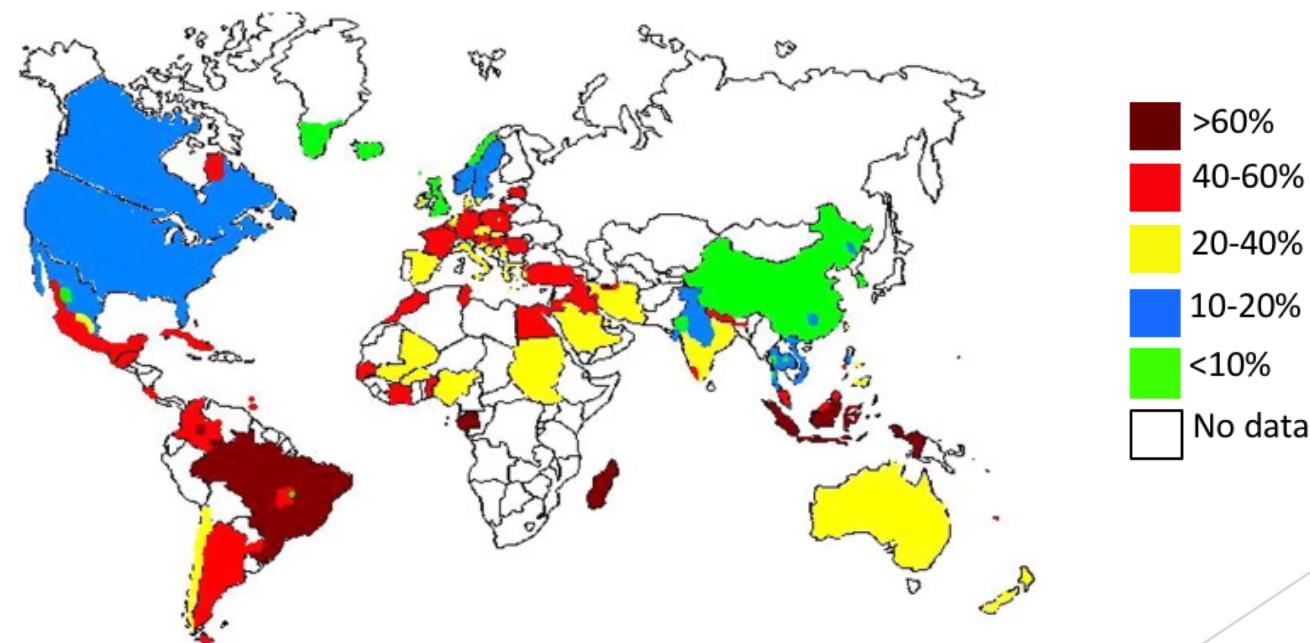


Computational Approach to Identifying Transcription Factors Involved in Lytic Cycle of *Toxoplasma gondii* Virulence Traits

Toxoplasmosis

- 1/3 world population infected
- Leading cause of death attributed to foodborne illness (Immune-compromised)



Toxoplasma Gondii

- ▶ T.gondii is:
 - ▶ Obligate Intracellular parasite
 - ▶ Eukaryotic organism like us (has nucleus)
 - ▶ Protozoan (single-celled organism as opposed to us)
 - ▶ Causes a disease known as toxoplasmosis
- ▶ T.gondii can be found in all sorts of animals such as:
 - ▶ Birds
 - ▶ Cows
 - ▶ Human
- ▶ Defenitive host
 - ▶ cats
- ▶ T gondii organelles responsible for infection
 - ▶ Rhopty
 - ▶ microneme

Toxoplasmosis

1. Foodborne

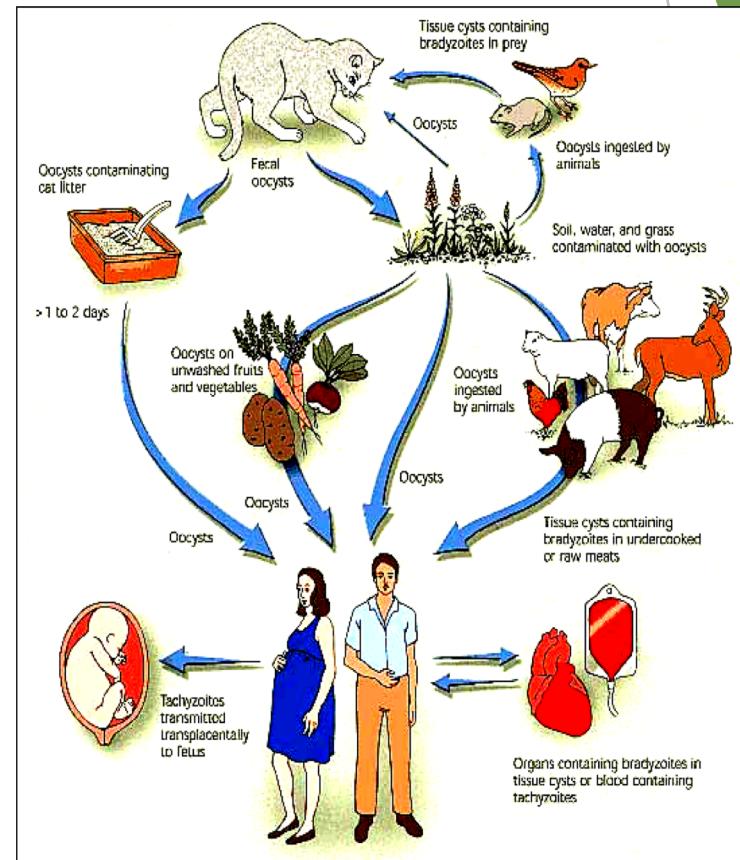
- ▶ Ingestion of uncooked meat,
- ▶ poorly washed fruit and veggies
- ▶ drinking water contaminated with oocyst
- ▶ Ingestion of soil that contains oocyst(animals)

2. Zoonotic

- ▶ Animal to human

3. Congenital

- ▶ Mother to child



Important Structures in T.Gondii Life Cycle

1. OOCYST

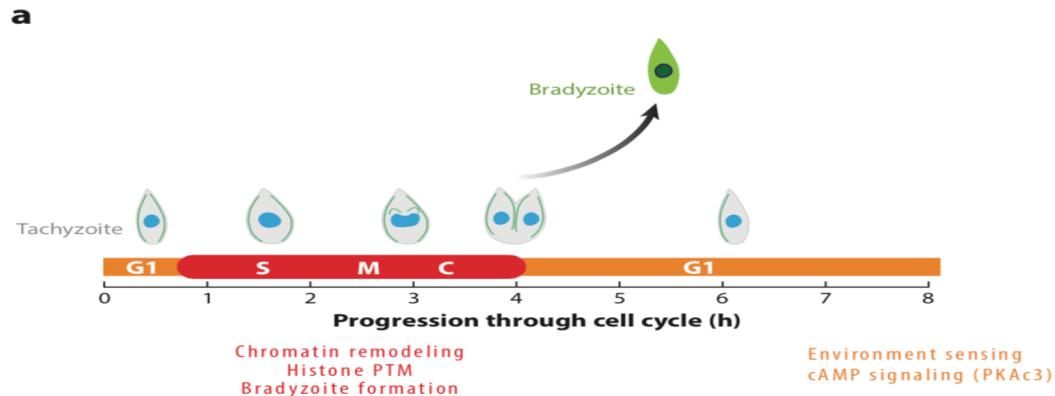
- ▶ a cyst that contains the **zygote or egg** of the parasite T gondii

1. Tachyzoites

- ▶ Tachyzoites are **Fast replicating** forms of T gondii (binary replication)
- ▶ In 6 to 12 hours
- ▶ invasive daughters from mother cell

