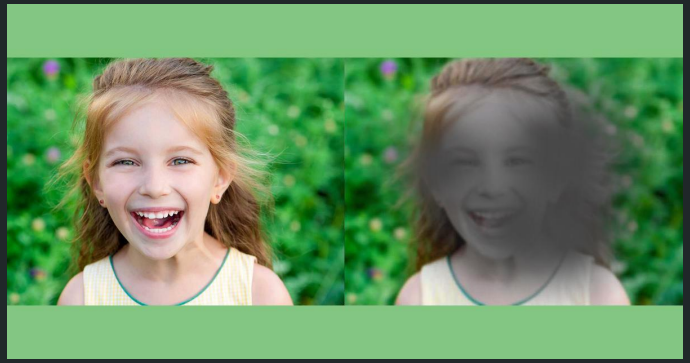
Baylor College of Medicine



D2K Data to Knowledge Lab Rice University

Inferring genomic signatures in age-related macular degeneration across different stages

Presented by: Minjun Park, Shryans Goyal, Yu Wu, Zach Moxley, Zishi Wang

Meet Rosa

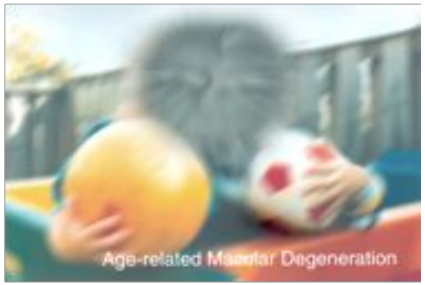
- 80 year old grandmother
- Loves observing nature
- But…..

Cannot See Properly!

The Curse of AMD

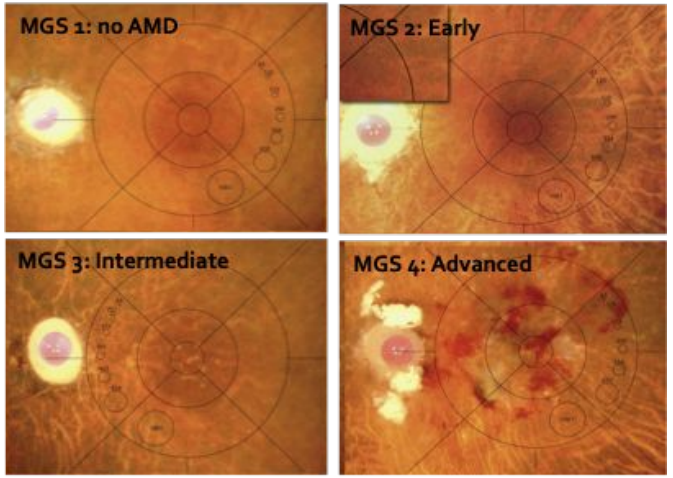
- AMD : Age-related macular degeneration

Normal vision



Loss of central vision

Ratnapriya et al., 2013, Genome Med; National Eye Institute

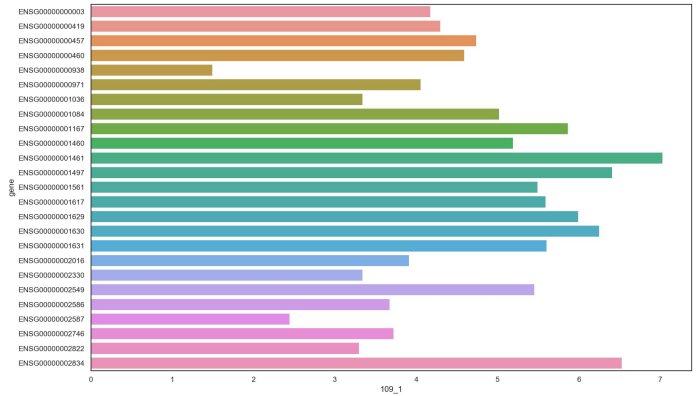


*Olsen and Feng, IOVS, 2004; Decanini et al., Am J Ophthalmol 2007

- AMD has 4 distinct disease stages

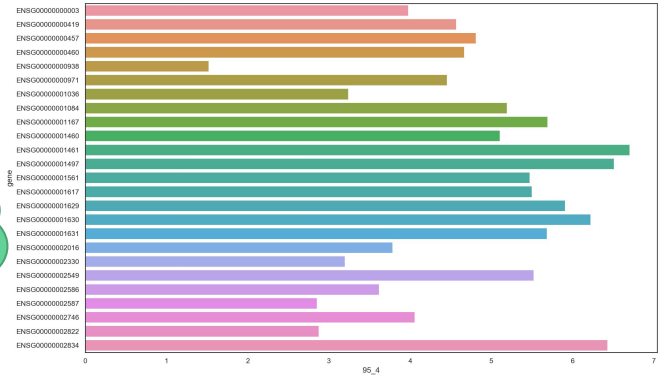
Why it Matters

- **Hard to prevent and treat!**
- **Hard to detect!**
- **Inevitable!** High magnitude of cases per year.



Normal (Stage 1 Gene Expression)

No Difference!



Advanced (Stage 4 Gene Expression)

Impact

Early detection → Identify Drug Targets → Protect Rosa!

