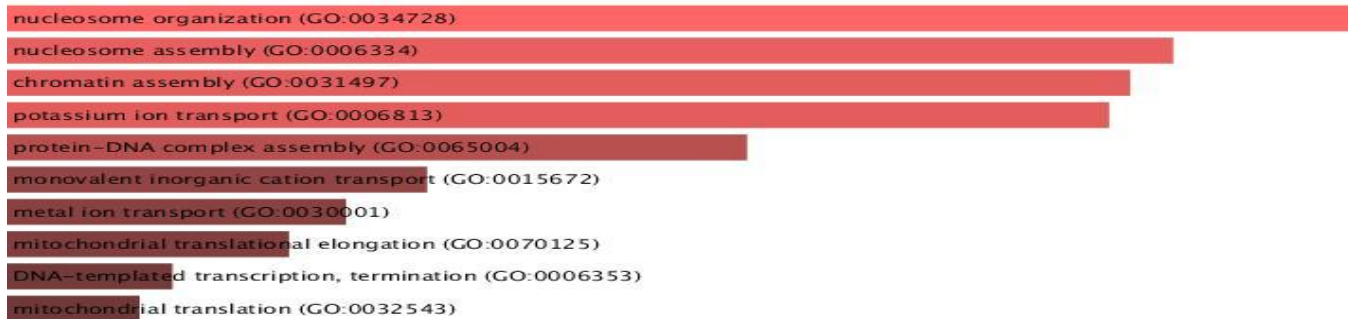
## b. Union Interactome Genes

We have 28880 Union Interactome genes so first we use python to remove duplicates and find all the unique union Interactome genes that came out out to be 2498 and then used EnrichR to perform enrichment analysis on them.

### Gene Ontologies

#### GO Biological Process 2018

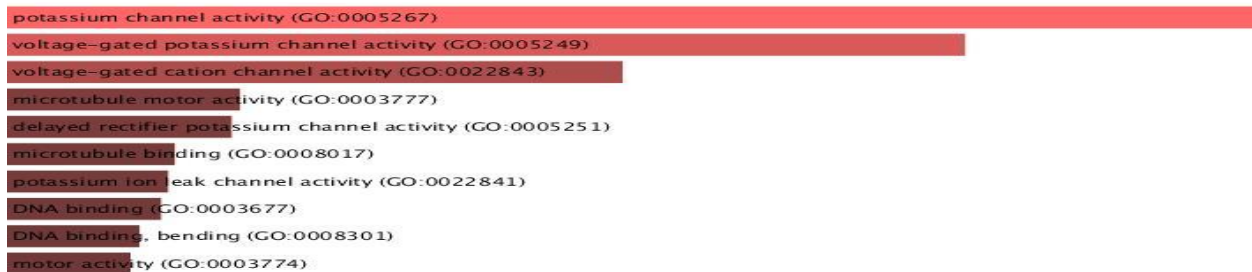
Term	Overlap	P-value	Adjusted P-value	Old P-value	Old Adjusted P-value	Odds Ratio	Combined Score	Genes	Column1	Column2	Column3	Column4	Column5	Column6	Column7	Column8	Column9	Column10
nucleosome organization (GO:0034728)	28/79	1.31E-07	6.68E-04	0	0	2.837713209	44.97541199	HIST1H2BN;SMARCD1;HIST1H2BM;HIST1H2BO;HIST1H2BL;HMG82;RSF1;SUPT6H;VPS72;HIST3H2BB;SOX9;BRD2;NFE2;PBRM1;H3F3C;ATRX;H										
nucleosome assembly (GO:0006334)	23/60	3.38E-07	8.62E-04	0	0	3.069121964	45.73199316	HIST1H2BN;BRD2;HIST1H2BM;HIST1H2BO;H3F3C;HIST1H2BL;ATRX;H3F3A;HMG82;RSF1;H2BFS;HIST2H2BF;HIST2H2BE;GRWD1;ANP32E;HIST										
chromatin assembly (GO:0031497)	24/65	4.27E-07	7.26E-04	0	0	2.956211123	43.35845525	HIST1H2BN;BRD2;HIST1H2BM;HIST1H2BO;H3F3C;HIST1H2BL;ATRX;H3F3A;HMG82;RSF1;H2BFS;HIST2H2BF;HIST2H2BE;CDAN1;GRWD1;ANP3										
potassium ion transport (GO:0006813)	23/61	4.79E-07	6.10E-04	0	0	3.018808489	43.93149465	KCNG1;KCNK6;KCNK2;KCNH3;KCNF1;KCNK7;KCNH4;KCNJ10;KCNK1;KCNK9;KCNJ8;KCNJ15;KCNJ16;KCNK16;KCNK17;KCNV1;KCNK1;KCNK2;KC										
protein-DNA complex assembly (GO:0065004)	28/91	3.41E-06	0.003484967	0	0	2.463509269	31.00928183	HIST1H2BN;HIST1H2BM;HIST1H2BO;HIST1H2BL;HMG82;RSF1;XPC;HIST3H2BB;SOX9;BRD2;CENPV;H3F3C;ATRX;H3F3A;H2BFS;GTF2H2;GTF2H										
monovalent inorganic cation transport (GO:0015629/104)	1.94E-05	0.016486507	0	0	2.232555275	24.22554074	KCNG1;KCNK2;KCNK6;KCNK7;KCNK1;KCNK9;SLC23A1;KCNV1;KCNMB2;KCNMB3;KCNK2;ASIC2;SLC36A1;NDUFA9;KCNH3;KCNF1;KCNH4;KCN											
metal ion transport (GO:0030001)	28/101	3.01E-05	0.021931257	0	0	2.21959746	23.1093748	KCNG1;KCNK2;KCNK6;KCNK7;KCNK1;KCNK9;SLC23A1;KCNV1;KCNMB2;KCNMB3;KCNK2;NDUFA9;KCNH3;KCNF1;KCNH4;KCNJ10;KCNJ8;SLC1										
mitochondrial translational elongation (GO:0070125/87)	4.10E-05	0.026151613	0	0	2.300691128	23.24155125	GFM1;MRPS36;MRPL18;MRPS10;MRPL36;MRPL57;MRPL32;MRPL10;MRPL42;MRPL20;MRPL40;MRPL1;MRPS28;GADD45GIP1;MRPS25;MRPL											
DNA-templated transcription, termination (GO:001281/106)	7.72E-05	0.043767412	0	0	2.114899467	20.02645901	SLBP;CSTF3;SRSF1;TTF2;TTF1;MED18;ZNRD1;PABPN1;PAPOLA;CSTF1;SYMPK;MTERF1;CAVIN1;CPSF3;CPSF2;THOC1;UPF38;THOC3;GTF2H2;G											