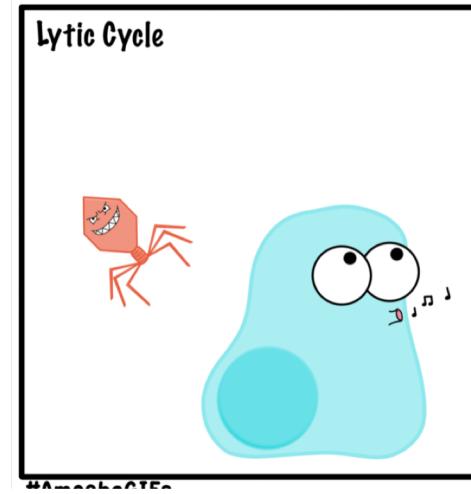
2. Bradyzoites

- ▶ As organism(**parasite egg**) continues to **develop** they become **Bradyzoites**
- ▶ Bradyzoites are **slow replicating** forms of T gondii
- ▶ Evade the immune system
- ▶ Resist common drug treatments
- ▶ Causing chronic infection



Lytic Cycle

- ▶ Pathogenesis
 - ▶ The origination and development of a disease
 - ▶ It is the direct result of the lytic cycle
- ▶ Virulence
 - ▶ The degree of damages caused by an organism to its host
 - ▶ Damage caused from Toxoplasma parasite to the host
- ▶ Virulence Trait
 - ▶ Cell behavior that contributes to virulence
 - ▶ Reinvansion
 - ▶ Replication
 - ▶ Controlling host cell expression



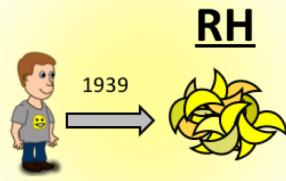
Problem Statement

- ▶ Parasite **virulence** and **pathogenesis** of its infection is the direct result of the parasites **lytic cycle**.
- ▶ **lack of evidence to link individual mutations to virulence trait**
- ▶ A specific **combinations of mutations** are responsible
- ▶ Differentially expression of **virulent genes** are responsible
- ▶ Virulence factor expression is tightly expressed, but the molecular mechanism controlling their regulation remained poorly understood.

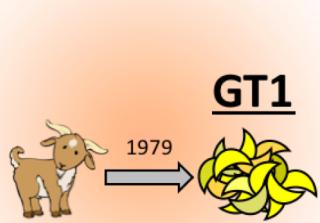
Data Analysis Pipeline Steps

- ▶ Perform RNA Sequencing
- ▶ Regression Analysis
- ▶ Clustering Trending Genes
- ▶ Enrichment Analysis
- ▶ Interaction map of AP2 Transcription Factors

T.Gondii Strains

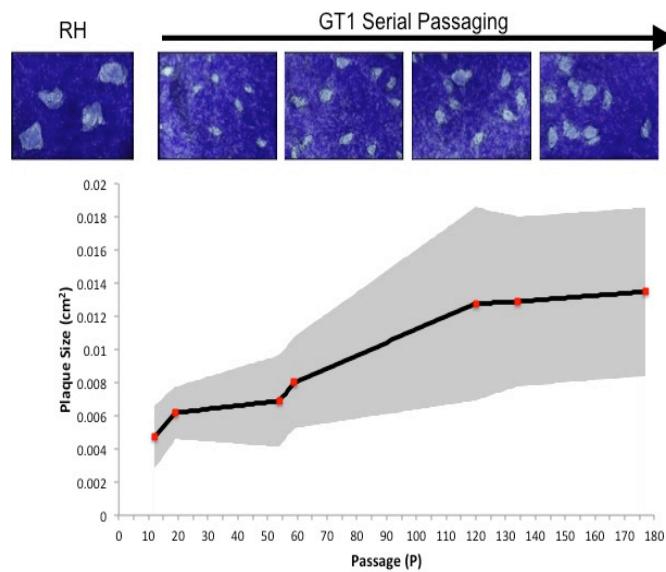
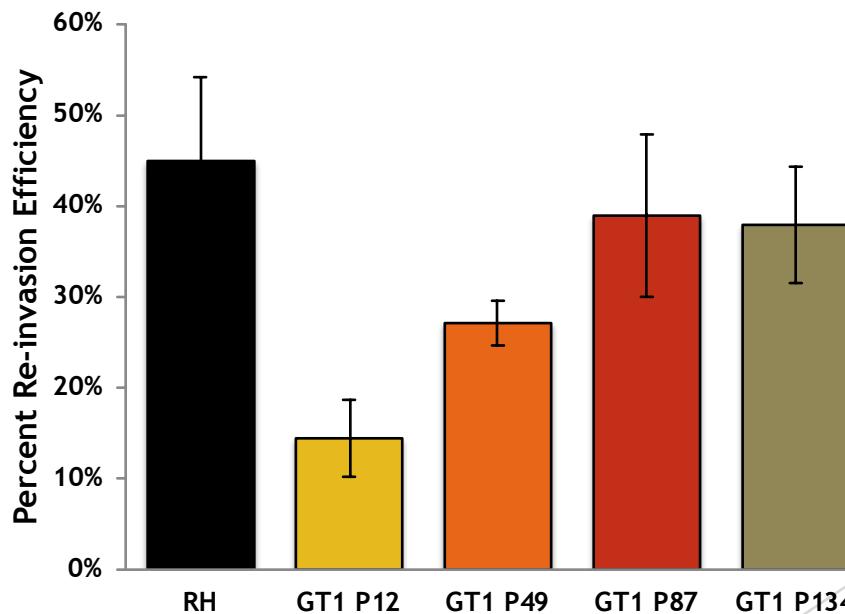
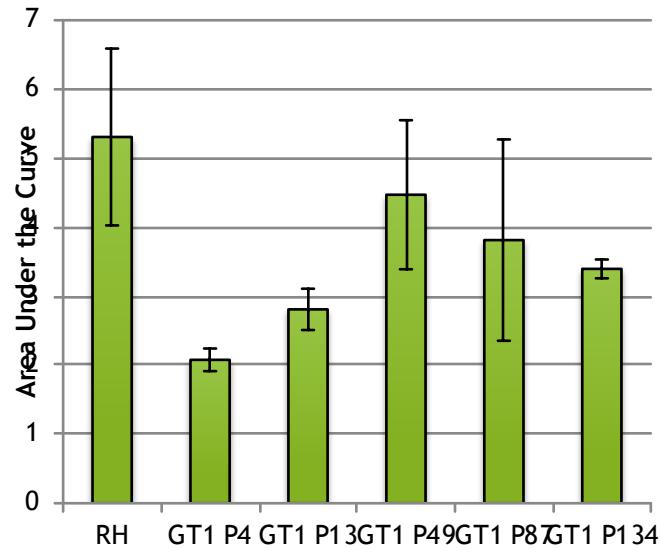


