

# GO Analyses on PCA Clustering

PCA on Bulk RNA Seq and Single Cell Data

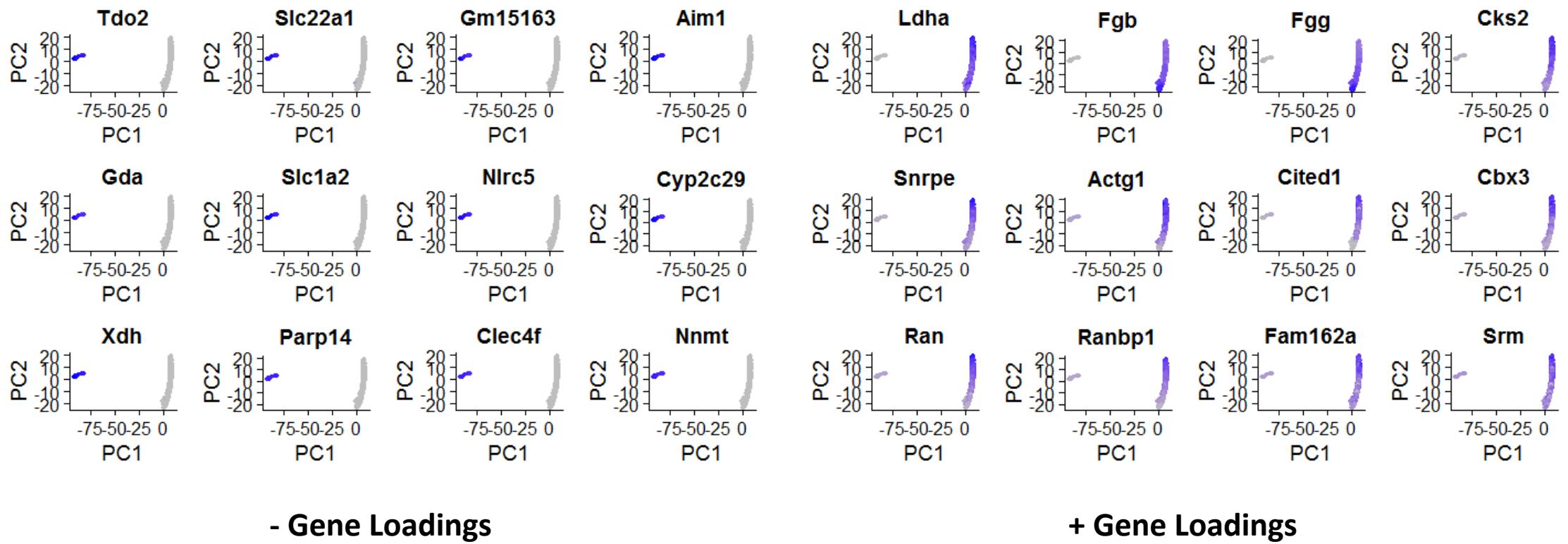
# Gene Loadings and Cluster Variation

- Genes with highest in magnitude loading scores were considered to attribute to variation in each principal component
- Loading scores were verified to match up with PC1 and PC2 coordinates
- 3 Main Analyses:
  - 50 Genes from PC1-
  - 50 Genes from PC1+
  - 100 Genes from PC2+&-

# Gene Loadings and Cluster Variation

- Analyses: BinGo, ClueGO, and DAVID
  - All P-Value and Frequency weighted results were only from ClueGO and BinGo
    - If GO term showed up twice, both P-values were multiplied, considered mutually exclusive
    - Frequency value was kept from ClueGO
  - GO term names of equal weights were used from all three GO analyses
  - All P-values $<0.05$ , Bonferroni correction test was used so that false positives did not occur

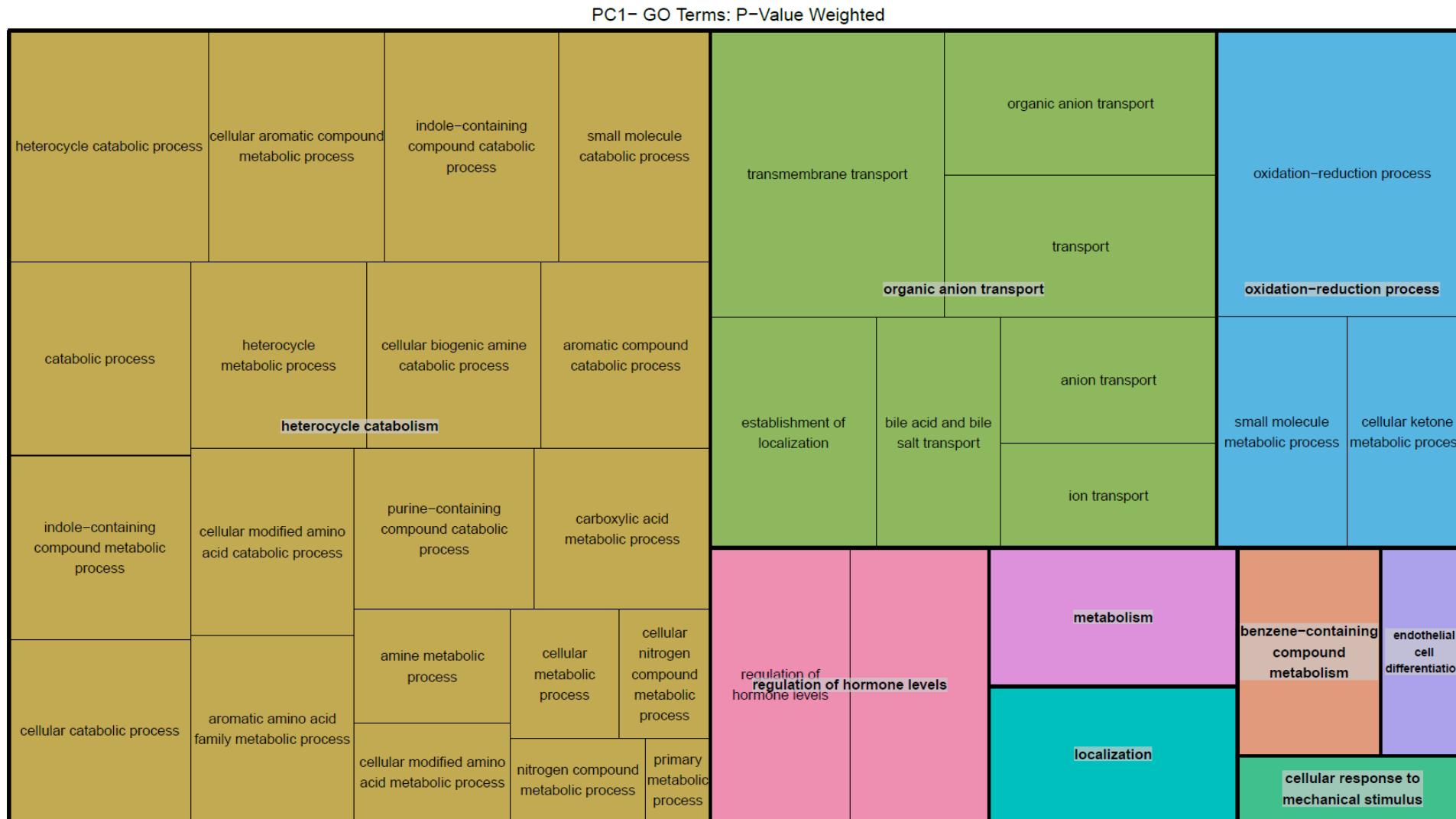
# PC1 Variation



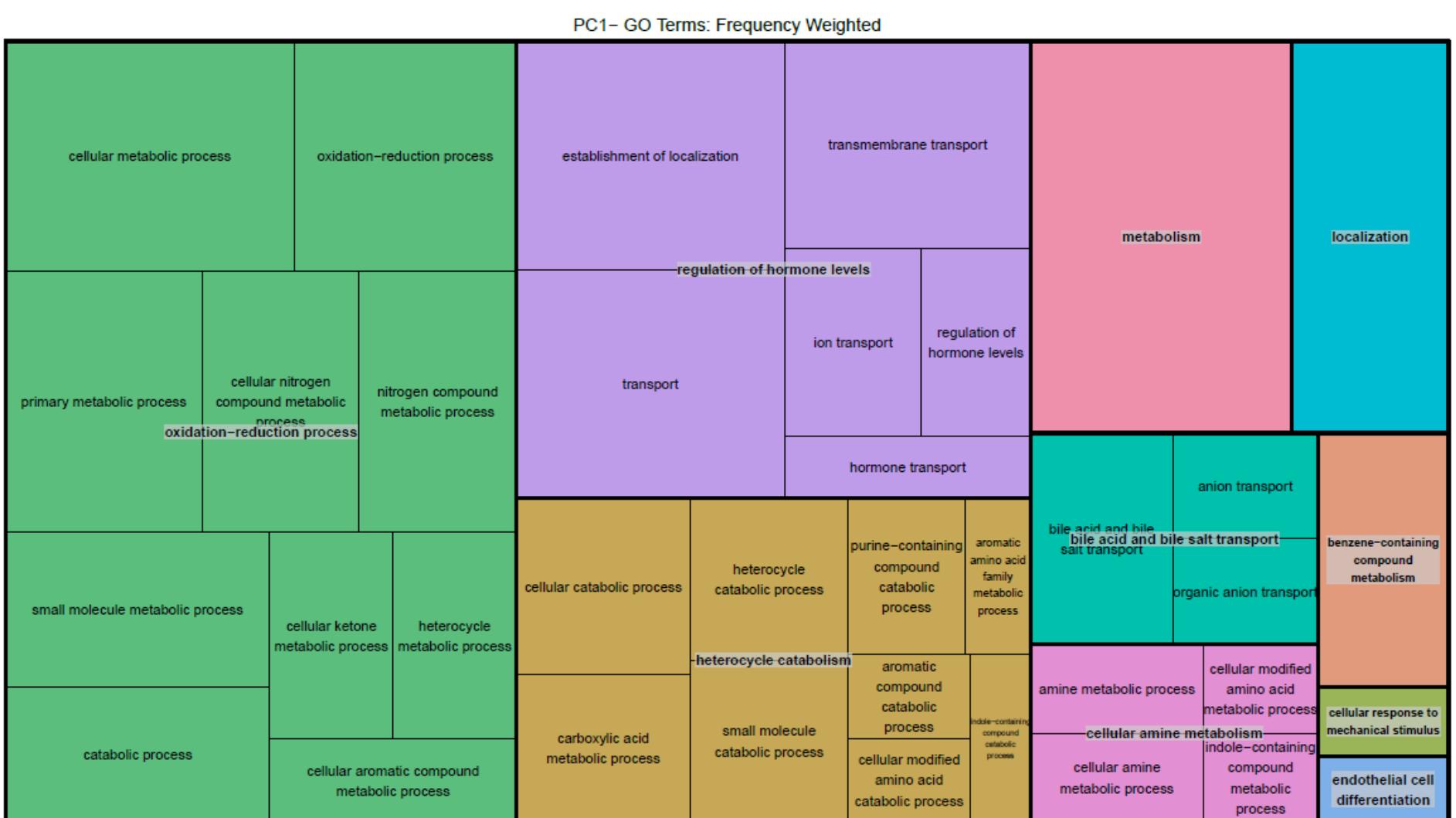
# PC1 Variable Gene- Dosed Cluster

- 50 Genes with largest in magnitude negative loading scores
- Terms that were in both PC1+ and PC1- were removed to see which genes are contributing to the variation of each cluster