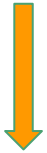
Data Preprocessing

453 patients in **4** disease stages

~ **58, 000** Genes

Low Variance filtering

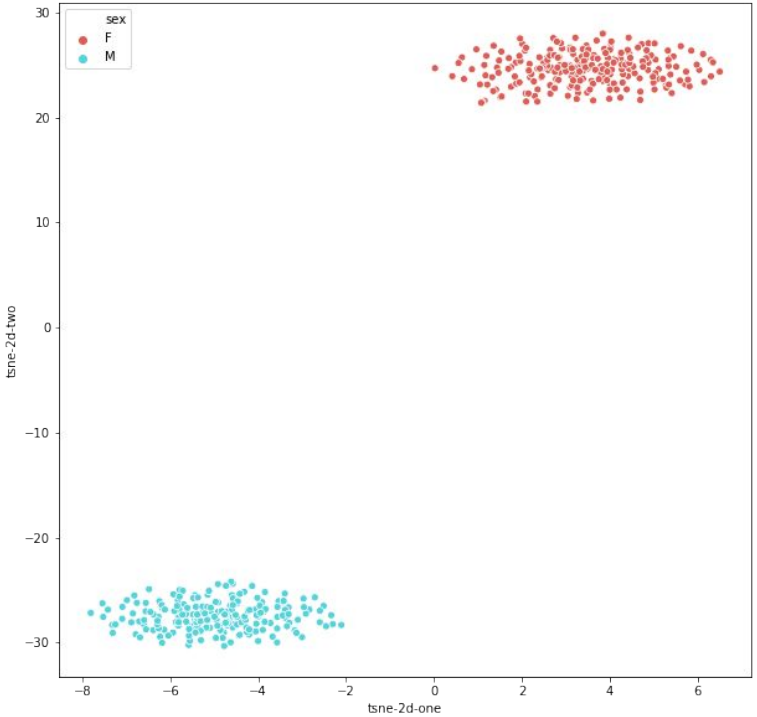
~ 37, 000 Genes

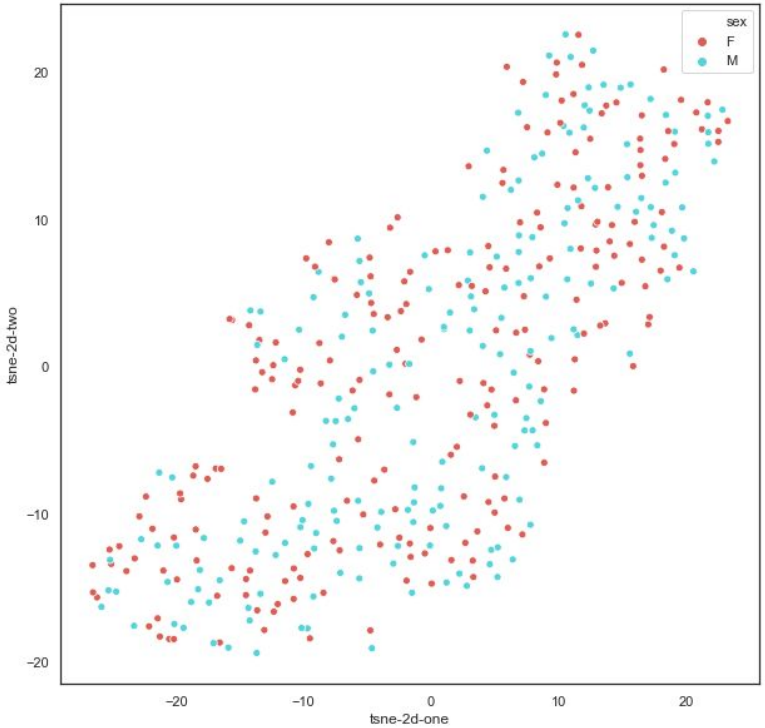


ANOVA

~ **4, 000** Genes

Dimension Reduction





Objective 1

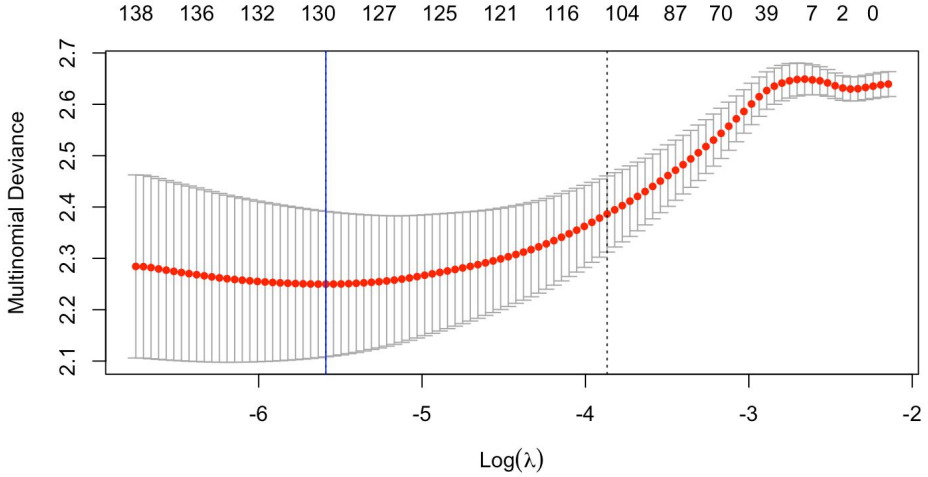
Extract set of genes that are most important to each stage of AMD

Modeling pipeline

- Feature Selections methods
  - LASSO Multinomial regression
  - Random forest
  - XGBoost
- 10-fold cross validations
- Bootstrapping of 500 iterations

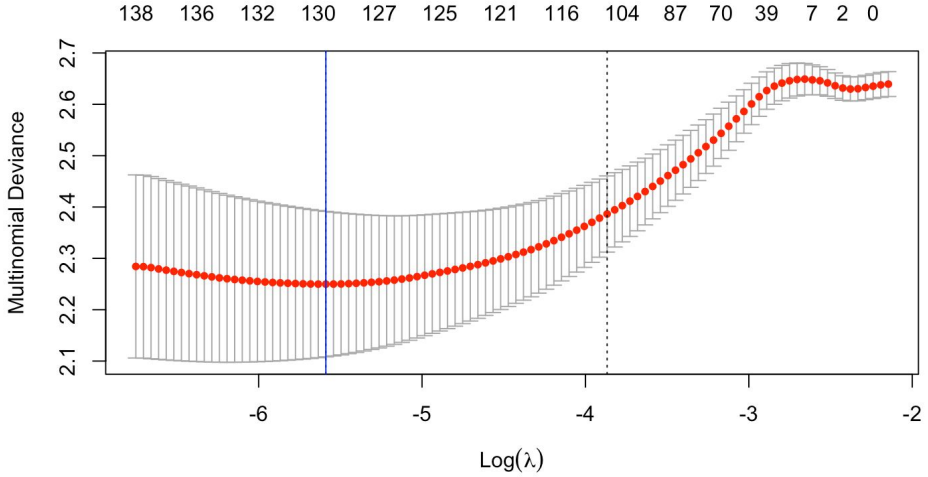
Modeling pipeline

- Feature Selections methods
  - LASSO Multinomial regression
  - Random forest
  - XGBoost
- 10-fold cross validations
- Bootstrapping of 500 iterations



