Different Alternative Types Simulated Data

David Gerard

2016-06-08

Abstract

I compare various competitors under the same alternative scenarios in Stephens (2016) but using simulated data rather than the GTEX data. The results are more muddled than in the GTEX simulations. Overall, it seems RUVASH does best. SUCCOTASH does much worse than in the GTEX simulations. This is different from this previous write-up in two ways: (1) I shrunk the variances when generating Z, α, and E, and (2) I used limma-shrunk variances for RUVASH. This resulted in improved performance of RUVASH.

Simulation Setup

I ran through 200 repetitions of generating data from the factor-augmented Gaussian regression model:

$$Y_{n\times p} = X_{n\times k}\beta_{k\times p} + Z_{n\times q}\alpha_{q\times p} + E_{n\times p} \tag{1}$$
$$E \sim N_{n\times p}(0, \Sigma \otimes I_n) \tag{2}$$
$$\Sigma = diag(\sigma_1^2, \ldots, \sigma_p^2) \tag{3}$$

under parameter conditions:

- $n \in \{10, 20, 40\}$,
- $p = 1000$,
- $k = 2$,
- $q = 5$
- $\beta_{1j} \overset{iid}{\sim} N(0, 1)$,
- β_{2j} being iid a mixture of normals with null proportion π_0,
- $Z_{ij} \overset{iid}{\sim} N(0, 1/4)$,
- $\alpha_{ij} \overset{iid}{\sim} N(0, 1/4)$,
- $X_{j1} = 1$ for all $j = 1, \ldots, n$,
- $X_{j2} = 1$ for $j = 1, \ldots, n/2$ and $X_{j2} = 0$ for $j = n/2, \ldots, n$,
- $\sigma_j^2 = 1/4$ for all $j = 1, \ldots, p$,
- $\pi_0 \in \{0.5, 0.9, 1\}$,
- The alternative distribution of β_{2j} being either spiky, near-normal, flattop, skew, big-normal, or bimodal, where these are the same alternatives defined in Stephens (2016) and the following table. New alternatives are generated every iteration.

Scenario	Alternative Distribution
Spiky	$0.4N(0, 0.25^2) + 0.2N(0, 0.5^2) + 0.2N(0, 1^2), 0.2N(0, 2^2)$
Near Normal	$2/3N(0, 1^2) + 1/3N(0, 2^2)$
Flattop	$(1/7)N(-1.5, .5^2) + N(-1, .5^2) + N(-.5, .5^2) + N(0, .5^2) + N(0.5, .5^2) + N(1.0, .5^2) + N(1.5, .5^2)$
Skew	$(1/4)N(-2, 2^2) + (1/4)N(-1, 1.5^2) + (1/3)N(0, 1^2) + (1/6)N(1, 1^2)$
Big-normal	$N(0, 4^2)$
Bimodal	$0.5N(-2, 1^2) + 0.5N(2, 1^2)$

At each iteration, I generated new values of X, Z, α, β, E, and thus also Y.

Questions on simulation settings

- Right now, I am simulating Z independently of X. Should I be generating Z to have some pre-specified correlation with X?
- I have $q = 5$ for all n. Should I increase q as n increases?
- I assume q is known. Should I estimate it every iteration, as I do in the GTEX simulations? Maybe SUCCOTASH and RUVASH are more robust to the choice of q? Or maybe the true q isn't the best q to use when doing estimation?
- I didn't choose the variances of β_{1j}, Z_{ij}, and E_{ij} carefully. Should I vary them so that proportion of variance explained by β_{2j} is different?
- Should I even include β_{1j} in the simulations?

