Transcriptomics analysis of Eosinophilic Esophagitis

By Harrison E. Smith

Introduction

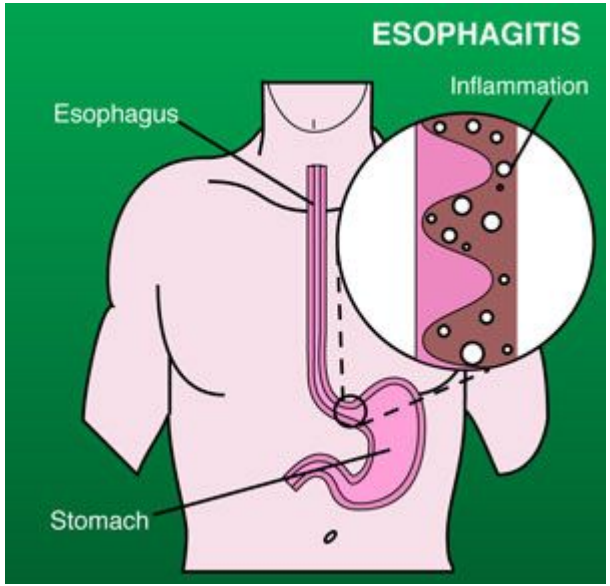
Dr. Calies Menard-Katcher of Gastroenterology

Allergic inflammation disease of the Esophagus, EoE

- 1 out of 2000 people
- White blood cell accumulation(eosinophil)
- Non fatal

Symptoms

- Narrowing of the esophagus
- pain or difficulty eating and breathing,
- nausea, vomiting



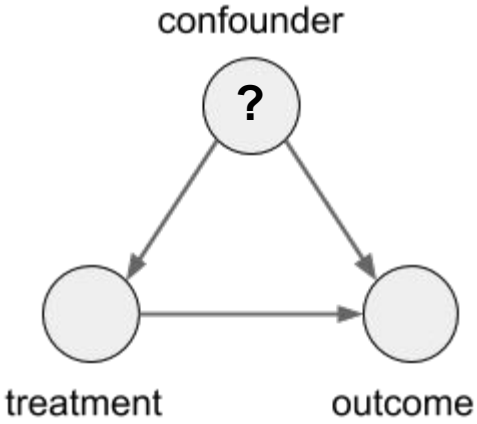
Introduction

Fibrostenotic EoE or FS EoE, is more severe

FS EoE increases severity and leads to permanent damage.

Previous RNA seq analysis (gene expression) was performed

Prior work fails to consider confounding variables



Goals

1. Observe and identify any differences in phenotype based on continuous and categorical variables
2. Adjust data for covariates
3. Discover how differences relate to pathways
4. Depict how results relate to previous research

methods

Perform DE analysis adjusting for inferred covariates

- R/ Rstudio
- DE analysis by DESeq2 w/ raw counts
- Surrogate variable analysis (SVA)

Compare results between different models

- ~Phenotype + Surrogate variables
- ~Phenotype + Age + Surrogate variables

Observe association of inferred covariates with metadata variables

Methods

Regroup samples by all EoE patients and STC use

STC is a steroid used for treatment in severe cases of EoE

STC may be a confounder in this study

Subjects :