GO Biological Process 2018 bar graph sorted by P-Value Ranking

#### GO Molecular Function 2018

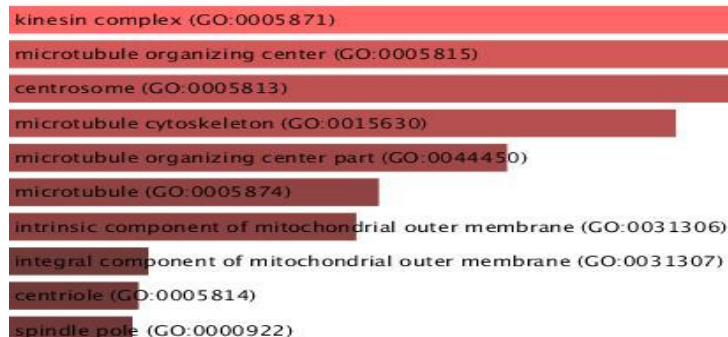
Term	Overlap	P-value	Adjusted P-value	Old P-value	Old Adjusted P-value	Odds Ratio	Combined Score	Genes	Column1	Column2	Column3
growth factor activity (GO:0008083)	Apr-69	7.86E-06	0.009048888	0	0	31.33568351	368.30395	GRN;BDNF;HGF;NGF			
MAP kinase kinase activity (GO:0004709)	Mar-78	4.03E-04	0.231918794	0	0	20.79002079	162.5074267	CSF1R;INSR;MAP3K5			
purine ribonucleoside triphosphate binding (GO:0035639)	5/396	7.68E-04	0.294525591	0	0	6.825006825	48.95006485	GCH1;INSR;ATP1A3;EIF4G1;MAP3K5			
transition metal ion binding (GO:0046914)	5/399	7.94E-04	0.228515781	0	0	6.773690984	48.35224585	PRKN;GCH1;SOD2;SOD1;SNCA			
Hsp70 protein binding (GO:0030544)	27-Feb	0.001135283	0.261342229	0	0	40.04004004	271.5064267	PRKN;SNCA			
ATP binding (GO:0005524)	4/255	0.001224613	0.234921648	0	0	8.479067303	56.85324999	INSR;ATP1A3;EIF4G1;MAP3K5			
adenyl ribonucleotide binding (GO:0032559)	4/279	0.001703855	0.280162483	0	0	7.749685169	49.40317199	INSR;ATP1A3;EIF4G1;MAP3K5			
protein kinase binding (GO:0019901)	5/495	0.002063109	0.296829821	0	0	5.46000546	33.76216852	PRKN;CSF1R;GCH1;INSR;GSTP1			
zinc ion binding (GO:0008270)	4/302	0.002273155	0.29071122	0	0	7.159477358	43.57677971	PRKN;GCH1;SOD1;SNCA			



