LabBook_18_08_2017

Claire Green

Similarity Network Fusion

Similarity Network Fusion (SNF) is a methodology whereby networks created from multiple datasets are fused into one network. The algorithm constructs what is called a similarity network for each of the data sets that and then these networks are then integrated into one single network using a nonlinear combination method. It can take any data type - continuous or binary - including data such as miRNA expression, RNA expression, and methylation data.

Link to paper - https://www.nature.com/nmeth/journal/v11/n3/full/nmeth.2810.html

In a quote from a paper by Gligorijevic, Malod-Dognin and Przulj of the Department of Computing, Imperial College London entitled "Integrative Methods for Analysing Big Data in Precision Medicine" (http://discovery.ucl.ac.uk/1519771/1/main.pdf) :

"To overcome scalability drawbacks of the previous ML clustering methods, that operate with high-dimensional gene × patient matrices, Wang et al. proposed a network-based method that integrates data represented by patient × patient matrices. This method, called Similarity Network Fusion (SNF), combines mRNA expression, DNA methylation and microRNA expression data for the same set of cancer patients. First, for each data type, it constructs a weighted network of patients, with nodes being patients and weighted links being similarities between patients. The similarities are computed based on their gene profiles for a particular data type. Second, it normalises weights of each network by taking into account the networks from all data types. Finally, it fuses all the networks into a single network by performing a diffusion of information within each network and across different networks. After the convergence of the diffusion process, the authors use a spectral clustering method on the final fused network to group patients into clusters. Unlike the previous methods, SNF is more scalable. Namely, instead of processing large-scale matrices constructed over a large number of genes, SNF method fuses much smaller matrices representing networks constructed over patients (i.e., samples), which makes the convergence faster. SNF is shown to be robust to noise and when applied on five different cancer types from TCGA database, it was shown to be effective in prediction of patient survival outcomes."