Type I strain RH, a highly virulent, lab-adapted strain



GT1 is a primary, non-lab-adapted, Type I isolate

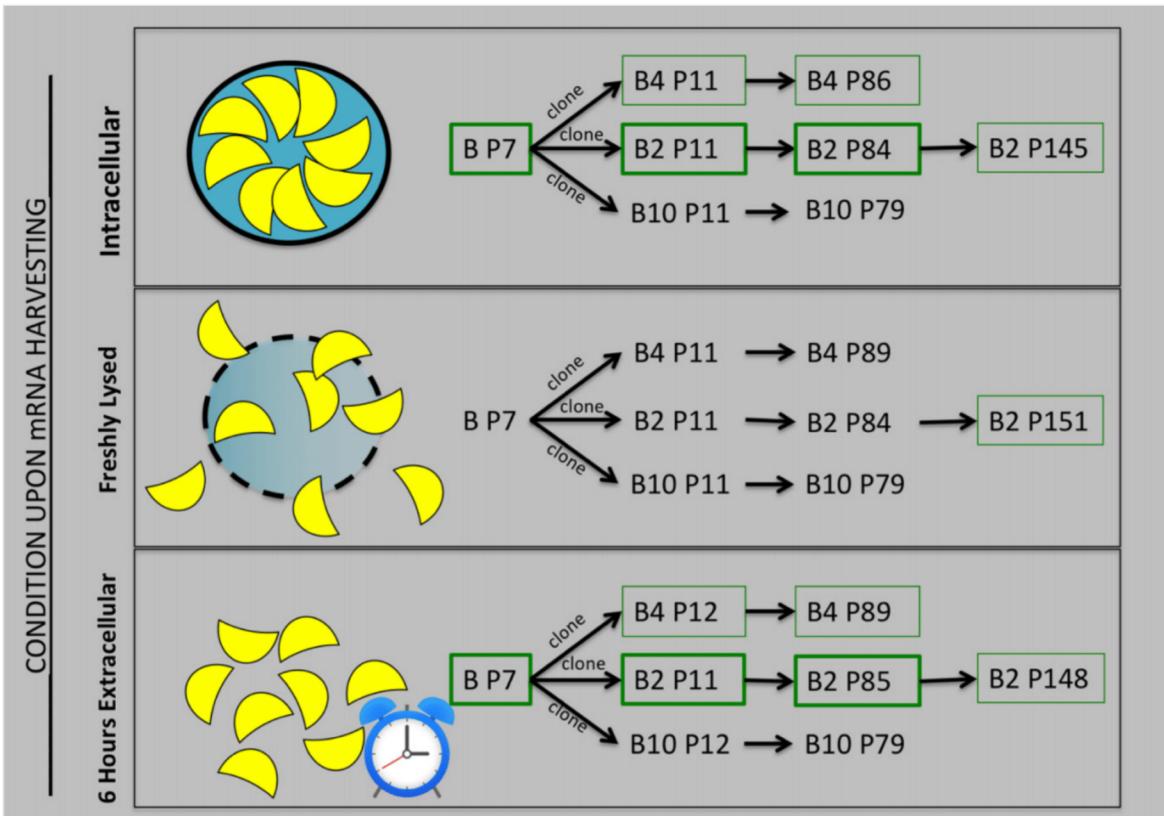
Phenotypes



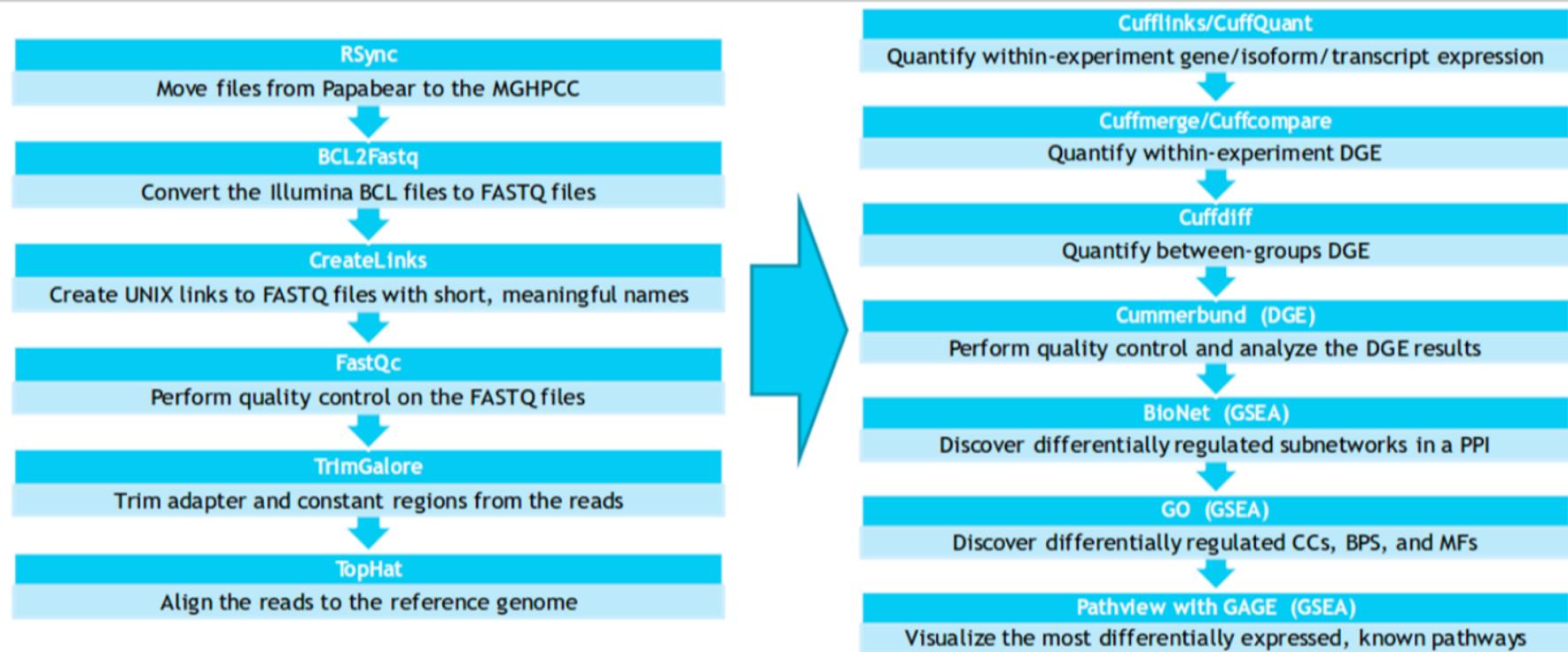
Preliminary Data

- ▶ Preliminary obtained data from non-lab-adapted GT1 strain from **BEI** resources
 - ▶ BEI Resources: a biological resource center for parasitologists
 - ▶ B
 - ▶ B2
 - ▶ B4
- ▶ The strain GT1 cultured and single cell cloned at different time passages
 - ▶ 4 Time points :
 - ▶ P7(initiation)
 - ▶ P11
 - ▶ P84
 - ▶ P148
- ▶ Conditions
 - ▶ 6 hours extracellular
 - ▶ Freshly lysed(GT1 has been forced to be intracellular)
 - ▶ Intracellular

- mRNA sequenced samples
- Conditions
- Serially time passaged lines



RNA-seq (Tailor Pipeline)



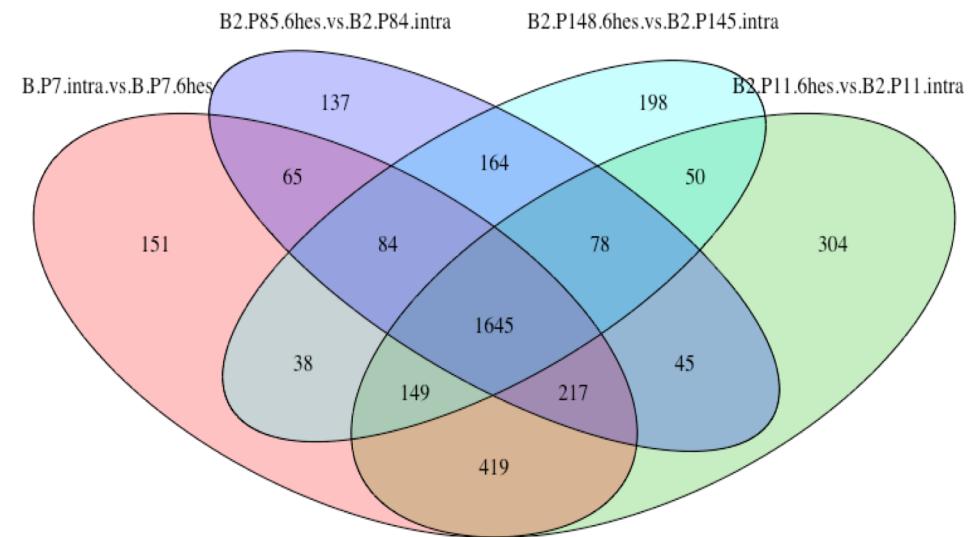
Results

1. Count the number of reads mapped to each gene(gene expression)
2. Normalize gene expression values(FPKM)
3. Differentially Expressed Genes for all the interesting contrast

