

# Gent data analysis

## Plasma + GTEx ICA application

December 3rd 2020 - Results overview

# DATA

## Used data for ICA approach

### PLASMA DATA

♦ Initial data:

- 38.003 genes
- 208 samples
- 24 cancer type + 1 Healthy



### PLASMA DATA

♦ Processed data:

- 19.575 genes
- 208 samples
- 24 cancer type + 1 Healthy

### GTEX DATA

♦ Initial data:

- 53.513 genes
- 22.722 samples
- 31 tissues (54 sub-tissues)



### GTEX DATA

♦ Processed data:

- ~20.000 genes
- 26 datasets (>100 samples)
- Average tissue expression dataset

## Used data for ICA approach

### PLASMA ANNOTATIONS

♦ Simplified and broken down:

- Data of collection: year only
- Age: age range 10 by 10
- Location: simplified (e.g. left breast -> breast)
- TNM: broken down to T, N and M separately
- Stage: simplified (e.g. IVA/IVB -> IV)
- Smoking history: approximated to 10 and binarised
- Family and medical history: binarised

## Whole analysis steps

### ICA DECOMPOSITION

- ♦ Apply ICA to all datasets: 50 and 100 components
- ♦ Obtain matrices: S (metagenes) and A (metasamples)
  - Metagenes: Number of IC x Genes
  - Metasample: Number of IC x Samples
- ♦ Apply Mutual Nearest Neighbours (MNN):
  - Obtain network of maximal correlational links
- ♦ Apply TOPPGENE (<https://toppgene.cchmc.org/enrichment.jsp>) analysis:
  - Obtain enrichment summary for each component
- ♦ Apply Metasample association using annotations:
  - Obtain T-test scores for each component

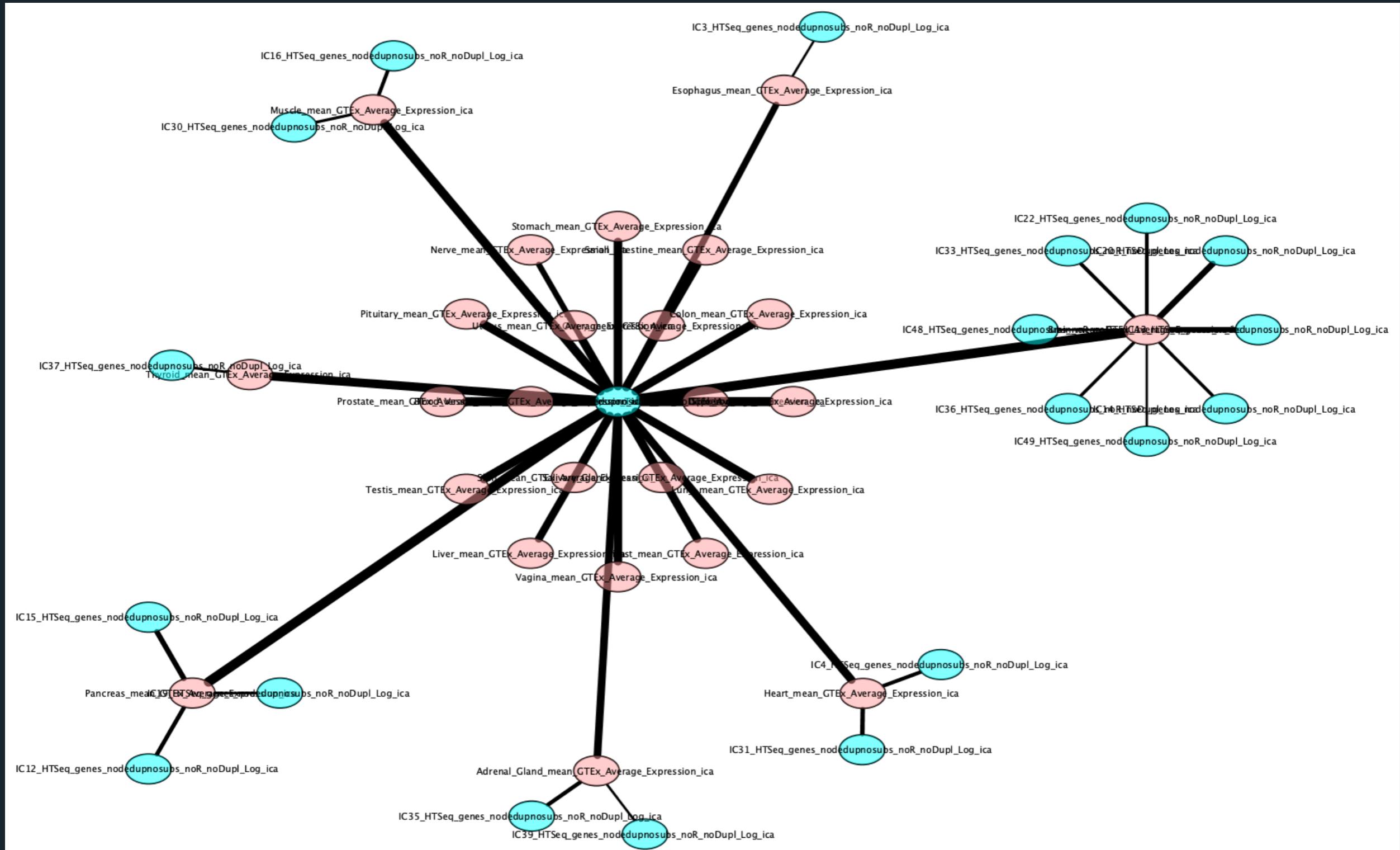
# RESULTS

## Results files available

### GITHUB IO

- ♦ Results available online:
  - ♦ URL: [https://nicolassompairac.github.io/Gent\\_Plasma\\_GTEx\\_ICA\\_analysis/](https://nicolassompairac.github.io/Gent_Plasma_GTEx_ICA_analysis/)
  - ♦ Browsable files
    - Metasamples/annotations correlations tables
    - TOPPGENE tables
  - ♦ Downloadable files:
    - Metagenes + Metasamples matrices
    - MNN graphs
    - This presentation
  - ♦ Not added: GTEx matrices and TOPPGENE results (for memory reasons)