Modeling pipeline

- Feature Selections methods
  - LASSO Multinomial regression
  - Random forest
  - XGBoost
- 10-fold cross validations
- Bootstrapping of 500 iterations

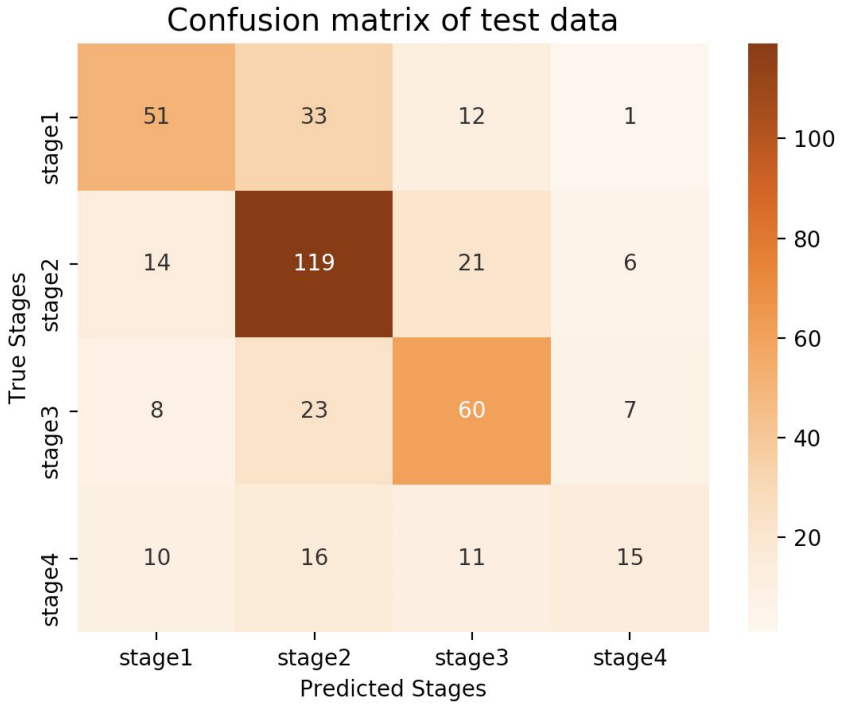


➢ **Identified important genes that overlap in all three models**

DE genes found: MOXD1, PMAIP1

Choosing the Best Model: **Multinomial Lasso Regression**

- **All Stages** Accuracy: 60.2%

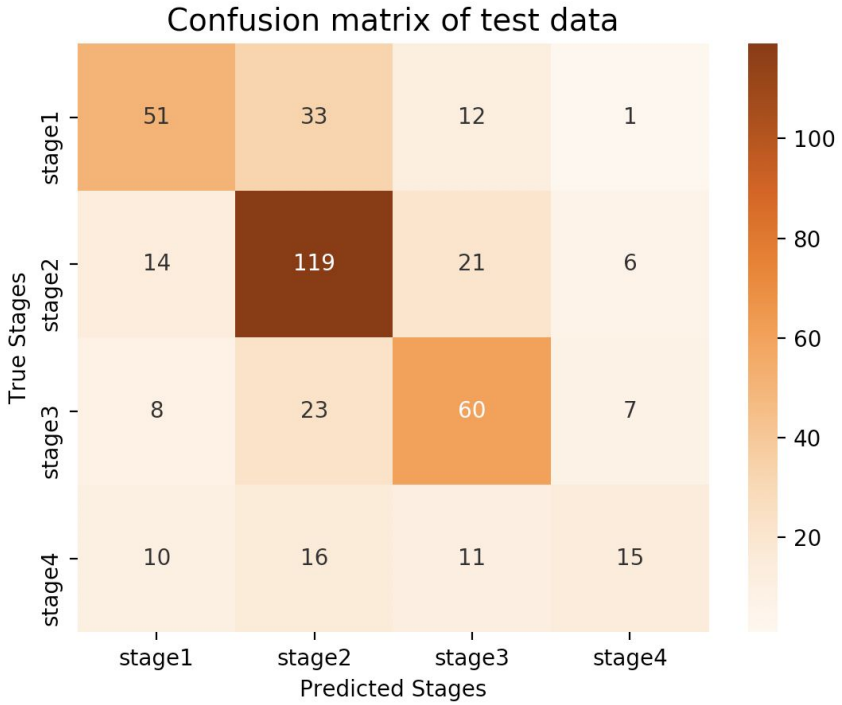


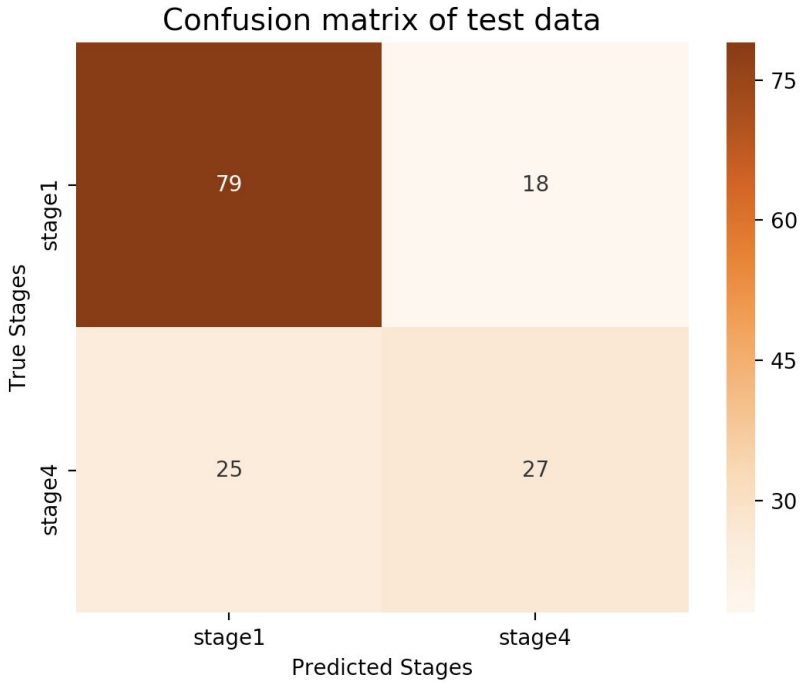
Choosing the Best Model: **Multinomial Lasso Regression**