Differentially Expressed Genes

- ▶ Comparison:
 - ▶ Intracellulars vs Extracellular samples

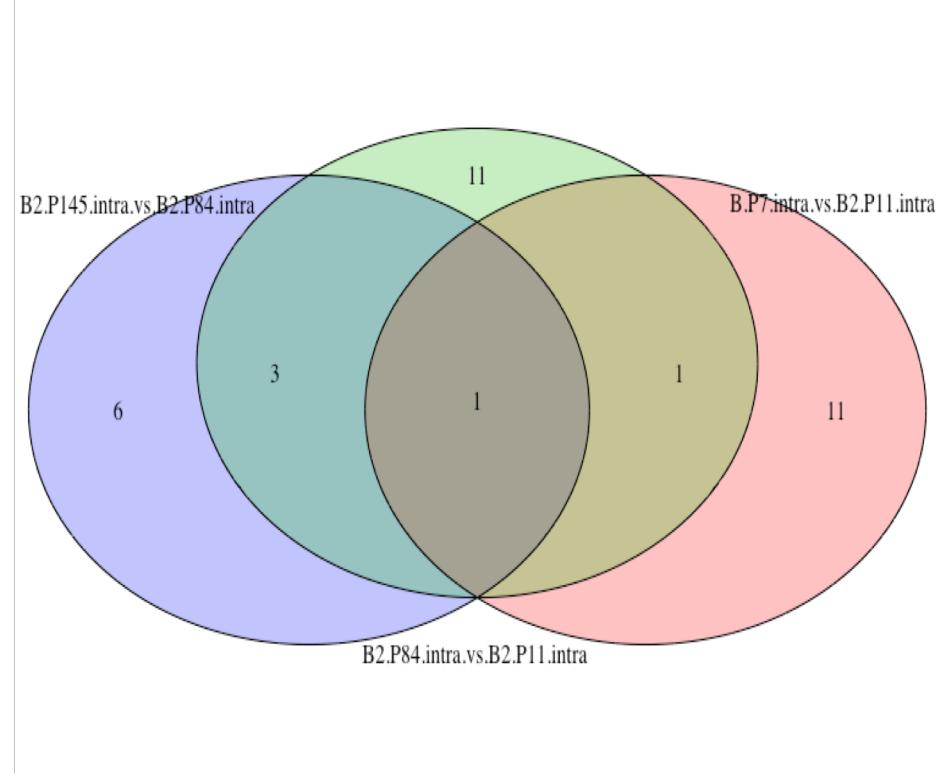
$|\log_2 FC| \geq 1, qvalue \leq 0.05$



Differentially Expressed Genes

- ▶ Comparison:
 - ▶ Intracellular samples over time passages

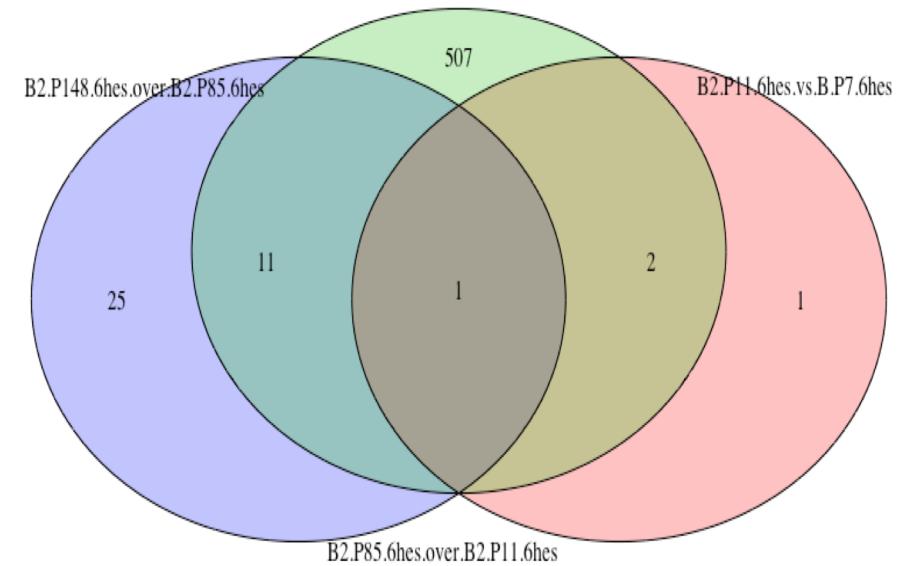
$|\log_2 FC| \geq 1, qvalue \leq 0.05$



Differentially Expressed Genes

- ▶ Comparison:
 - ▶ Extracellular samples over time passages

$|\log_2 FC| \geq 1, qvalue \leq 0.05$



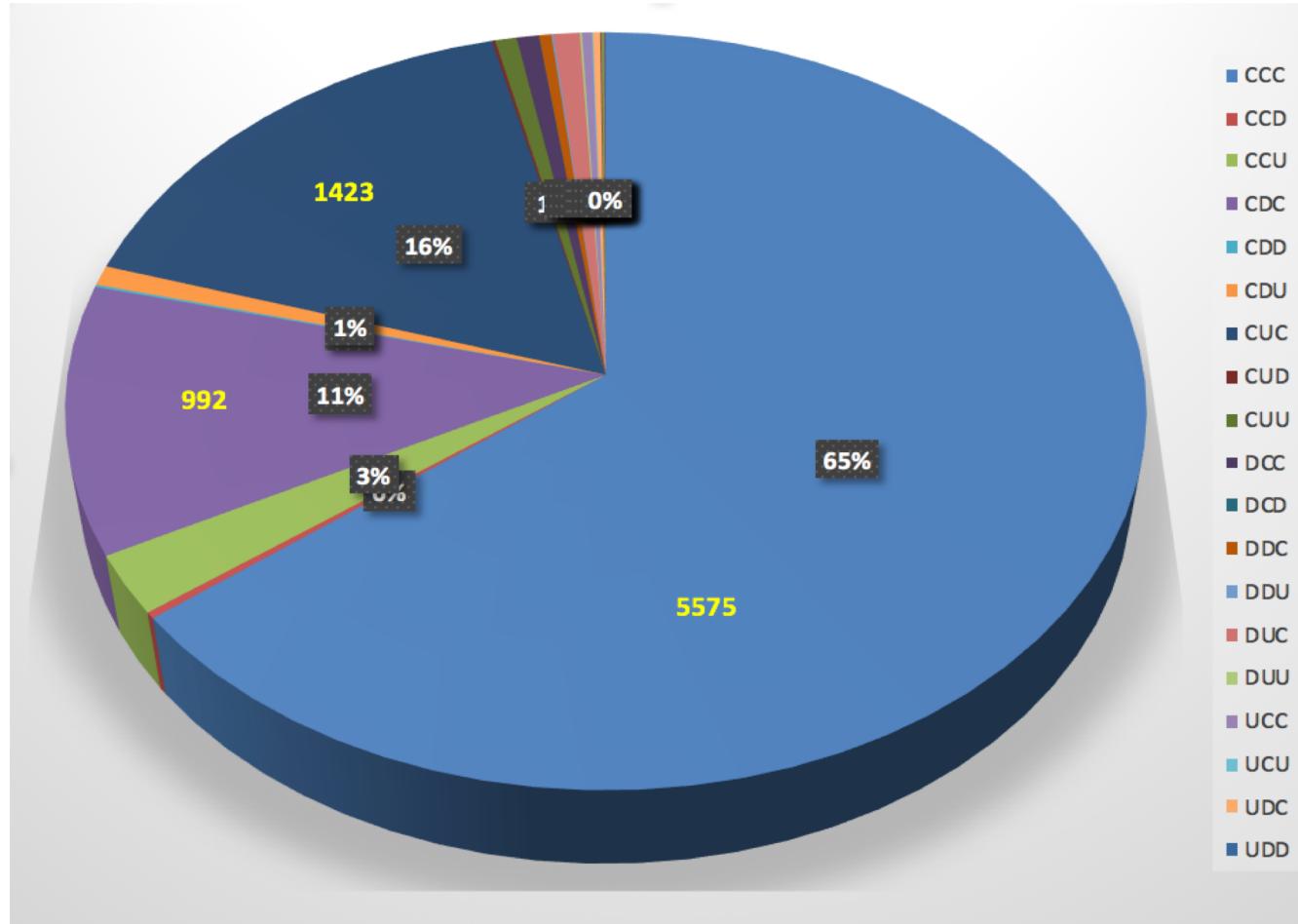
Clustering Genes based on their Trends

Time Points	P-7(initial/7hrs)	P-11(hours)	P-84(hours)	P-148(hours)
Trending Pattern	Constant	Down-Regulated	Up-Regulated	

