Batch-effect Assessment on Network Using GTEx

BST 550 Project Presentation

Yun Zhang

12-16-2015

Batch-effect on Gene Network

▶ Technical effect: when, where

▶ Biological effect: types of death (if used **ventilator**), **gender**, ethnicity

▶ Mixture of cell types: especially for models that rely on correlation structure

▶ How much of the reported findings is due to batch-effects?

▶ Does an edge in a gene network really represent a biological mechanism? Or is it just due to similar composition of cell types?

GTEx Data: Data Preprocessing

- Focused on the **lung** tissue samples:
  - 133 samples
- RNA-seq
- Variance-stabilizing transformation on the count data
- Filtered out low expression and small variation genes
 - Sample data 1: 175 genes
  - Sample data 2: 2157 genes

Analysis Plan

- Weighted Gene Co-expression Network Analysis (WGCNA)
  - Fit a correlation network
  - Gene modules and their corresponding eigengenes
- Investigate potential batch-effects
  - Gender
  - Ventilator
- Adjust known/unknown batch-effects using existing methods
  - ComBat – for known effects
  - Surrogate Variable Analysis (SVA) – for unknown effects
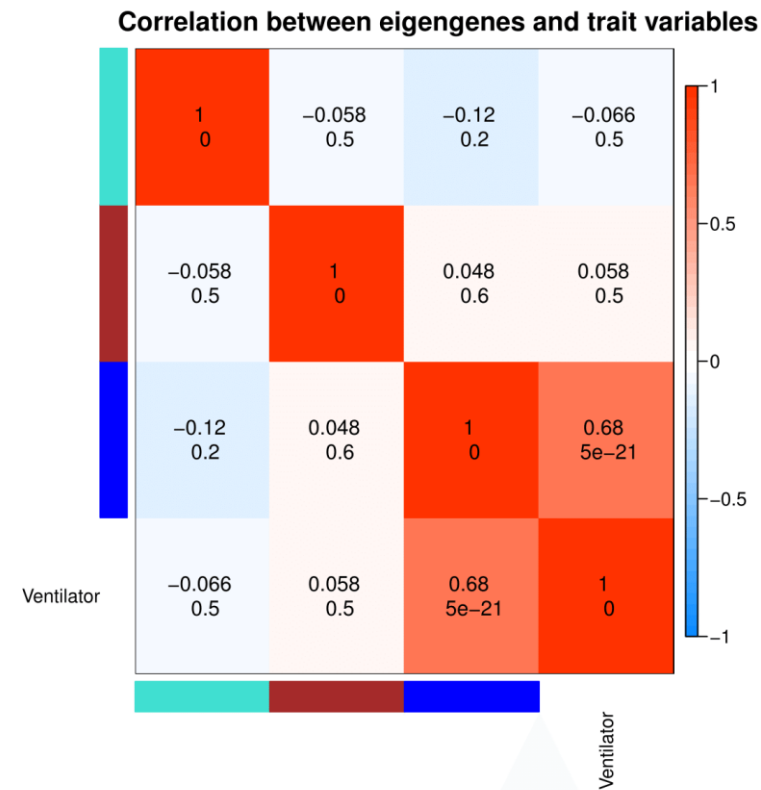
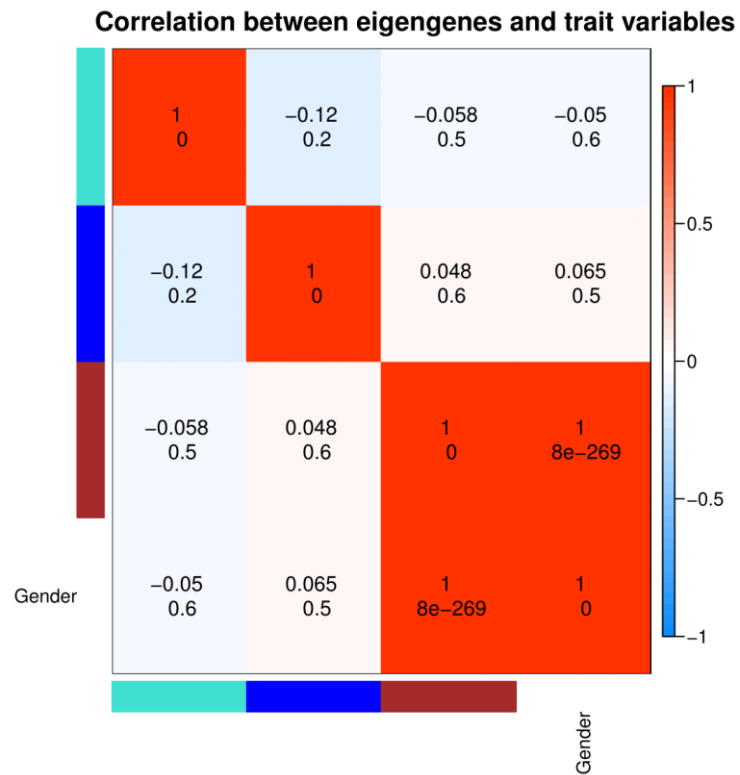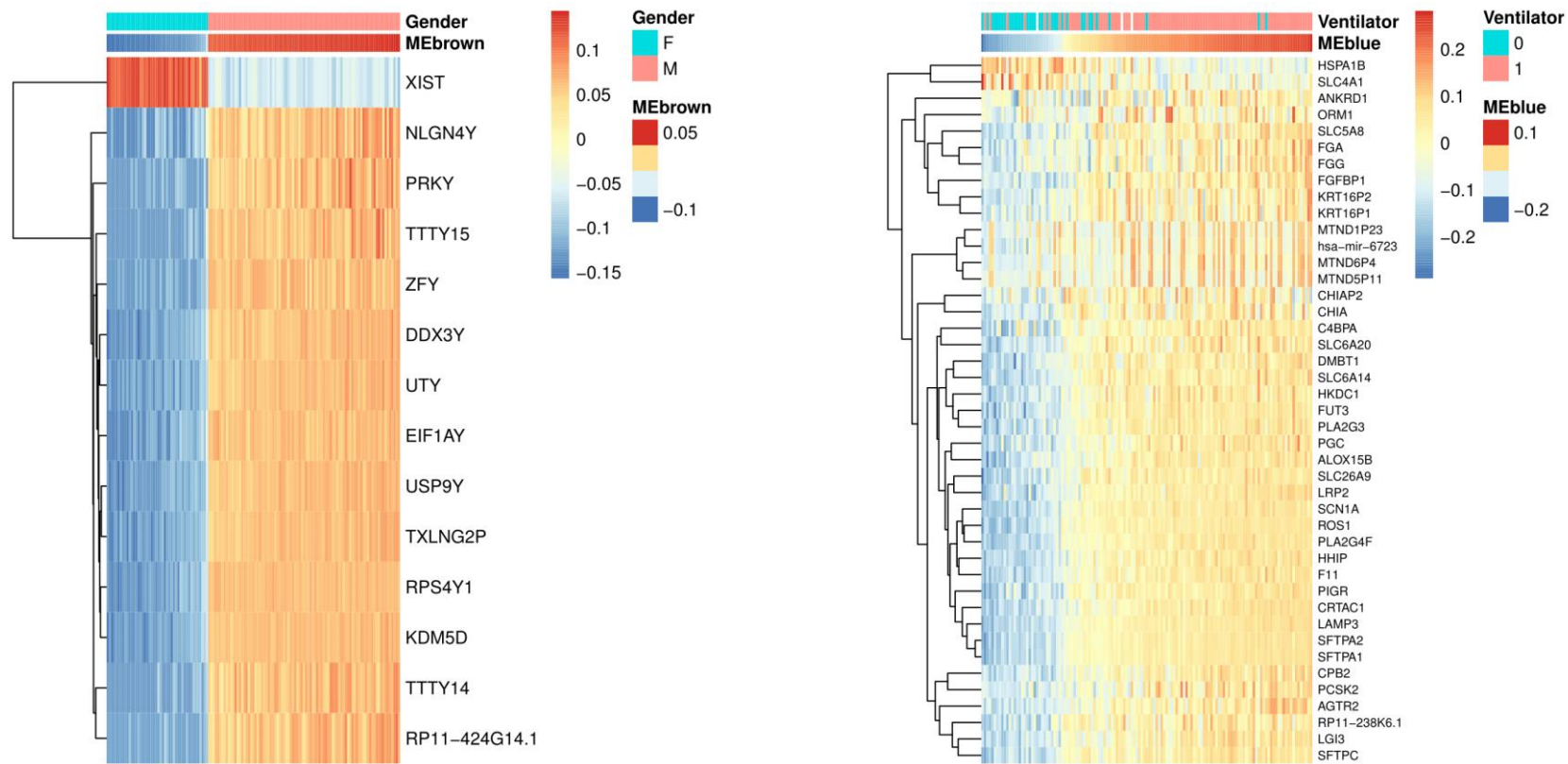# Network by WGCNA for Sample Data 1
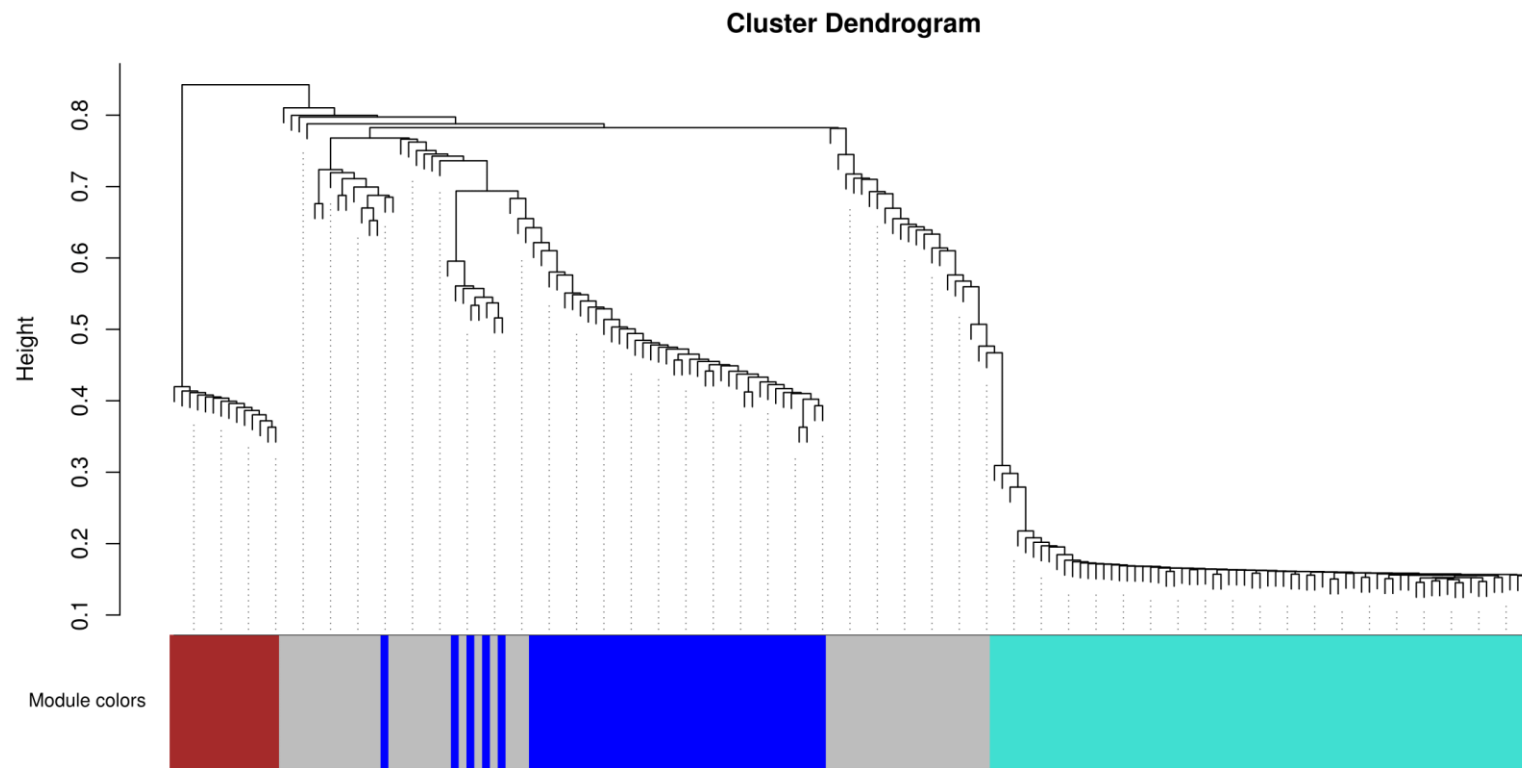
An Actual View of the Network
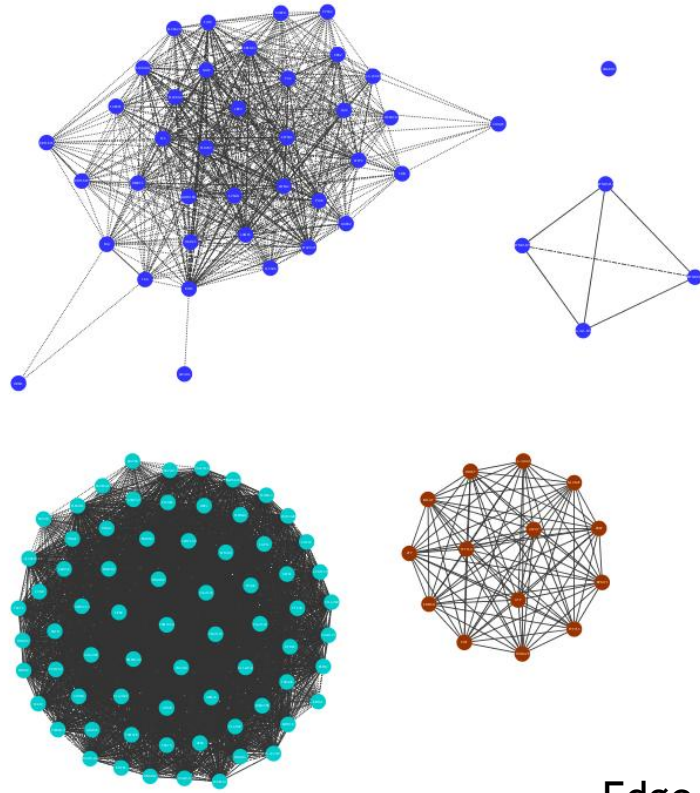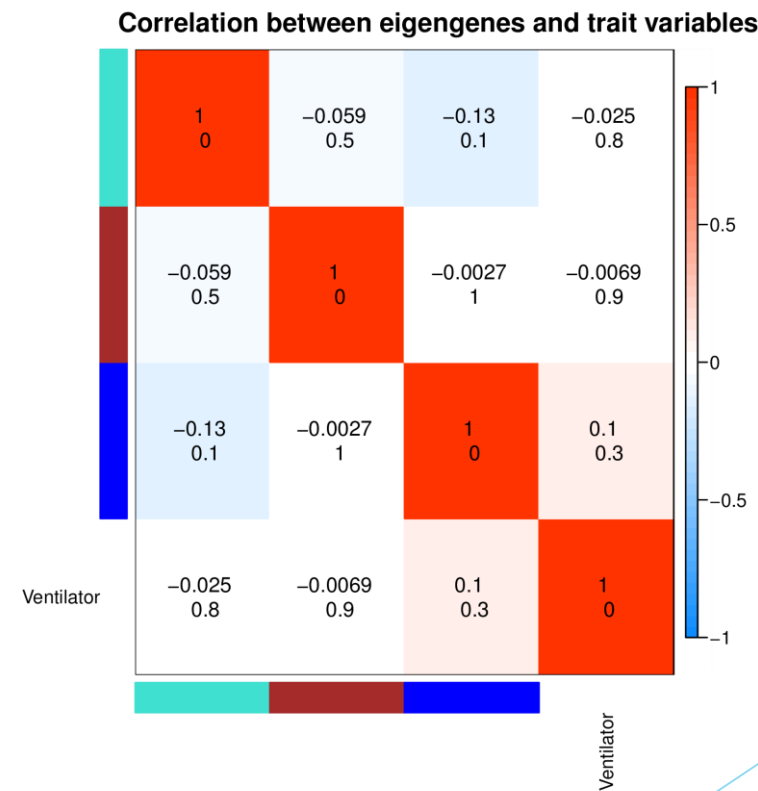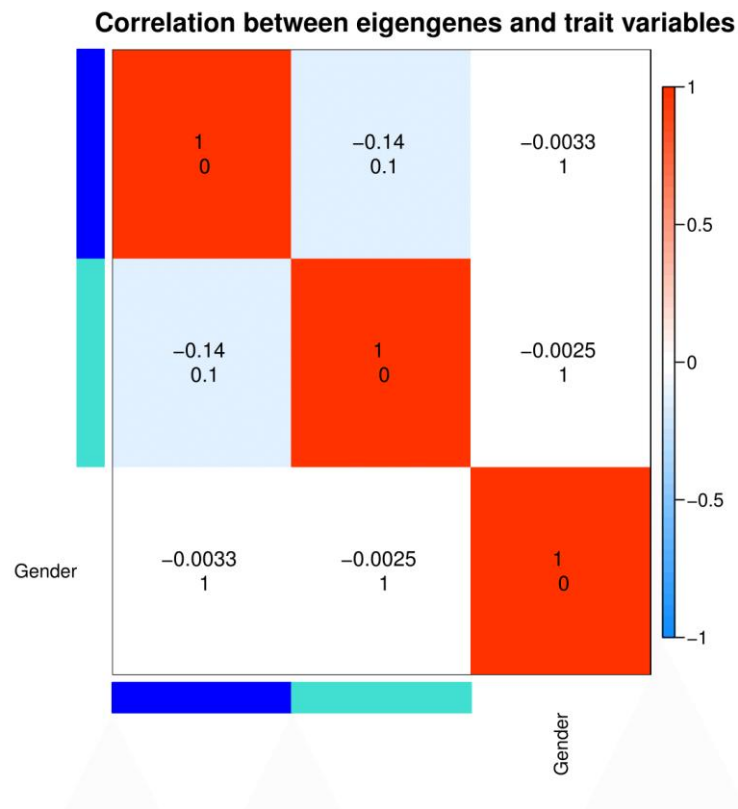