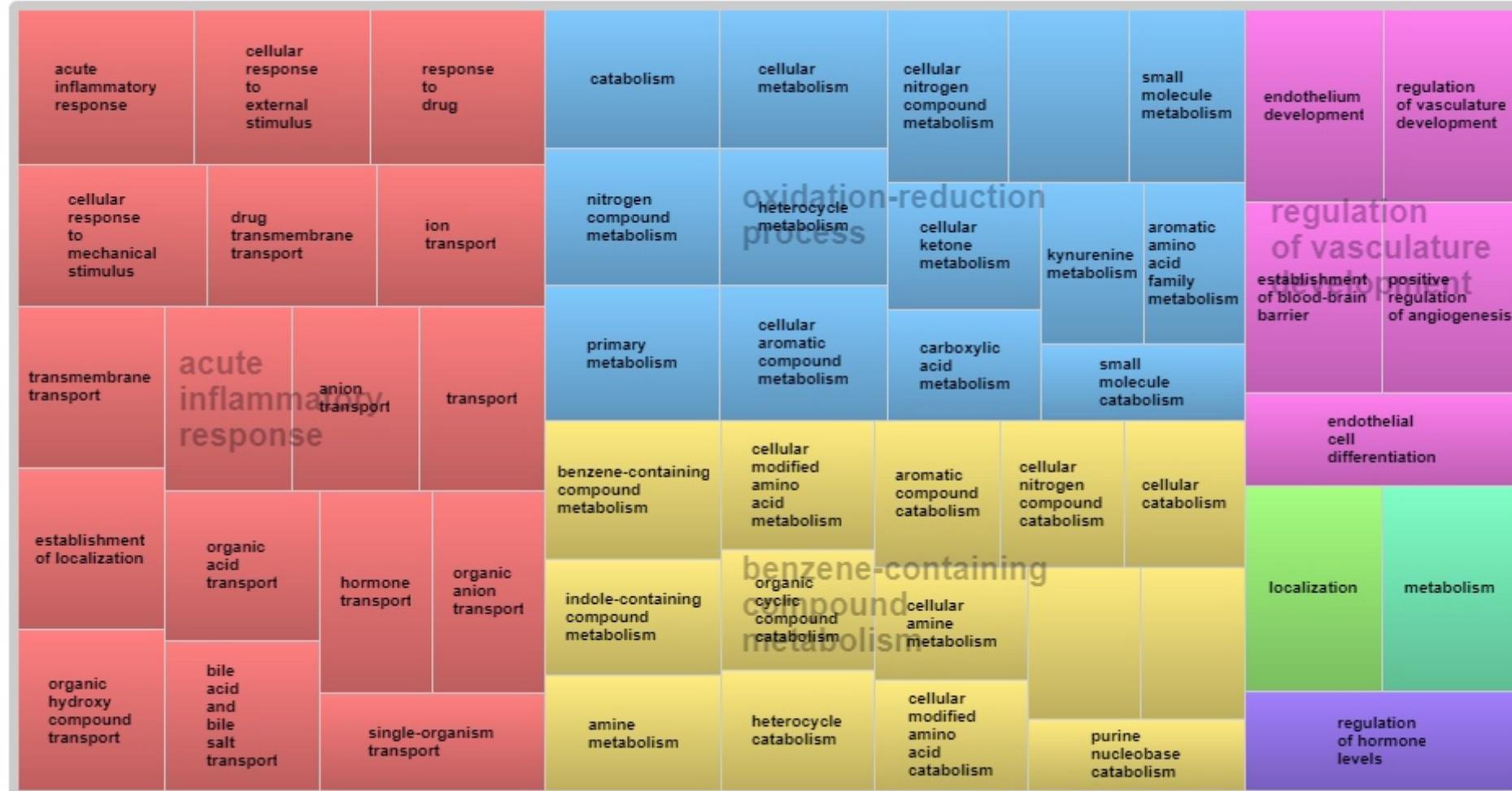
# PC1- w/ P-vals, P-Value Weighted



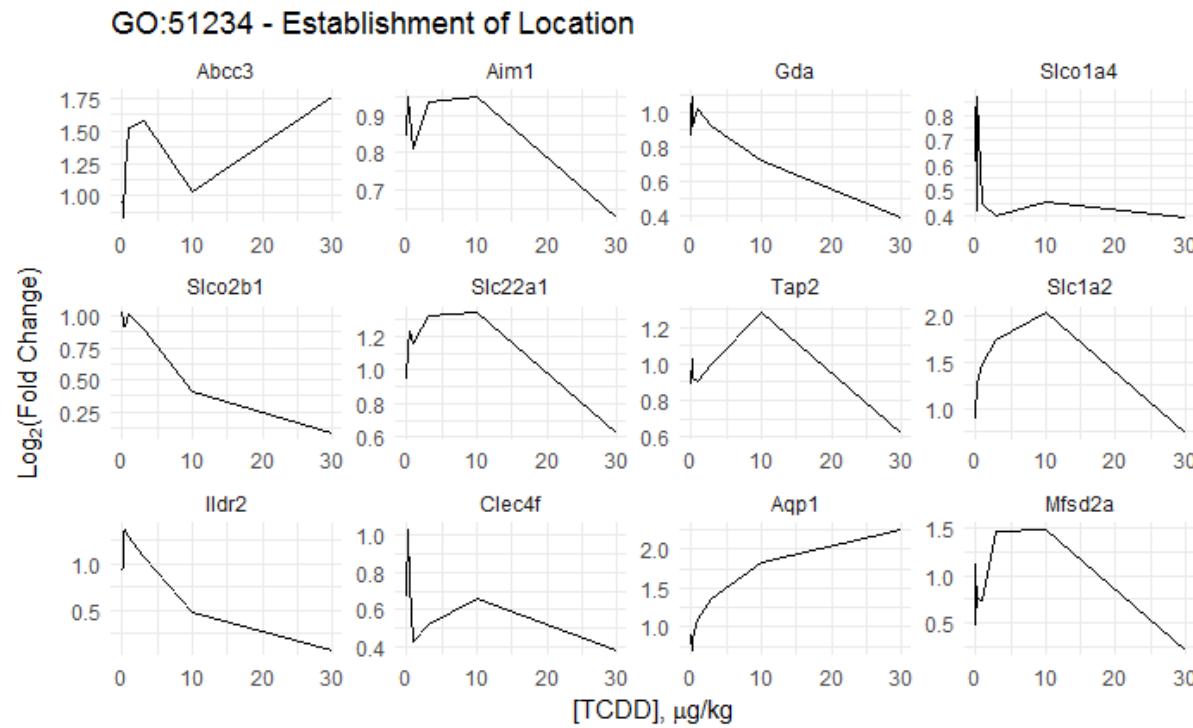
# PC1- w/ P-vals, Weighted Gene Frequency



# PC1- Results: DAVID, BinGo, & ClueGO



# PC1- Results: Establishment of Location Genes



# PC1-: GOSummaries: Top 100

## PC1- GO Terms: Top 100



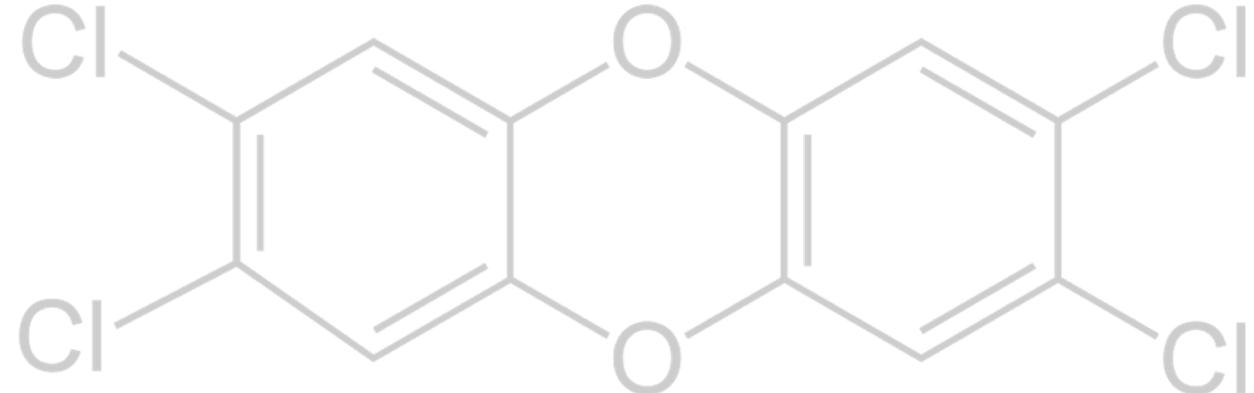
# PC1-: GOSummaries: Top 50

## PC1- GO Terms: Top 50



# PC1 Variable Gene- Dosed Cluster Observations

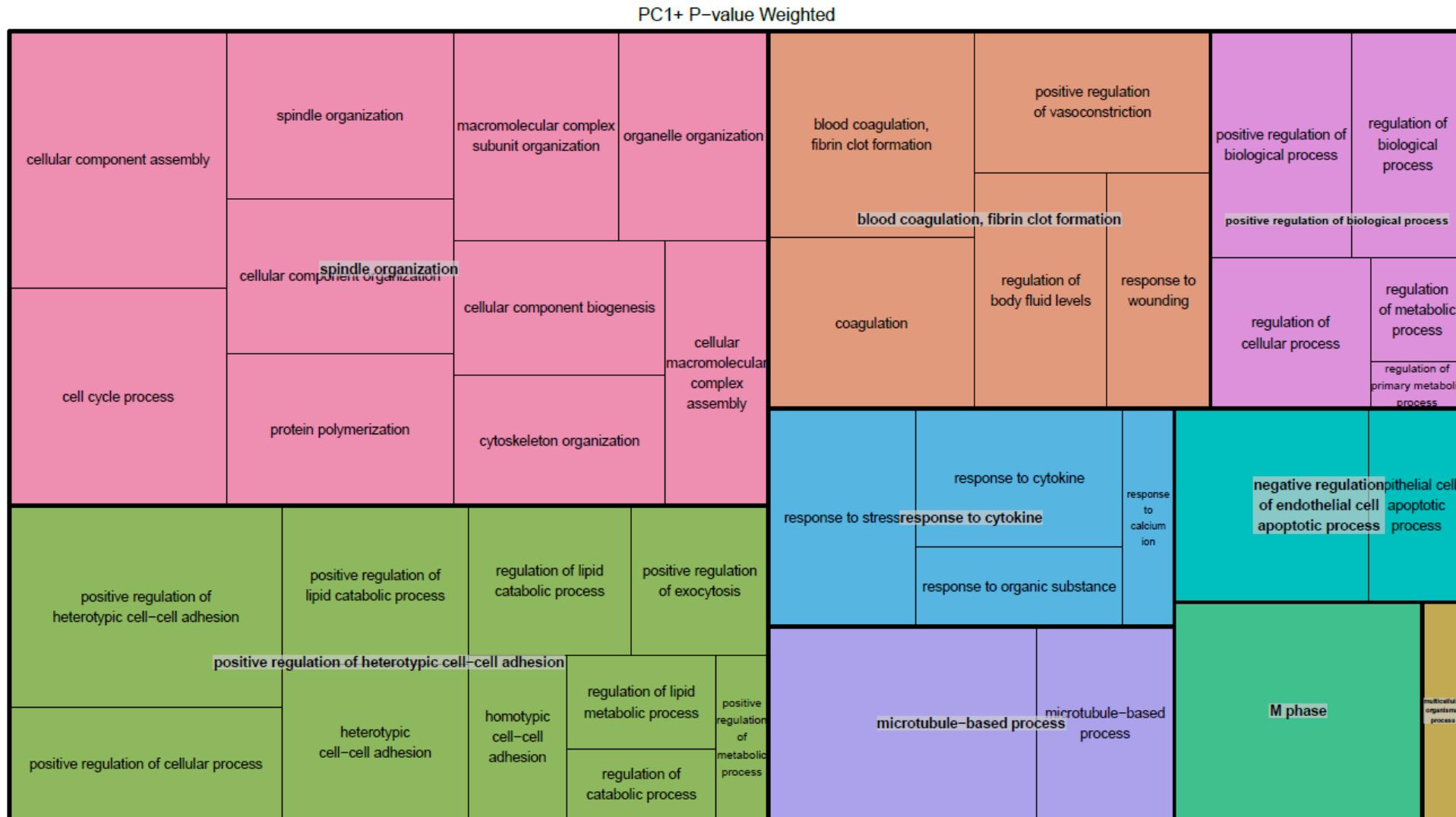
- Gene networks involved in TCDD catabolism attribute to the variation in PC1 of the dosed cluster
- TCDD could potentially have aromatic and anionic catabolic intermediates