GO Molecular Function 2018 bar graph sorted by P-Value Ranking

## GO Cellular Component 2018

Term	Overlap	P-value	Adjusted P-value	Old P-value	Old Adjusted P-value	Odds Ratio	Combined Score	Genes	Column1	Column2	Column3	Column4
axon (GO:0030424)	7/141	6.43E-09	2.87E-06	0	0	26.83534598	506.1845305	BDNF;INSR;HTT;ATP1A3;NGF;SLC6A3;SNCA				
cytosolic ribosome (GO:0022626)	5/124	3.14E-06	7.00E-04	0	0	21.79598954	276.1755995	RP58;RPL14;RPL23A;RPL6;EIF4G1				
lysosome (GO:0005764)	7/422	1.06E-05	0.001569033	0	0	8.966312284	102.7449924	HLA-DRB5;GRN;GBA;HLA-DRA;CP;GFAP;SNCA				
cytosolic part (GO:0044445)	5/159	1.06E-05	0.001179843	0	0	16.99813021	194.7373549	RP58;RPL14;RPL23A;RPL6;EIF4G1				
lytic vacuole (GO:0000323)	5/183	2.09E-05	0.001861725	0	0	14.76886723	159.166042	GRN;GBA;HLA-DRA;GFAP;SNCA				
cytoplasmic vesicle (GO:0031410)	5/215	4.51E-05	0.003355079	0	0	12.57071025	125.7805082	PRKN;GCH1;BDNF;NGF;SOD1				
protein kinase complex (GO:1902911)	6-Feb	4.97E-05	0.003167713	0	0	180.1801802	1785.433183	INSR;MAP3K5				
cytosolic large ribosomal subunit (GO:0022625)	Mar-69	2.81E-04	0.015650904	0	0	23.50176263	192.1998586	RPL14;RPL23A;RPL6				
MHC class II protein complex (GO:0042613)	14-Feb	2.99E-04	0.014808301	0	0	77.22007722	626.6920272	HLA-DRB5;HLA-DRA				

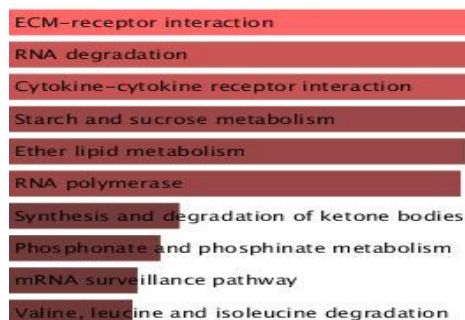


GO Cellular Component bar graph sorted by P-Value Ranking

## Pathways

## KEGG 2019 Human

Term	Overlap	P-value	Adjusted P-value	Old P-value	Old Adjusted P-value	Odds Ratio	Combined Score	Genes	Column1	Column2	Column3
MAPK signaling pathway	7/295	9.97E-07	3.07E-04	0	0	12.82638571	177.2447597	CSF1R;BDNF;INSR;HGF;NGF;TNF;MAP3K5			
Huntington disease	6/193	1.35E-06	2.09E-04	0	0	16.80436914	227.0689546	GPX1;TBP;BDNF;HTT;SOD2;SOD1			
Amyotrophic lateral sclerosis (ALS)	Apr-51	2.33E-06	2.39E-04	0	0	42.39533651	549.9032843	GPX1;TNF;SOD1;MAP3K5			
Fluid shear stress and atherosclerosis	5/139	5.50E-06	4.23E-04	0	0	19.44390434	235.4868527	NQO1;EDN1;GSTP1;TNF;MAP3K5			
Asthma	31-Mar	2.53E-05	0.001557139	0	0	52.31037489	553.7349885	HLA-DRB5;HLA-DRA;TNF			
Hematopoietic cell lineage	Apr-97	3.04E-05	0.00155882	0	0	22.29033157	231.8677421	CSF1R;HLA-DRB5;HLA-DRA;TNF			
Allograft rejection	Mar-38	4.70E-05	0.002068879	0	0	42.6742532	425.2462871	HLA-DRB5;HLA-DRA;TNF			
Graft-versus-host disease	Mar-41	5.92E-05	0.002278793	0	0	39.55174687	385.0270567	HLA-DRB5;HLA-DRA;TNF			
Ras signaling pathway	5/232	6.48E-05	0.002216753	0	0	11.64958062	112.3553944	CSF1R;BDNF;INSR;HGF;NGF			
Type I diabetes mellitus	Mar-43	6.83E-05	0.002105146	0	0	37.71213074	361.6927084	HLA-DRB5;HLA-DRA;TNF			