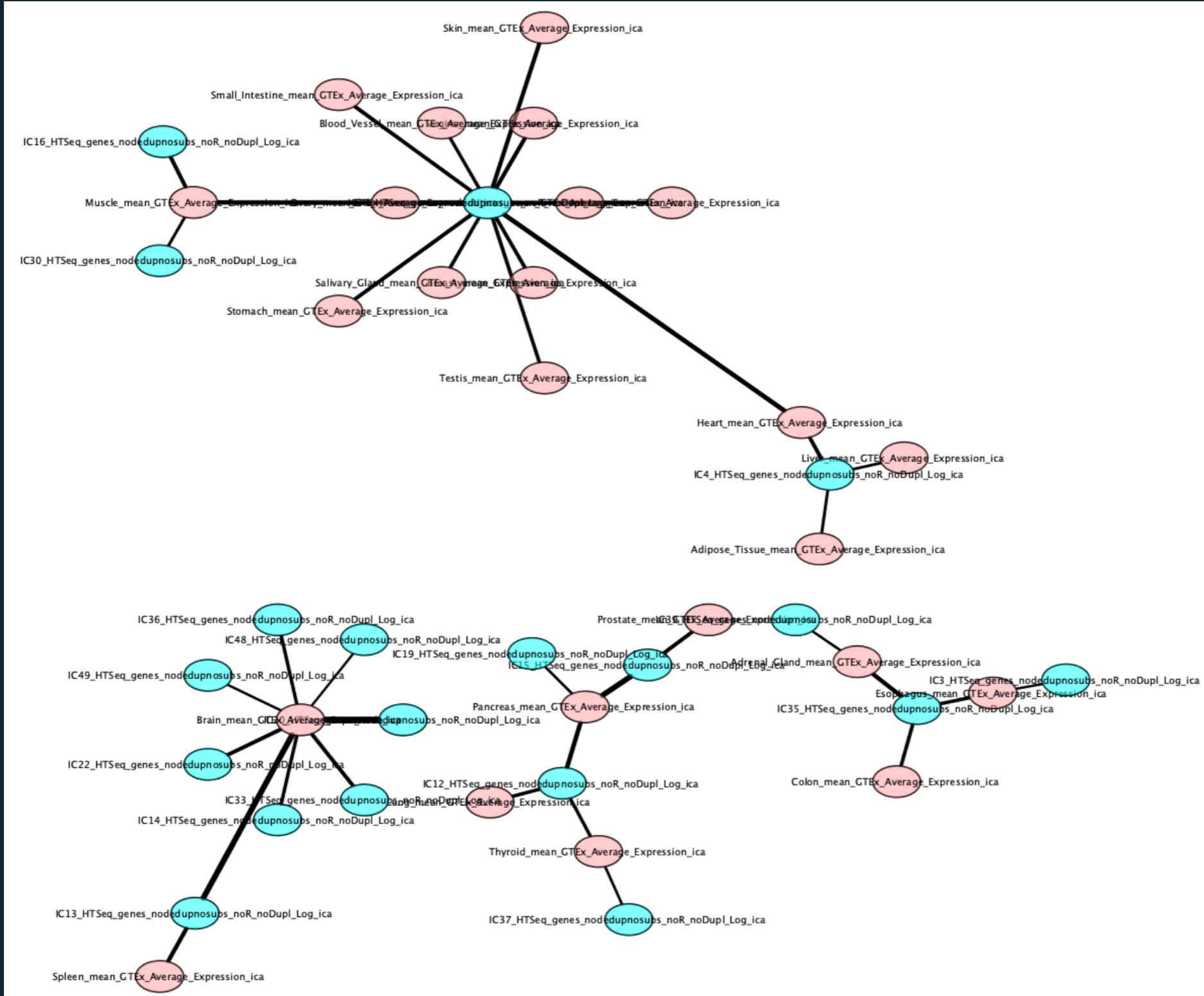
## Plasma/Mean Tissue - 50 ICs



Edge: Pearson correlation (proportional to width)