# PC1 Variable Gene+: Single Cell Cluster

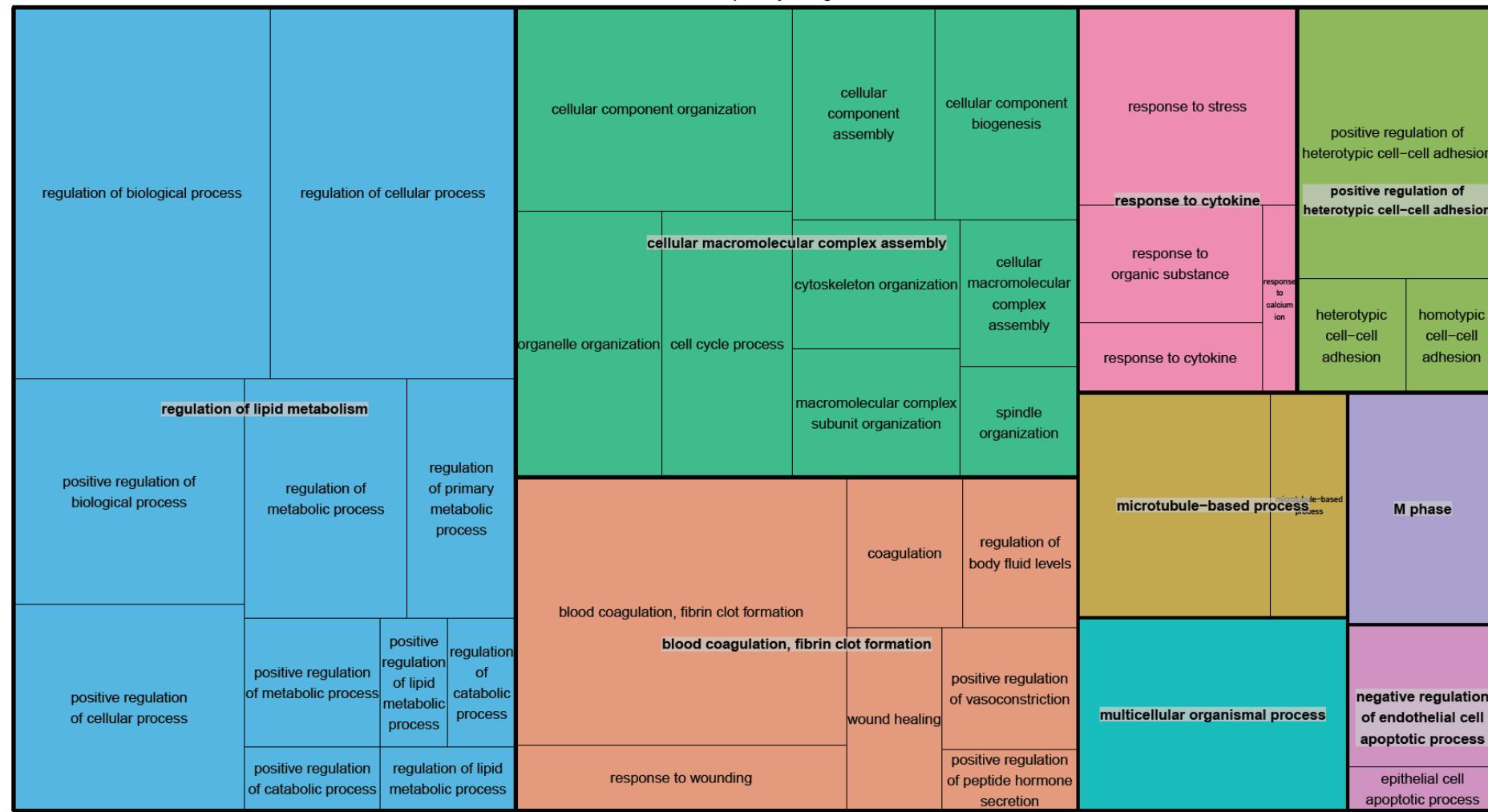
- 50 Genes with largest in magnitude positive loading scores
- Terms that were in both PC1+ and PC1- were removed to see which genes are contributing to the variation of each cluster

# PC1+ Results w/ P-vals, P-Value Weighted



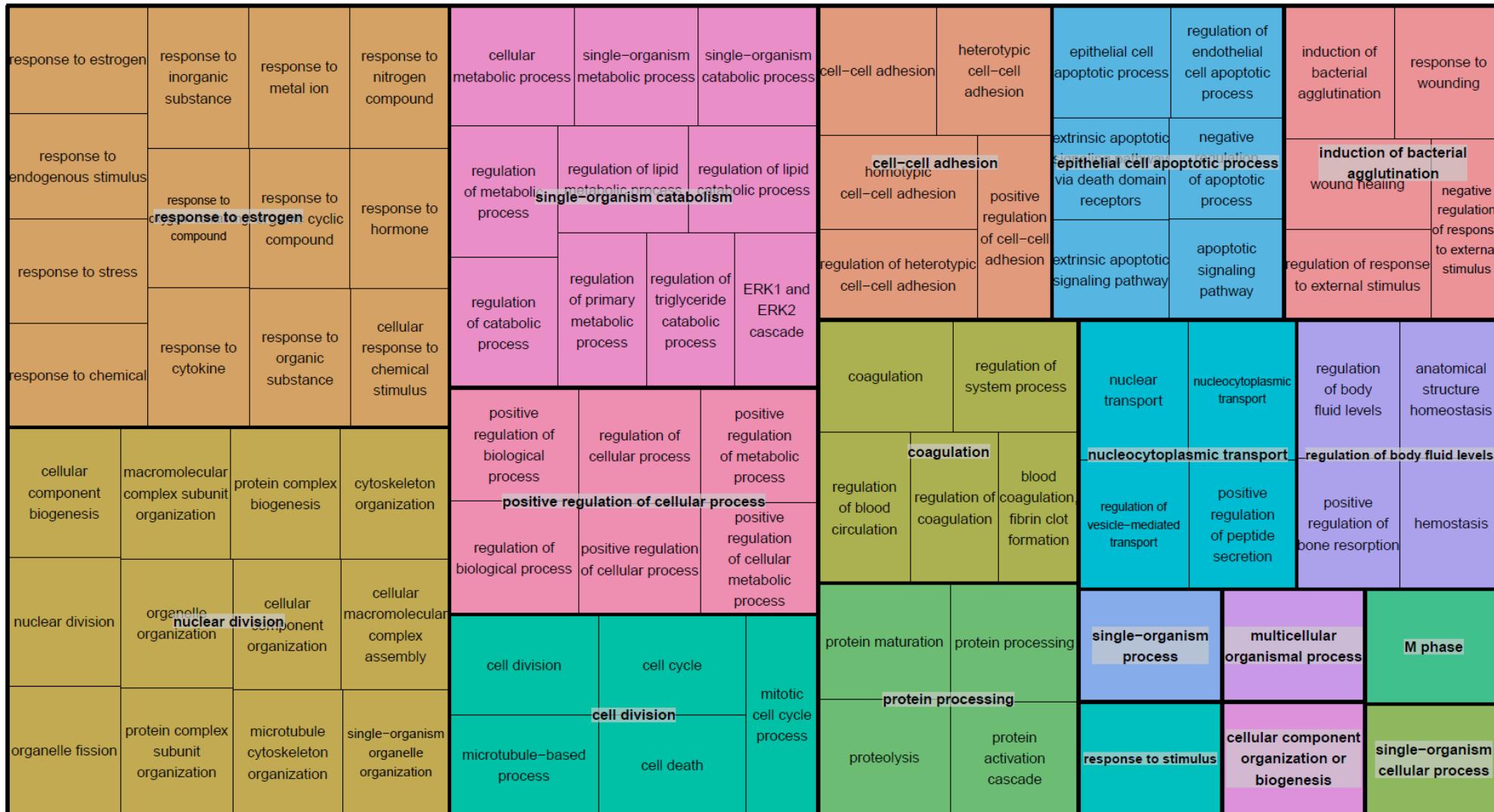
# PC1+ Results w/ Dvolc Frequency Weighted

PC1+: Frequency Weighted



# PC1+ Results: DAVID, BinGo, & ClueGO

PC1+ GO Terms: DAVID, BinGo, and ClueGO

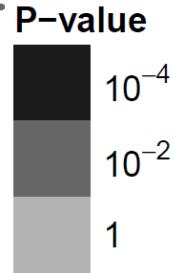


# PC1+: GOSummaries Word Cloud

## PC1 Plus GO Terms

response to calcium ion      blood coagulation, fibrin clot f...  
plasminogen activation      fibrinolysis      positive regulation of substrate...  
**platelet aggregation**      negative regulation of endotheli...  
positive regulation of heterotyp...

positive regulation of exocytosis      regulation of apoptotic signalin...  
negative regulation of extrinsic...  
**positive regulation of vasoconst...**  
induction of bacterial agglutina...  
response to inorganic substance      platelet degranulation  
response to interleukin-1



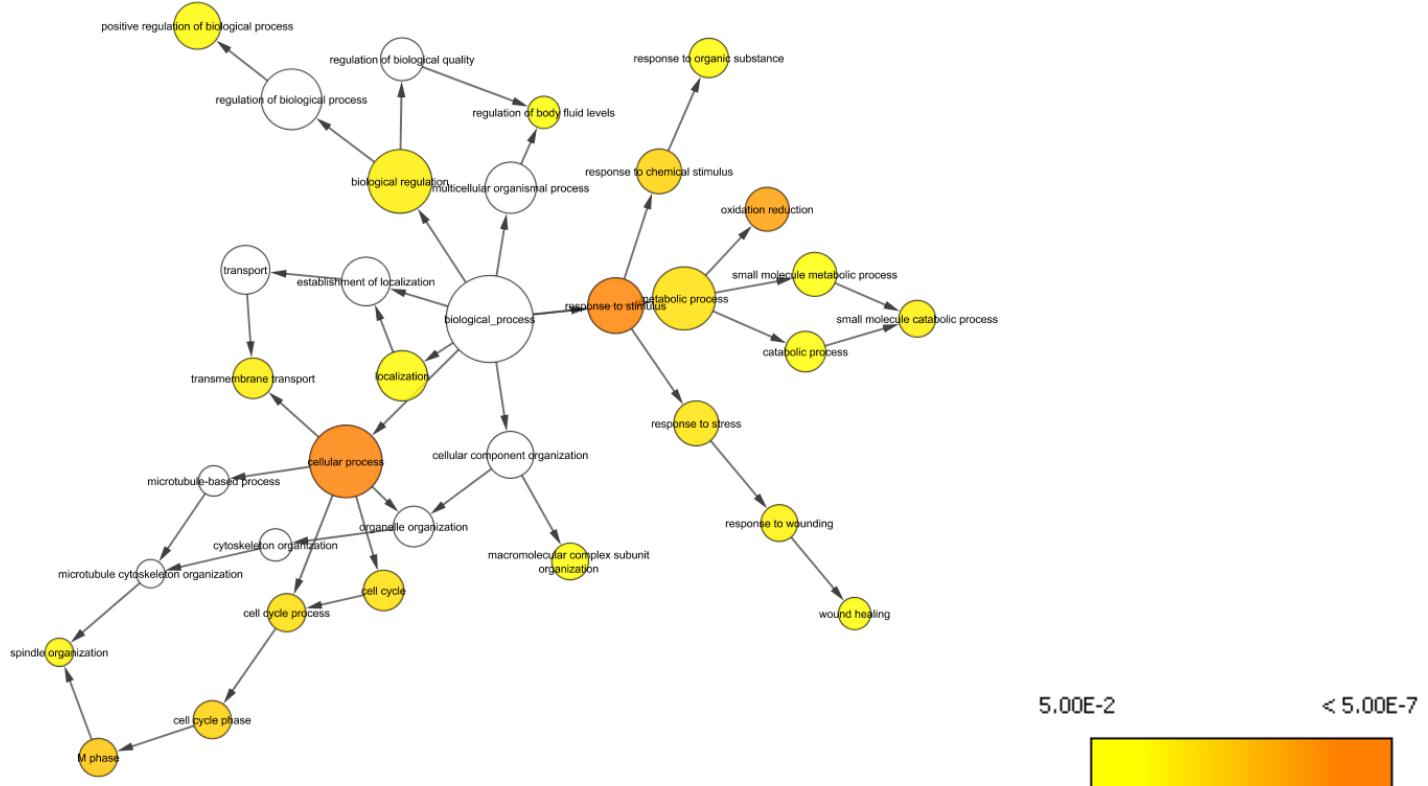
# PC1 Variable Gene+ Observations

- Some GO terms associated with inflammation are present, could have happened because of dissection
  - Fibrin clot formation and blood coagulation
  - Apoptotic processes
- Single organism processes are present