Control - EoE with STC - EoE w/o STC medication

Design = ~Phenotype2 + Age + Surrogate variables

Results

Tableone

	Control	EoE	FS-EoE	p	test
n	0	11	9		
Gender = Male (%)	-	9 (81.8)	9 (100.0)	0.479	exact
Age (mean (SD))	-	12.15 (3.98)	15.38 (2.62)	0.052	
Food_impaction = Yes (%)	-	0 (0.0)	3 (33.3)	0.074	exact
stricture = Yes (%)	-	0 (0.0)	9 (100.0)	<0.001	exact
dilation = Yes (%)	-	1 (9.1)	9 (100.0)	<0.001	exact
EREFi (mean (SD))	-	3.91 (1.22)	3.00 (1.22)	0.115	
EREFf (mean (SD))	-	0.18 (0.40)	1.67 (1.12)	0.001	
Dysphagia_Severity (mean (SD))	-	5.82 (4.85)	9.56 (3.09)	0.061	
Fibrotic Score (mean (SD))	-	1.18 (1.17)	4.00 (0.71)	<0.001	
peak_eos/hpf (mean (SD))	-	65.00 (35.81)	75.44 (39.53)	0.543	
min_eos/hpf (mean (SD))	-	17.18 (19.19)	37.33 (24.99)	0.056	
average_eos/hpf (mean (SD))	-	38.09 (20.95)	55.44 (29.34)	0.14	
Medication-PPI = Yes (%)	-	7 (63.6)	3 (33.3)	0.37	exact
Medication-STCs = Yes (%)	-	2 (18.2)	7 (77.8)	0.022	exact