Edge visibility: correlation > 0.5

Suspects of Gene Modules Induced by Batch-Effect

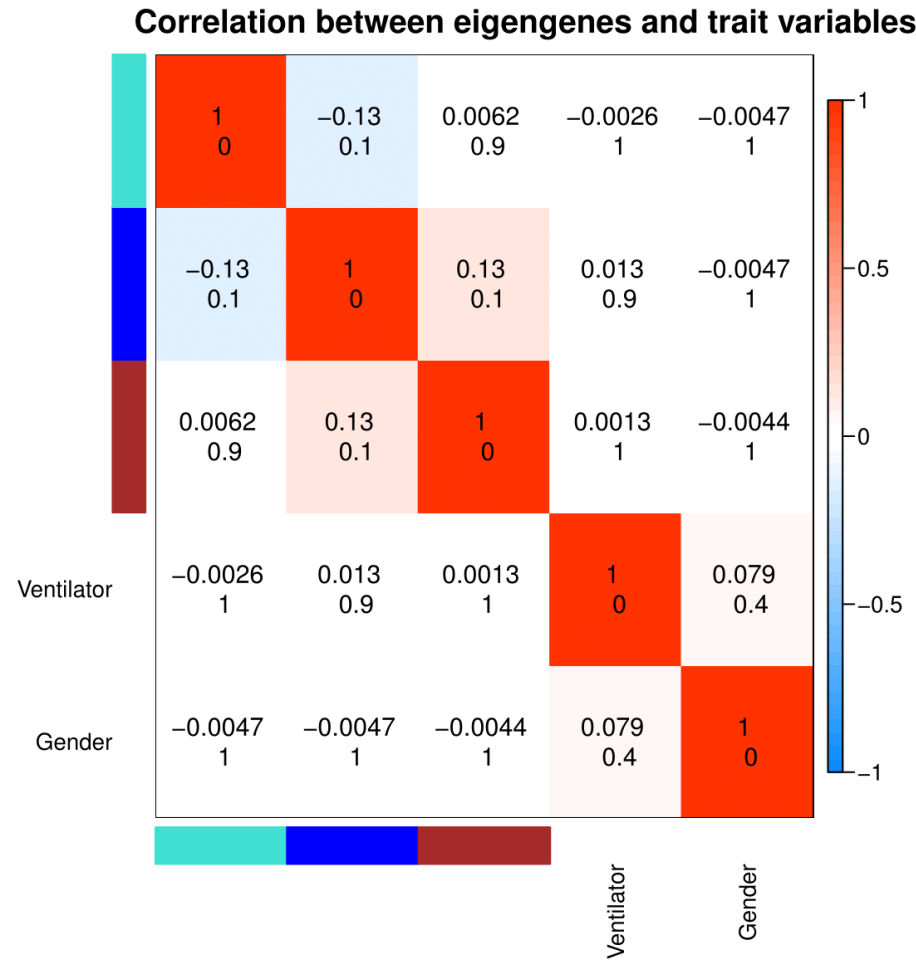Correlation between eigengenes and trait variables