Script

```
##### SIMILARITY NETWORK FUSION #####

library(SNFtool)
library(gdata)

## First, set all the parameters:
K = 20;      # number of neighbors, usually (10~30)
alpha = 0.5;     # hyperparameter, usually (0.3~0.8)
T = 10;      # Number of Iterations, usually (10~20)

#Read in network nodes
DEG_PPI <- read.table("~/Documents/PhD/TDP-43/TDP-43_Code/Results/PPI_Network/DEG_PPI_Genes.txt")
DEG_PPI <- DEG_PPI$V1

#Extract PPI network genes from each dataset
```

```r
C9 <- read.csv("~/Documents/PhD/TDP-43/TDP-43_Code/Results/GeneExpression/noMedian/C9_unique.csv")
rownames(C9) <- C9$Gene.Symbol
C9 <- C9[,12:19]
C9 <- subset(C9, rownames(C9) %in% DEG_PPI)

VCP <- read.csv("~/Documents/PhD/TDP-43/TDP-43_Code/Results/GeneExpression/noMedian/VCP_unique.csv")
rownames(VCP) <- VCP$Gene.Symbol
VCP <- VCP[,12:18]
VCP <- subset(VCP, rownames(VCP) %in% DEG_PPI)

FTLD <- read.csv("~/Documents/PhD/TDP-43/TDP-43_Code/Results/GeneExpression/noMedian/ftld_unique.csv")
rownames(FTLD) <- FTLD$Gene.Symbol
FTLD <- FTLD[,17:32]
FTLD <- subset(FTLD, rownames(FTLD) %in% DEG_PPI)

sALS <- read.csv("~/Documents/PhD/TDP-43/TDP-43_Code/Results/GeneExpression/noMedian/sals_unique.csv")
rownames(sALS) <- sALS$Gene.Symbol
sALS <- sALS[,12:18]
sALS <- subset(sALS, rownames(sALS) %in% DEG_PPI)

# PET <- read.csv("~/Documents/PhD/TDP-43/TDP-43_Code/Results/GeneExpression/TDP-43_DEseq2/PET_results_
# rownames(PET) <- PET$hgnc_symbol
# PET <- PET[,19:35]
# PET <- subset(PET, rownames(PET) %in% DEG_PPI)
#
# RAV <- read.csv("~/Documents/PhD/TDP-43/TDP-43_Code/Results/GeneExpression/TDP-43_DEseq2/RAV_results_
# rownames(RAV) <- RAV$hgnc_symbol
# RAV <- RAV[,18:30]
# RAV <- subset(RAV, rownames(RAV) %in% DEG_PPI)

#Find the gene names that all datasets have in common
DEG_com <- Reduce(intersect, list(rownames(C9), rownames(VCP), rownames(FTLD),
                                  rownames(sALS)))

#Subset each dataset with these common names so they are all the same size
C9 <- subset(C9, rownames(C9) %in% DEG_com)
VCP <- subset(VCP, rownames(VCP) %in% DEG_com)
FTLD <- subset(FTLD, rownames(FTLD) %in% DEG_com)
sALS <- subset(sALS, rownames(sALS) %in% DEG_com)
# PET <- subset(PET, rownames(PET) %in% DEG_com)
# RAV <- subset(RAV, rownames(RAV) %in% DEG_com)

#Order by genename alphabetically
Data1 <- C9
Data1 <- Data1[order(rownames(Data1)),]
Data2 <- VCP
Data2 <- Data2[order(rownames(Data2)),]
Data3 <- FTLD
Data3 <- Data3[order(rownames(Data3)),]
Data4 <- sALS
Data4 <- Data4[order(rownames(Data4)),]
# Data5 <- PET
# Data5 <- Data5[order(rownames(Data5)),]
```

```r
# Data6 <- sALS
# Data6 <- Data6[order(rownames(Data6)),]


## Calculate the pair-wise distance; If the data is continuous, we recommend to use the function "dist2"
Dist1 = dist2(as.matrix(Data1),as.matrix(Data1));
Dist2 = dist2(as.matrix(Data2),as.matrix(Data2));
Dist3 = dist2(as.matrix(Data3),as.matrix(Data3));
Dist4 = dist2(as.matrix(Data4),as.matrix(Data4));
# Dist5 = dist2(as.matrix(Data5),as.matrix(Data5));
# Dist6 = dist2(as.matrix(Data6),as.matrix(Data6));

#Returns an affinity matrix that represents the neighborhood graph of the data points.
W1 = affinityMatrix(Dist1, K, alpha)
W2 = affinityMatrix(Dist2, K, alpha)
W3 = affinityMatrix(Dist3, K, alpha)
W4 = affinityMatrix(Dist4, K, alpha)
# W5 = affinityMatrix(Dist5, K, alpha)
# W6 = affinityMatrix(Dist6, K, alpha)

#Combine affinity matrices into fused network
W = SNF(list(W1,W2,W3,W4), K, T)

#name rows and columns
rownames(W) <- colnames(W) <- rownames(Data1)

## spectral clustering
C = 8 # number of clusters
group = spectralClustering(W, C);    # the final subtypes information
colors <- c("red", "orange", "yellow", "green", "blue", "violet", "magenta", "brown")
col <- getColorsForGroups(group, colors)
col <- data.frame(rownames(W), col)

## Generate file for Cytoscape
#Remove doubled values
df <- W
df[lower.tri(df, diag = TRUE)] <- NA

#Turn into vector
df <- as.matrix(df)
dfvec <- unmatrix(df)
#Remove any NA values
dfvec <- na.omit(dfvec)
dfvec <- as.data.frame(dfvec)

#Split row names into two and assign to columns
library(stringr)
genenames <- as.data.frame(str_split_fixed(rownames(dfvec), ":", 2))
dfvec$gene1 <- genenames$V1
dfvec$gene2 <- genenames$V2

#reorder columns?me
dfvec <- dfvec[,c(2,3,1)]
```

```r
#Take subset of edges if required
cyt <- subset(dfvec, dfvec > 0.003)

setwd("/Users/clairegreen/Documents/PhD/TDP-43/TDP-43_Code/Results/PPI_Network/SimilarityNetworkFusion/
write.csv(dfvec, file = "networkfile.csv", row.names = F, quote = F) #whole network file
write.csv(cyt, file = "0.003_networkfile.csv", row.names = F, quote = F) #subsetted file
write.csv(col, file = "colourmodules.csv", row.names = F, quote = F) #node table assigning cluster colo

#Generate gene list files for each of the modules
setwd("/Users/clairegreen/Documents/PhD/TDP-43/TDP-43_Code/Results/PPI_Network/SimilarityNetworkFusion/
for (i in 1:length(colors)){
  ref = colors[i]
  colour = subset(col, col == ref)
  write.csv(colour, file = paste(ref, ".csv", sep = ""), row.names = F, quote = F)
}
```

Datasets

I used only the microarray datasets because inclusion of the RNA seq datasets created clusters whose functional enrichment was less significant and more unclear. The strongest signal comes with using microarray only.