- **All Stages** Accuracy: 60.2%

Choosing the Best Model: **Multinomial Lasso Regression**

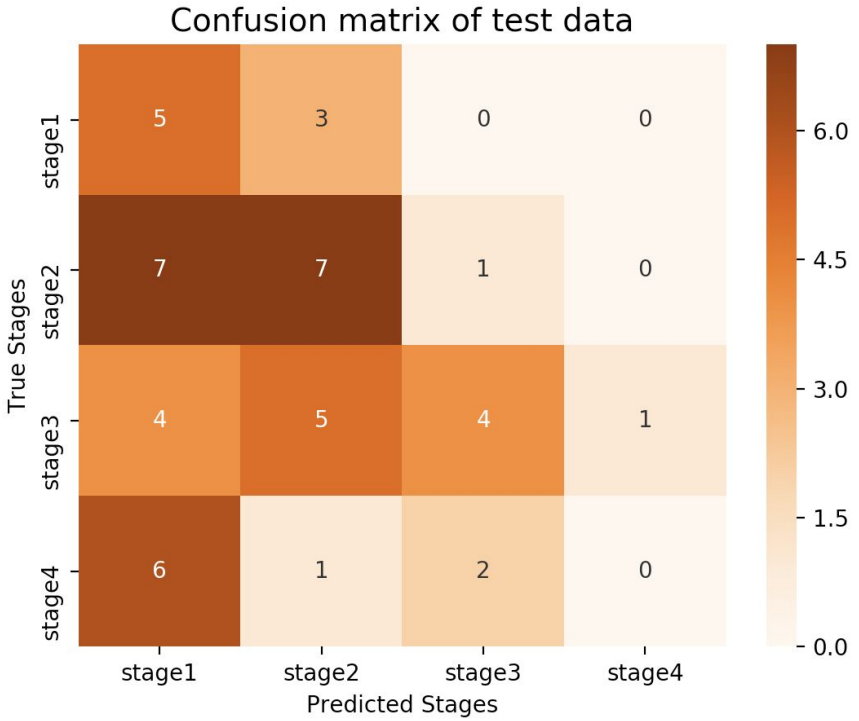
- **All Stages** Accuracy: 60.2%
- **Stage 1 vs Stage 4** Accuracy: 71.1%

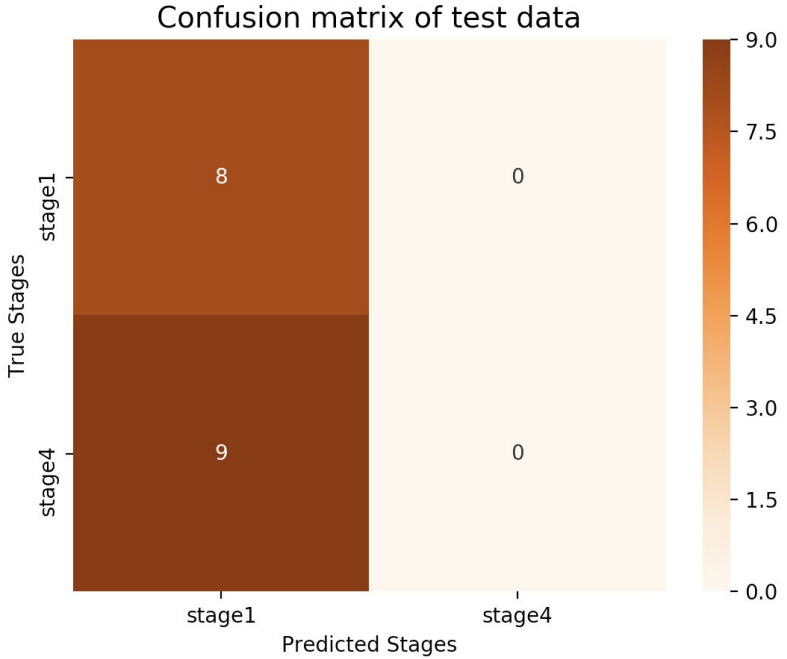




Hold-out data performance

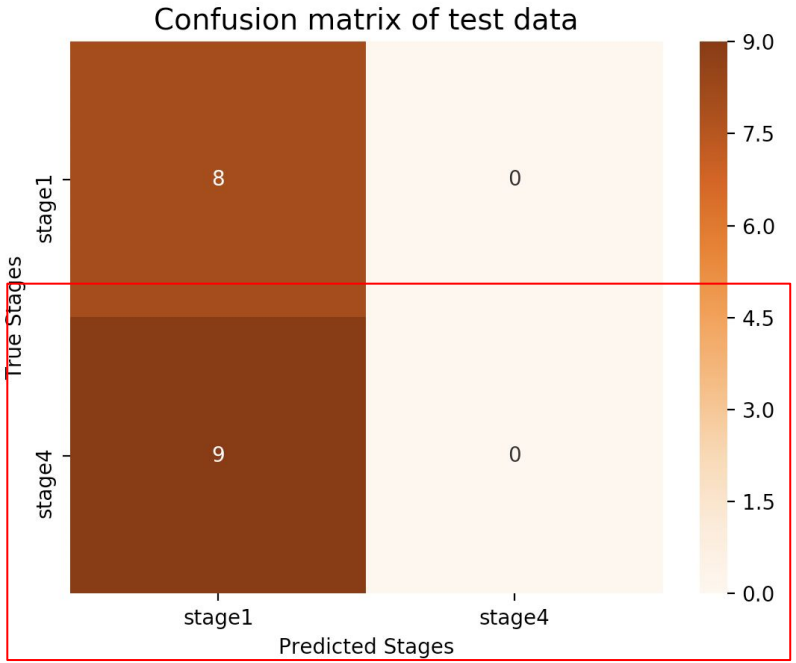
- **46 Patients** Prediction accuracy: 34.8% and 47.0%





