

- \*no need to post on canvas if Google Form
- \*no meeting next week (Veteran's Day)
- \*Dataset update
- \*report grading
- \*not recording

Week 5



# DESeq2

## ◦ cleaning vs. normalization

fixing /  
checking  
quality

align

post-mapping  
"quantify"  
comparisons

make cross-sample comparisons

"apples vs apples"

"apples vs oranges"

## ◦ Differential Expression (DE, DGEA) genes

### ◦ Fold Change? log scale?

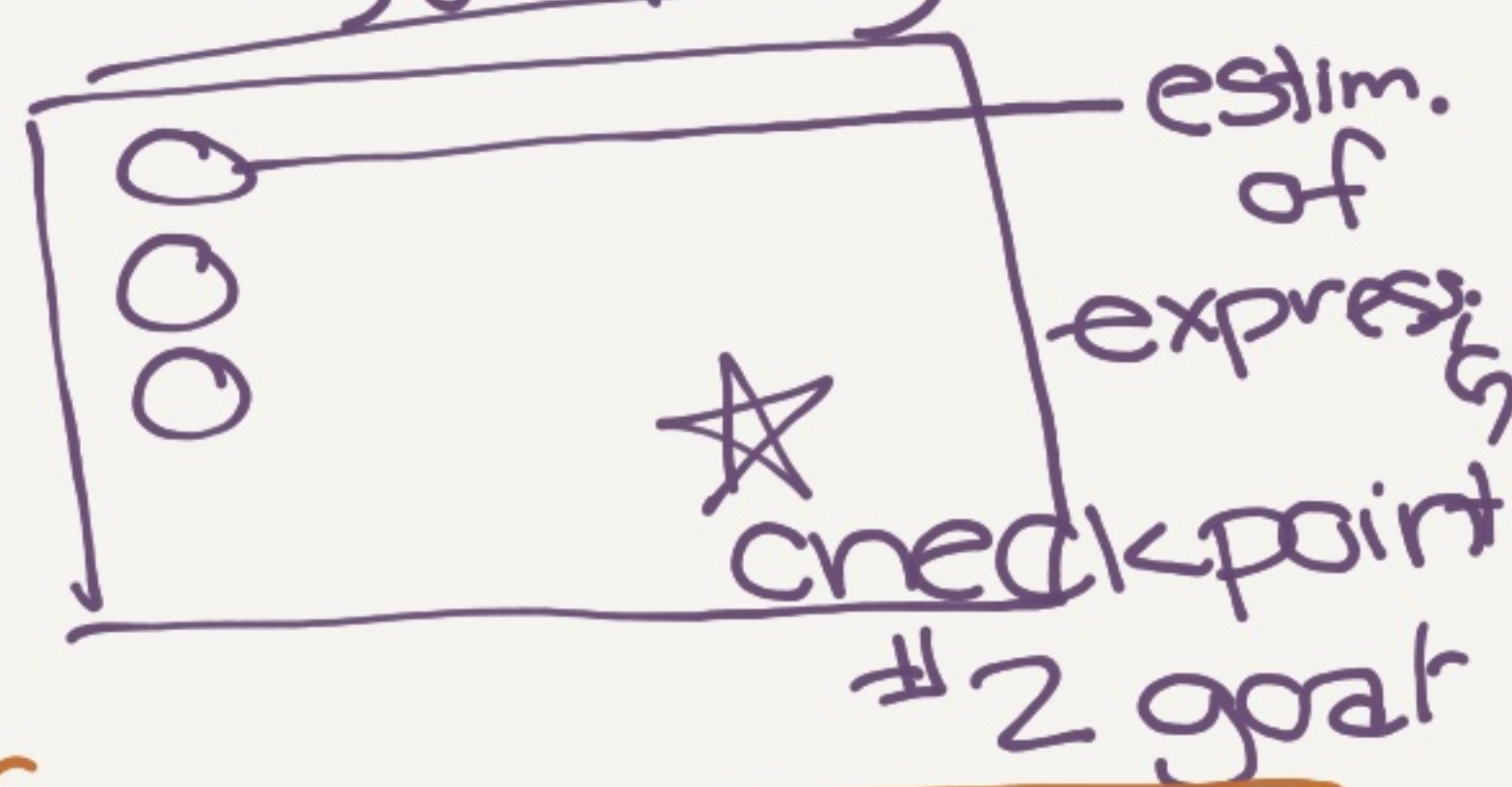
★ 50 → 100  $FC = 2$   
 $\log_2 FC = 1$