Results

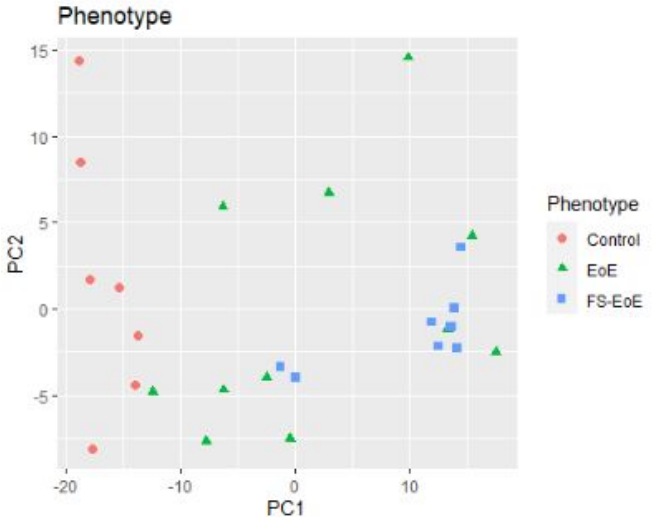
PCA

Top genes with highest CV

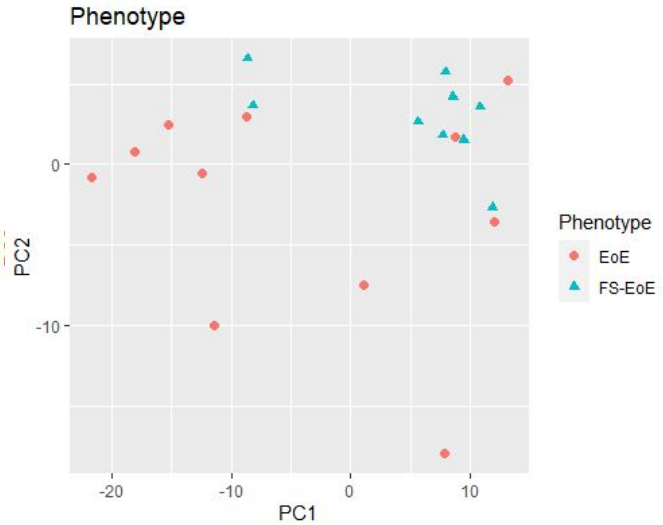
Clustering of unique phenotype groups

Some EoE and FS EoE overlap

All samples



No Control



Results

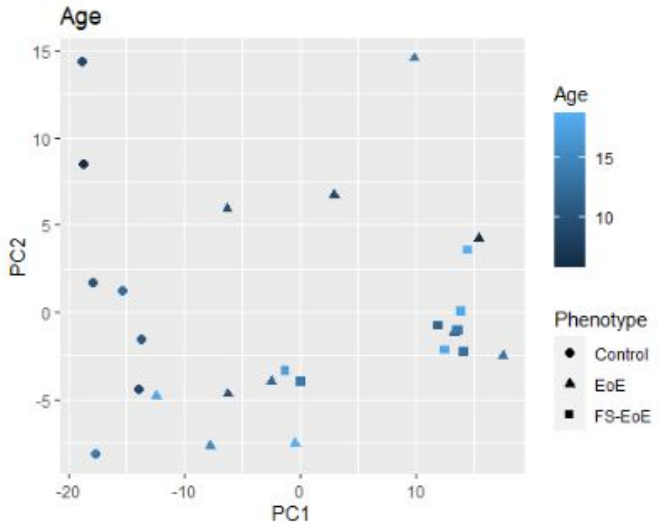
PCA

Colored by Age

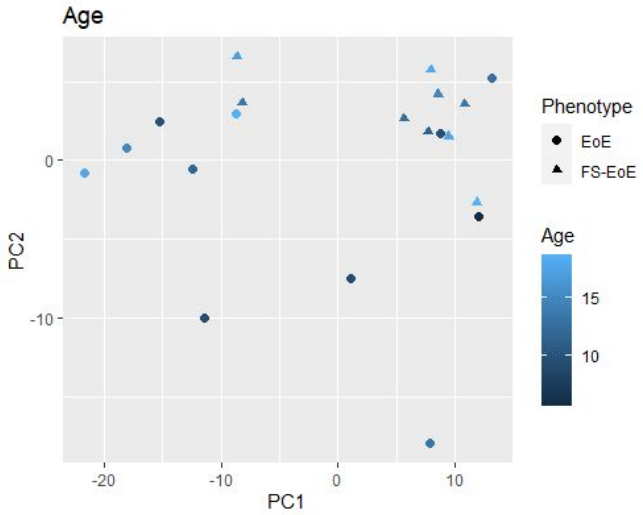
Gradient

Potential indication of association?

All samples



No Control



Results

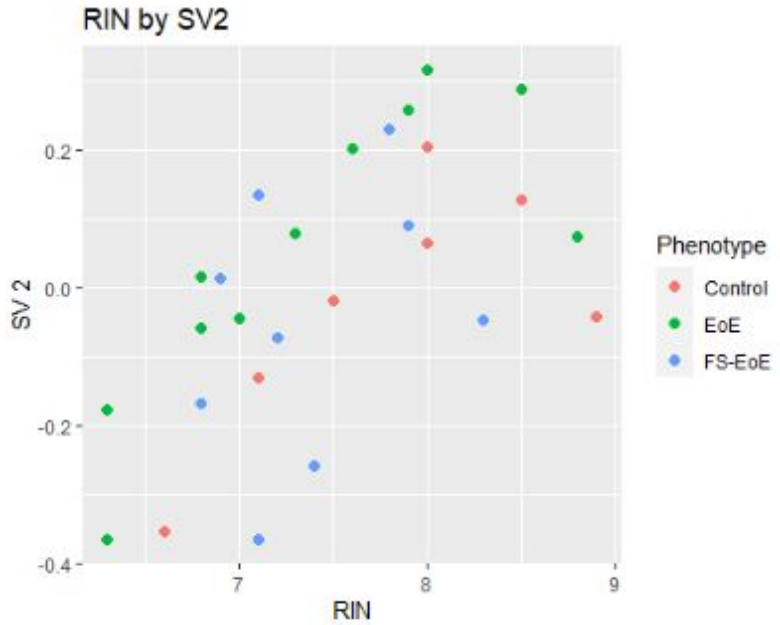
Potential associations between surrogate variables and metadata