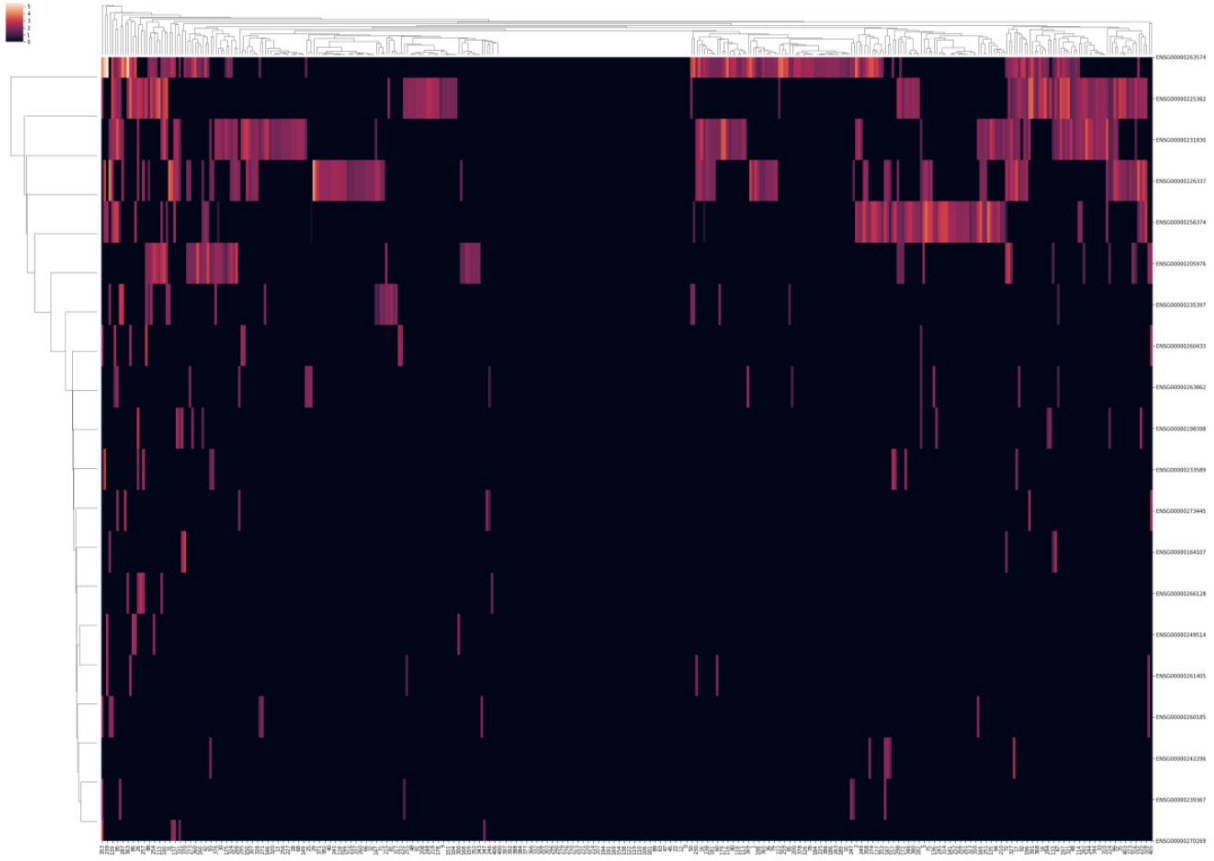
Hold-out data performance

- **46 Patients** Prediction accuracy: 34.8% and 47.0%



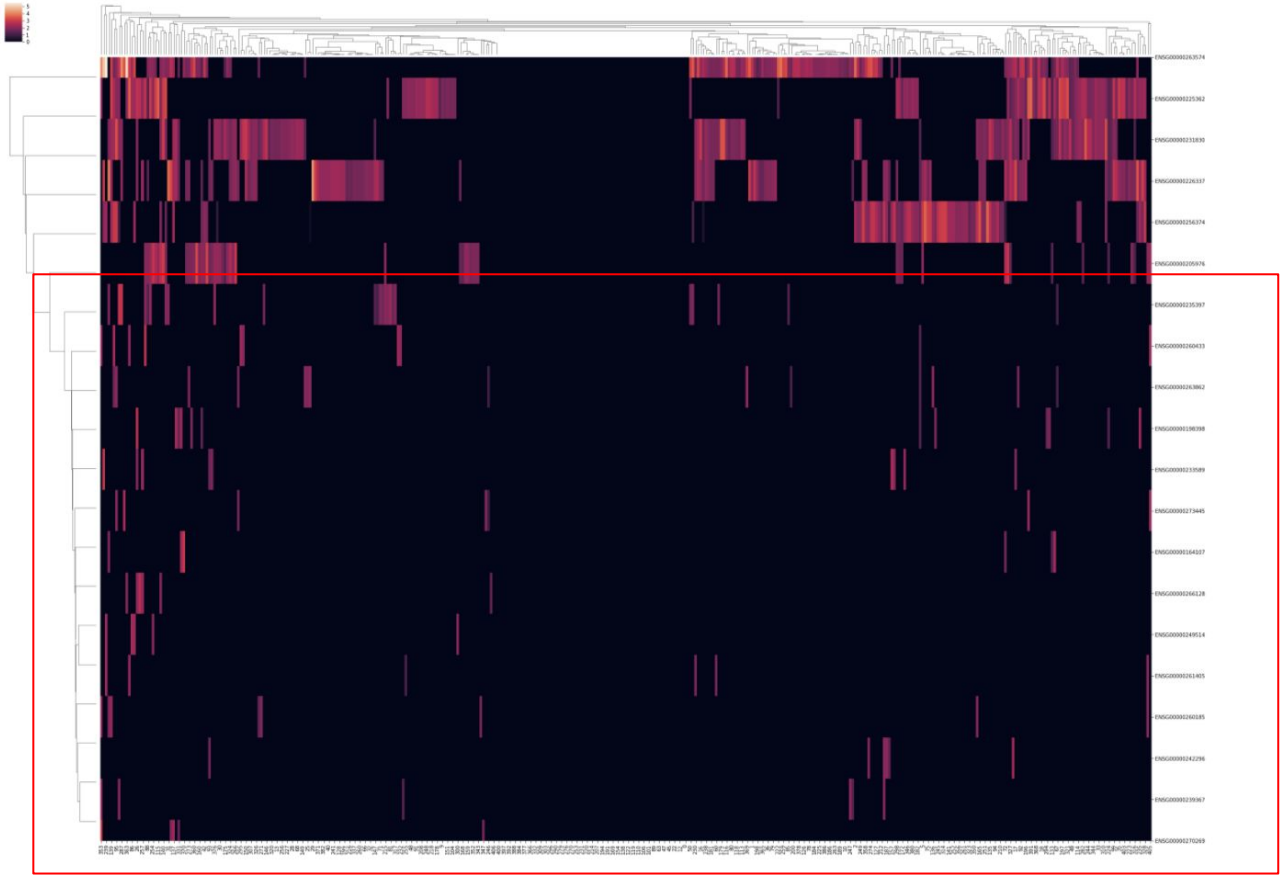
Cluster heatmap of 20 important genes

- Stage 4

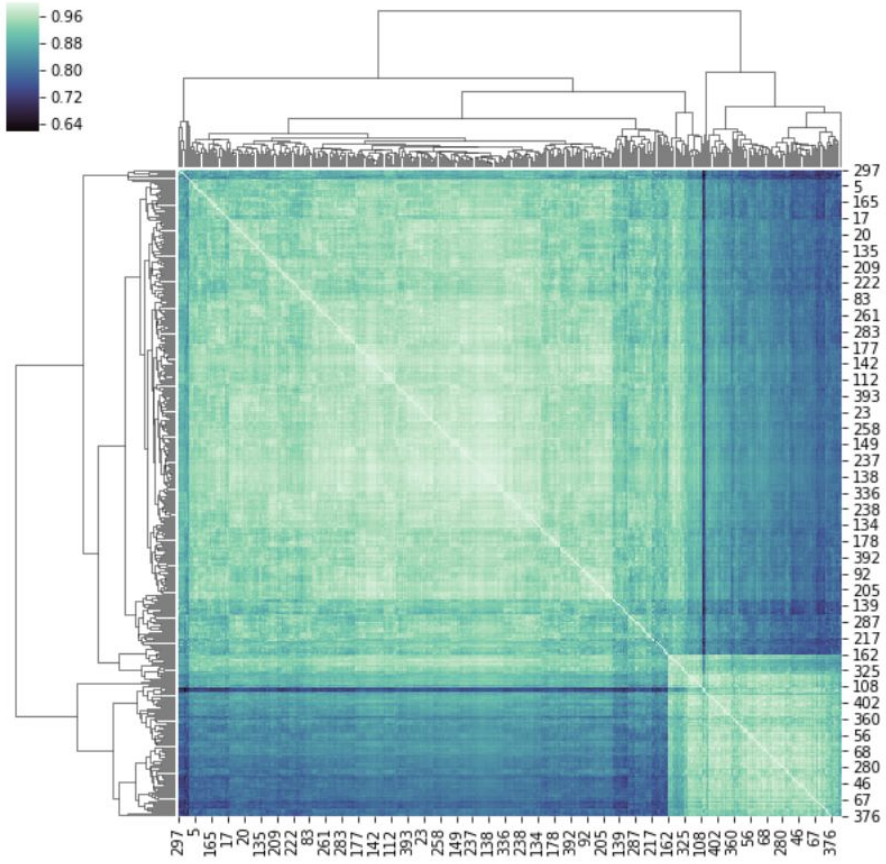


Cluster heatmap of 20 important genes

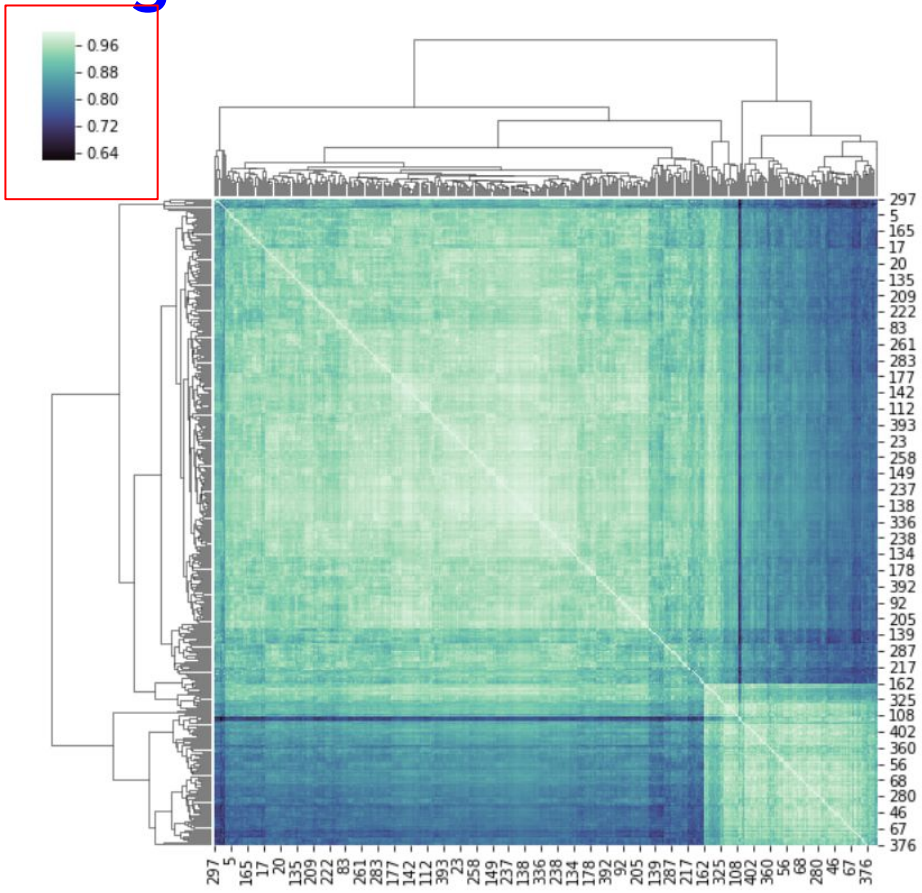
- Stage 4



Stage 4 Correlation Matrix important genes