Looking at the **expression level** of each sample at each time point and compare it with the next time point to see if **expression level** went **Up**, **Down** or stayed **Constant**

Clustering result

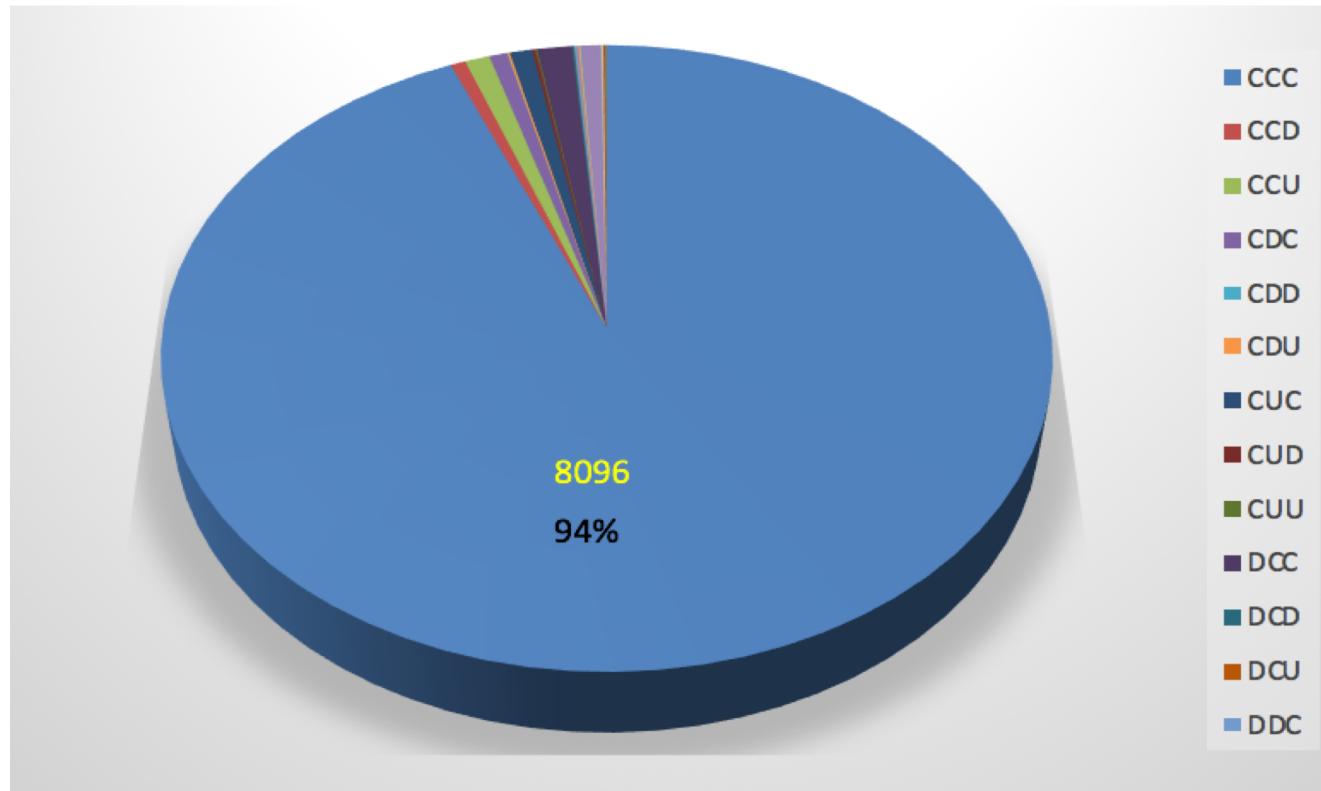
Genes in each trend for B2 extracellular samples



Interesting Trending Patterns: CUC, CDC

Clustering Results

Genes in each trend for B2 intracellular samples



Interesting Trending Patterns: CUC, CDC

Regression Analysis

- ▶ Developing **Regression Models** to associate gene expression values to the measured phenotypes(**Plaque size, reinvasion, Extracellular survival**)
 - ▶ Challenges:
 - ▶ The **number of samples** is low vs a very **high dimensional feature space** (~8,000 genes)
 - ▶ **Regularization** might be able to eliminate all redundant features.
 - ▶ Methods:
 - ▶ Combination of Elastic Net regression and bootstrap can fit well

Regression Analysis

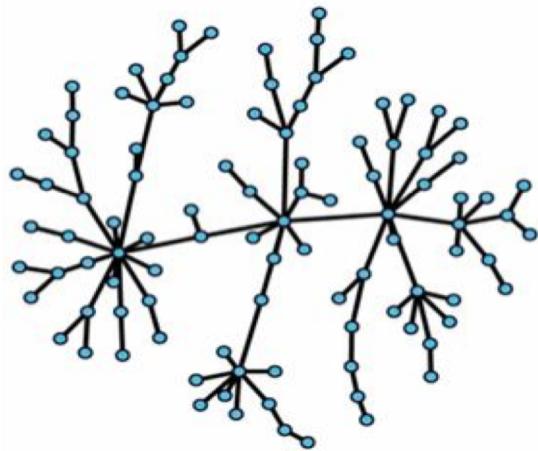
RNA-Seq Expression Values (FPKMs)

Apply regression analysis with L1 regularization (Bootstrap For feature selection)

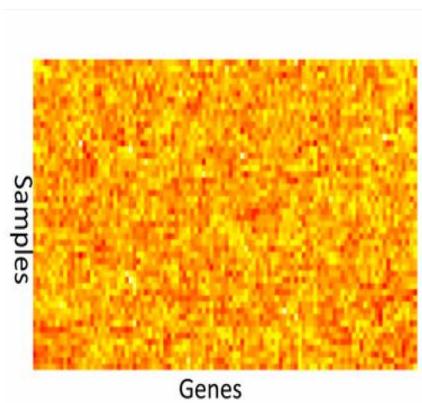
Fit a line between samples with respect to phenotype

- Find the genes that are associated with a specific phenotype
- 30 to 70 genes selected based on coefficient and phenotype

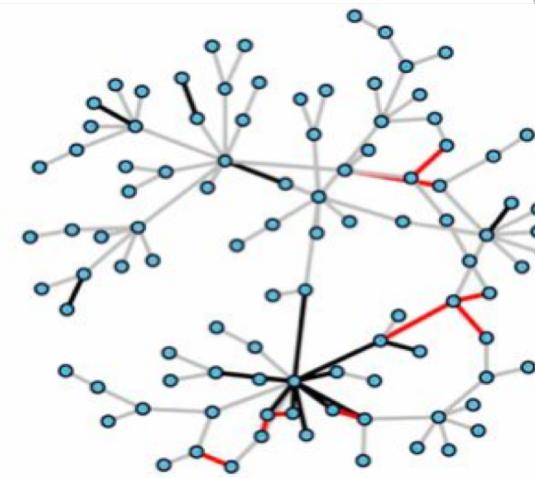
Gaussian Graphical Model



True Network



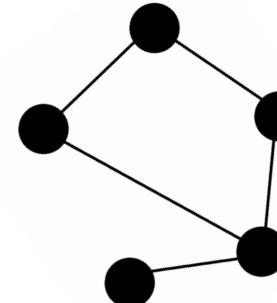
Gene Expression Data



Glasso Inferred network

Gaussian Graphical Model

- ▶ Undirected Graph
 - ▶ Vertex: Random Variable
 - ▶ Graph: Visual way of understanding the joint distribution between vertices
- ▶ Absence of an edge: Corresponding random variables are conditionally independent given other variables
- ▶ Sparse Graph: small number of edges