Clustering Results

Blue	Brown	Green	Magenta	Orange	Red	Violet	Yellow
AGO3	ADAMTS1	ACTB	AHCYL1	ABCA5	AASDHPPT	ABI2	A2M
CSTA	ANXA1	ANXA5	ANP32B	ABCD3	ACOT7	ABL1	ACTN1
DACH1	CD44	CFL1	ANXA2	ABI1	ACSL6	ACAT2	ALOX5
DDX3Y	CHST15	CLU	B2M	ACAT1	ACTR3	ACTN2	CDKN1A
FCER1G	CLIC4	LAMP1	BNIP3	ACOT8	ADD1	ADNP	CEBPB
FRAT1	COL1A1	MAP2	BTF3	ACOX1	ADD2	AHCY	CHGB
GAB1	COL1A2	MSN	CIRBP	ACTR2	ADSL	AHCYL2	CLIC1
GNAL	DAP3	RPL37A	CLTC	ADAM17	AHSA1	AIP	CSTB
HMGXB4	FYN	RTN4	COX5B	ADAM9	AKAP12	AKAP6	CYFIP2
MEIS2	GADD45A	SPP1	DDX17	ADSS	AKR1B1	ALAD	EPAS1
PAK2	GOLIM4	TMSB10	DPYSL2	AGFG1	AKR7A2	ALDH7A1	EZR
PSMD11	GPNMB	TMSB4X	EEF1A1	AIMP1	ALAS1	ALMS1	FKBP5
RNF8	HDAC4	TUBA1A	EEF1A2	AKR1C1	ALDH9A1	ANKRD27	GADD45B
SULT1A1	HEY1	TUBA1B	EEF1D	AMD1	ALDOA	ANXA11	HIPK2
	ITGB1BP1	TUBB2A	EEF2	ANKRD10	AMPH	AP2B1	HLA-B
	KAT6A	VDAC3	EIF3E	APPBP2	ANK2	APLP2	HLA-C
	KCTD12	YWHAB	EIF3L	ARAF	ANKMY2	APTX	ID2
	MAN2A1		FEZ1	ARFGEF2	ANXA7	ARF6	IFITM2
	MAP4		FEZ2	ARHGAP24	AP2M1	ARHGAP17	IGFBP5
	NBPF1		FHL1	ARID4A	AP3S1	ARHGEF7	ISG15
	NOTCH2NL		GAPDH	ARL1	APP	ARNT2	ITGB1
	PLSCR1		HIPK3	ARL15	ARGLU1	ARRB1	KIF5B
	PTP4A3		HLA-A	ARMC8	ARL6IP1	ASB13	LAPTM5
	RECK		HMGB1	ASXL2	ARL6IP5	ASB8	LPL
	SKIL		HNRNPH1	ATF2	ARMC1	ATG4B	MAPK1
	SSFA2		HSP90AA1	ATF7IP	ARPC2	ATG9A	NAIP
	TIMP2		HSPB1	ATG12	ATP5L	AUTS2	NEDD9
	TOR1AIP2		HSPD1	ATM	ATP6V0C	B4GALT3	PGRMC1
	YES1		HUWE1	ATP5C1	ATP6V1A	BAZ2B	PHLDA1
	ZEB2		KIF1B	ATP8B1	ATP6V1B2	BCL10	PLEKHA2
			KIF3A	ATRX	ATP6V1E1	BCL2	PLSCR4
			KTN1	ATXN1	ATP6V1H	BHLHE40	PTGER3
			LDHA	ATXN7	AUH	BRAP	RBPJ
			MAFF	BBS10	BAG3	BRD4	RPL3
			MORF4L1	BCAT1	BAG6	BTG1	SORBS1
			MORF4L2	BLZF1	BANF1	BZRAP1	SPARC
			MT1X	BPGM	BASP1	C14orf1	SRSF6
			MT2A	BRWD1	BCAS2	C16orf45	STAT3
			MYBPC1	C1orf109	BCL6	C6orf120	TNFAIP3
			MYL6	C8orf33	BCLAF1	CALD1	TXNIP
			MYOT	CALR	BIRC2	CALU	
			NACA	CAMK2B	BNIP3L	CAMKK2	

Blue	Brown	Green	Magenta	Orange	Red	Violet	Yellow
			NACAP1	CAND1	BTBD1	CAPN3	
			NDUFB8	CANX	C1QBP	CASP9	
			NPM1	CASK	CACYBP	CAST	
			PABPC1	CBLB	CAPNS1	CCDC130	
			PGM1	CBX5	CAPZA1	CCDC59	
			PRKCA	CCNA2	CBX1	CCDC90B	
			PRNP	CCNB1	CBX3	CCNC	
			PTMA	CD24	CBX4	CCNL1	
			QKI	CD2AP	CCDC6	CCSER2	
			RAC1	CD47	CCNH	CD40	
			RAP1B	CDC14B	CD63	CD55	
			RHOA	CDH11	CDH2	CD81	
			RPL12	CDK12	CDKN1B	CDC123	
			RPL15	CDK13	CHCHD2	CDC34	
			RPL18	CDK19	CKB	CDC42	
			RPL19	CDK2	CLNS1A	CDH22	
			RPL22	CDK6	CNOT7	CDKN1C	
			RPL23A	CDK8	COPS2	CDKN2C	
			RPL24	CENPN	COPS3	CEBPZ	
			RPL30	CENPU	COPS4	CELSR2	
			RPL32	CEP57	COPS5	CEP170B	
			RPL35	CEP70	COPS6	CHCHD3	
			RPL38	CFLAR	COPS8	CHD4	
			RPL6	CGGBP1	CORO1C	CIAO1	
			RPL7	CHD1	COX5A	CLCN6	
			RPS10	CHFR	CREG1	CLK1	
			RPS12	CLK4	CRKL	CLP1	
			RPS14	CMTM6	CSDE1	CNOT1	
			RPS16	CNOT3	CSNK2B	CNOT2	
			RPS17	CNOT4	CTBP1	COMMD8	
			RPS23	CNOT6	CUL3	COPG1	
			RPS24	CNOT8	CUL4A	COPS7A	
			RPS28	COIL	CUL5	CPNE3	
			RPS29	COL13A1	CYCS	CPSF4	
			RPS4X	CPOX	DAZAP2	CPSF6	
			RPS6	CPPED1	DCTN2	CPSF7	
			RPS7	CREB1	DDB1	CRCP	
			SAT1	CRY1	DDOST	CREBBP	
			Sep-02	CSAD	DDX1	CRK	
			Sep-07	CSTF1	DDX24	CRLF3	
			Sep-09	CTAGE5	DDX3X	CSK	
			SERPINI1	CTNNAL1	DDX42	CSNK1A1	
			SKP1	CTSD	DHX15	CSNK2A1	
			SOD1	CTSK	DLAT	CSNK2A2	
			SPTBN1	CTSZ	DNTTIP2	CSRP2	
			SRP14	CTTN	DPP8	CTBP2	
			TFRC	CUL2	DRG1	CTNNA1	
			TPI1	CUL4B	DSTN	CTNNB1	
			TPM1	CXorf56	DYNC1H1	CTSH	
			TSC22D1	CYLC2	DYNC1I2	CYB5A	
			TUBB2B	CYLD	DYNLRB1	CYB5B	
			TXNL1	CYP2E1	DYNLT3	CYFIP1	