# PC1 +&- Variable Genes

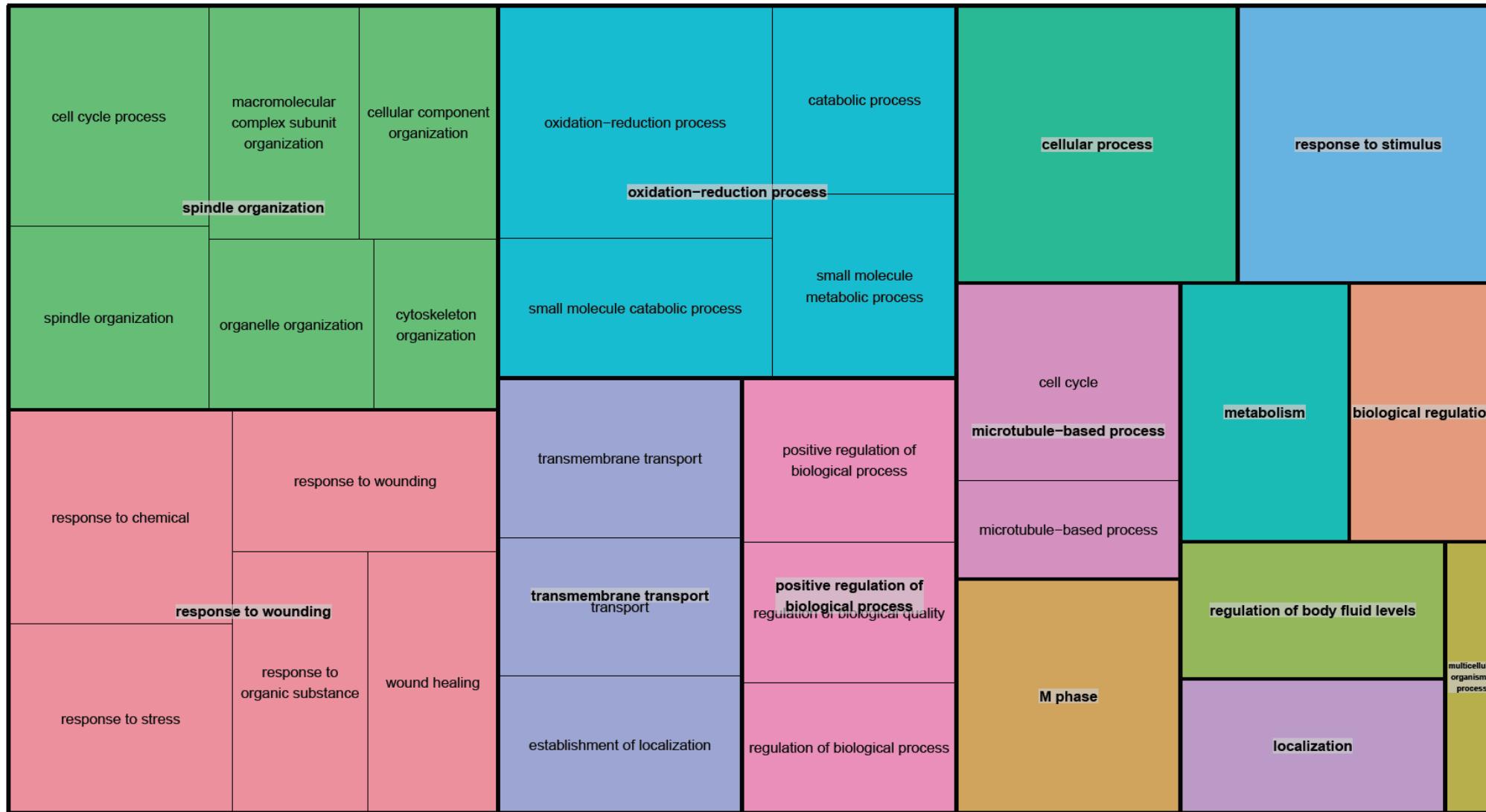
- Top 100 genes were found by taking the highest in magnitude of the gene loading values
- BiNGO is only used
  - Hypergeometric test
  - $P < 0.05$ , Bonferroni Family-Wise Error Rate correction

# PC1 +&- : BiNGO Network



# DC1 +& - · BiNGO Treemap

PC1 +& -: BiNGO P-Value Weighted

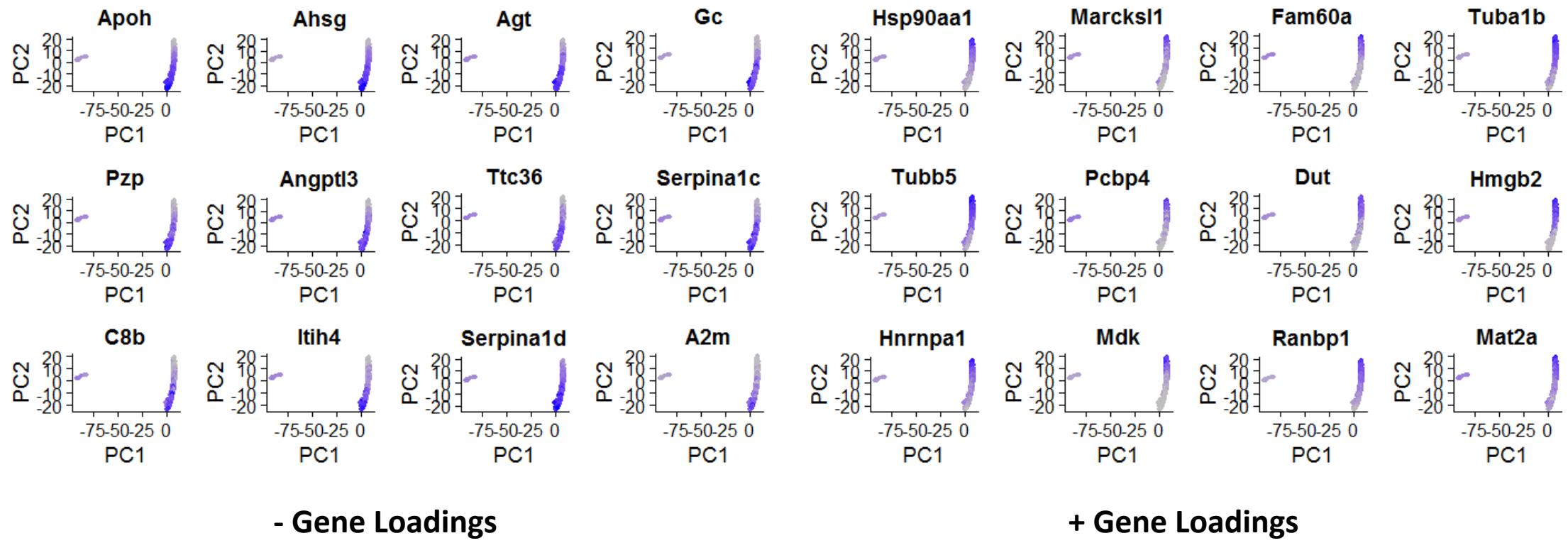


## PC2 +&- Single Cell Cluster

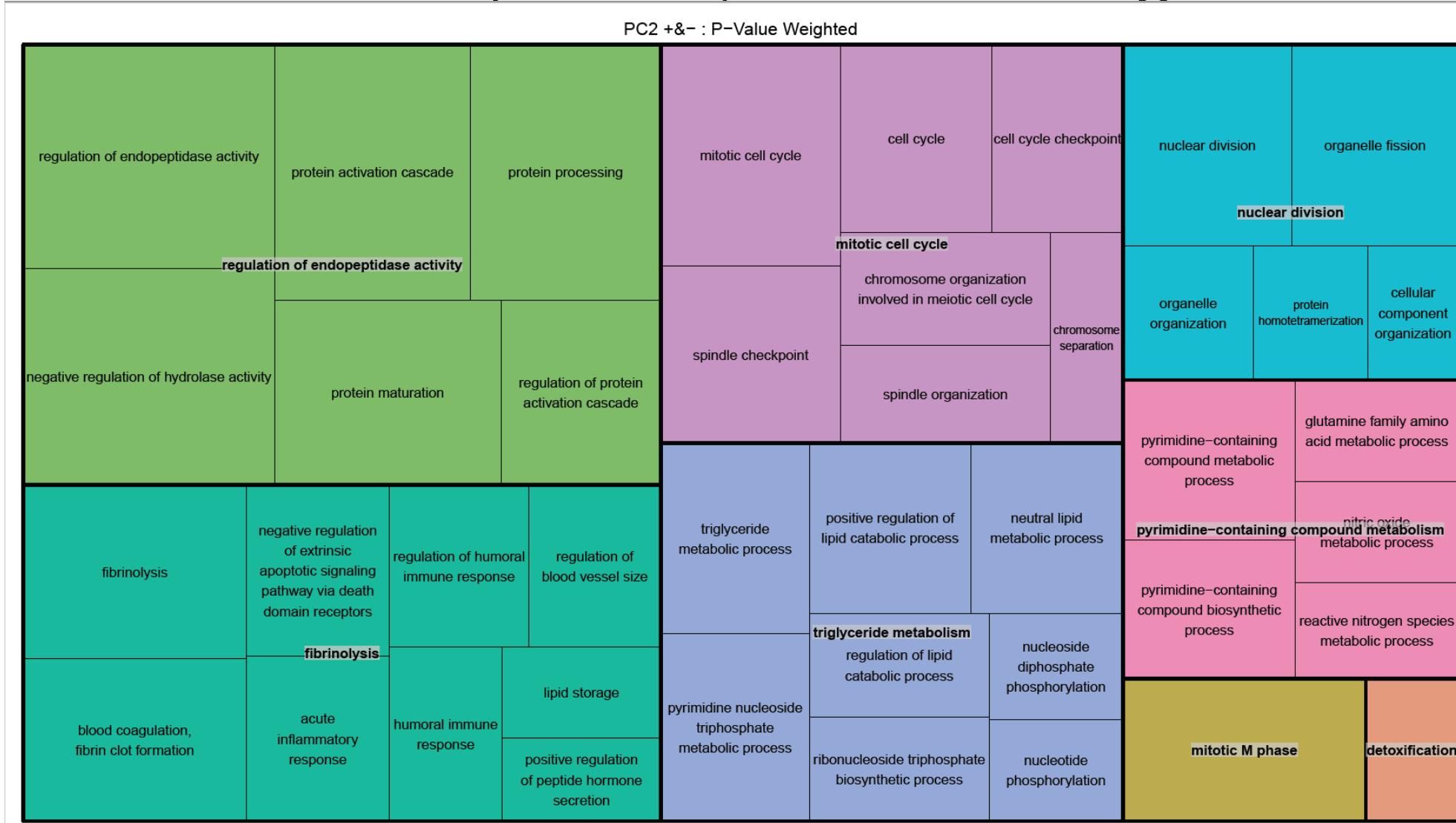
- 100 genes with highest magnitude loading scores underwent analyses
  - 50 + loading scores and 50 – loading scores