KEGG Human bar graph sorted by P-Value Ranking



**PART – II (DATA ANALYSIS)**

A table with Local measures of SGI, UI, II

	betweenness centrality	degree centrality	closeness centrality	eigenvalues centrality	betweennessDxegreeCentarilityRatio
<b>O60260</b>	1.000000	1.000000	1.000000	1.000000	0.968860
<b>Q99497</b>	0.682791	0.473684	0.886792	0.459067	1.396561
<b>P00441</b>	0.243808	0.368421	0.643836	0.189366	0.641156
<b>P37840</b>	0.239992	0.578947	0.839286	0.684375	0.401623
<b>P37837</b>	0.209532	0.105263	0.610390	0.079568	1.928571
<b>Q99683</b>	0.118735	0.263158	0.610390	0.161200	0.437143
<b>Q9Y484</b>	0.108646	0.105263	0.456311	0.013780	1.000000
<b>P0DMV8</b>	0.067257	0.157895	0.681159	0.310564	0.412698
<b>P38646</b>	0.051801	0.105263	0.681159	0.200141	0.476786

Table 8: SGI (Seed Gene Interactome)

	betweenness centrality	degree centrality	closeness centrality	eigenvalues centrality	betweennessDxegreeCentarilityRatio
<b>P37840</b>	0.035846	0.213483	0.558282	0.345049	0.143623
<b>O60260</b>	0.033785	0.303371	0.587097	0.460932	0.095256
<b>Q99683</b>	0.011763	0.202247	0.551515	0.331685	0.049750
<b>P10636</b>	0.005628	0.157303	0.541667	0.287743	0.030605
<b>P42858</b>	0.005530	0.146067	0.538462	0.252081	0.032383
<b>Q99497</b>	0.004982	0.134831	0.535294	0.240944	0.031603
<b>P04350</b>	0.004369	0.134831	0.538462	0.253306	0.027715
<b>Q9UL15</b>	0.004209	0.146067	0.538462	0.279073	0.024649
<b>P20226</b>	0.003355	0.123596	0.529070	0.193145	0.023221

Table 9: UI (Union Interactome)

	betweenness centrality	degree centrality	closeness centrality	eigenvalues centrality	betweennessDxegreeCentarilityRatio
<b>P37840</b>	0.036699	0.220930	0.564417	0.354113	0.144059
<b>O60260</b>	0.034876	0.313953	0.593548	0.473320	0.096338
<b>Q99683</b>	0.014702	0.209302	0.564417	0.337932	0.060920
<b>O14976</b>	0.010053	0.058140	0.525714	0.116122	0.149952
<b>P09211</b>	0.009908	0.093023	0.528736	0.146683	0.092368
<b>P21397</b>	0.009118	0.069767	0.522727	0.117836	0.113338
<b>P09488</b>	0.008816	0.081395	0.525714	0.131032	0.093930
<b>P05181</b>	0.008372	0.058140	0.525714	0.104580	0.124876
<b>P27338</b>	0.007645	0.058140	0.519774	0.092517	0.114043
<b>P10636</b>	0.005810	0.162791	0.547619	0.294857	0.030952

Table 10: II (Intersection Interactome)