Red: Mean tissue expression  
 Blue: Plasma IC

# Plasma/Mean Tissue - 50 ICs (No hub IC)

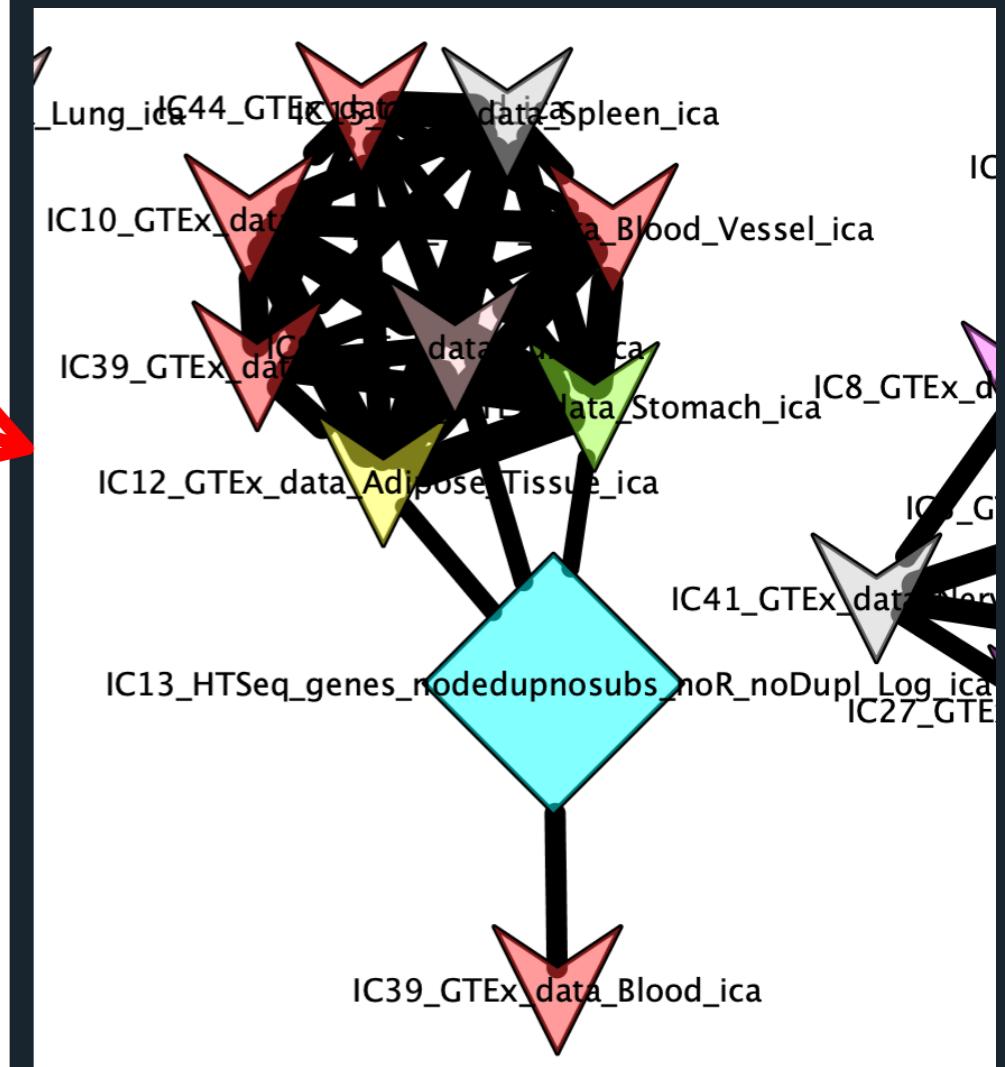
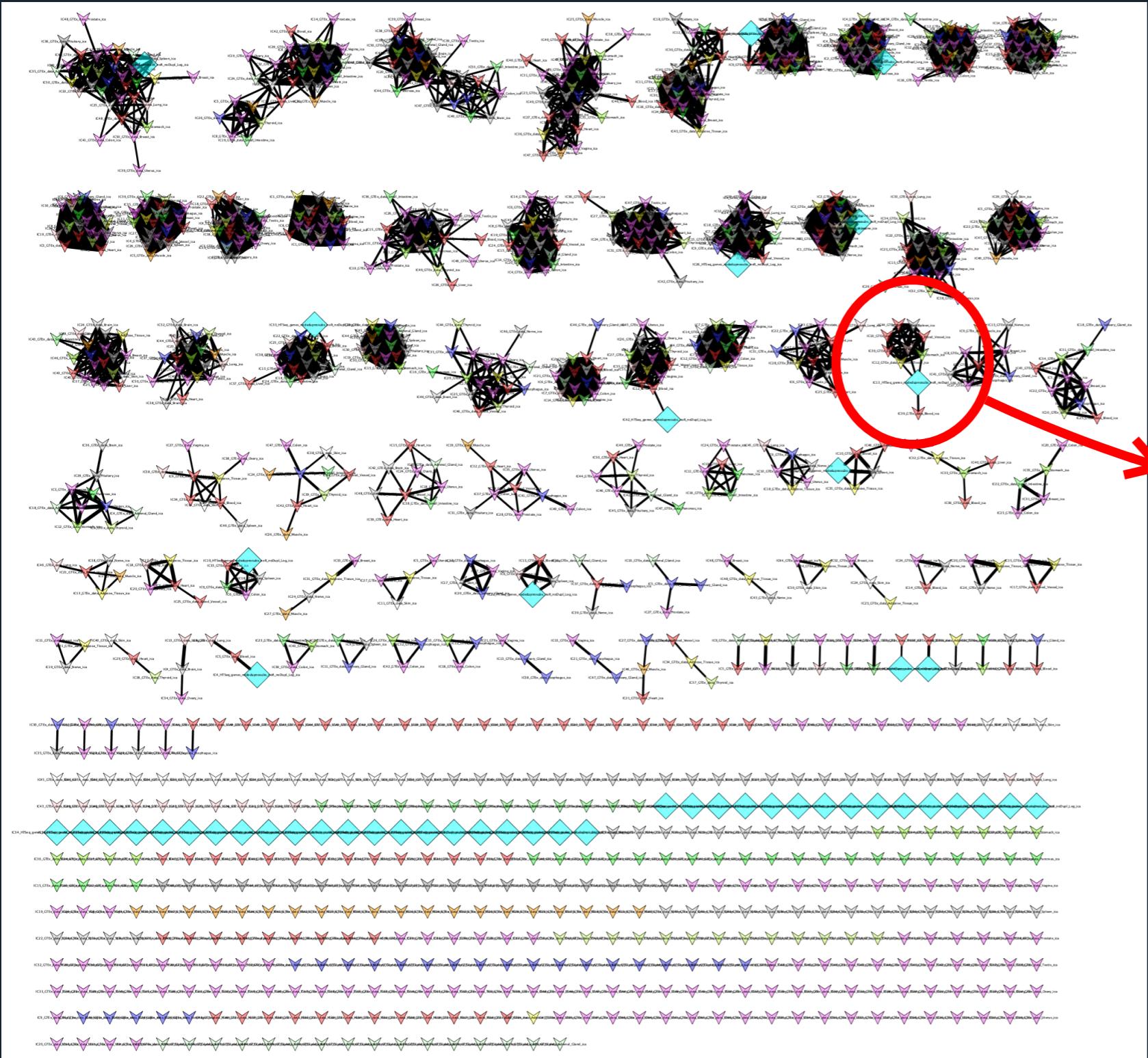


Edge: Pearson correlation (proportional to width)