RIN & SV2

RIN scores represent the quality of samples

Exact batch effects are unknown

Safe to adjust for SV2



Results

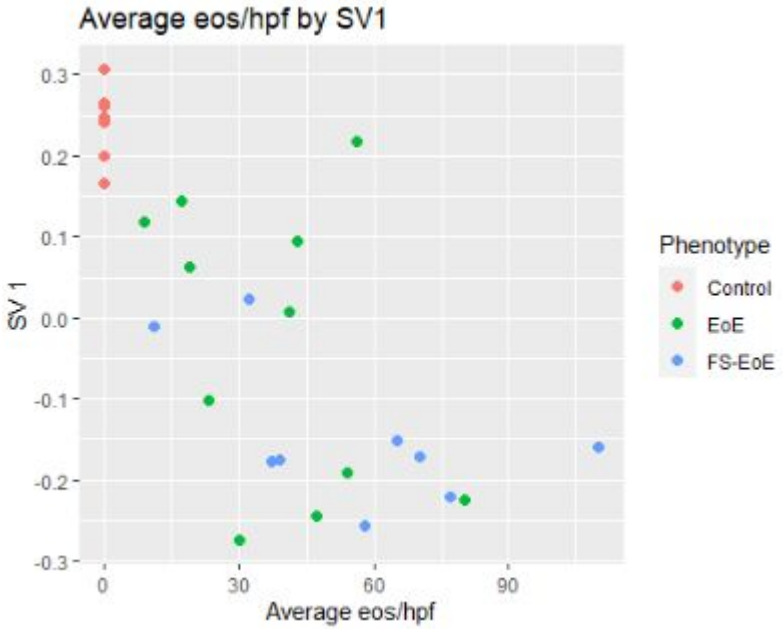
Potential associations between surrogate variables and metadata

eos/hpf Eosinophils/high power field

Diagnostic measurement for EoE

Appears to have an association with **SV1**

If this is reflective of the phenotype maybe adjusting for SV1 would be limiting gene output.