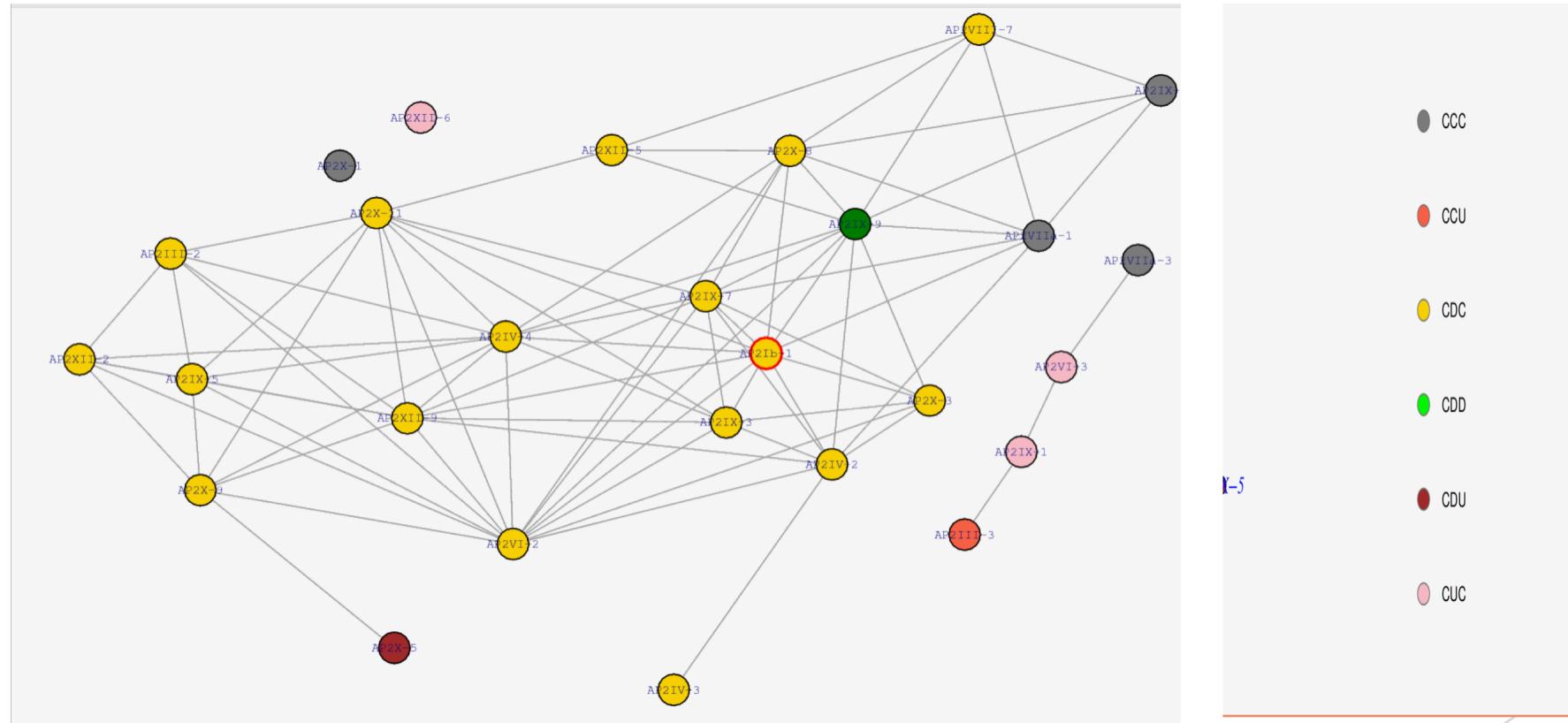
Glasso Network inference

- ▶ Sparse undirected graphical models are estimated using L1(Lasso) regularization
- ▶ Network Inference
 - ▶ Find a graphical model showing statistical interactions between different components of a system
 - ▶ The interactions can be modeled through an underlying Inverse covariance matrix Σ^{-1}
 - ▶ If the (i, j) entry of Σ^{-1} is 0, then the variables i and j are conditionally independent, given other variables.
 - ▶ For a given empirical Σ

$$\min_S \quad \text{tr}[S\Sigma] - \log(\det(S)) + \lambda\|S\|_1$$

Glasso Result

AP2s transcriptionfactors interaction Map



ApiAP2 Transcription Factors

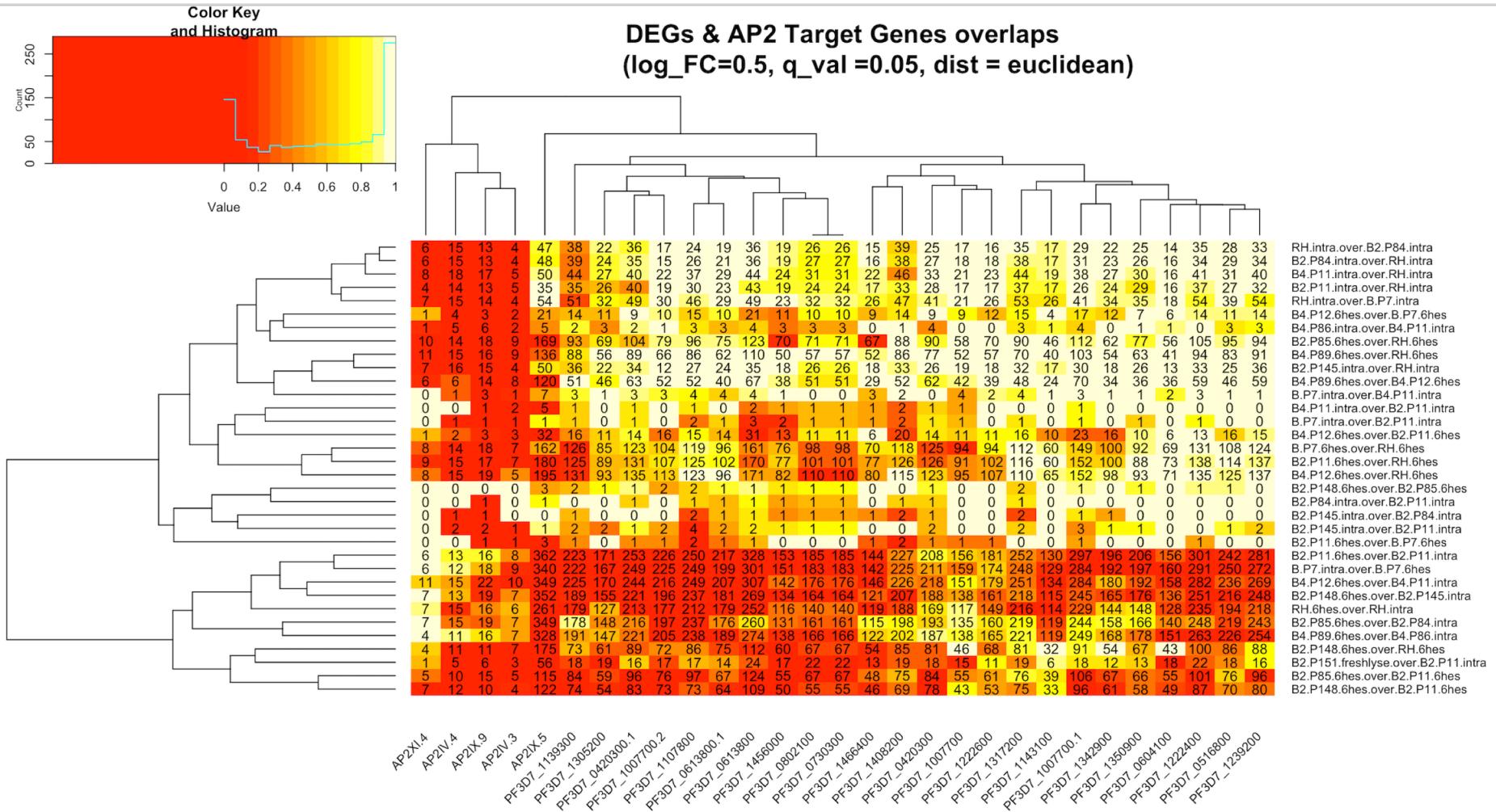
DNA specific sequence regulators of gene expression in toxoplasma