Red: Mean tissue expression  
Blue: Plasma IC

MNN

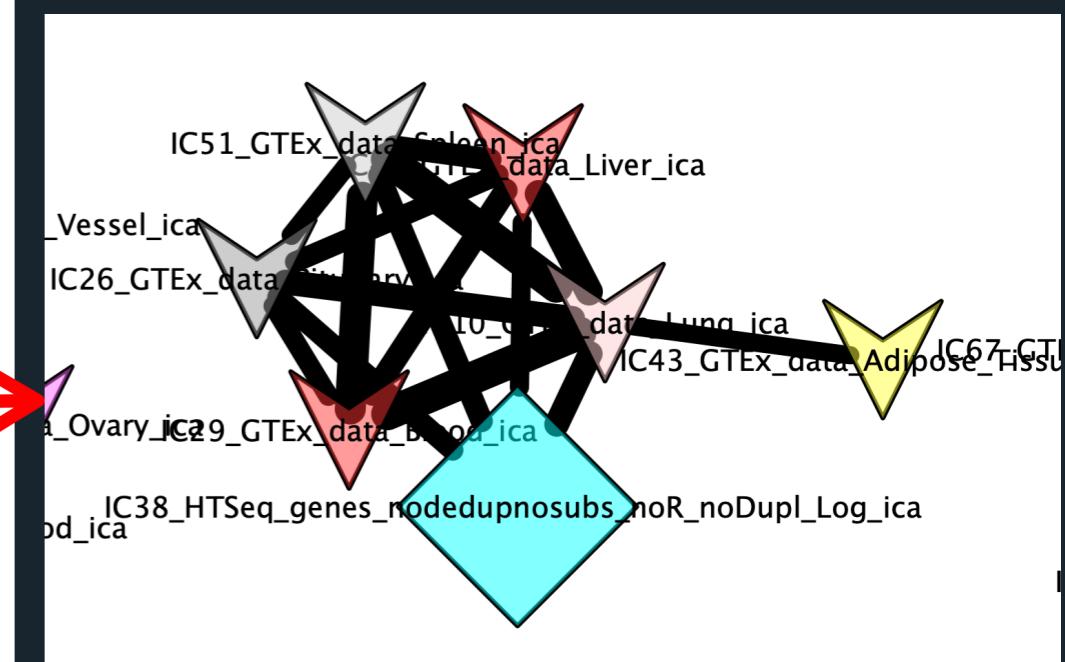
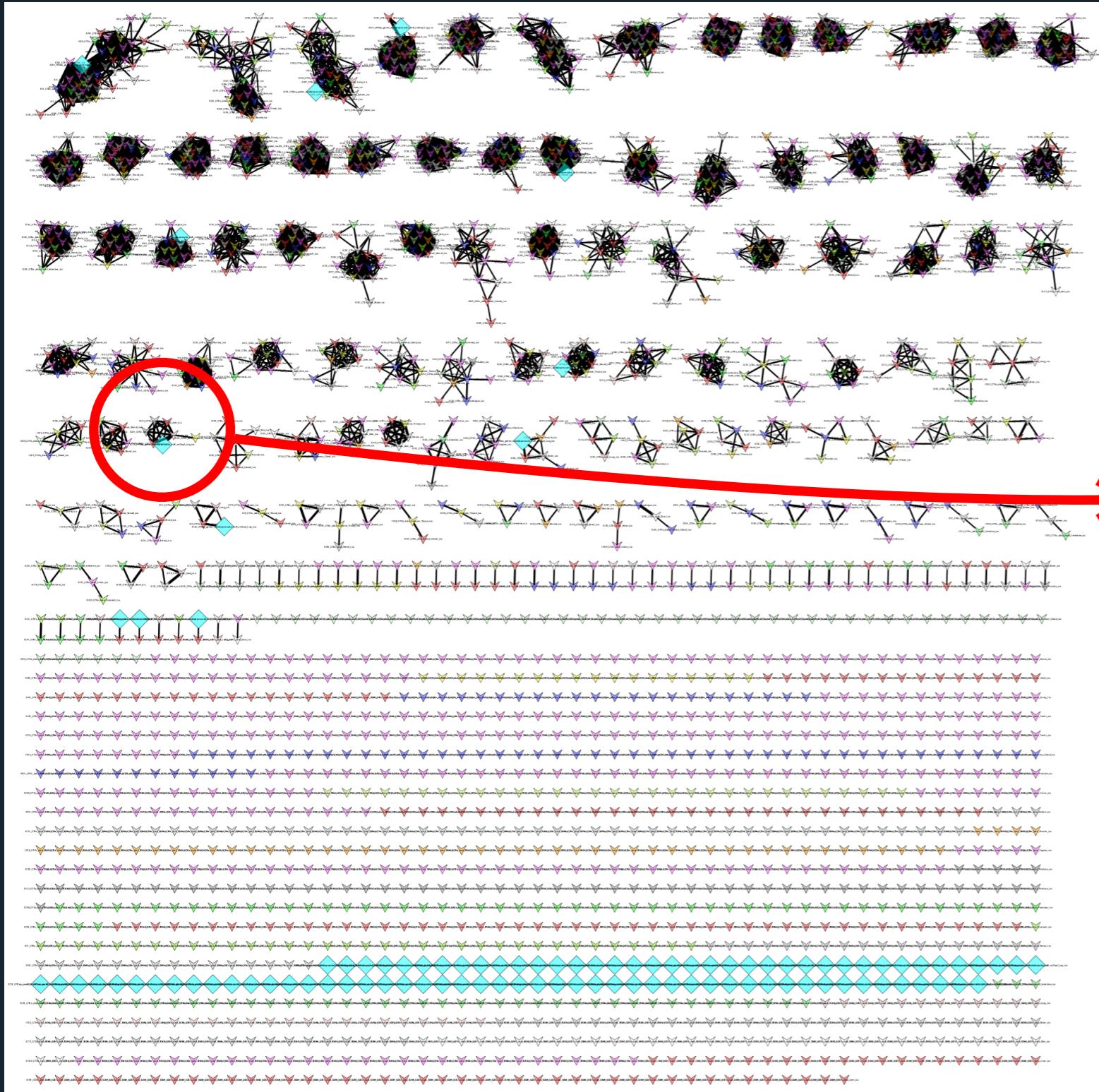
# Plasma/All Tissues - 50 ICs



- Plasma IC  
- GTEx IC

MNN

# Plasma/All Tissues - 100 ICs



## Plasma TOPPGENE table - 50ICs

Detailed annotations

Top enrichments

POSITIVE SIDE	ANNOTATION	*	NEGATIVE SIDE	ANNOTATION
IC1+ (18 genes)	13: Coexpression: <a href="#">Up-regulated genes detecting gender effects in global expression profiling studies.</a> 14: Coexpression Atlas: <a href="#">ratio EmbryoidBody-fibro vs EmbryoidBody-derived-from CD34-iPSC top-relative-expression-ranked 100</a>	*	IC1- (12 genes)	none
IC2+ (106 genes)	6: Domain: <a href="#">IG LIKE</a>	*	IC2- (107 genes)	none
IC3+ (195 genes)	1: GO: Molecular Function: <a href="#">serine hydrolase activity</a> 2: GO: Biological Process: <a href="#">steroid metabolic process</a> 3: GO: Cellular Component: <a href="#">blood microparticle</a> 4: Human Phenotype: <a href="#">Abnormality of the common coagulation pathway</a> 5: Mouse Phenotype: <a href="#">abnormal blood coagulation</a> 6: Domain: <a href="#">Cyt P450 CS</a> 7: Pathway: <a href="#">Complement and coagulation cascades</a> 9: Interaction: <a href="#">GDPD1 interactions</a> 11: Transcription Factor Binding Site: <a href="#">V\$HNF1_Q6</a> 13: Coexpression: <a href="#">Liver selective genes</a> 14: Coexpression Atlas: <a href="#">liver</a> 18: Disease: <a href="#">Adverse reaction to drug</a>	*	IC3- (42 genes)	none
IC4+ (374 genes)	1: GO: Molecular Function: <a href="#">actin binding</a> 2: GO: Biological Process: <a href="#">platelet degranulation</a> 3: GO: Cellular Component: <a href="#">platelet alpha granule</a> 4: Human Phenotype: <a href="#">Increased mean platelet volume</a> 5: Mouse Phenotype: <a href="#">abnormal blood coagulation</a> 7: Pathway: <a href="#">Platelet degranulation</a> 9: Interaction: <a href="#">MYH9 interactions</a> 11: Transcription Factor Binding Site: <a href="#">V\$SRF_01</a> 13: Coexpression: <a href="#">Human StemCell Macaulay07 516genes</a> 14: Coexpression Atlas: <a href="#">PP MEG top-relative-expression-ranked 500 k-means-cluster#4</a> 16: MicroRNA: 18: Disease: <a href="#">Malaria</a>	*	IC4- (3 genes)	none