Blue	Brown	Green	Magenta	Orange	Red	Violet	Yellow
			UBA52	DAPP1	DYRK1A	DAAM1	
			UBC	DCP2	ECHS1	DARS	
			VAPA	DDX39A	EIF1B	DBN1	
			YBX1	DENND4A	EIF2S2	DCAF7	
			YWHAZ	DERL1	ELL2	DCTD	
				DHX9	ELP3	DCUN1D1	
				DNM1L	ENSA	DDX21	
				DR1	EPM2AIP1	DEDD	
				DST	EPRS	DHX29	
				E2F3	EPS15	DIDO1	
				ECT2	ERH	DLG1	
				EDRF1	ESD	DNAJB1	
				EED	FABP5	DNPEP	
				EFEMP1	FAM96B	DOCK9	
				EFHC1	FBXL5	DPYSL3	
				EGLN3	FBXW7	DUSP3	
				EIF2S3	FDFT1	DUSP7	
				EIF3H	FH	DVL1	
				EIF3M	FKBP3	DVL2	
				EIF4E2	FLOT1	DYNC1LI1	
				ENAH	FMR1	DYNLT1	
				EPS8	FOXO3	DZIP3	
				ERBB3	FXR1	EFEMP2	
				ERBB4	GABARAP	EIF2AK2	
				ERCC6	GABARAPL2	EIF2S1	
				EXOSC8	GARS	EIF4B	
				FANCC	GAS7	EIF4ENIF1	
				FBXO3	GDI1	EP300	
				FBXW2	GLRX5	EPHB6	
				FEM1C	GNAQ	ERC1	
				FOXN3	GORASP2	ERCC5	
				FTSJ1	GOT2	ETS2	
				FUBP3	GPI	EWSR1	
				GART	GSPT1	F8	
				GBP2	HADHA	FAF1	
				GCH1	HERC1	FAF2	
				GEMIN2	HERPUD1	FEN1	
				GFPT1	HINT1	FIBP	
				GFPT2	HK1	FKBP4	
				GGH	HNRNPA2B1	FLNC	
				GIN1	HNRNPK	FLOT2	
				GNA13	HNRNPU	FXR2	
				GNE	HPRT1	GABARAPL1	
				GNL1	HSP90AB1	GAS2L1	
				GPBP1L1	HSPA4	GATM	
				GPD2	HSPA5	GGA1	
				GSAP	HSPE1	GIGYF2	
				GTF2F1	HSPH1	GIPC1	
				GTF3C3	HTATSF1	GIT1	
				GTPBP4	IMMT	GLI3	
				GULP1	IRS1	GLRX3	
				H2AFV	ITPR1	GMNN	

Blue	Brown	Green	Magenta	Orange	Red	Violet	Yellow
				HABP4	JUN	GMPS	
				HCFC2	KCMF1	GNA11	
				HCLS1	KIAA0368	GNB5	
				HEATR3	KPNB1	GNG5	
				HELZ	KRAS	GNG7	
				HIP1	LAMTOR5	GOLGA2	
				HIPK1	LARP1	GOLGA3	
				HNMT	LONRF1	GOLGA4	
				HNRNPA0	LRRC47	GOLGA7	
				HNRNPM	MAGED1	GPN1	
				HOXA3	MAGED2	GRB10	
				HUS1	MAP2K1	GRB2	
				IKBKAP	MAP2K4	GRK5	
				IKBKB	MAP3K5	GRPEL1	
				ILF2	MAPK6	GSK3B	
				ILF3	MAPRE2	GSTP1	
				ING2	MAPT	GTF2B	
				IQGAP1	MARK4	GTF3C1	
				ITGA6	MAT2B	H2AFX	
				ITPR2	MED7	H2AFY	
				IVNS1ABP	METTL9	HAUS7	
				KCNAB1	MRPL20	HDAC2	
				KCND2	MTHFD2	HDAC3	
				KCNE4	MYH10	HES1	
				KCTD20	MYL6B	HGH1	
				KDM5A	NAA50	HLTF	
				KDM6A	NAP1L1	HMOX2	
				KHDRBS3	NARS	HNRNPA1	
				KIN	NCAM1	HNRNPC	
				KLF3	NCKAP1	HNRNPL	
				KMT2A	NCOR2	HNRNPUL1	
				KPNA3	NDEL1	HOXB5	
				KRIT1	NDFIP1	HRAS	
				KRR1	NDUFB11	HSP90B1	
				L3MBTL1	NDUFS1	HSPA2	
				LAMA4	NDUFS6	HSPA9	
				LAMTOR3	NECAP1	ICE1	
				LATS1	NONO	IFRD1	
				LCP2	NPEPPS	IGF2R	
				LGR4	NREP	IK	
				LIG4	NRIP1	IL17RA	
				LIMA1	NUAK1	IMPDH1	
				LIMK2	OAT	ING4	
				LMF1	OGT	IP6K2	
				LMO1	OXCT1	IPO8	
				LNPEP	PAFAH1B1	IRAK1	
				LPCAT4	PAK1	ITGB5	
				LRBA	PAM	ITSN1	
				LRCH3	PANK2	ITSN2	
				LRIF1	PAPSS1	JMJD1C	
				LRRC14	PCBP1	JMJD6	
				LYN	PCBP2	KALRN	