- A figure of the SGI and of the I-LCC networks

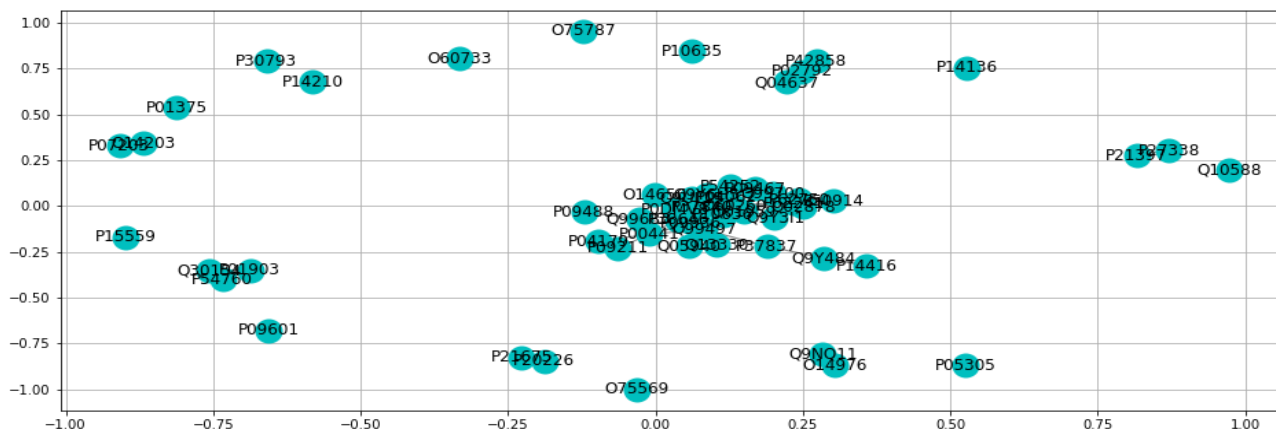


Fig 4. SGI Seed Genes Interactions graph

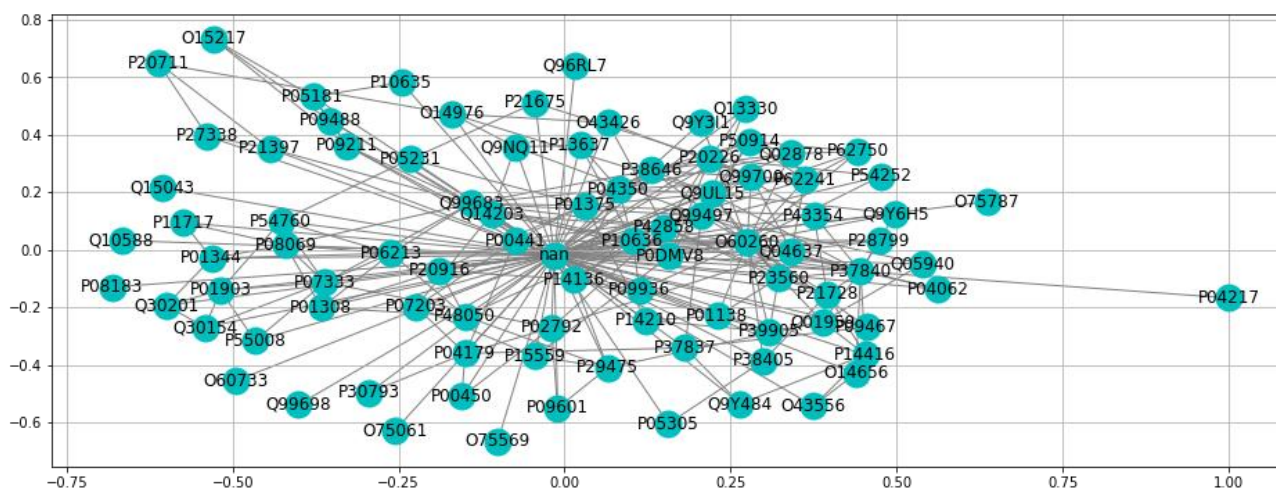


Fig 5. I-LCC Network graph

- A summary table of the putative disease modules found with MCL for Union interactome and intersection interactome genes is given below.

	Algorithm	index_module	n_seed_genes	n_genes	seedToGene_Ratio	p_value
0	markov_chain	0	30	90	0.333333	0.0

Fig 6: MCL for Intersection interactome gene

	Algorithm	index_module	n_seed_genes	n_genes	seedToGene_Ratio	p_value
0	markov_chain	0	30	91	0.32967	0.0

Fig 7: MCL for union interactome gene

### a. Putative disease proteins using the DIAMOND

For finding putative disease proteins for our seed gene list we used DIseAse MOdule Detection (DIAMOND) algorithm which is derived from a systematic analysis of connectivity patterns of disease proteins in the Human Interactome. For doing this analysis we provided the Diamond python script as input 2 txt files names seed gene list and protein protein interaction list from Bio Grid DB (which we obtained from previous analysis). The files are present in the folder data/diamond folder with names PPI.txt and genes.txt. We restricted our putative protein list to 200. So, as a 3rd input to the Diamond script we gave 200 and we got the top 200 putative protein list which we saved in data/diamond/first\_200\_nodes.txt file. We applied following command for getting top 200 nodes.

```
python ./DIAMOND.py BioGridPPI.txt SeedGene.txt 200
```

From the gene ids of top 200 nodes we performed overrepresented pathway analysis and overrepresented GO analysis manually from Enrichr website [6] and we included the results in Data/Enrichr\_Diamond folder.