Higher stability

ICs arranged by stability in descending order

Lower stability

List of significant genes

# ANNOTATIONS

# Plasma metasamples - 50ICs

## Meta Sample Annotation for HTSeq\_genes\_nodedupnosubs\_noR\_noDupl\_Log

- HTSeq\_genes\_nodedupnosubs\_noR\_noDupl\_Log\_A\_associations\_num.xls
- HTSeq\_genes\_nodedupnosubs\_noR\_noDupl\_Log\_A\_associations\_num\_info.xls
- HTSeq\_genes\_nodedupnosubs\_noR\_noDupl\_Log\_A\_associations\_cat\_info.xls
- HTSeq\_genes\_nodedupnosubs\_noR\_noDupl\_Log\_A\_associations\_cat.xls
- HTSeq\_genes\_nodedupnosubs\_noR\_noDupl\_Log\_A\_associations\_slf\_info.xls
- HTSeq\_genes\_nodedupnosubs\_noR\_noDupl\_Log\_A\_associations\_slf.xls
- HTSeq\_genes\_nodedupnosubs\_noR\_noDupl\_Log\_A\_annotation.xls

### Table of IC associations with themselves.

The values are Spearman correlation between a pair IC metasamples.

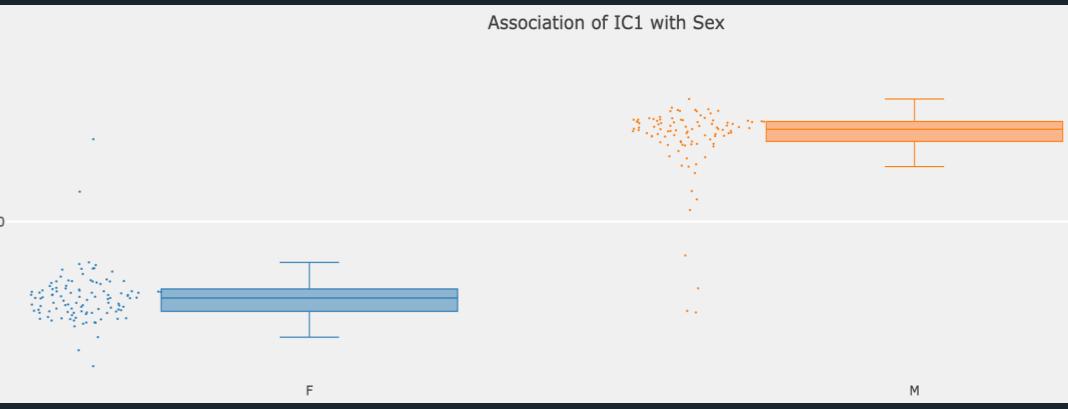
VAL	IC1	IC2	IC3	IC4	IC5	IC6	IC7	IC8	IC9	IC10
IC1	1			0.38		0.26				
IC2		1	0.58		0.32		0.06	0.2	0.13	0.5
IC3		0.58	1		0.28	0.11				0.0
IC4	0.38			1		0.45				
IC5		0.32	0.28		1	0.12			0.01	0.
IC6	0.26		0.11	0.45	0.12	1				
IC7		0.06					1	0.14	0.04	0.0
IC8		0.2					0.14	1	0.56	0.6
IC9		0.13			0.01		0.04	0.56	1	0.5
IC10		0.52	0.08		0.1		0.08	0.65	0.55	
IC11	0.12	0.11	0.4	0.16	0.32	0.62				
IC12	0.84	0.71		0.37	0.04					0

### Table of IC associations with categorical variables.

The values are -log10 p-values of the maximal t-test between any pairs of categories.

The tooltip for each number shows information on the score in the format <cat1>/<cat2><actual t-test score>, where cat1 and cat2 are two categories between which the association was observed.

VAL	IC1	IC2	IC3	IC4	IC5	IC6	IC7	IC8	IC9	IC10	IC11	IC12	IC13	IC14	IC15	IC16	IC17	IC18	IC19	
Plate_nr										3.11										
Disease	12.86	5.28	3.95	5.66	4.19	3.26				4.96	4.2	3.66	4.1		3.06	3.52		4.53	4.5	
Date_of_collection	22.78																			
Sex	81.06																			
Age_range		3.18																		
Specimen_type	4.58	3.01	4.48	3.27						4.19	3.52	3.32	3.35		3.18	0.37	0.37	0.37	0.37	
Location_approx	14.41	4.27	4.3	3.8	3.35					4.35	4.8	3.99	5.17	3.34	3.57		3.16	3.16	3.75	4
Grade																				
T																				
N																				
M																				
Stage	5.35																			