Blue	Brown	Green	Magenta	Orange	Red	Violet	Yellow
				LZTFL1	PCDH17	KAT2B	
				MAGI1	PCMT1	KAT7	
				MAP3K7	PDCD10	KAT8	
				MAP4K3	PDHB	KATNB1	
				MAP4K5	PEBP1	KBTBD2	
				MAPK8	PFDN5	KDM5C	
				MARK3	PFKM	KHDRBS1	
				MBIP	PFN2	KIAA0141	
				MCC	PGAM1	KLF6	
				MCL1	PGK1	KLHDC3	
				MCM4	PHYH	KLHL12	
				MCMBP	PIK3R1	KLHL22	
				MDC1	PJA2	KPNA1	
				MED14	PLCB1	LAGE3	
				MED23	PLS3	LAMP2	
				MED28	PMPCB	LASP1	
				MED4	POLR2I	LMO2	
				MFN1	POLR2L	LPP	
				MGEA5	PPP1CB	LRPPRC	
				MIA3	PPP1CC	LRRC40	
				MIS18A	PPP2CA	MACF1	
				MITF	PPP2CB	MADD	
				MLF1	PPP2R2A	MAML1	
				MLLT10	PPP2R2B	MAPK14	
				MLLT3	PPP3CA	MAPK3	
				MOB1A	PPP6C	MAPRE1	
				MON2	PRDX1	MAPRE3	
				MRFAP1L1	PRDX6	MBD1	
				MRPL42	PRKCI	MBP	
				MSH3	PRPS1	MCTS1	
				MYOF	PRPSAP1	MDH2	
				NAA15	PSAP	MEAF6	
				NCK1	PSMA1	MED12	
				NCOA2	PSMA2	MED8	
				NCOA3	PSMB1	MEN1	
				NCOR1	PSMB3	MFAP1	
				NEK1	PSMB4	MFN2	
				NFIL3	PSMB6	MGAT4B	
				NFYB	PSMC2	MLF2	
				NLK	PSMC5	MLH3	
				NNT	PSMC6	MLPH	
				NOL11	PSMD1	MRPL23	
				NPTX1	PSMD12	MRPL40	
				NUDT21	PSMD2	MRPS11	
				NUP54	PSMD4	MRPS7	
				NUP85	PSMD6	MTCL1	
				NXT2	PTGES3	MTHFD1	
				OFD1	PTPN11	MUS81	
				PAF1	PUM1	MYL12A	
				PAPOLA	RAB1A	NANS	
				PAPOLG	RAB21	NAP1L4	
				PAXIP1	RABAC1	NCKIPSD	

Blue	Brown	Green	Magenta	Orange	Red	Violet	Yellow
				PCNA	RAD21	NCOA6	
				PCSK5	RAD23B	NDRG1	
				PDCD5	RAN	NEDD4L	
				PDE4D	RANBP6	NFX1	
				PDE4DIP	RANBP9	NGLY1	
				PDS5A	RAP2A	NINL	
				PDS5B	RB1	NKRF	
				PEX7	RBBP7	NMT1	
				PFAS	RCAN2	NOL3	
				PHC3	RCN2	NOTCH2	
				PHF3	RDX	NOTCH4	
				PHKB	RERE	NPLOC4	
				PIBF1	RGL1	NRAS	
				PIGA	RGS2	NRG2	
				PIPOX	RND3	NSMAF	
				PKD2	RNF10	NUDC	
				PKN2	RNF11	NUP133	
				PLAA	RNF19A	NUP214	
				PLCB4	RNPS1	NUP43	
				PLCG2	RPA1	NXN	
				PLOD2	RSRC2	OSBP	
				PNN	RUVBL2	PA2G4	
				PNO1	SAMM50	PAAF1	
				POGLUT1	SAR1A	PACSIN2	
				POLH	SARAF	PAICS	
				POLI	SARS	PALLD	
				POT1	SCRN1	PARP1	
				PPARGC1A	SDCBP	PCM1	
				PPP3R1	SDHA	PDCD6IP	
				PPP4R2	SEC23A	PDCL	
				PPP6R3	SEC24B	PDE4A	
				PRDM10	SEC61B	PDE4B	
				PRDX2	SEPHS2	PDGFB	
				PRDX3	Sep-06	PDIA3	
				PRDX4	SERINC1	PDLIM2	
				PREP	SF3B5	PECAM1	
				PRKAA1	SLC25A6	PEX14	
				PRKAB2	SMARCA1	PFKL	
				PRKCB	SMARCA2	PGLS	
				PRKDC	SNCA	PGRMC2	
				PROSER1	SNRNP25	PIAS1	
				PRPF3	SNRPA1	PIAS3	
				PRPF40A	SNRPD2	PIK3R3	
				PRPF4B	SNW1	PIM1	
				PRPS2	SNX3	PIN1	
				PSEN1	SPOCK1	PKP4	
				PSMF1	SRRM1	PLAT	
				PSPH	SSBP1	PNISR	
				PTBP3	SSR1	POLD3	
				PTCD3	STAU1	POLR1C	
				PTPN12	STK39	POLR2C	
				PTPN13	STXBP1	POP5	