## 1 Enrichment Analysis for Putative disease proteins using the DIAMOND

### 3.1 Gene Ontologies

#### GO Biological Process 2018

Term	Overlap	P-value	Adjusted P-value	Old P-value	Old Adjusted P-value	Odds Ratio	Combined Score	Genes	Column1	Column2	Column3	Column4	Column5	Column6
negative regulation of neuron death (GO:1901215)	Aug-83	2.18E-12	1.11E-08	0	0	52.10029306	1398.997857	PRKN;BDNF;GBA;NGF;SOD2;SOD1;EIF4G1;SNCA						
negative regulation of apoptotic process (GO:0043066)	12/485	3.86E-11	9.85E-08	0	0	13.37419894	320.6861355	NR4A2;PRKN;CSF1R;GPX1;BDNF;INSR;HGF;GSTP1;NGF;SOD2;SOD1;SNCA						
negative regulation of apoptotic signaling pathway (GO:2001234)	Jun-65	2.00E-09	3.40E-06	0	0	49.8960499	999.4662755	NR4A2;PRKN;GPX1;HGF;GSTP1;HTT						
regulation of cysteine-type endopeptidase activity involved in apoptotic process (GO:0043281)	Jun-93	1.78E-08	2.27E-05	0	0	34.87358326	622.3517543	GPX1;HGF;NGF;TNF;MAP3K5;SNCA						
positive regulation of cellular component organization (GO:0051130)	6/106	3.91E-08	3.99E-05	0	0	30.59663437	521.9068184	GPX1;BDNF;INSR;GBA;TNF;SNCA						
negative regulation of neuron apoptotic process (GO:0043524)	May-63	1.06E-07	9.05E-05	0	0	42.9000429	688.8155233	PRKN;BDNF;NGF;SOD2;SOD1						
regulation of apoptotic process (GO:0042981)	11/815	1.56E-07	1.13E-04	0	0	7.295639197	114.3641363	PRKN;CSF1R;INSR;HGF;GSTP1;HTT;NGF;TNF;SOD1;MAP3K5;SNCA						
regulation of neuron death (GO:1901214)	May-72	2.09E-07	1.33E-04	0	0	37.53753754	577.3222043	PRKN;GBA;EIF4G1;MAP3K5;SNCA						
regulation of JUN kinase activity (GO:0043506)	Apr-34	4.42E-07	2.50E-04	0	0	63.59300477	930.537146	EDN1;GSTP1;TNF;MAP3K5						
regulation of MAP kinase activity (GO:0043405)	May-87	5.42E-07	2.76E-04	0	0	31.06554831	448.2375281	EDN1;UCHL1;GSTP1;GBA;TNF						

negative regulation of neuron death (GO:1901215)

negative regulation of apoptotic process (GO:0043066)

negative regulation of apoptotic signaling pathway (GO:2001234)

regulation of cysteine-type endopeptidase activity involved in apoptotic process (GO:0043281)

positive regulation of cellular component organization (GO:0051130)

negative regulation of neuron apoptotic process (GO:0043524)

regulation of apoptotic process (GO:0042981)

regulation of neuron death (GO:1901214)

regulation of JUN kinase activity (GO:0043506)

regulation of MAP kinase activity (GO:0043405)

#### GO Molecular Function 2018

Term	Overlap	P-value	Adjusted P-value	Old P-value	Old Adjusted P-value	Odds Ratio	Combined Score	Genes	Column1	Column2	Column3
growth factor activity (GO:008083)	Apr-69	7.86E-06	0.009048888	0	0	31.33568351	368.30395	GRN;BDNF;HGF;NGF			
MAP kinase kinase activity (GO:0004709)	Mar-78	4.03E-04	0.231918794	0	0	20.79002079	162.5074267	CSF1R;INSR;MAP3K5			
purine ribonucleoside triphosphate binding (GO:0035639)	5/396	7.68E-04	0.294525591	0	0	6.825006825	48.95006485	GCH1;INSR;ATP1A3;EIF4G1;MAP3K5			
transition metal ion binding (GO:0046914)	5/399	7.94E-04	0.228515781	0	0	6.773690984	48.35224585	PRKN;GCH1;SOD2;SOD1;SNCA			
Hsp70 protein binding (GO:0030544)	27-Feb	0.001135	0.261342229	0	0	40.04004004	271.5064267	PRKN;SNCA			
ATP binding (GO:0005524)	4/255	0.001225	0.234921648	0	0	8.479067303	56.85324999	INSR;ATP1A3;EIF4G1;MAP3K5			
adenyl ribonucleotide binding (GO:0032559)	4/279	0.001704	0.280162483	0	0	7.749685169	49.40317199	INSR;ATP1A3;EIF4G1;MAP3K5			
protein kinase binding (GO:0019901)	5/495	0.002063	0.296829821	0	0	5.46000546	33.76216852	PRKN;CSF1R;GCH1;INSR;GSTP1			
zinc ion binding (GO:0008270)	4/302	0.002273	0.29071122	0	0	7.159477358	43.57677971	PRKN;GCH1;SOD1;SNCA			
protein tyrosine kinase activity (GO:0004713)	3/147	0.002517	0.289714502	0	0	11.0314396	66.0194223	CSF1R;INSR;HGF			

## Group no. 12, Topic (Parkinson Disease), Tahreem, Chandraprakash

growth factor activity (GO:0008083)  
 MAP kinase kinase kinase activity (GO:0004709)  
 purine ribonucleoside triphosphate binding (GO:0035639)  
 transition metal ion binding (GO:0046914)  
 Hsp70 protein binding (GO:0030544)  
 ATP binding (GO:0005524)  
 adenylyl ribonucleotide binding (GO:0032559)  
 protein kinase binding (GO:0019901)  
 zinc ion binding (GO:0008270)  
 protein tyrosine kinase activity (GO:0004713)