Gene Expression Profile of Suspicious Modules

Batch-Effect Adjustment Using ComBat (One Batch-Effect A Time)

Correlation between eigengenes and trait variables

Batch-Effect Adjustment Using ComBat (Combined Batch-Effects)

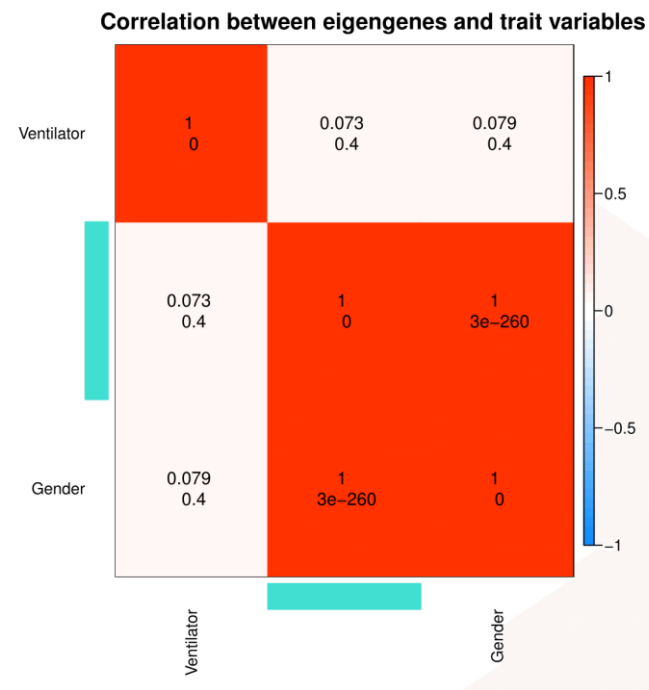Correlation between eigengenes and trait variables

Batch-Effect Adjustment Using SVA (Bad Example)

Adjust for gender
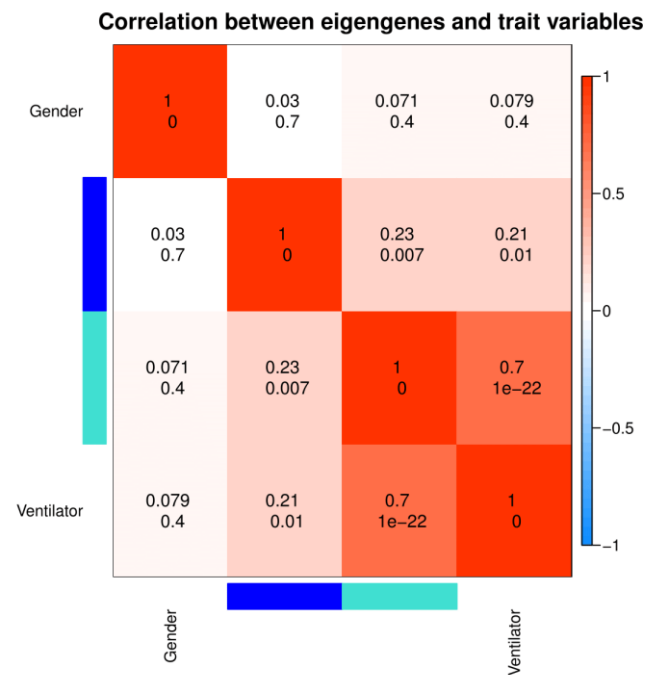
Adjust for ventilator

Conclusion

- All currently existing batch-effect removal methods focus on adjusting batch-effects for differential expression

- We conducted the first assessment of batch-effect adjustment on network

- Sample data 2 shows similar results as sample data 1

- ComBat can effectively remove batch-effects for correlation network; however, there is technical difficulty of using SVA to adjust the batch-effects

- Future work:
    - Look at other types of network: information theory, Bayesian inference
    - Batch-effect methods specifically for network, e.g. improvement of SVA