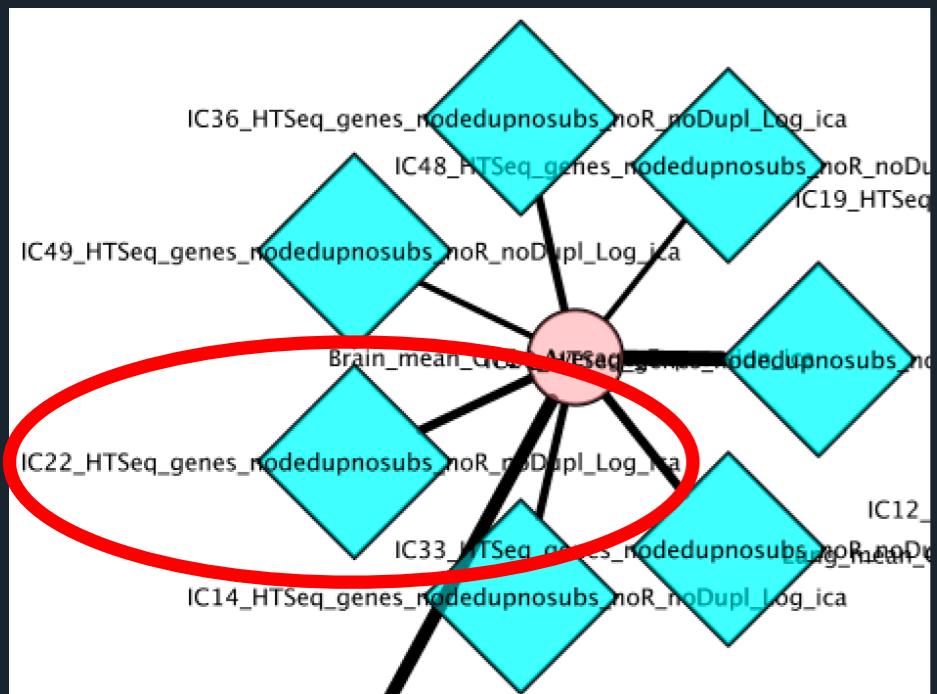
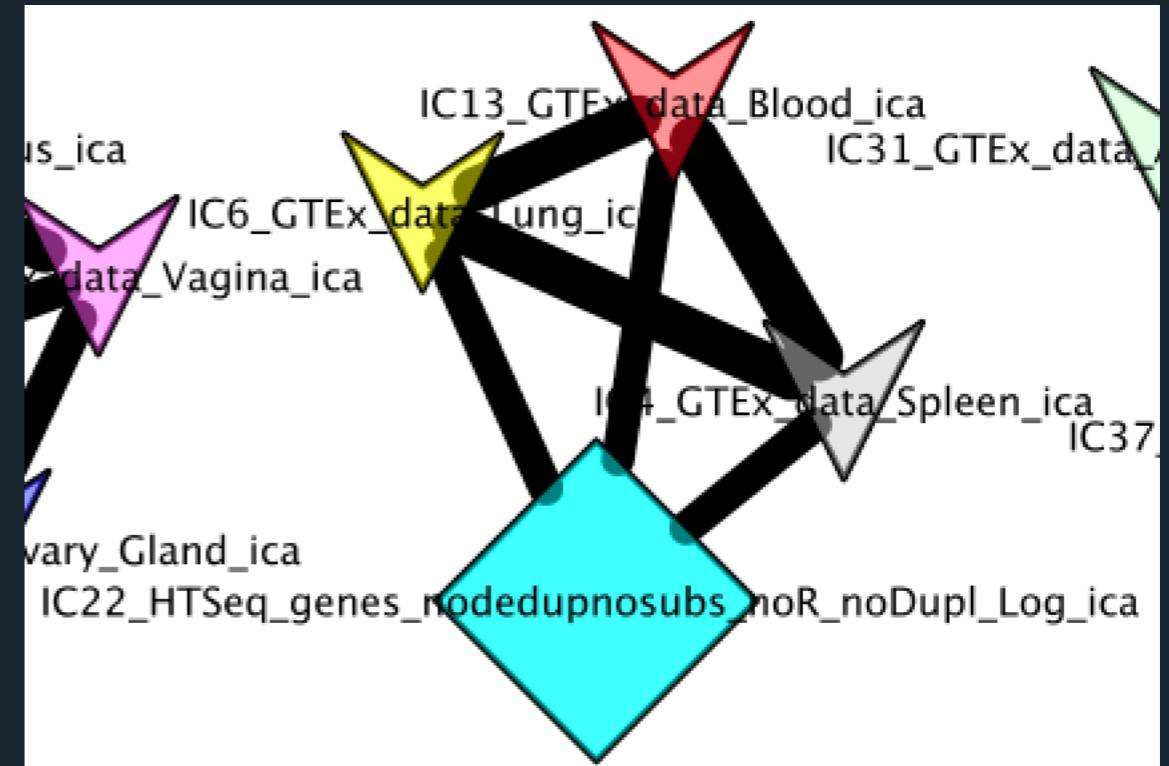


## EXAMPLE

# Glimpse at some results

### LIVE EXAMPLES

- ♦ IC33-50IC: Response to Virus
- ♦ Immune related components:
  - ♦ 50ICs: IC13, IC19, IC22, IC36, IC42
  - ♦ 100ICs: IC27, IC38, IC43 (interferon)
- ♦ IC22-50IC/ (Brain related metasample annot)



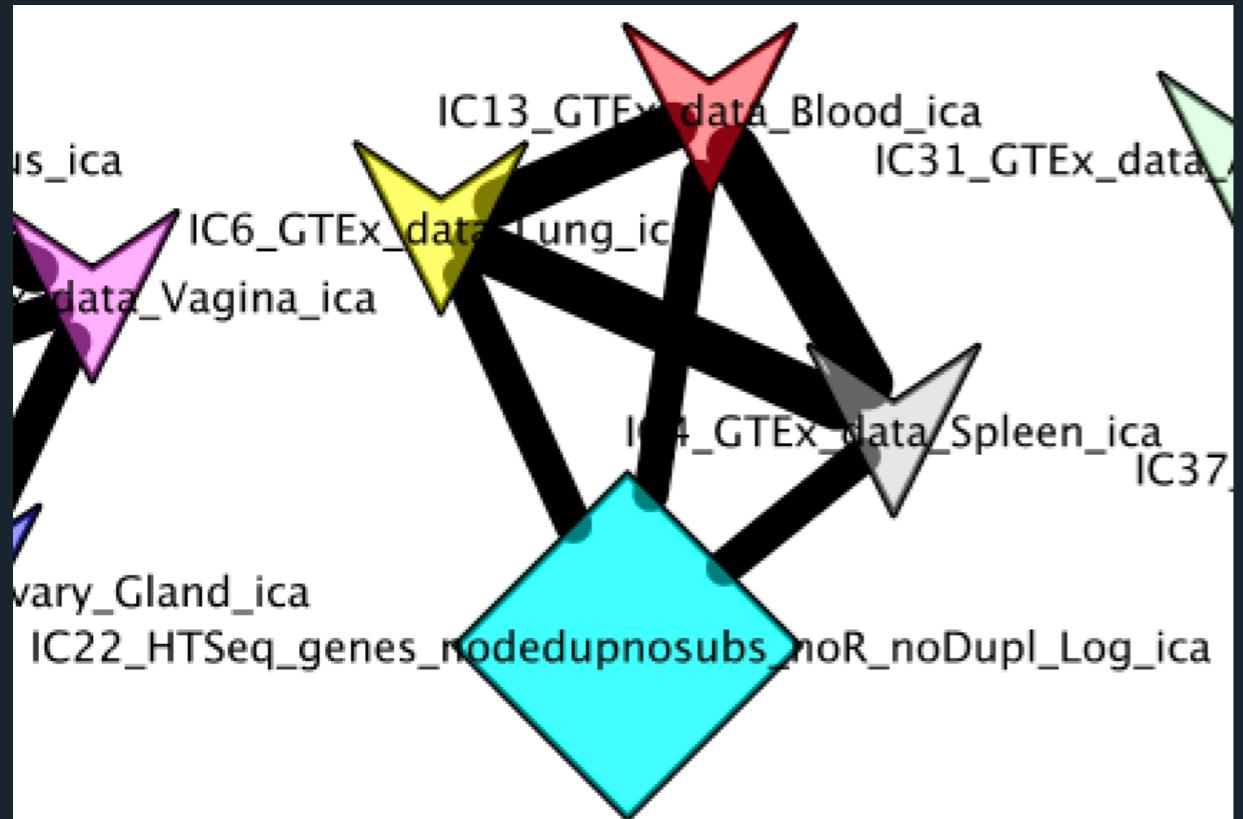
### Plasma

IC22+ (300 genes)