## GO Cellular Component 2018

Term	Overlap	P-value	Adjusted P-value	Old P-value	Old Adjusted P-value	Odds Ratio	Combined Score	Genes	Column1	Column2	Column3	Column4
axon (GO:0030424)	7/141	6.43E-09	2.87E-06	0	0	26.83534598	506.1845305	BDNF;INSR;HTT;ATP1A3;NGF;SLC6A3;SNCA				
cytosolic ribosome (GO:0022626)	5/124	3.14E-06	7.00E-04	0	0	21.79598954	276.1755995	RPS8;RPL14;RPL23A;RPL6;EIF4G1				
lysosome (GO:0005764)	7/422	1.06E-05	0.001569033	0	0	8.966312284	102.7449924	HLA-DRB5;GRN;GBA;HLA-DRA;CP;GFAP;SNCA				
cytosolic part (GO:0044445)	5/159	1.06E-05	0.001179843	0	0	16.99813021	194.7373549	RPS8;RPL14;RPL23A;RPL6;EIF4G1				
lytic vacuole (GO:0000323)	5/183	2.09E-05	0.001861725	0	0	14.76886723	159.166042	GRN;GBA;HLA-DRA;GFAP;SNCA				
cytoplasmic vesicle (GO:0031410)	5/215	4.51E-05	0.003355079	0	0	12.57071025	125.7805082	PRKN;GCH1;BDNF;NGF;SOD1				
protein kinase complex (GO:1902911)	6-Feb	4.97E-05	0.003167713	0	0	180.1801802	1785.433183	INSR;MAP3K5				
cytosolic large ribosomal subunit (GO:0022625)	Mar-69	2.81E-04	0.015650904	0	0	23.50176263	192.1998586	RPL14;RPL23A;RPL6				
MHC class II protein complex (GO:0042613)	14-Feb	2.99E-04	0.014808301	0	0	77.22007722	626.6920272	HLA-DRB5;HLA-DRA				
large ribosomal subunit (GO:0015934)	Mar-72	3.18E-04	0.014198047	0	0	22.5225252	181.3600166	RPL14;RPL23A;RPL6				

axon (GO:0030424)  
 cytosolic ribosome (GO:0022626)  
 lysosome (GO:0005764)  
 cytosolic part (GO:0044445)  
 lytic vacuole (GO:0000323)  
 cytoplasmic vesicle (GO:0031410)  
 protein kinase complex (GO:1902911)  
 cytosolic large ribosomal subunit (GO:0022625)  
 MHC class II protein complex (GO:0042613)  
 large ribosomal subunit (GO:0015934)

## 3.2 Pathways

## KEGG 2019 Human

Term	Overlap	P-value	Adjusted P-value	Old P-value	Old Adjusted P-value	Odds Ratio	Combined Score	Genes	Column1	Column2	Column3
MAPK signaling pathway	7/295	9.97E-07	3.07E-04	0	0	12.82638571	177.2447597	CSF1R;BDNF;INSR;HGF;NGF;TNF;MAP3K5			
Huntington disease	6/193	1.35E-06	2.09E-04	0	0	16.80436914	227.0689546	GPX1;TBP;BDNF;HTT;SOD2;SOD1			
Amyotrophic lateral sclerosis (ALS)	Apr-51	2.33E-06	2.39E-04	0	0	42.39533651	549.9032843	GPX1;TNF;SOD1;MAP3K5			
Fluid shear stress and atherosclerosis	5/139	5.50E-06	4.23E-04	0	0	19.44390434	235.4868527	NQO1;EDN1;GSTP1;TNF;MAP3K5			
Asthma	31-Mar	2.53E-05	0.001557139	0	0	52.31037489	553.7349885	HLA-DRB5;HLA-DRA;TNF			
Hematopoietic cell lineage	Apr-97	3.04E-05	0.00155882	0	0	22.29033157	231.8677421	CSF1R;HLA-DRB5;HLA-DRA;TNF			
Allograft rejection	Mar-38	4.70E-05	0.002068879	0	0	42.6742532	425.2462871	HLA-DRB5;HLA-DRA;TNF			
Graft-versus-host disease	Mar-41	5.92E-05	0.002278793	0	0	39.55174687	385.0270567	HLA-DRB5;HLA-DRA;TNF			
Ras signaling pathway	5/232	6.48E-05	0.002216753	0	0	11.64958062	112.3553944	CSF1R;BDNF;INSR;HGF;NGF			
Type I diabetes mellitus	Mar-43	6.83E-05	0.002105146	0	0	37.71213074	361.6927084	HLA-DRB5;HLA-DRA;TNF			

MAPK signaling pathway  
 Huntington disease  
 Amyotrophic lateral sclerosis (ALS)  
 Fluid shear stress and atherosclerosis  
 Asthma  
 Hematopoietic cell lineage  
 Allograft rejection  
 Graft-versus-host disease  
 Ras signaling pathway  
 Type I diabetes mellitus

Note: The first 30 genes coming from the DIAMOnD tool can be found in Appendix B, Table 3

## Notes and comments

Important notes regarding the entire project are presented below.