- ▶ 68 APiAP2 TFs are annotated in *T. gondii* genome on Toxodb.org
- ▶ APiAP2 TFs control formation of **cyst(bradyzoites stage)**
- ▶ More than one TF might be responsible for regulation of a gene

Total Ap2 Target genes

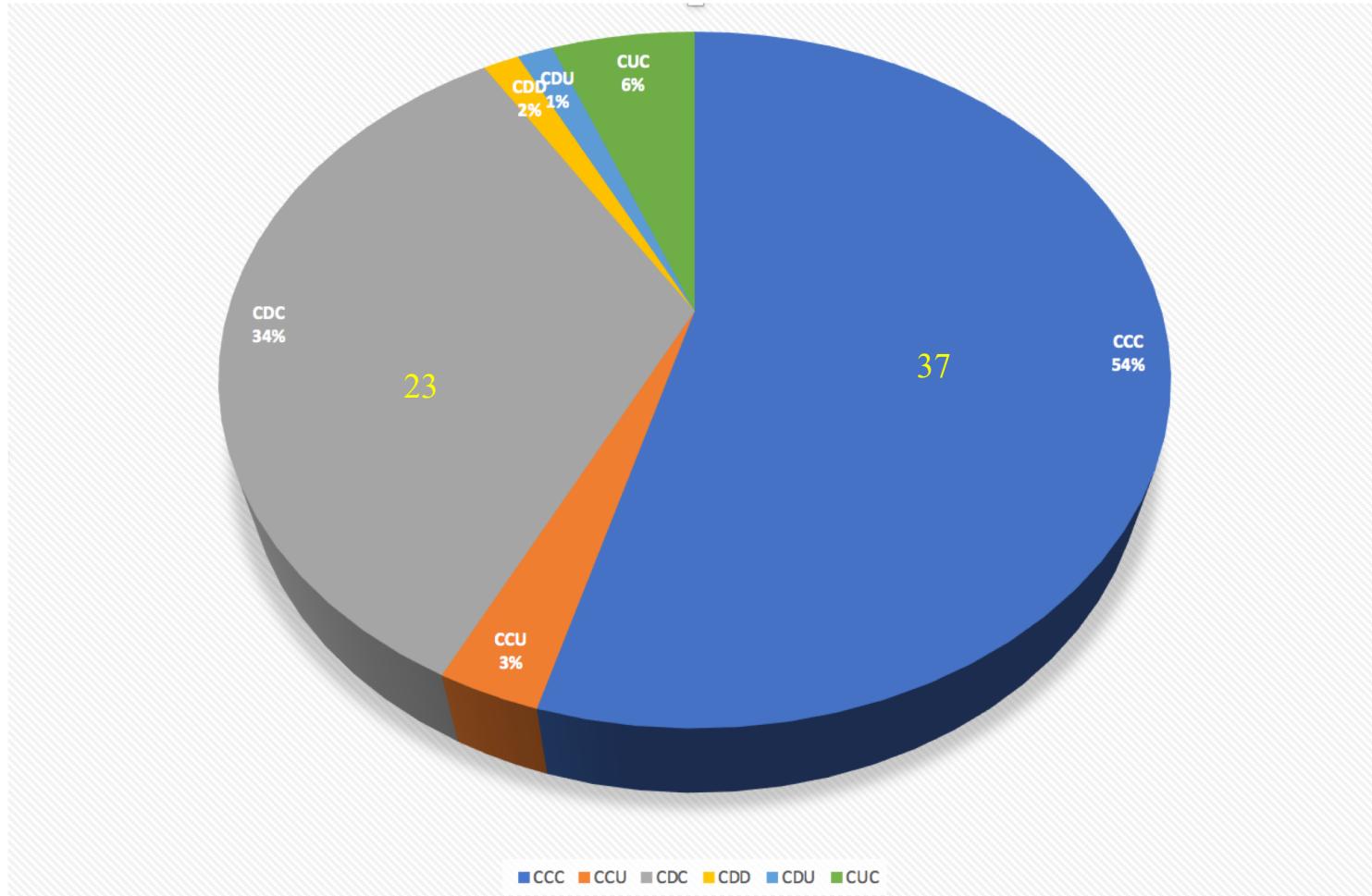
AP2IX.5	537
AP2IV-3	13
AP2XI-4	21
AP2IV-4	15
AP2IX-9	34