Stage 4 Correlation Matrix important genes



Objective 2

Explore genomic pathways and their association with AMD stages

Objective 3

Explore pathway networks

What is a pathway?

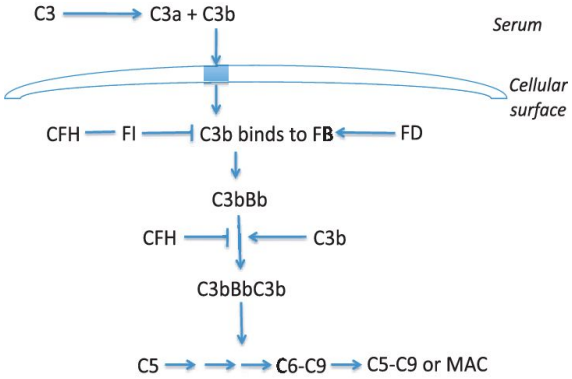
A pathway is a linked **series of actions** among genes in a cell that **lead to** a certain product or a **change**.

Genes affected in disease

CFH
C3
C5
.
.

C9

Complement pathway



What is a pathway?

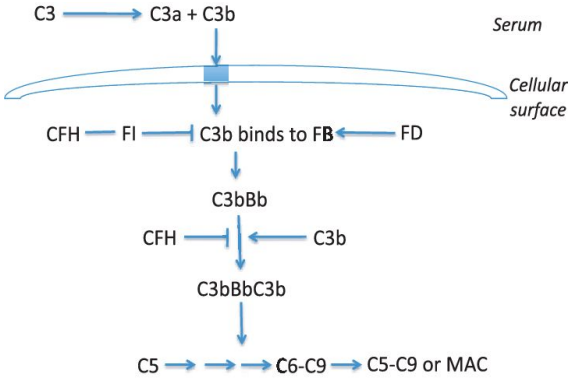
A pathway is a linked **series of actions** among genes in a cell that **lead to** a certain product or a **change**.

Genes affected in disease

CFH
C3
C5
.
.

C9

Complement pathway



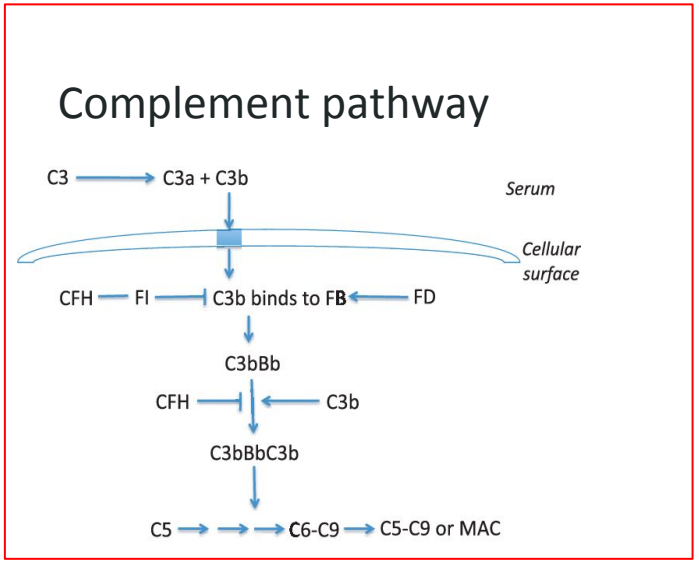
What is a pathway?

A pathway is a linked **series of actions** among genes in a cell that **lead to** a certain product or a **change**.