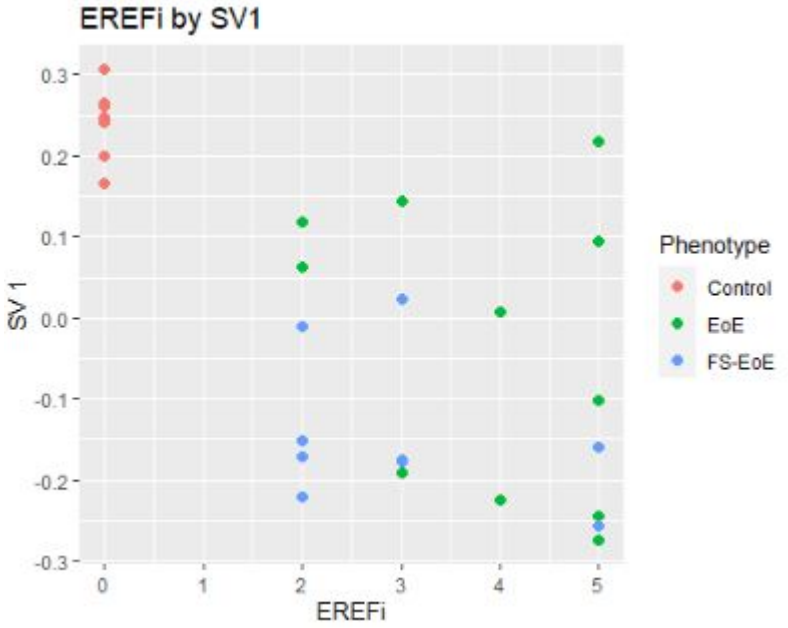
Results

Potential associations between surrogate variables and metadata

EREFi

Diagnostic measurement for inflammation

Also seems to have association with **SV1**



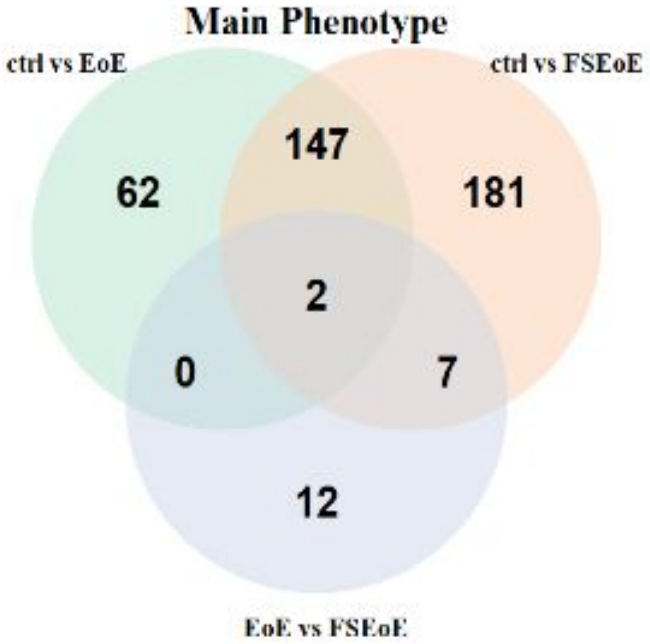
Results

Model 1 adjusts for the 4 found surrogate variables

Reduction of found DE genes by 90-95%

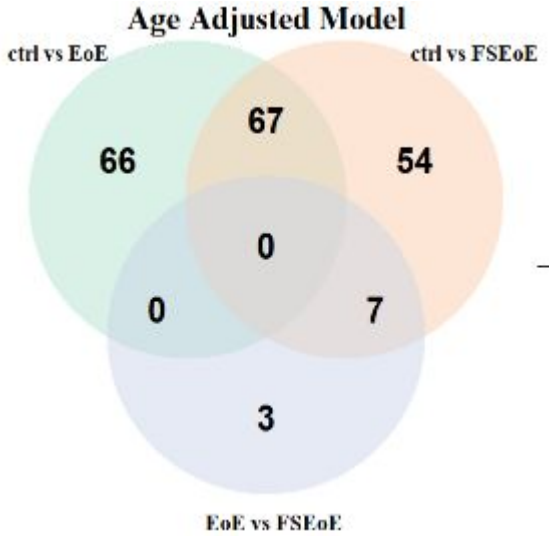
(~3.7k **vs** ~200)

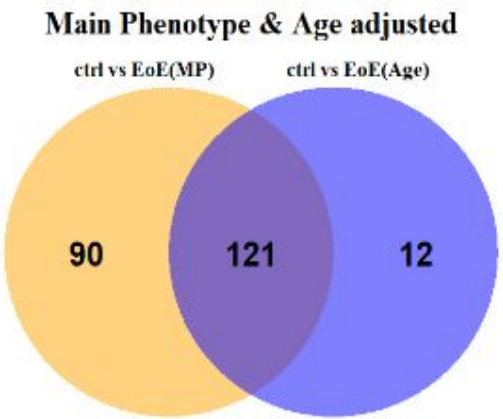
Could we be ruling out essential genes of interest?

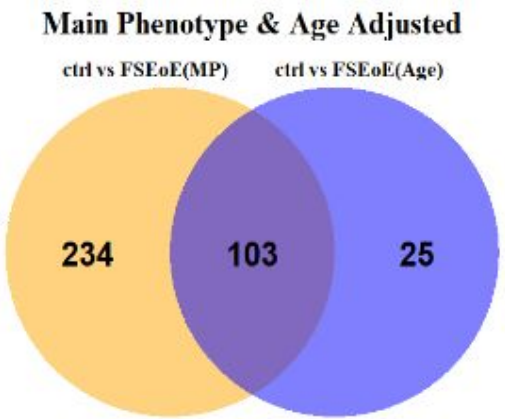


Results

Age + SV adjusted model







Discussion

More evidence that Age is confounding?

Are we creating false negatives when adjusting for SV1?

- How can we be sure?
- Overlap SV1 genes with EoE related genesets?

How and what can we extrapolate from the new STC model?

- Subset genes in first model
- Do pathway analysis with subset for validation