Enrichment Analysis



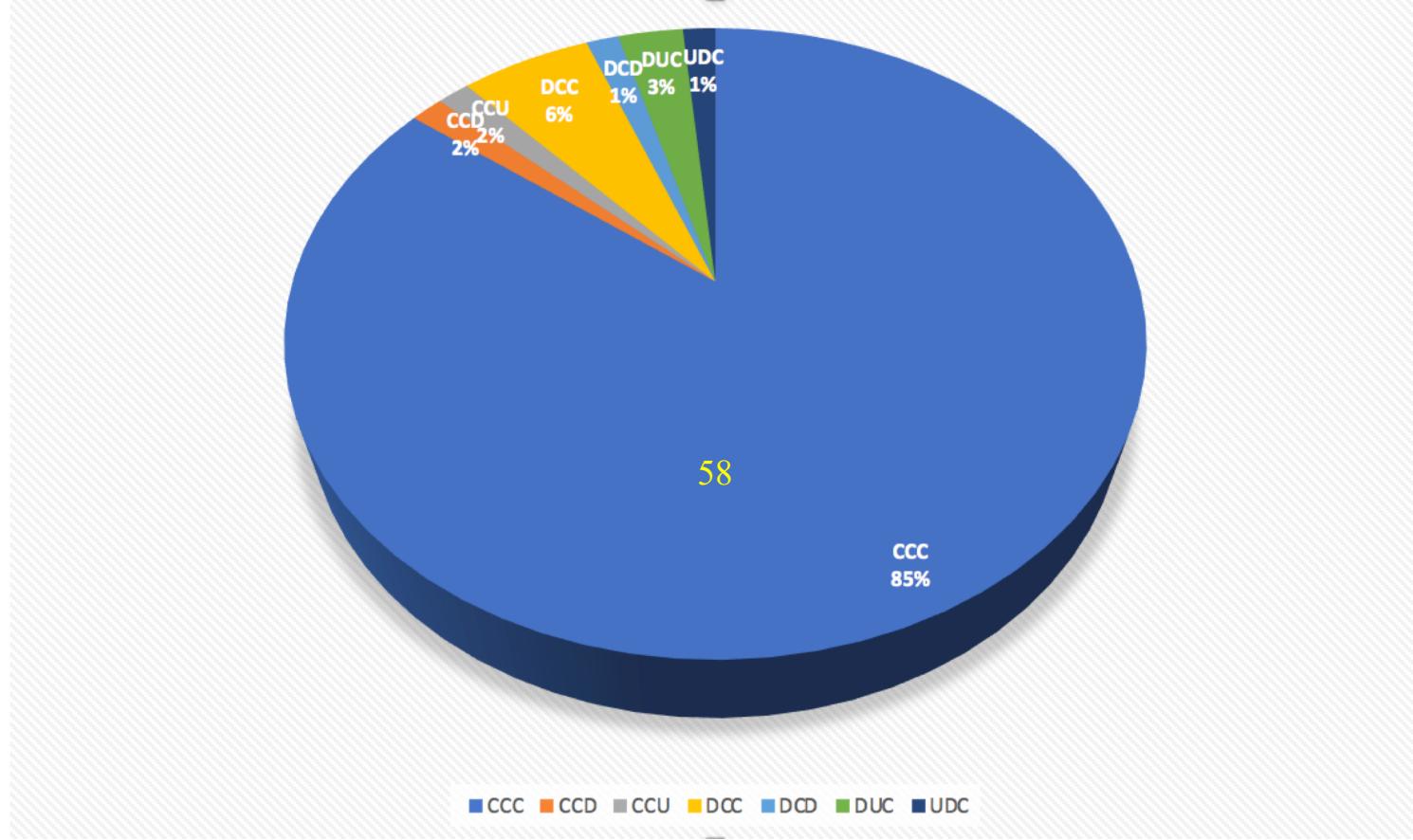
Clustering Results

AP2s in each cluster for B2 6 hours extracellular samples



Clustering Results

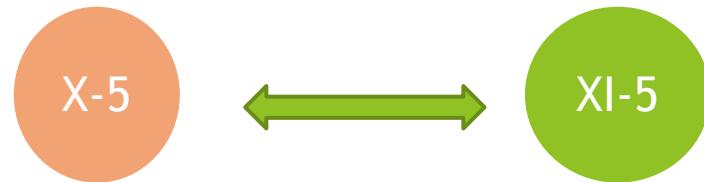
AP2s in each cluster for B2 intracellular samples



AP2XI-5

- ▶ AP2XI-5
 - ▶ Targets more than 300 genes promoters
 - ▶ Actively controls the transcription of those genes
 - ▶ Is constitutively expressed during cell cycle

AP2X-5



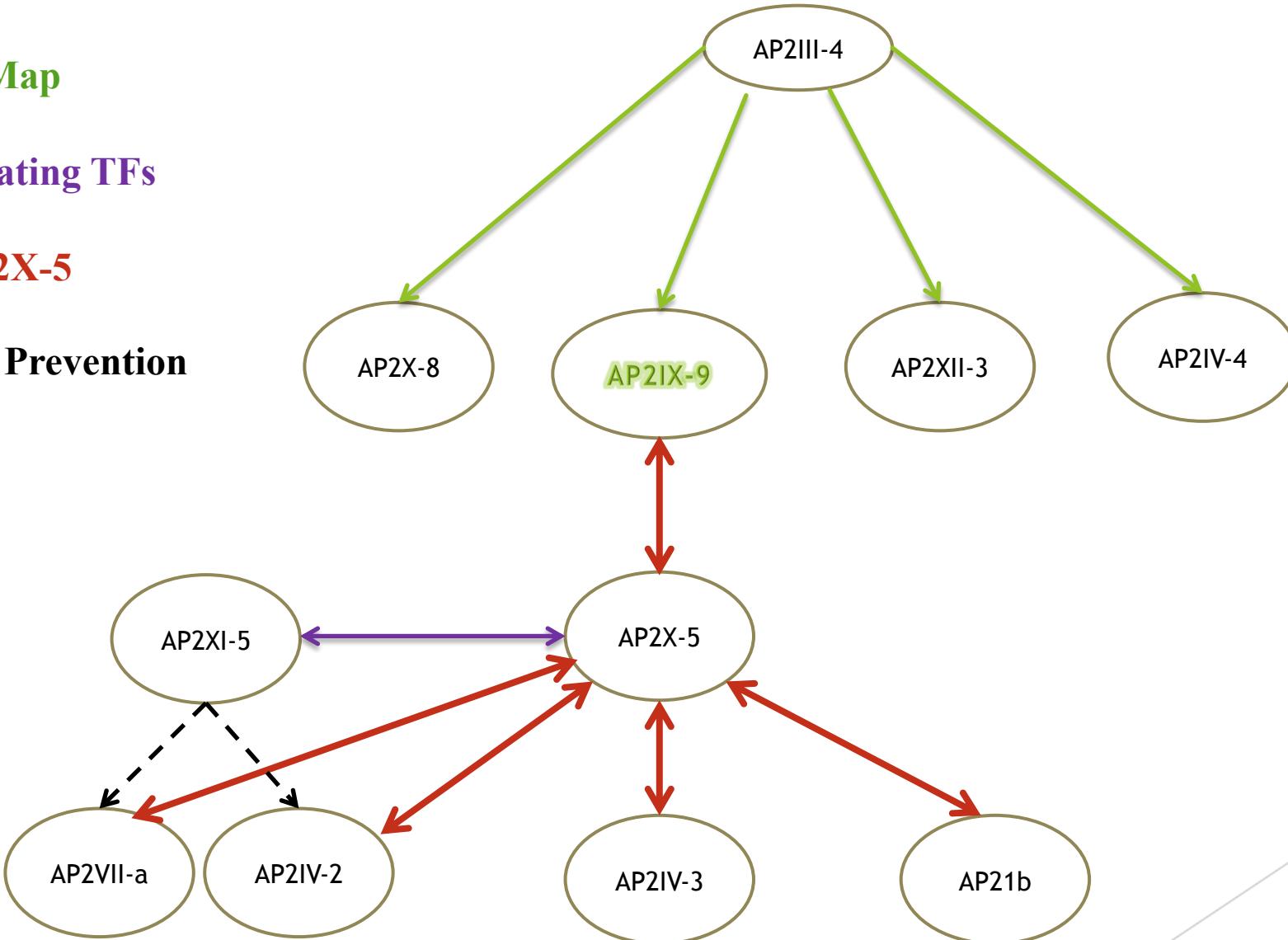
- ▶ AP2X-5
 - ▶ Interacts with AP2XI-5
 - ▶ Is cell cycle regulated TF(S/M phase)

AP2IX-9

- ▶ AP2IX-9
 - ▶ Acts as a repressor at the bradyzoites developmental stage(chronic toxoplasmosis)
 - ▶ Overexpression → **reduces** tissue cyst formation in Bradyzoite
 - ▶ Disruption → increases tissue cyst formation in bradyzoites

Significant TFs

- Glasso Map
- Coregulating TFs
- iKD AP2X-5
- Binding Prevention



AP2 Name	Phenotype			Trend	Expression level	Expression Cycle	motif
	reinvasion	Plaque size	Extra survival				
AP2III-4	✓	✓	✓	CDC			
AP2IX-9				CDD	high		CAGTGT
AP2X-8				CDC	high		
AP2XII-3		✓		CDC			
AP2IV-4				CDC	high		
AP2X-5						Cell cycle	
AP2Ib-1		✓	✓	CDC	high	Constitutive	
AP2XI-5						Constitutive	
AP2IV2				CDC			
AP2IV3				CDC		G1	
AP2XI-4				CCC		S/M	

Future Work

- ▶ Replicating previous analysis for more samples(will be sequence in near future)
- ▶ Applying Chip-seq analysis on the significant Ap2 transcription factors

Acknowledgments

Dr. Kourosh Zarringhalam

Dr. Marc Gubbles

Gubble's Lab: Vincent Primo

Kourosh's Lab: Amir Vajdi

References

- ▶ Walker, R., Gissot, M., Huot, L., Alayi, T. D., Hot, D., Marot, G., . . . Tomavo, S. (2013). Toxoplasma Transcription Factor TgAP2XI-5 Regulates the Expression of Genes Involved in Parasite Virulence and Host Invasion. *Journal of Biological Chemistry*, 288(43), 31127-31138. doi:10.1074/jbc.m113.486589
- ▶ Lesage, K. M., Huot, L., Mouveaux, T., Courjol, F., Saliou, J., & Gissot, M. (2018). Cooperative binding of ApiAP2 transcription factors is crucial for the expression of virulence genes in *Toxoplasma gondii*. *Nucleic Acids Research*, 46(12), 6057-6068. doi:10.1093/nar/gky373
- ▶ Lesage, K. M., Huot, L., Mouveaux, T., Courjol, F., Saliou, J., & Gissot, M. (2018). Cooperative binding of ApiAP2 transcription factors is crucial for the expression of virulence genes in *Toxoplasma gondii*. *Nucleic Acids Research*, 46(12), 6057-6068. doi:10.1093/nar/gky373

Thanks!