- 2: GO: Biological Process:[neutrophil degranulation](#)
- 3: GO: Cellular Component:[specific granule](#)
- 5: Mouse Phenotype:[abnormal mitosis](#)
- 7: Pathway:[Neutrophil degranulation](#)
- 9: Interaction:[FZR1 interactions](#)
- 11: Transcription Factor Binding Site:[V\\$E2F Q3](#)
- 13: Coexpression:[Genes down-regulated in comparison of peripheral blood mononuclear cells \(PBMC\) from patients with acute E. coli infection versus PBMC from patients with acute S. pneumoniae infection.](#)
- 14: Coexpression Atlas:[B cells, B.Pt.AA4-.BM, CD138+ AA4.1-CD43+, Bone marrow, avg-2](#)
- 16: MicroRNA:[Functional MTI \(Weak\)](#)

## EXAMPLE

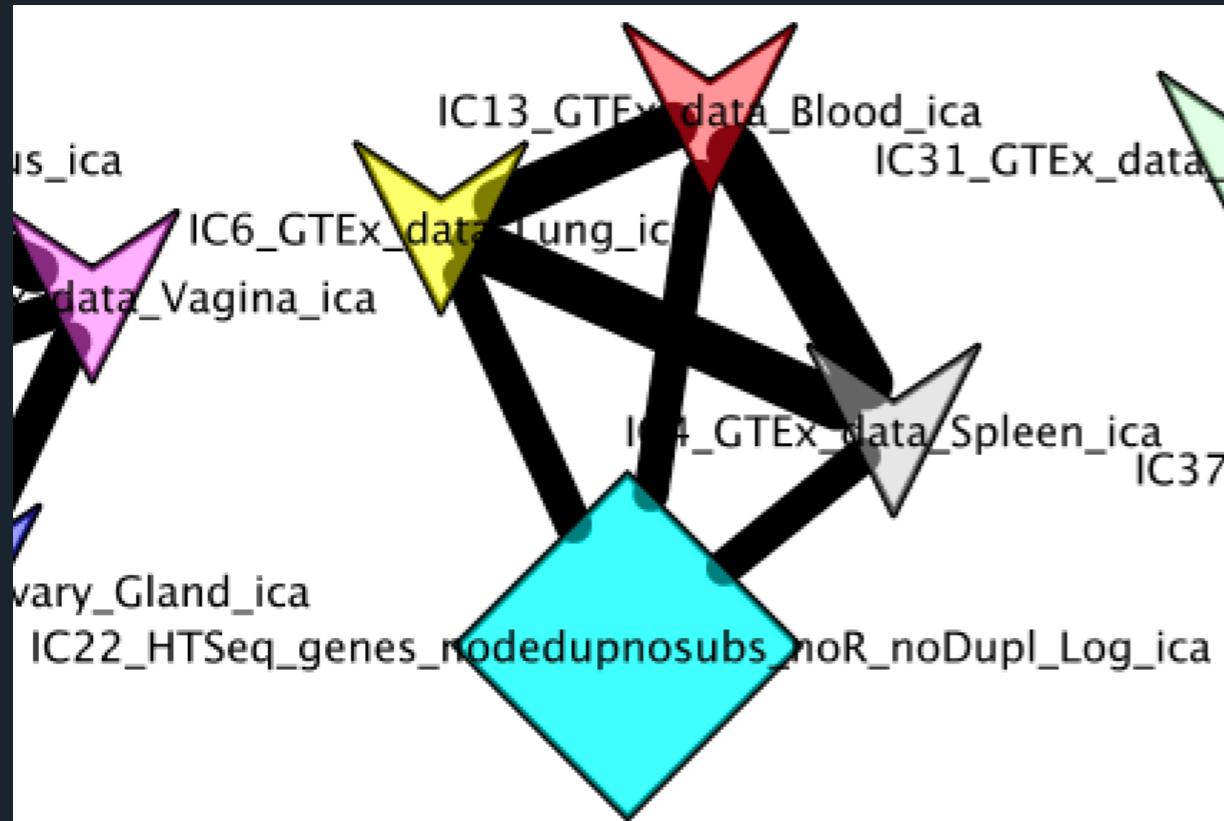
# Glimpse at some results



<b>Blood</b> <a href="#">IC13+ (102 genes)</a>	1: GO: Molecular Function: <a href="#">serine-type endopeptidase activity</a> 2: GO: Biological Process: <a href="#">neutrophil degranulation</a> 3: GO: Cellular Component: <a href="#">specific granule</a> 5: Mouse Phenotype: <a href="#">altered susceptibility to bacterial infection</a> 7: Pathway: <a href="#">Neutrophil degranulation</a> 13: Coexpression: <a href="#">Genes down-regulated in comparison of peripheral blood mononuclear cells (PBMC) from patients with acute E. coli infection versus PBMC from patients with acute S. pneumoniae infection.</a> 14: Coexpression Atlas: <a href="#">bone marrow</a>
<b>Lung</b> <a href="#">IC6+ (99 genes)</a>	1: GO: Molecular Function: <a href="#">serine-type endopeptidase activity</a> 2: GO: Biological Process: <a href="#">neutrophil degranulation</a> 3: GO: Cellular Component: <a href="#">secretory granule lumen</a> 5: Mouse Phenotype: <a href="#">altered susceptibility to bacterial infection</a> 7: Pathway: <a href="#">Neutrophil degranulation</a> 13: Coexpression: <a href="#">Genes down-regulated in comparison of peripheral blood mononuclear cells (PBMC) from healthy donors versus PBMCs from infant with acute RSV infection.</a> 14: Coexpression Atlas: <a href="#">bone marrow</a>
<b>Plasma</b> <a href="#">IC22+ (300 genes)</a>	2: GO: Biological Process: <a href="#">neutrophil degranulation</a> 3: GO: Cellular Component: <a href="#">specific granule</a> 5: Mouse Phenotype: <a href="#">abnormal mitosis</a> 7: Pathway: <a href="#">Neutrophil degranulation</a> 9: Interaction: <a href="#">FZR1 interactions</a> 11: Transcription Factor Binding Site: <a href="#">V\$E2F_Q3</a> 13: Coexpression: <a href="#">Genes down-regulated in comparison of peripheral blood mononuclear cells (PBMC) from patients with acute E. coli infection versus PBMC from patients with acute S. pneumoniae infection.</a> 14: Coexpression Atlas: <a href="#">B cells, B.Pl.AA4-.BM, CD138+ AA4.1-CD43+, Bone marrow, avg-2</a> 16: MicroRNA: <a href="#">Functional MTI (Weak)</a>
<b>Spleen</b> <a href="#">IC4+ (127 genes)</a>	1: GO: Molecular Function: <a href="#">serine-type endopeptidase activity</a> 2: GO: Biological Process: <a href="#">neutrophil degranulation</a> 3: GO: Cellular Component: <a href="#">secretory granule lumen</a> 5: Mouse Phenotype: <a href="#">altered susceptibility to bacterial infection</a> 6: Domain: <a href="#">Trypsin</a> 7: Pathway: <a href="#">Neutrophil degranulation</a> 13: Coexpression: <a href="#">Genes down-regulated in comparison of peripheral blood mononuclear cells (PBMC) from patients with acute E. coli infection versus PBMC from patients with acute S. pneumoniae infection.</a> 14: Coexpression Atlas: <a href="#">bone marrow</a>