# PC2 Variation

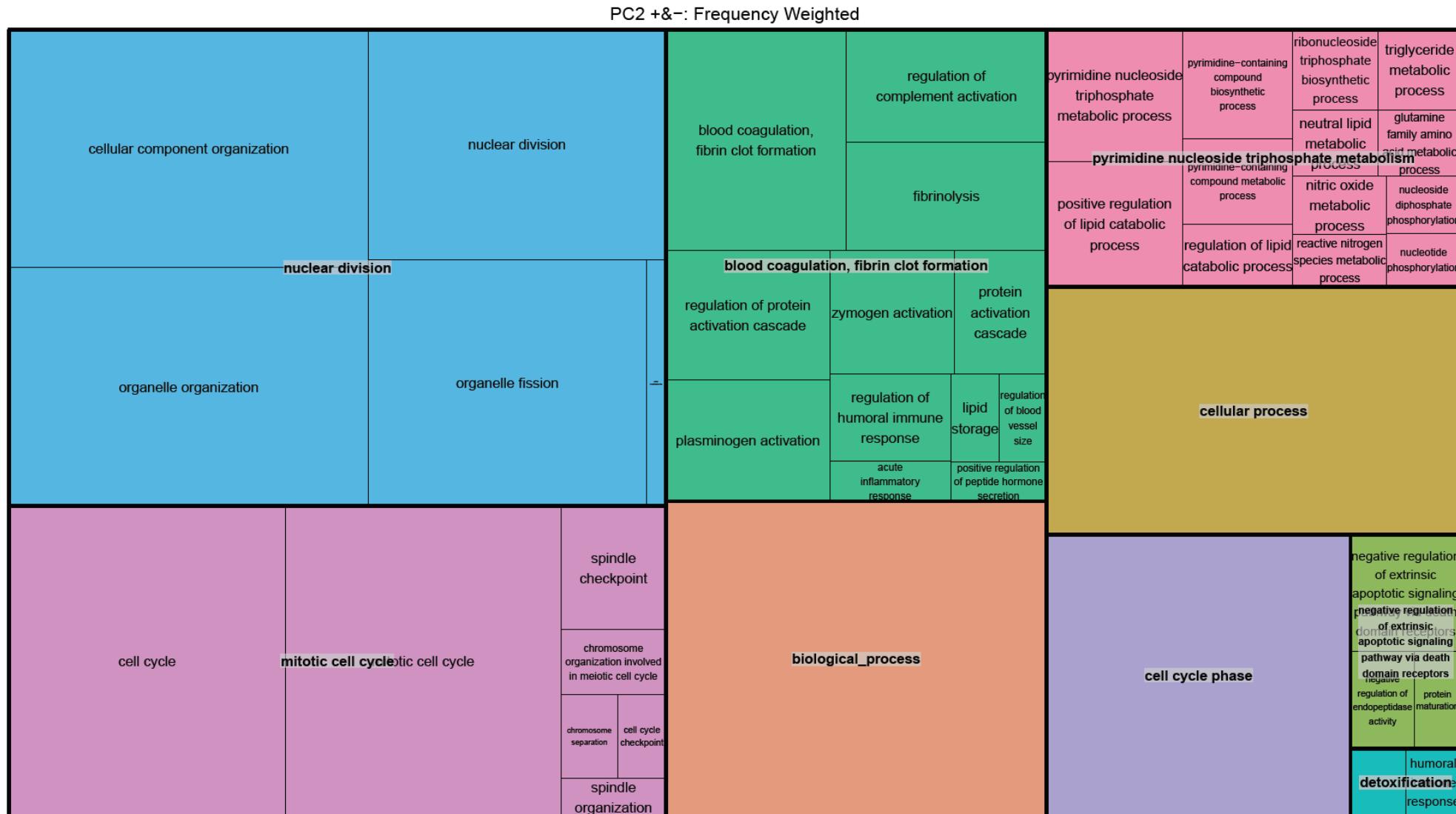
+ & - Gene loadings were combined for GO analyses.



# PC2 +&- w/ P-vals, P-Value Weighted



# PC2 +&- w/ Pvals, Frequency Weighted

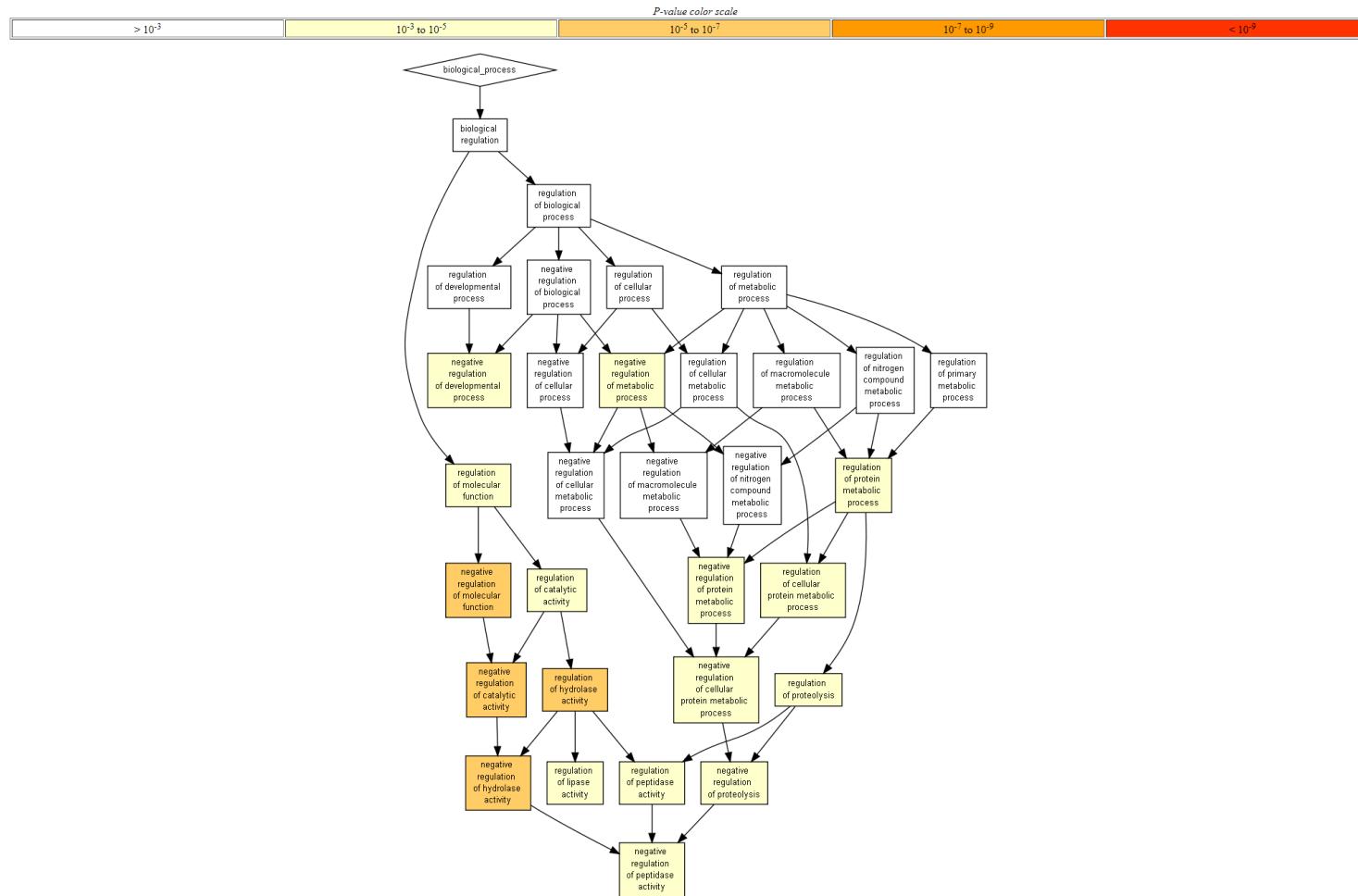


# PC2 +&-: DAVID, BinGo, & ClueGO

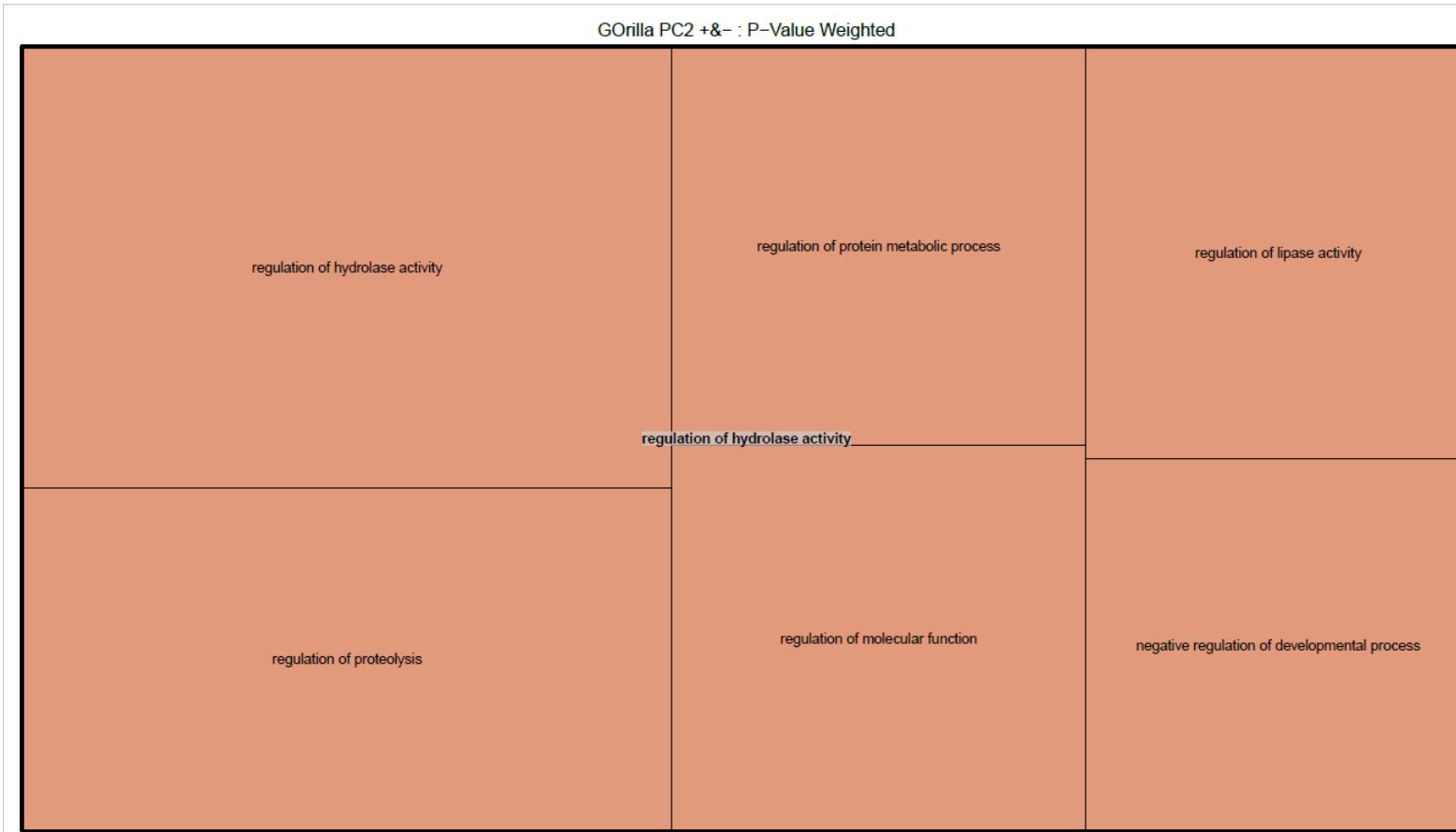
PC2 +&-: DAVID, BinGo, and ClueGO Names

response to endogenous stimulus	response to metal ion	response to nitrogen compound	response to oxygen-containing compound	response to wounding	acute inflammatory response	response to lipid	reactive nitrogen species metabolic process	pyrimidine-containing compound metabolic process	DNA replication	pyrimidine-containing compound biosynthetic process	protein homotetramerization	protein complex biogenesis	protein complex assembly	lipid localization	nuclear transport					
humoral immune response	response to chemical	response to organic cyclic compound	response to other organism	proteolysis		defense response		inflammatory response		nitric oxide metabolic process	carbohydrate metabolic process	organonitrogen compound metabolic process	organophosphate catabolic process	nucleo <sup>lipid</sup> transport	lipid localization					
				regulation of response to external stimulus		extrinsic apoptotic signaling pathway via death domain receptors		single-organism carbohydrate catabolic process	single-organism catabolic process					lipid storage						
		response to steroid hormone	humoral immune response	regulation of acute inflammatory response		regulation of defense response					cellular component organization	nuclear division	single-organism organelle organization							
detoxification	response to drug	response to humoral immune response	steroid hormone response	regulation of response to stress	regulation of acute inflammatory response		regulation of defense response		regulation of proteolysis		regulation of molecular function	positive regulation of biological process	negative regulation of molecular function	regulation of hydrolase activity						
					regulation of extrinsic apoptotic signaling pathway		regulation of cellular protein metabolic process		regulation of proteolysis											
					regulation of wound healing		regulation of response to stress		regulation of cellular protein metabolic process											
response to external stimulus	response to inorganic substance	response to organic substance	wound healing	regulation of response to stress	regulation of reproductive process		regulation of animal organ regeneration		regulation of coagulation		regulation of angiogenesis	positive regulation of biological process	negative regulation of molecular function	negative regulation of molecular function						
					regulation of multi-metabolic process		female pregnancy regulation of coagulation		positive regulation of multicellular organismal process											
					regulation of protein activation cascade		regulation of endopeptidase activity		negative regulation of apoptotic process											
response to stress	protein maturation	response to hormone	protein activation cascade	regulation of protein activation cascade	regulation of positive regulation of blood circulation		blood coagulation, fibrin clot formation		regulation of body fluid levels		regulation of homeostatic process	positive regulation of homeostatic process	single-organism process	metabolic process						
					regulation of nucleoside triphosphate biosynthetic process		regulation of ribonucleoside triphosphate biosynthetic process		regulation of lipid metabolic process											
					regulation of nucleoside diphosphate phosphorylation		regulation of ribonucleoside phosphorylation		regulation of nucleotide-containing small molecule metabolic process											
catabolic process	organic substance metabolic process	cellular metabolic process	steroid metabolic process	neutral lipid metabolic process	positive regulation of lipid metabolic process		positive regulation of lipid metabolic process		cell division		cell death	spindle organization	chromosome organization involved in meiotic cell cycle	regulation of body fluid levels						
					oxidation-reduction process		triglyceride process		cellular metabolic process											
					lipid oxidation-reduction process		triglyceride process		oxidation-reduction process											
carbohydrate derivative metabolic process	phosphorus metabolic process	phosphate-containing compound metabolic process	lipid metabolic process	triglyceride process	nucleoside diphosphate phosphorylation		ribonucleoside biosynthetic process		cell cycle phase transition		cell death	spindle assembly checkpoint	chromosome organization involved in meiotic cell cycle	regulation of body fluid levels						
					nucleoside diphosphate phosphorylation		ribonucleoside biosynthetic process		microtubule-based process											
					nucleoside triphosphate phosphorylation		carboxylic acid metabolic process		ERK1 and ERK2 cascade											
primary metabolic process	single-organism metabolic process	oxidation-reduction process	small molecule metabolic process	pyrimidine nucleoside triphosphate metabolic process	nucleotide phosphorylation		carboxylic acid metabolic process		cell cycle		mitotic cell cycle	positive regulation of mitotic cell cycle	reproduction	cellular process						
					nucleotide phosphorylation		carboxylic acid metabolic process		ERK1 and ERK2 cascade											
					nucleotide triphosphate phosphorylation		carboxylic acid metabolic process		ERK1 and ERK2 cascade											