★ 100 → 50  $FC = 0.5$   
 $\log_2 FC = -1$

comparable

Why log scale?  
(bias toward large diff)

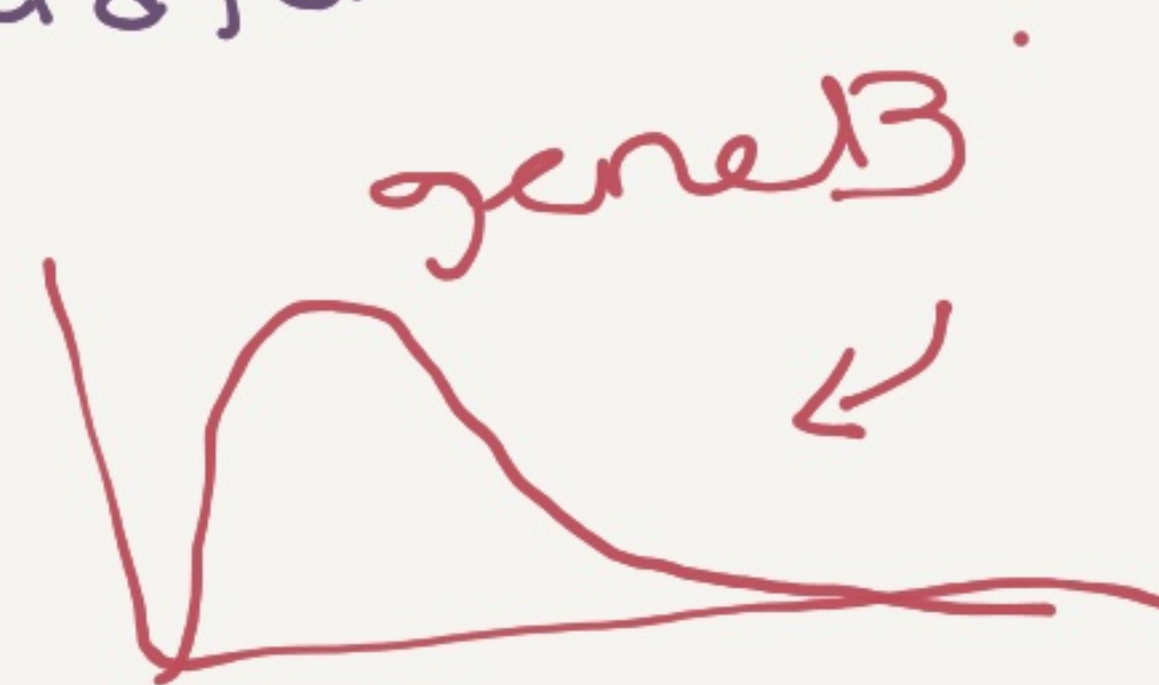
control cases  
30 → 60 ( $FC = 2$ )  
80 → 20 ( $FC = -0.75$ )  
all between 0 + 1  
(ratio of the difference)



## ◦ what are priors? Shrinkage estimators?

"Empirical Bayes" - estimating distribution from the data; updating analysis

"prior distribution estimated from the data"