Blue	Brown	Green	Magenta	Orange	Red	Violet	Yellow
				PTPN4	SUB1	PPFIA1	
				RAB14	SUMO1	PPFIBP1	
				RAB27A	SUMO3	PPIB	
				RAB27B	SYNJ1	PPIF	
				RAB9A	TALDO1	PPIG	
				RAD17	TAX1BP1	PPM1B	
				RAPGEF2	TBC1D4	PPME1	
				RBBP8	TBCA	PPP1R12A	
				RBKS	TCEAL4	PPP1R12B	
				RBM12	TCEB1	PPP1R9A	
				RBM3	TCERG1	PPP2R2D	
				RBM48	TDP2	PPP2R5D	
				RCHY1	TERF2IP	PPP6R1	
				RCOR1	TIPRL	PQBP1	
				REL	TKT	PRKAA2	
				REPS1	TMED2	PRKAR2A	
				RFC3	TMED9	PRMT2	
				RGS14	TMEM30A	PTEN	
				RIT1	TMOD1	PTK2	
				RNF14	TNKS	PTN	
				RNF2	TNPO1	PTPRA	
				ROCK1	TOMM20	PTPRM	
				RORB	TSG101	PUF60	
				RPIA	TSNAX	PUM2	
				RRM1	TSPYL5	R3HDM2	
				RSRC1	TTC1	RAB15	
				RWDD1	TTC19	RAB1B	
				SCAF11	TUBA4A	RABIF	
				SCRN3	TUBB	RALBP1	
				SEC23B	TUBB4B	RALGDS	
				SEC23IP	TUFM	RANBP2	
				SEC24A	TXNDC9	RANGRF	
				SEC63	UBE2A	RAP1A	
				SENP7	UBE2D3	RASSF4	
				SFMBT1	UBE2N	RBBP4	
				SFXN1	UBL5	RBBP6	
				SH2B3	UBR4	RBM39	
				SHC1	UCHL1	RBM5	
				SHMT2	UFC1	RCBTB2	
				SIAH1	UGP2	RGS3	
				SKP2	UQCRB	RIMS1	
				SLC11A2	UQCRC2	RING1	
				SLC25A17	USP11	RLF	
				SLC26A2	USP14	RNF114	
				SLC35A2	USP7	RNF41	
				SLC35A3	VDAC2	RPA2	
				SMAD4	VPS26A	RPAP1	
				SMAD6	VPS35	RPH3A	
				SMC3	WARS	RPN1	
				SMC5	WASF1	RPS6KA2	
				SMU1	WASL	RQCD1	
				SNAP23	WBP11	RRP1B	

Blue	Brown	Green	Magenta	Orange	Red	Violet	Yellow
				SNAPC1	WDR11	RSL1D1	
				SNAPC3	XPNPEP1	RTCA	
				SNRPD1	XRCC6	RUNX1T1	
				SNTB2	YTHDF1	RXRB	
				SREK1	YTHDF2	RYK	
				SRI	YTHDF3	SAFB	
				SRP72	YWHAE	SAMHD1	
				SRPK1	YWHAH	SAP18	
				SRPK2	YWHAQ	SART3	
				SRSF10	YY1	SASH1	
				SS18	ZBTB18	SAV1	
				SSR3	ZC3H15	SCYL2	
				SSX2IP	ZDHHC17	SDF4	
				STAG2		SEC16A	
				STK3		SEL1L	
				STK4		SEPHS1	
				STRN		Sep-08	
				STRN3		SERBP1	
				STX17		SERPINB9	
				SWAP70		SET	
				TAF4		SETD2	
				TARS		SF1	
				TBXA2R		SF3A1	
				TCF12		SF3A3	
				TDRD7		SF3B1	
				TERF2		SF3B2	
				TET3		SFPQ	
				TGFBR2		SFTPB	
				TGOLN2		SFXN3	
				THADA		SGCD	
				THBS1		SH3GLB1	
				THRA		SH3GLB2	
				THUMPD2		SIK3	
				TLE3		SIRT2	
				TLN1		SKAP2	
				TMEM165		SKIV2L2	
				TMEM38B		SMAD1	
				TMOD2		SMAD2	
				TNKS2		SMARCC1	
				TOLLIP		SMARCD3	
				TPP2		SMARCE1	
				TRA2B		SMNDC1	
				TRAF5		SMURF2	
				TRAK1		SNAPC5	
				TRDMT1		SNRPB	
				TRIM16		SNX11	
				TRIM2		SNX6	
				TRIM24		SON	
				TRIM33		SORBS2	
				TRIM38		SORT1	
				TRIM5		SOS1	
				TRIP11		SOS2	