# PC2 +&-: Gorilla Network



# PC2 +&-: Gorilla Treemap

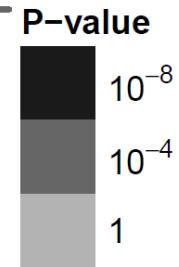


# PC2+&-: GOSummaries Word Cloud

## PC2+&- GO Terms: Top 100

protein processing regulation of response to extern...  
negative regulation of endopeptid...  
**negative regulation of hydrolase...**  
platelet degranulation acute inflammatory response  
blood coagulation, fibrin clot f...  
regulation of inflammatory response  
triglyceride metabolic process tyrosine catabolic process

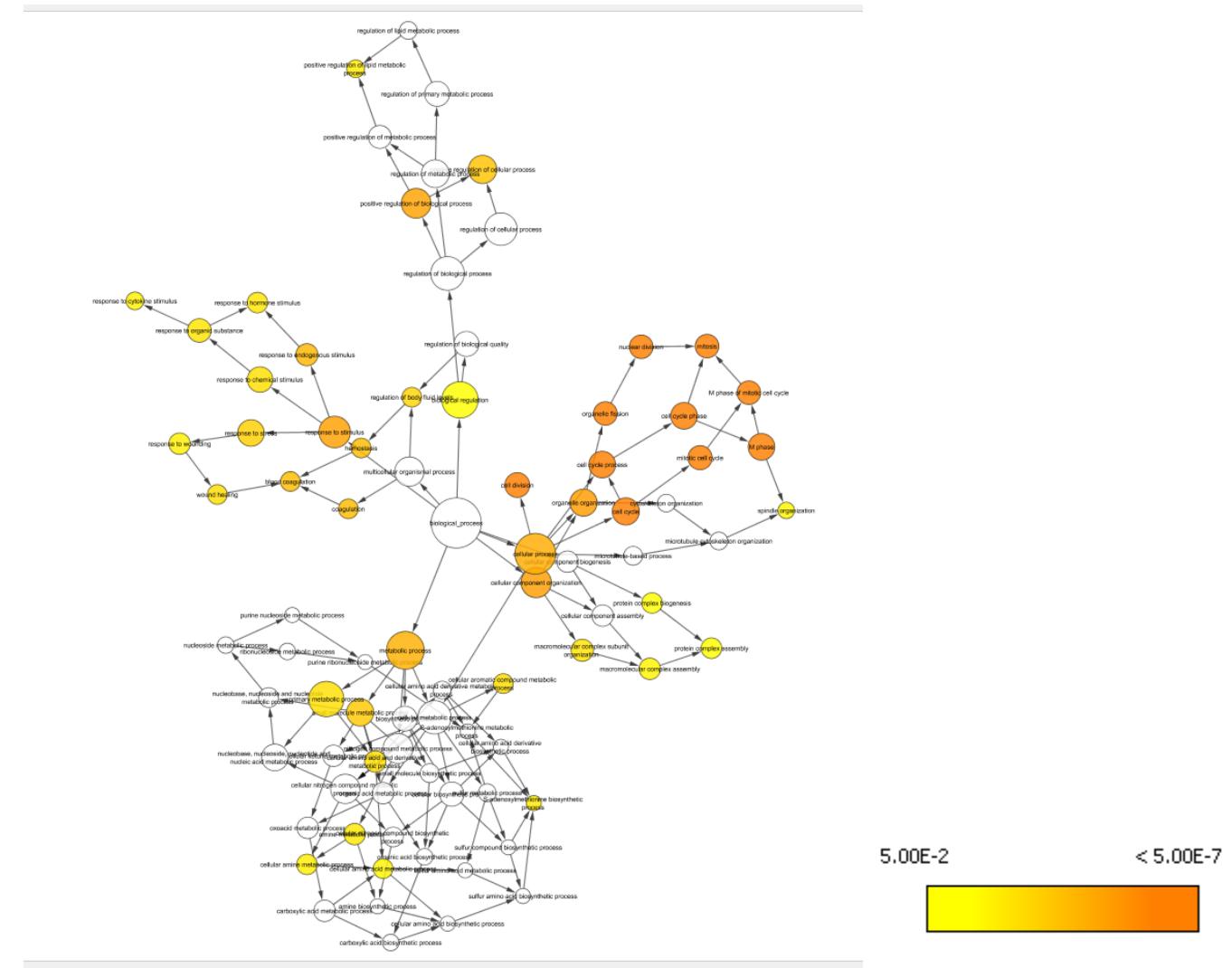
DNA metabolic process  
cell cycle phase transition  
**cell division**  
**mitotic cell cycle process**  
regulation of cell cycle process nuclear chromosome segregation  
cell cycle checkpoint microtubule-based process nuclear division



# PC2 +&- Variable Genes

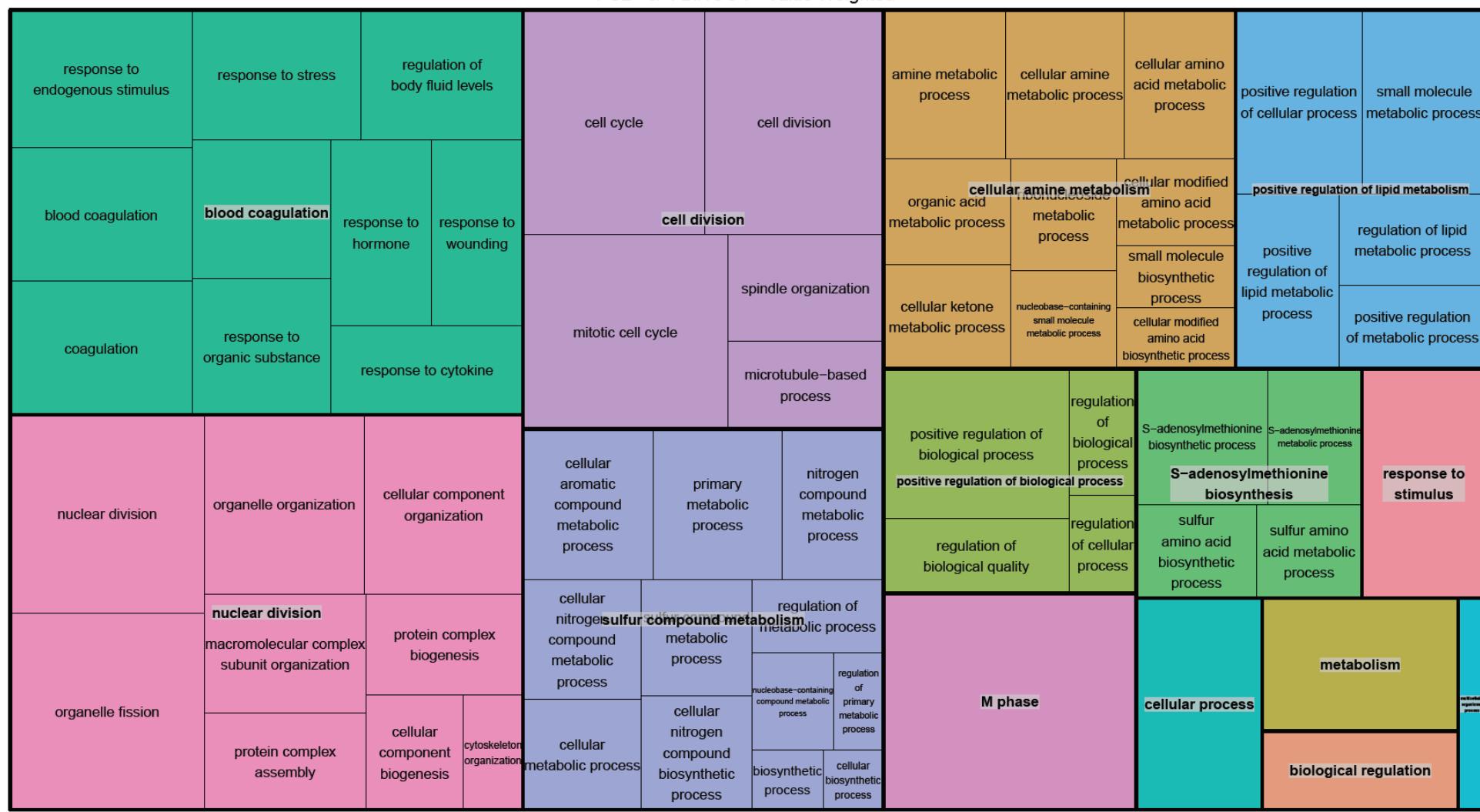
- Top 100 genes were found by taking the highest in magnitude of the gene loading values
- BiNGO is only used
  - Hypergeometric test
  - $P < 0.05$ , Bonferroni Family-Wise Error Rate correction