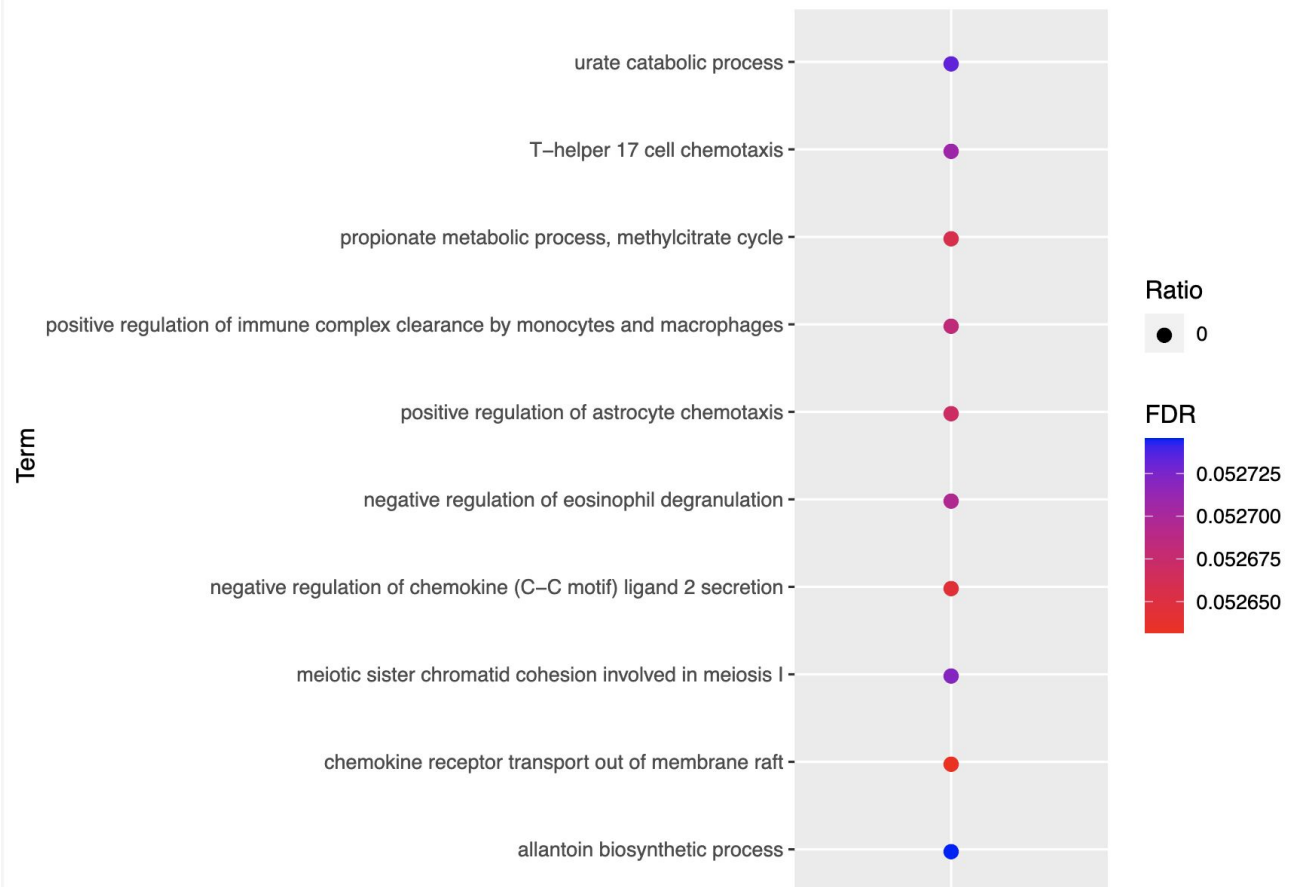
Genes affected in disease

CFH
C3
C5
.
.

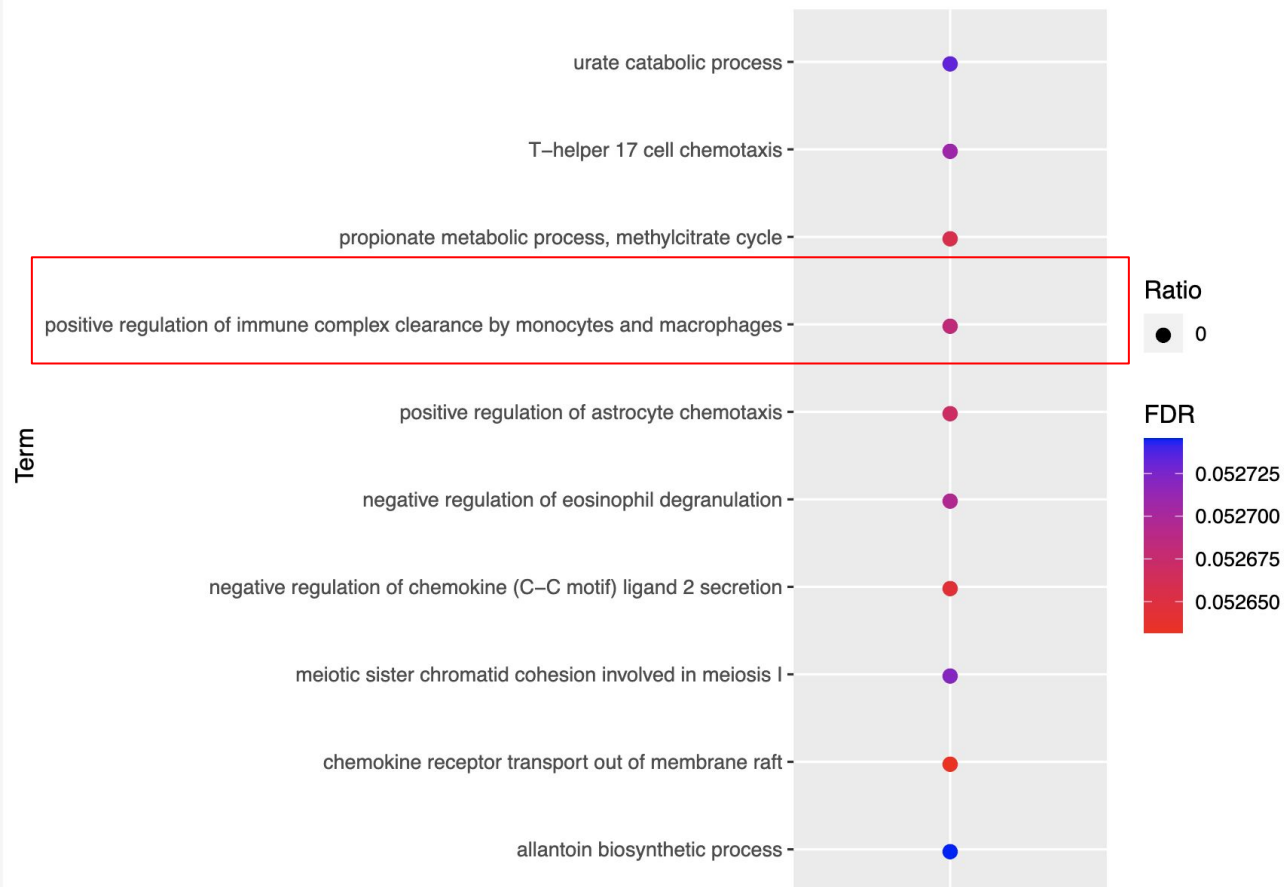
C9



GO Enrichment Analysis



GO Enrichment Analysis



Next step: Pathway Networks

- Network comprises **multiple pathways** that are already known

  to exist and interact with each other

- **Holistic view** of pathways with other experimental datasets

Thank You!