## EXAMPLE

# Glimpse at some results

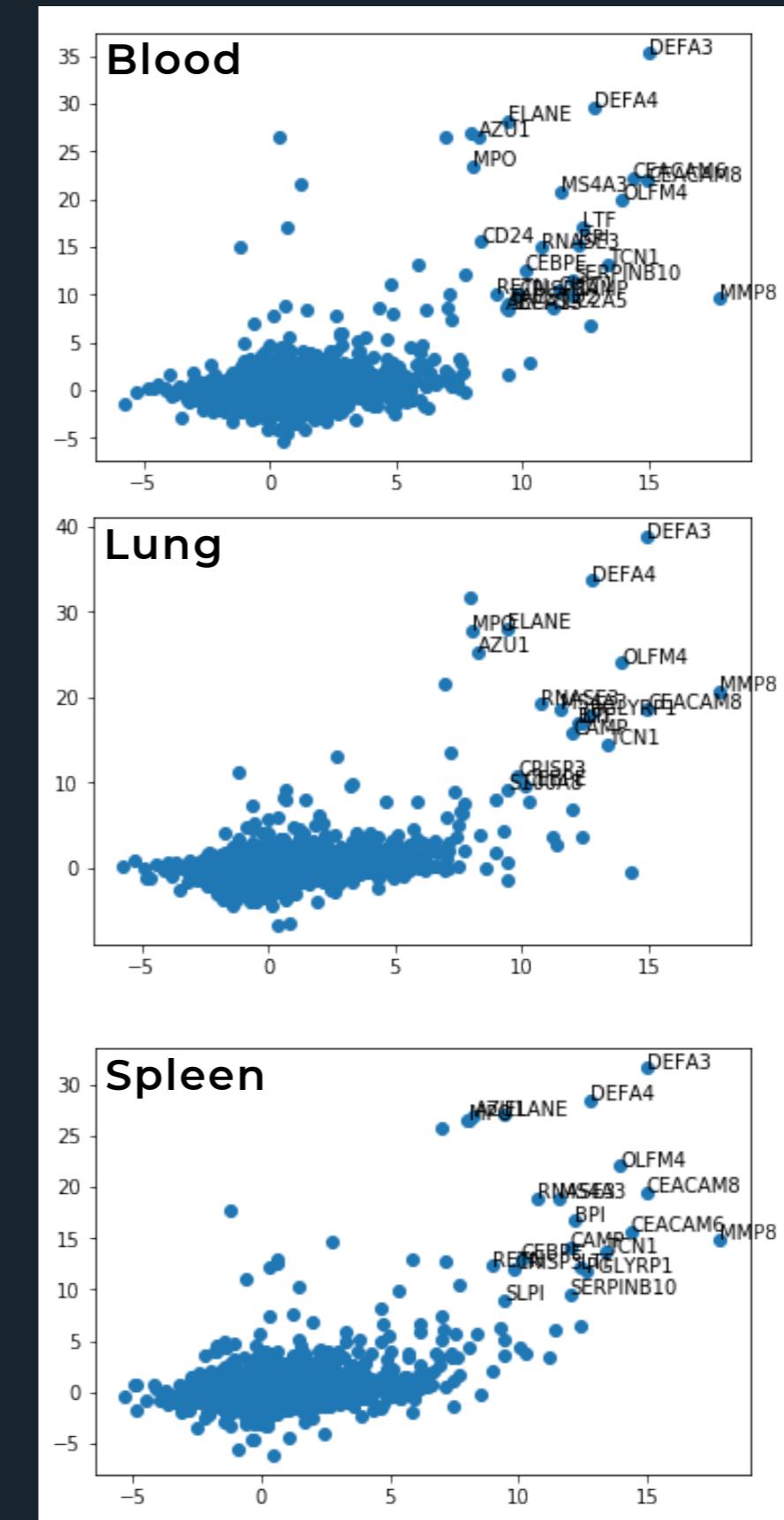


### Plasma

[IC22+ \(300 genes\)](#)

- 2: GO: Biological Process:[neutrophil degranulation](#)
- 3: GO: Cellular Component:[specific granule](#)
- 5: Mouse Phenotype:[abnormal mitosis](#)
- 7: Pathway:[Neutrophil degranulation](#)
- 9: Interaction:[FZR1 interactions](#)
- 11: Transcription Factor Binding Site:[V\\$E2F\\_Q3](#)
- 13: Coexpression:[Genes down-regulated in comparison of peripheral blood mononuclear cells \(PBMC\) from patients with acute E. coli infection versus PBMC from patients with acute S. pneumoniae infection.](#)
- 14: Coexpression Atlas:[B cells, B.Pl.AA4-.BM, CD138+ AA4.1-CD43+, Bone marrow, avg-2](#)
- 16: MicroRNA:[Functional MTI \(Weak\)](#)

Pearson correlation ~ 0.5



Top genes intersection:

AZU1  
BPI  
CAMP  
CEACAM8  
CEBPE  
CRISP3  
DEFA3  
DEFA4  
ELANE  
LTF  
MMP8  
MPO  
MS4A3  
OLFM4  
RNASE3  
TCN1