# PC2+&-: BiNGO Network



# PC2+&-: BiNGO Treemap

PC2 +&-: BiNGO P-Value Weighted

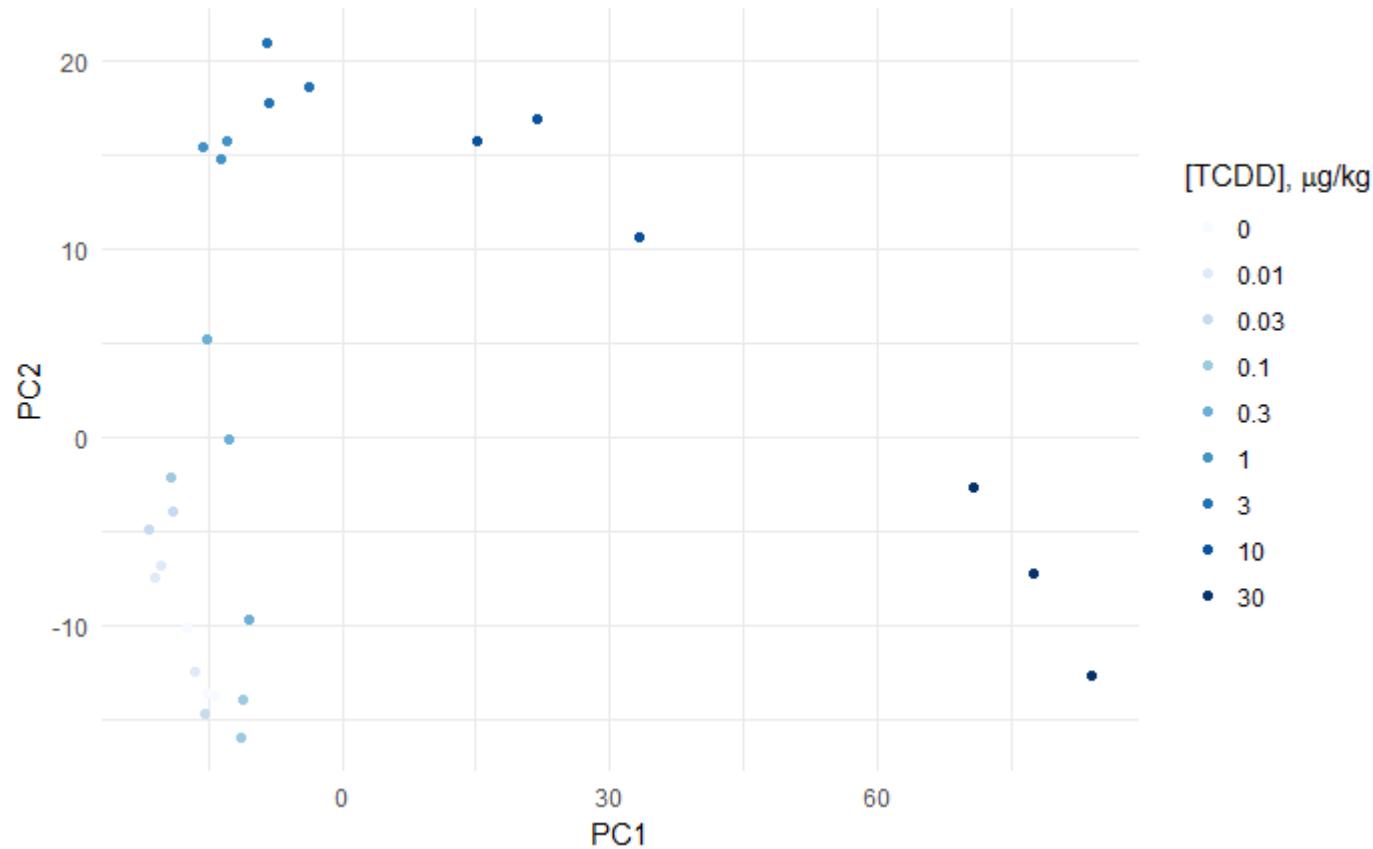


PCA on Bulk RNA Seq Data Only

# Bulk RNA Seq Data

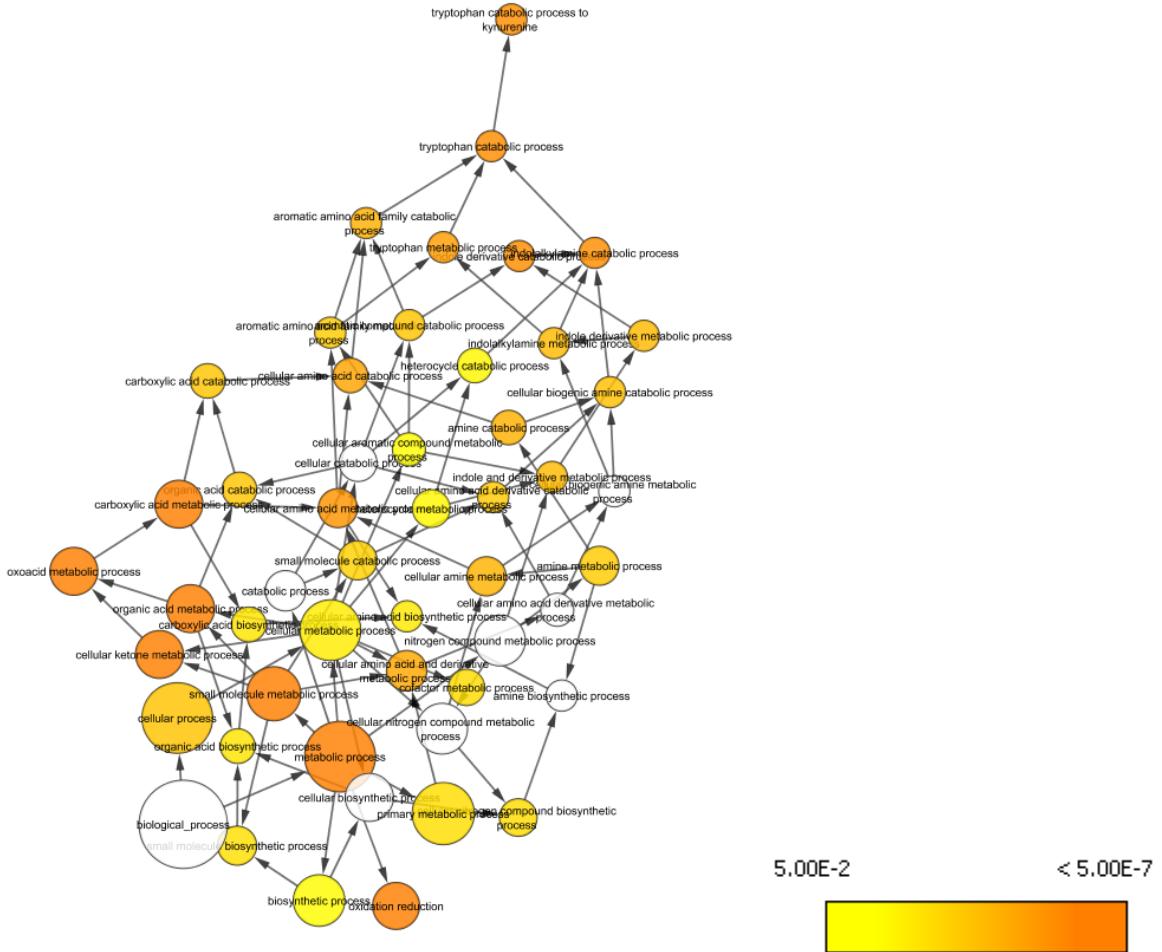
- The top 100 highest in magnitude gene loading values for the variable genes for both PC1 and PC2 were used to identify the variation between both principal components