Blue	Brown	Green	Magenta	Orange	Red	Violet	Yellow
				TRIT1		SP3	
				TRPC1		SPEN	
				TSN		SPG21	
				TUBG1		SPOP	
				TUBGCP3		SRRM2	
				TWIST1		SRSF1	
				TXNL4B		SRSF4	
				UBA2		SRSF7	
				UBE2B		SSBP3	
				UBE2D1		STAT5B	
				UBE2D2		STMN1	
				UBE2J1		STX16	
				UBE2V2		STX2	
				UBXN4		SUN1	
				UBXN7		SUPT16H	
				UNG		SUPT5H	
				UPF2		SUZ12	
				USP1		SYNCRIP	
				USP25		TAB2	
				USP3		TARDBP	
				USP34		TBC1D2B	
				USP53		TBK1	
				USPL1		TBP	
				UTP14A		TCEA2	
				UTRN		TCEAL1	
				VAMP4		TCF25	
				VDAC1		TCF3	
				VLDLR		TCF4	
				WBP4		TDG	
				WIPF1		TERF1	
				WIPI2		TGFB1I1	
				WNK1		THUMPD1	
				WRN		TIA1	
				WWP2		TIMM13	
				XBP1		TLE1	
				XIAP		TMBIM4	
				YOD1		TMCO1	
				ZBTB1		TMEM11	
				ZBTB10		TMX1	
				ZBTB7A		TNFAIP1	
				ZCCHC10		TNIK	
				ZCCHC6		TNIP1	
				ZCCHC8		TNIP2	
				ZFC3H1		TNPO3	
				ZFPM2		TNRC6B	
				ZFYVE26		TOB1	
				ZNF12		TOMM40	
				ZNF24		TOP1	
				ZNF264		TPD52L2	
				ZNF292		TPM3	
				ZNF440		TPP1	
				ZNF573		TRA2A	

Blue	Brown	Green	Magenta	Orange	Red	Violet	Yellow
				ZNF593		TRAK2	
				ZNF83		TRAPPC2L	
				ZWINT		TRIM3	
						TRIM32	
						TRIM8	
						TSPYL2	
						TSR1	
						TTLL12	
						TXN2	
						U2AF2	
						U2SURP	
						UBE2D4	
						UBE2G2	
						UBE2H	
						UBE2I	
						UBE2K	
						UBE2L3	
						UBE2L6	
						UBE2Z	
						UBE3A	
						UBE4B	
						UBQLN4	
						UNC13B	
						UPF3A	
						UPF3B	
						USP13	
						USP4	
						VAMP2	
						VAMP3	
						VASP	
						VCL	
						VCP	
						VKORC1	
						VPS37C	
						VPS72	
						WDR26	
						WDR48	
						WDTC1	
						WWP1	
						XPC	
						XPO1	
						XPOT	
						XRCC5	
						YARS	
						YEATS2	
						YY1AP1	
						ZBTB5	
						ZFP36	
						ZMPSTE24	
						ZNF544	
						ZNF587	
						ZNF638	

Blue	Brown	Green	Magenta	Orange	Red	Violet	Yellow
						ZPR1	

Cluster Enrichment Results - EnrichR

Functional enrichment was conducted in EnrichR (http://amp.pharm.mssm.edu/Enrichr/). This allows you to conduct enrichment of a number of different gene lists. The links below follow to the results page for each cluster.

Blue - http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=2lphv

Enrichment unclear, gene roles suggest signalling/Cell cycle

Brown - http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=2lpid

Extracellular matrix

Green - http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=2lpie

Cytoskeleton/cell cycle

Magenta - http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=2lpil

Ribosome

Orange - http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=2lpis

Gene expression (this contains a few things, I think I might need to increase the number of modules)