## ◦ key takeaways

★ Diff. Expression Analysis

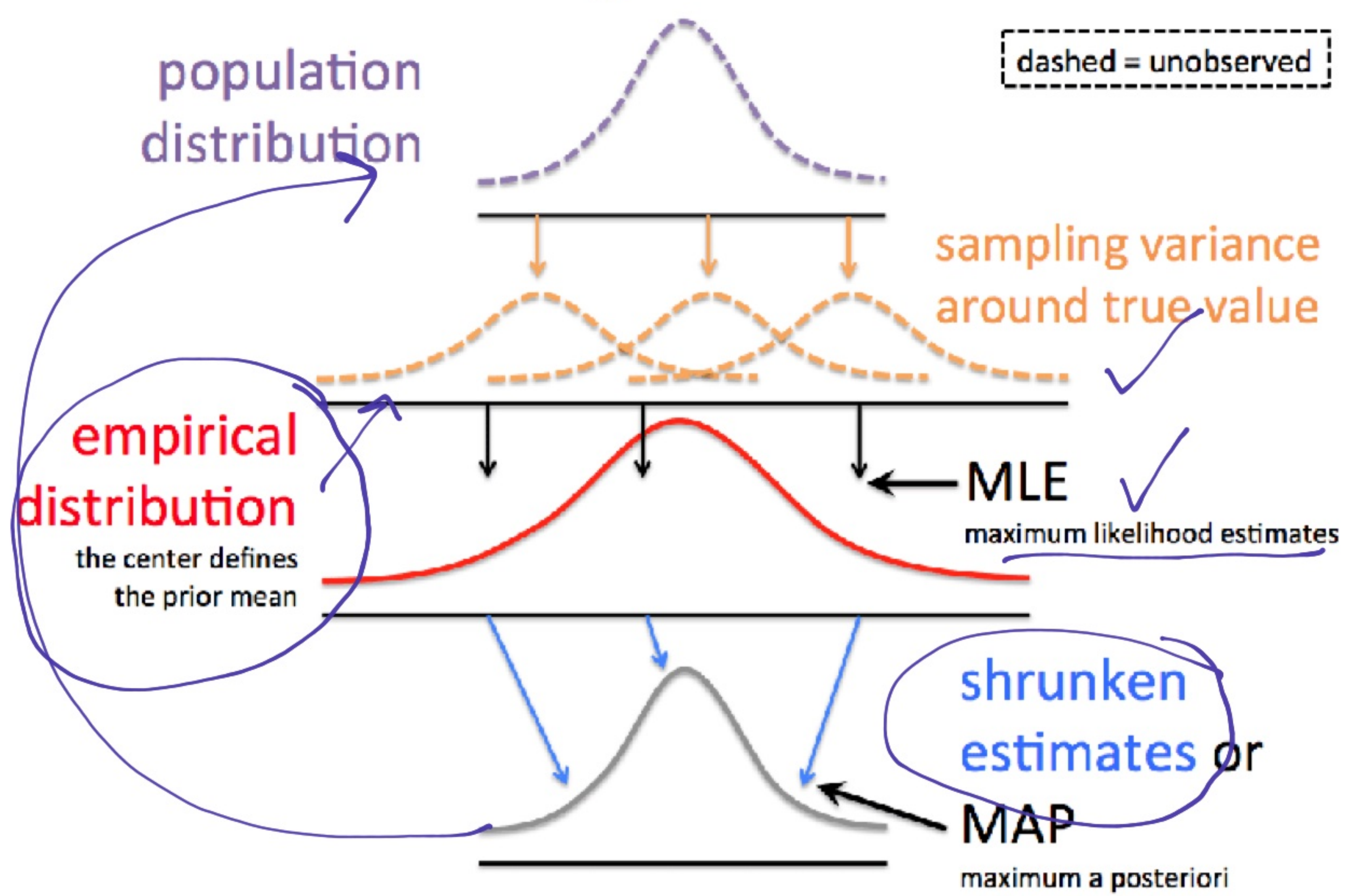
★ statistics each gene

★ Empirical Bayes

★ Normalize data



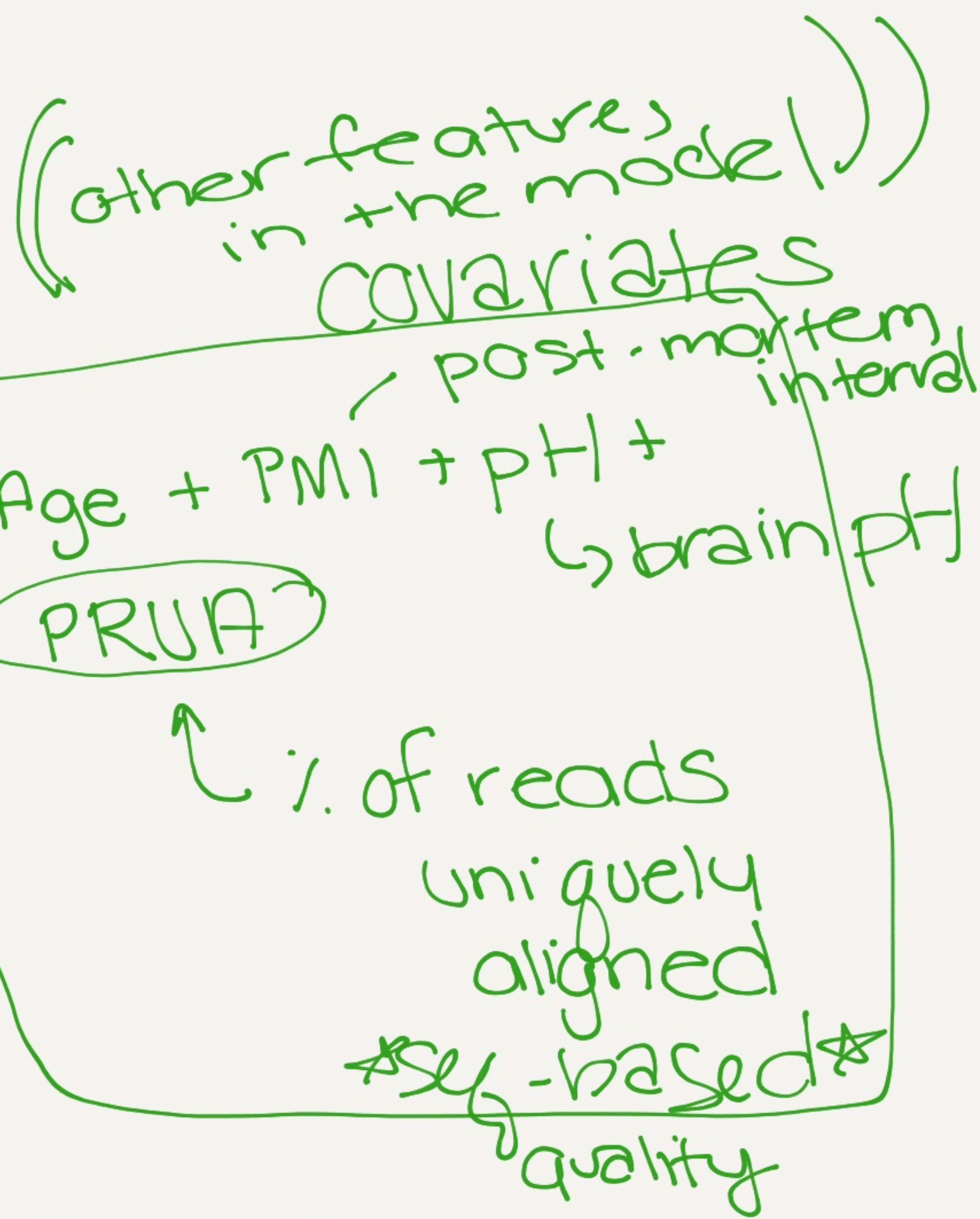
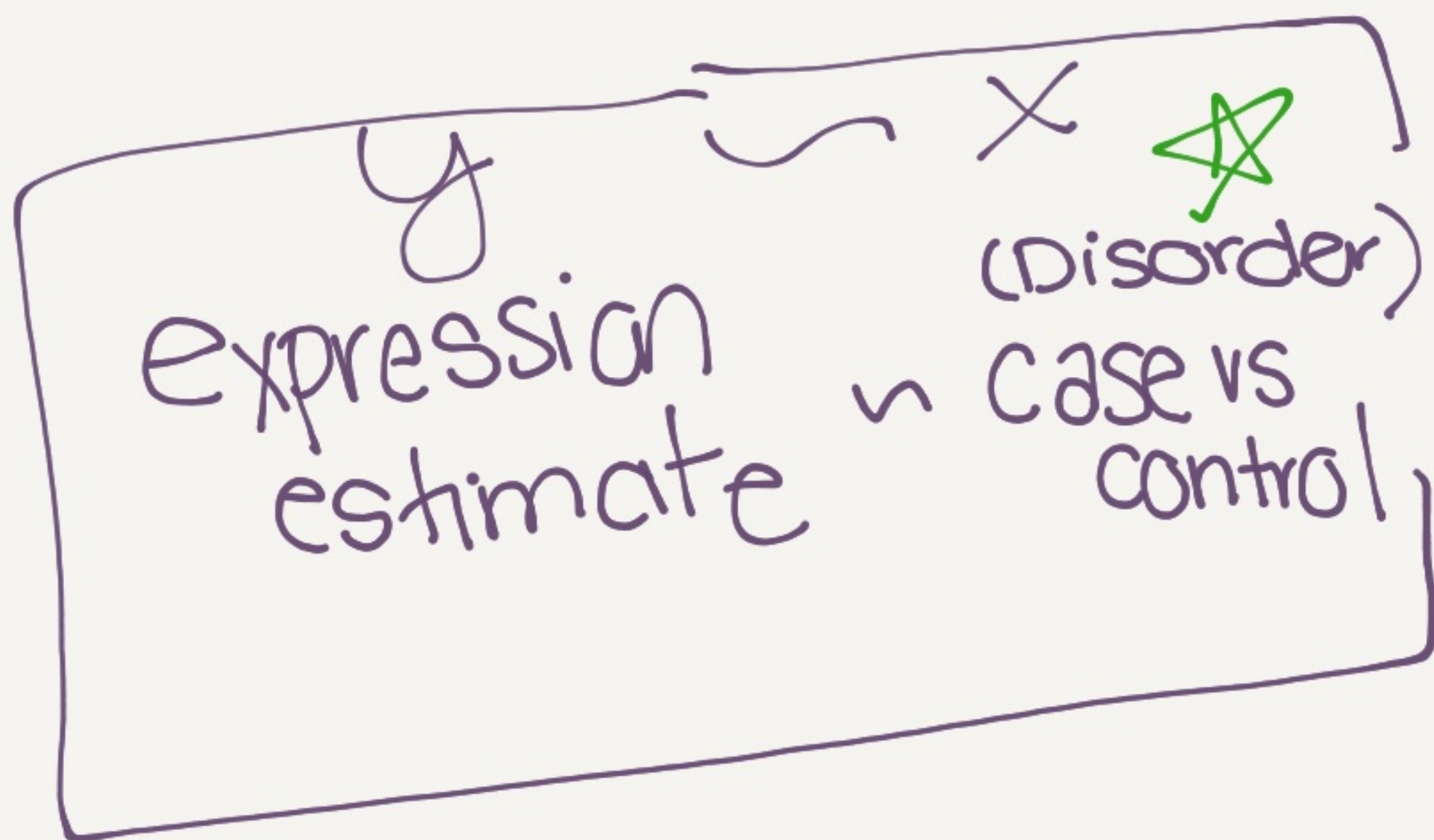
## Shrinkage estimation





# Replication Paper

## Sequencing analysis



DESeq stabilizes estimates

$H_0$ : No difference in LFC in cases vs. controls

Log Fold change

$LFC == \emptyset$  (gene not affected by treatment) \*

$H_a$ :

① expression ~ case-control \*

② expression ~ case-control + Age



# Replication Paper

## Results

AnCg  
DEPFC  
nAcc

control vs

MDD  
SCZ  
BPD

24 each

(281 samples)

(57,005 genes)

(20,000)

AnCg

→ SCZ vs control

87 genes

(FDR < 0.05)

↳ 935 GO terms (FDR < 0.05)

↳ 9 in DLPFC  
3/9 overlap

→ nAcc:  $\emptyset$

→ MDD vs control:  $\emptyset$  genes ✓

→ BPD vs control:  $\emptyset$  genes ✓

Cross-disorder comparison

SCZ + BPD = largest overlap