# PCA on Bulk RNA Seq Data

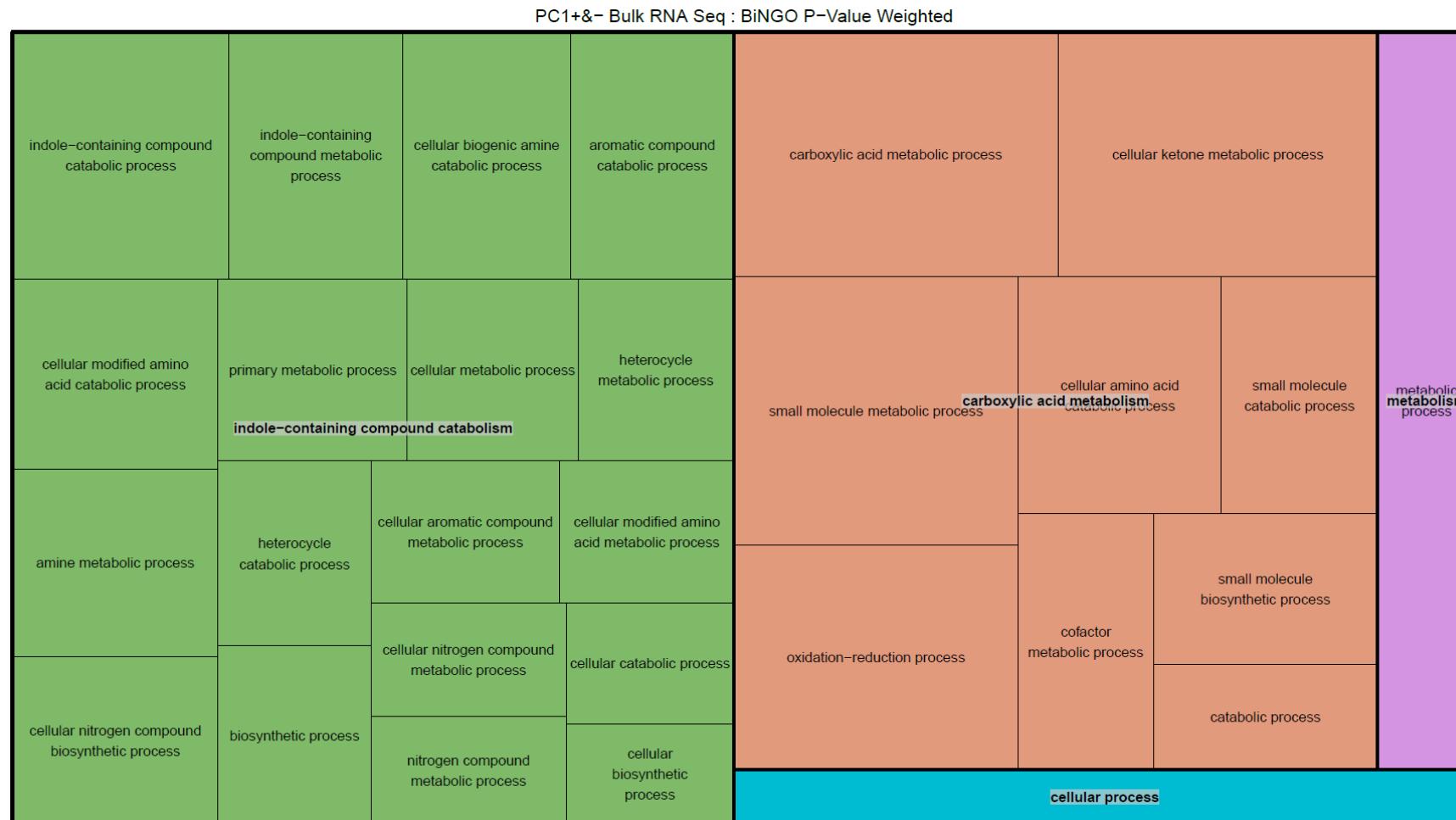


- It appears that the samples vary with concentration along the PC1 axis

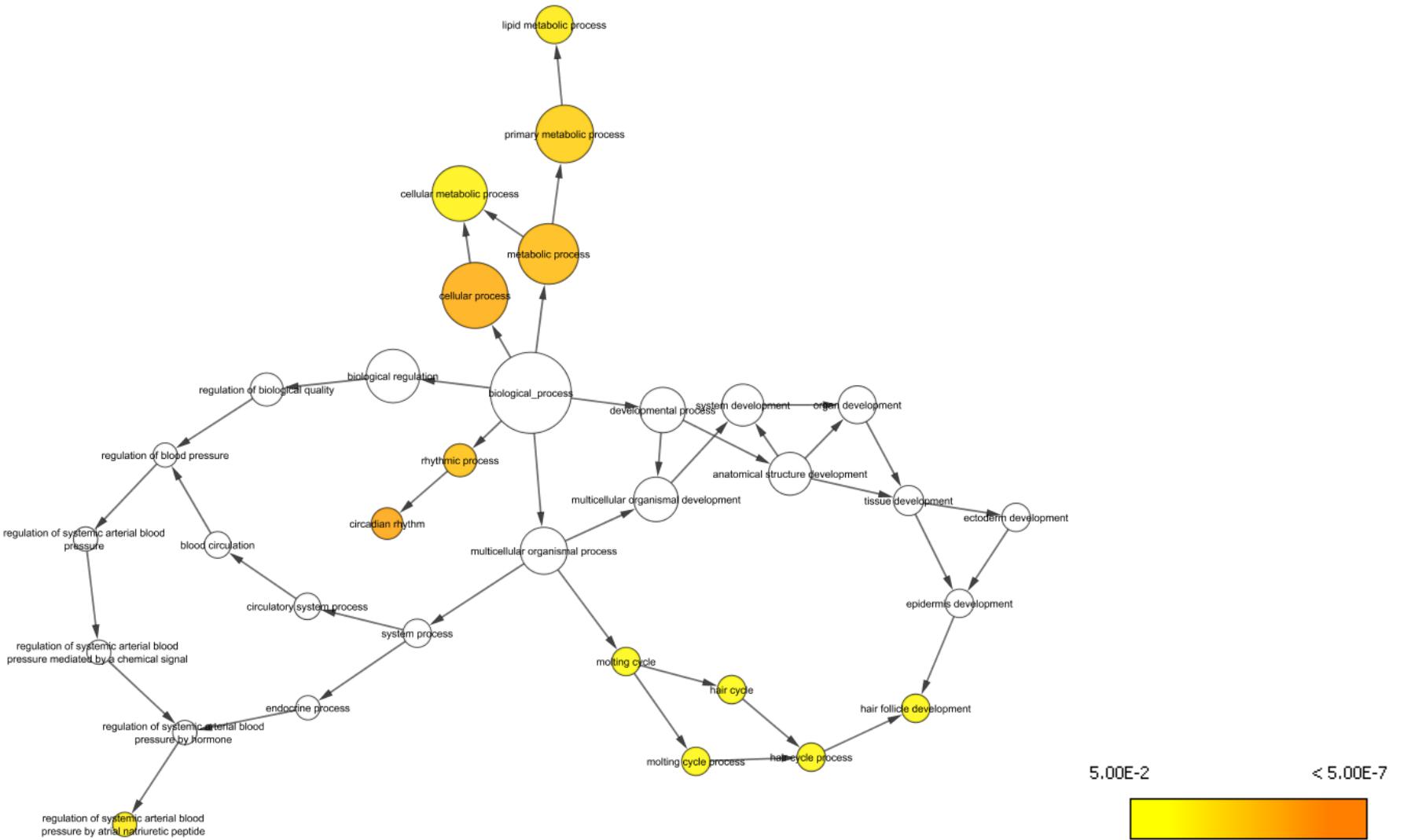
# PC1+&-: BiNGO Network Map



# PC1+&-: BiNGO Treemap



# PC2+&-: BiNGO Network Map



# PC2+&-: BiNGO Treemap

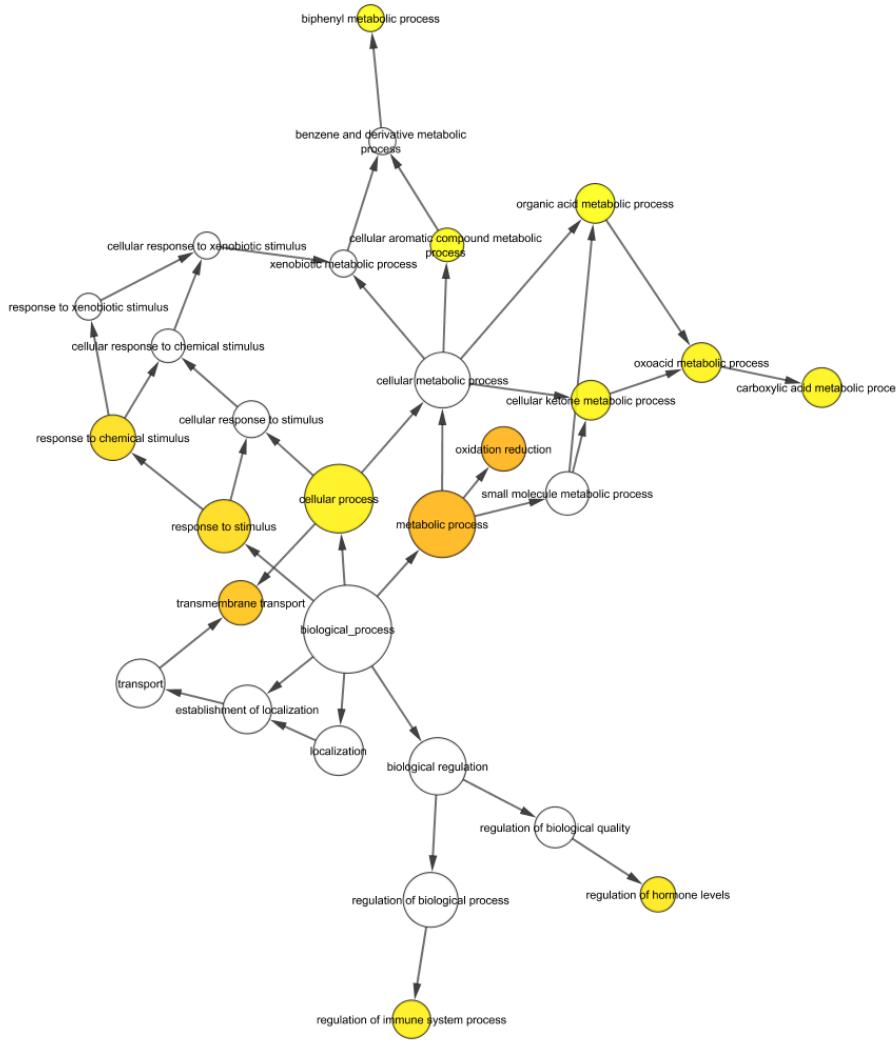
PC2+&- Bulk RNA Seq : BiNGO P-Value Weighted



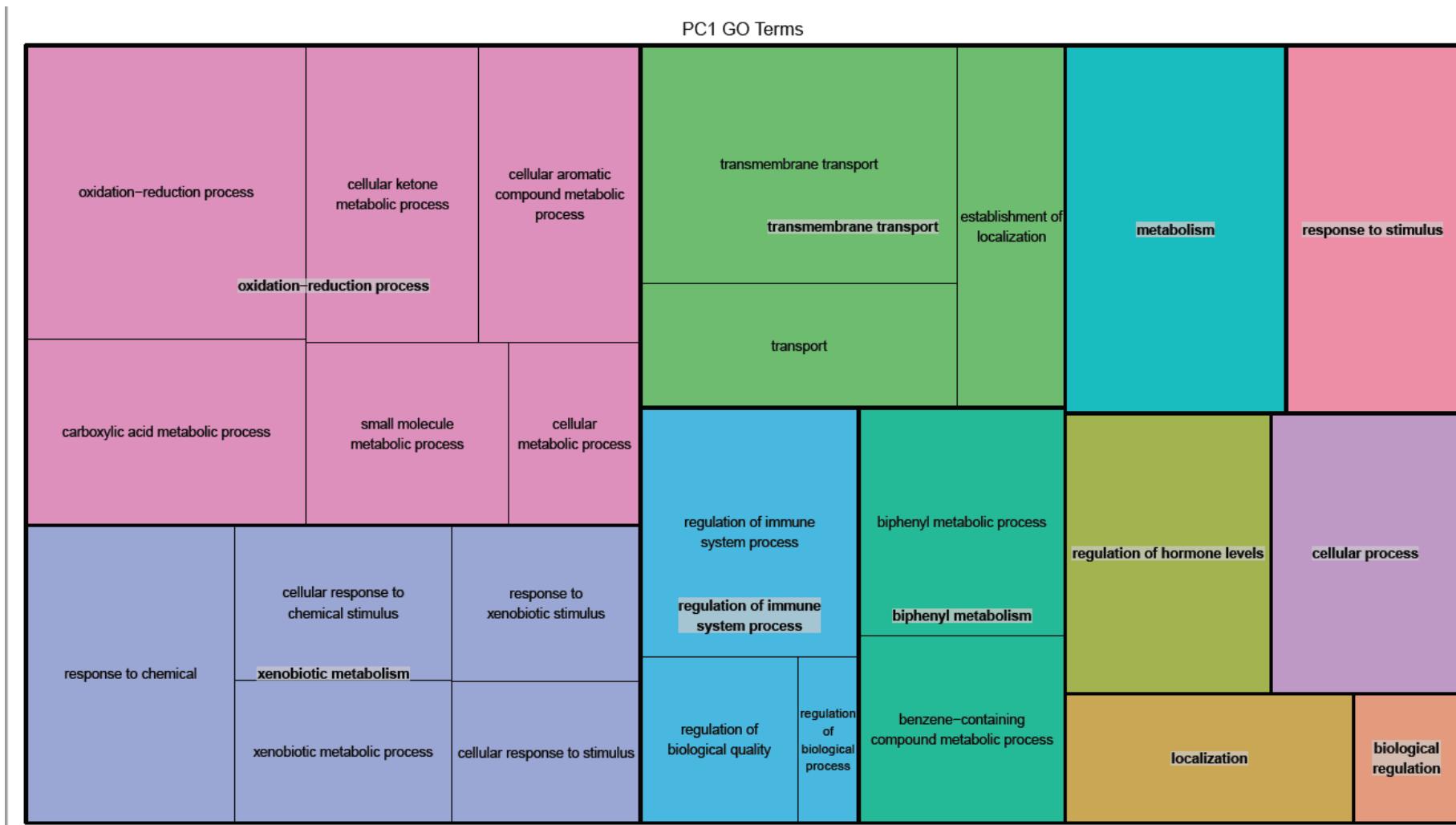
# Conclusions

- Further analysis of gene regulation could be undergone to investigate regulatory networks
- Key genes for enzymes and other proteins involved could have their expression analyzed to validate GO analysis

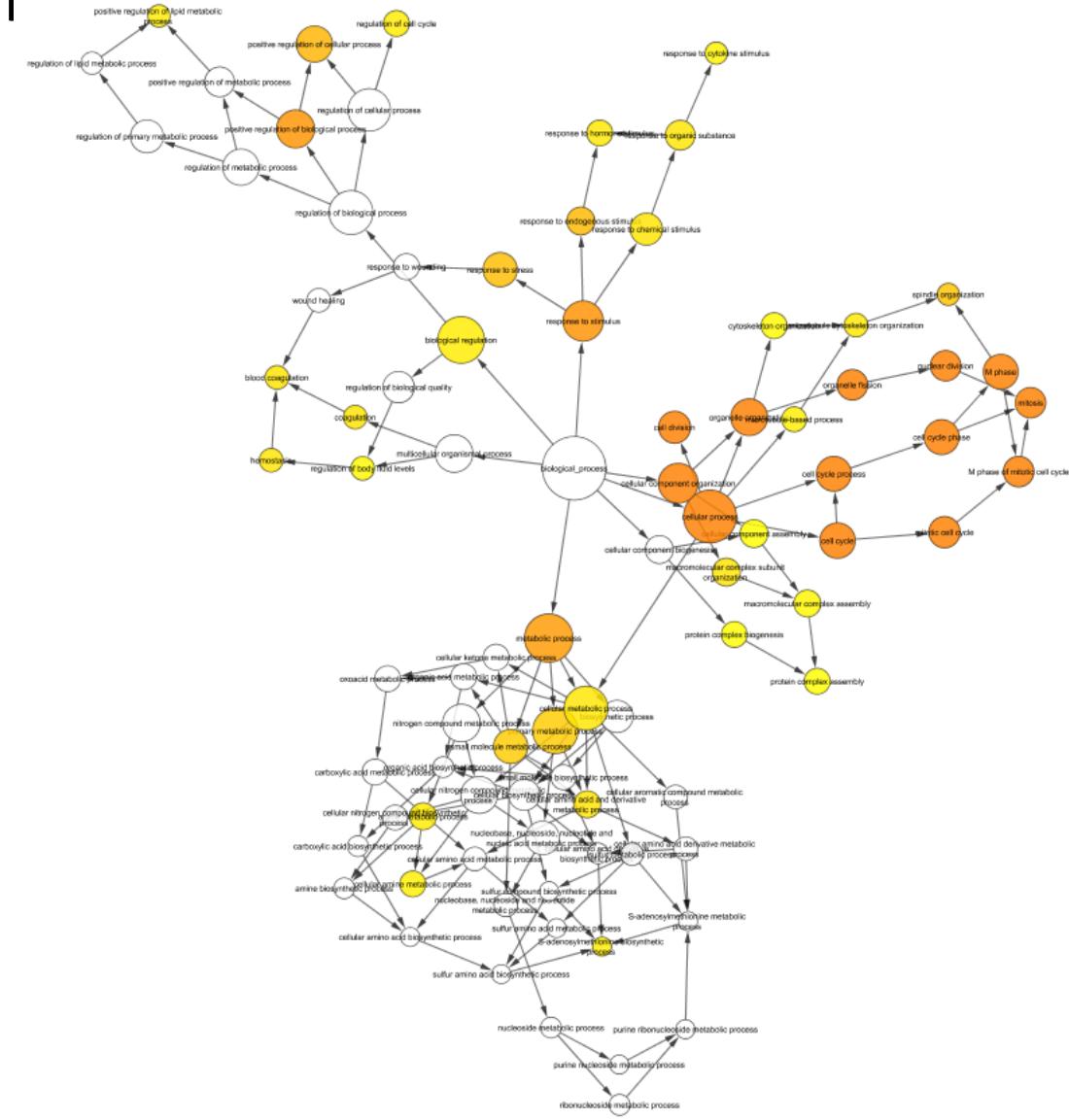
# Absolute PC1 Top 100: Network Map



# Absolute PC1 Top 100: Tree map



# Absolute PC2 Top 100 Network Map



# Absolute DC2 Top 100+ Term Map

PC2 GO Terms