- We find 109 seed genes from DisGeNet database out of which only 90 were found in Uniprot.
- We couldn't extract the information regarding "type of Interaction" from Biogrid but IID has this information.
- Data from different sources created NAN and duplicate values in interaction data.
- There's a substantial difference between the number of interactions returned by "Bio Grid" and "IID".
- Some of the interactions are repeated more than once, as it is possible under the presence of different environment.
- Using different Gene Symbol may return different interactions.
- We also observed that the mapping between gene id, gene name and Uniprot id doesn't always result in the same output always. One common data set coupled with some standardization would be an ideal solution for these kinds of discrepancies.
- The Software and the relevant tables are present in the attachment folder.

## Appendix A:

20 highest ranking genes (for betweenness):

### 1. Largest Connected Component of Interaction Interactome (LCC I)

	betweenness centrality	degree centrality	closeness centrality	eigenvalues centrality	betweennessDxegreeCentralityRatio
P37840	0.036699	0.220930	0.564417	0.354113	0.144059
O60260	0.034876	0.313953	0.593548	0.473320	0.096338
Q99683	0.014702	0.209302	0.564417	0.337932	0.060920
O14976	0.010053	0.058140	0.525714	0.116122	0.149952
P09211	0.009908	0.093023	0.528736	0.146683	0.092368
P21397	0.009118	0.069767	0.522727	0.117836	0.113338
P09488	0.008816	0.081395	0.525714	0.131032	0.093930
P05181	0.008372	0.058140	0.525714	0.104580	0.124876
P27338	0.007645	0.058140	0.519774	0.092517	0.114043
P10636	0.005810	0.162791	0.547619	0.294857	0.030952
P42858	0.005708	0.151163	0.544379	0.257868	0.032751
Q99497	0.005143	0.139535	0.541176	0.246397	0.031962
Q9UL15	0.005046	0.151163	0.547619	0.285687	0.028950
P04350	0.004510	0.139535	0.544379	0.259192	0.028030
P20226	0.003464	0.127907	0.534884	0.197138	0.023485
P38646	0.002877	0.127907	0.541176	0.266297	0.019510
P14136	0.002727	0.127907	0.534884	0.224577	0.018492
P07333	0.002666	0.127907	0.534884	0.193759	0.018079
P00441	0.002651	0.116279	0.534884	0.208897	0.019773
P06213	0.002362	0.127907	0.534884	0.222009	0.016012

## 2. Largest Connected Component of Union Interactome (LCC U)

	betweenness centrality	degree centrality	closeness centrality	eigenvalues centrality	betweennessDxegreeCentarlityRatio
P37840	0.035846	0.213483	0.558282	0.345049	0.143623
O60260	0.033785	0.303371	0.587097	0.460932	0.095256
Q99683	0.011763	0.202247	0.551515	0.331685	0.049750
P10636	0.005628	0.157303	0.541667	0.287743	0.030605
P42858	0.005530	0.146067	0.538462	0.252081	0.032383
Q99497	0.004982	0.134831	0.535294	0.240944	0.031603
P04350	0.004369	0.134831	0.538462	0.253306	0.027715
Q9UL15	0.004209	0.146067	0.538462	0.279073	0.024649
P20226	0.003355	0.123596	0.529070	0.193145	0.023221
P38646	0.002787	0.123596	0.535294	0.260079	0.019291
P14136	0.002642	0.123596	0.529070	0.219748	0.018284
P07333	0.002583	0.123596	0.529070	0.190240	0.017875
P06213	0.002288	0.123596	0.529070	0.217399	0.015832
P01375	0.002175	0.112360	0.526012	0.201591	0.016554
P00441	0.002175	0.112360	0.526012	0.205061	0.016554
P02792	0.001968	0.101124	0.522989	0.154229	0.016646
P04179	0.001870	0.101124	0.522989	0.145350	0.015814
P0DMV8	0.001845	0.112360	0.529070	0.242511	0.014045
P21728	0.001806	0.101124	0.532164	0.217854	0.015273
Q01959	0.001550	0.101124	0.526012	0.195239	0.013109



## Appendix B:

Putative disease proteins using the DIAMOnD tool Top 30 nodes after applying DIAMOnD tool are given below.

Rank	Diamond Nodes
1	4914
2	1994
3	5478
4	7818
5	5885
6	5701
7	267
8	4343
9	9924
10	9531
11	3320
12	10273
13	3676
14	3308
15	10482
16	2335
17	7157
18	409
19	1956
20	9820
21	7337
22	9933
23	26986
24	5702
25	7412
26	16341
27	7186
28	149041
29	23435
30	2099