Red - http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=2lpj4

Proteasome degradation

Violet - http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=2lpj5

Very similar to orange. Highest enrichment for mRNA splicing

Yellow - http://amp.pharm.mssm.edu/Enrichr/enrich?dataset=2lpj8

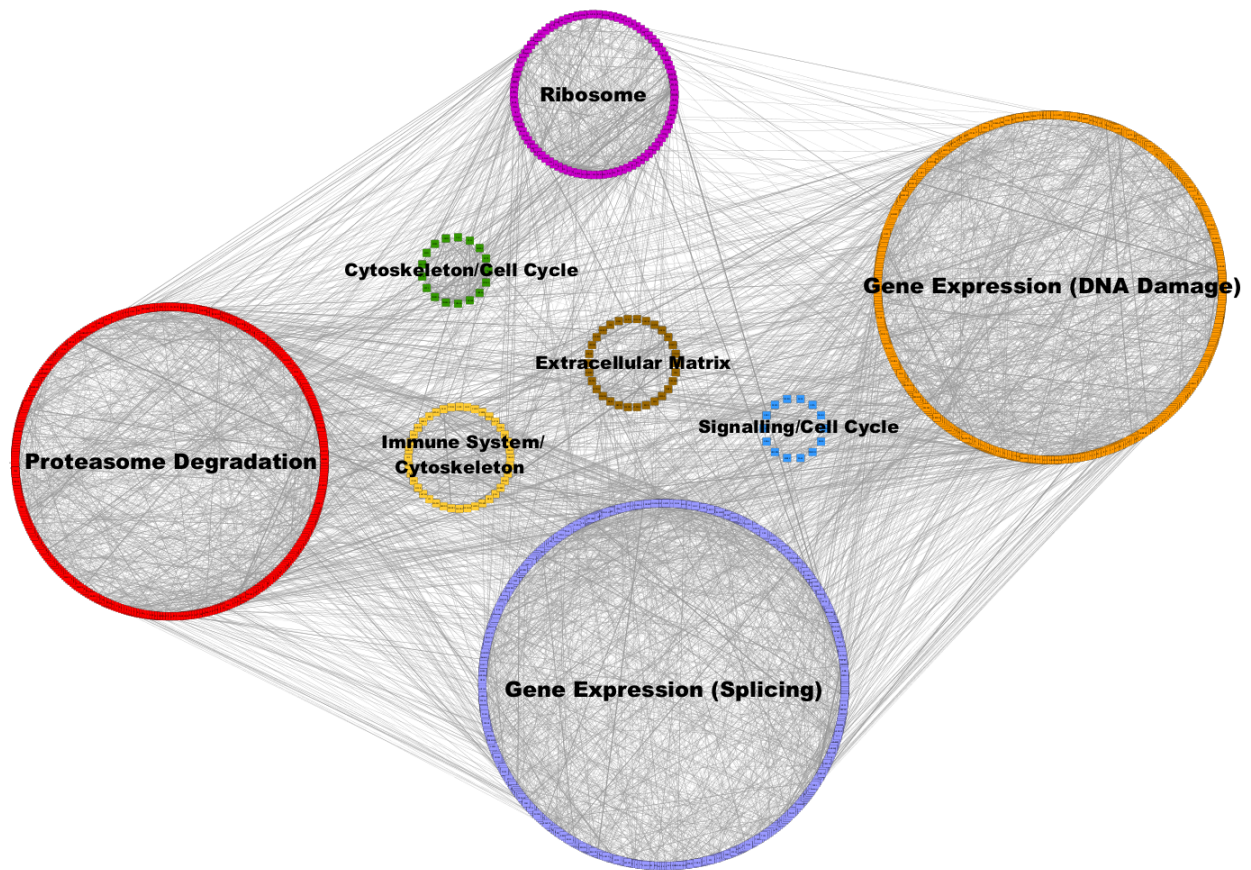Immune system/cytoskeleton

#Visualisation in Cytoscape

Figure 1: SNF Clusters: edge values > 0.003. Definitions in brackets are highest enriching sub-function

Looking at the functional enrichment I believe that I should probably re-run the experiment with more clusters. The two gene expression clusters I believe could be split into more modules as they enrich with multiple processes.

Benchmark Enrichment

To see if the clusters could enrich with genes known to be associated with neurodegeneration, I performed a hypergeometric test on the clusters versus my collected gene lists. In the table below are the BH adjusted p values. Subnetwork 28 is an internally generated list of genes whose expression correlates with TDP-43 pathology. Taylor and Pasterkamp are lists from two papers investigating protein-protein interactions of TDP-43.

Disease Gene List	Blue	Brown	Green	Magenta	Orange	Red	Violet	Yellow
GWAS Central ALS	0.3231	0.4460	0.0844	0.5083	0.0844	0.9424	0.9424	0.4460
NeuroX (GWS)	1	1	1	0.2169	1	1	1	1
GWAS Central AD	1	1	0.0092	1	0.1932	1	0.5188	1
Subnetwork 28	0.0098	1	1	0.2139	1	1	0.3356	1
Cirulli	1	1	1	0.0349	0.3875	0.0151	0.0019	1
AD Malacards	1	1	0.0039	0.0064	0.0082	0.0064	0.0673	0.0098
ALS Malacards	1	1	0.0016	0.0022	0.1677	0.0016	0.0074	0.0063
Pasterkamp	1	1	1	3.59E-09	0.0010	9.58E-07	5.01E-08	1
Taylor	1	0.0128	0.0258	7.25E-31	0.0248	2.53E-10	0.0008	0.1050
ALSOD	1	1	1	0.2477	0.2477	0.0036	0.0366	0.0076
Parkinsons Malacards	1	1	0.0163	0.2419	0.2419	0.0093	0.2624	1

Network Density

R script - "SNF_NetDens.R"

I wanted to find out whether the clusters that had been created by SNF were more intraconnected than interconnected. To do this you calculate what is called network density - the number of actual connections divided by the number of potential connections. The total potential connections is every node in a network connected to every other node. To understand the network, I calculated the total density of the whole network, the density of each individual cluster, the density of all intra-cluster interactions, and the density of all extra-cluster interactions.

The results are as follows:

Network Selection	Density
Red	0.0110183
Orange	0.007997736
Yellow	0.07435897
Green	0.5661765
Blue	0.0989011
Violet	0.005686658
Magenta	0.07173779
Brown	0.09425287
Whole network	0.004021154
Extra-cluster	0.002171221
Intra-cluster	0.009238546

Something I noticed is that the density values are usually higher with smaller clusters, suggesting it's not independent of cluster size. Despite this, the connections within clusters are always more dense than the connections between clusters.