Methods

The confounder adjustment methods I look at in this write-up are:

- OLS + qvalue.
- OLS + ASH
- SUCCOTASH using normal mixtures and heteroscedastic PCA as the factor-analysis method. This is the two-step version that does variance inflation.
- The robust regression version of CATE using PCA as the factor analysis method + qvalue.
- SVA + qvalue.
- RUVASH with normal likelihood. I used limma-variances here.
- RUV4 (GLS version) using the variance inflation. I used limma-shrunk variances here.
- Negative control version of CATE using PCA as the factor analysis method + qvalue.
- RUV2 + qvalue.
- RUV4 + qvalue.
- RUV4 + ASH (without variance inflation).

Results

Estimates of π_0

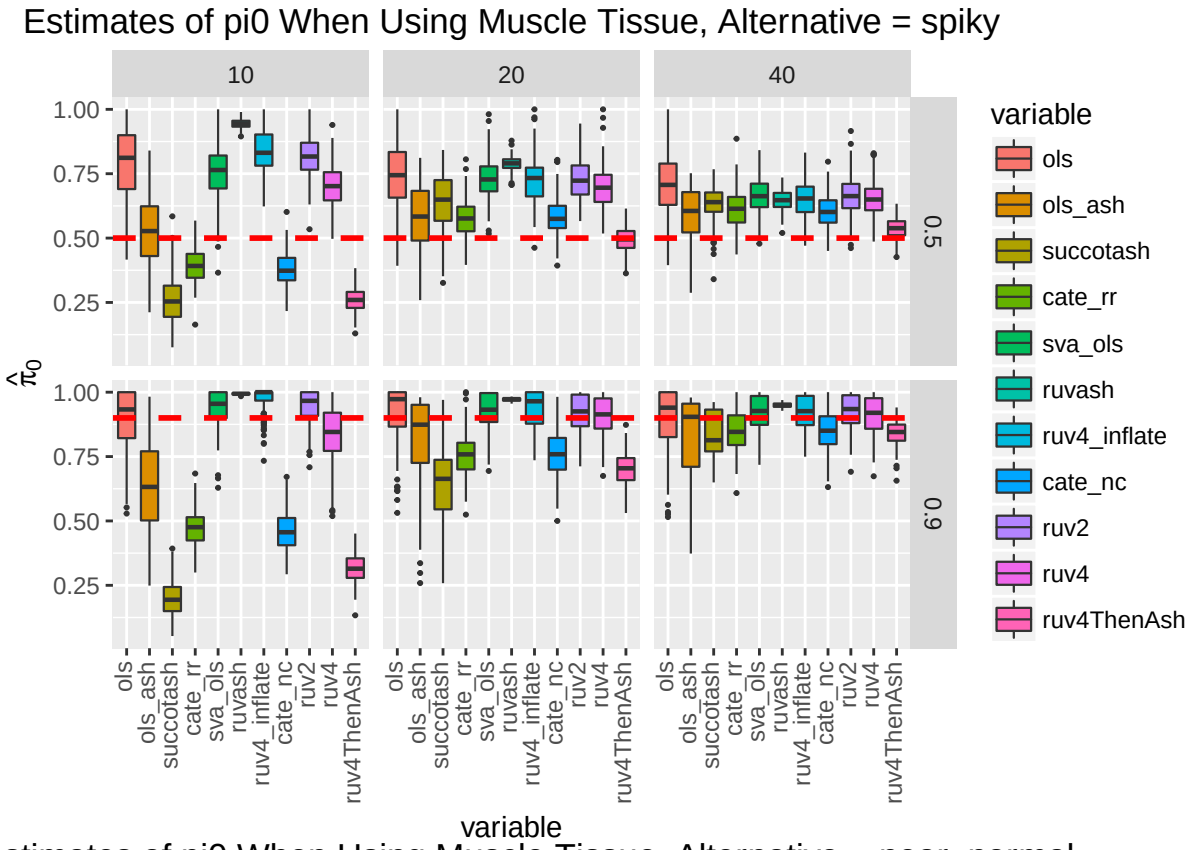
- When $n = 40$, the clear winner is RUVASH.
- SUCCOTASH does well when $n = 40$ and $\pi_0 = 0.5$, but not so well when $\pi_0 = 0.9$.

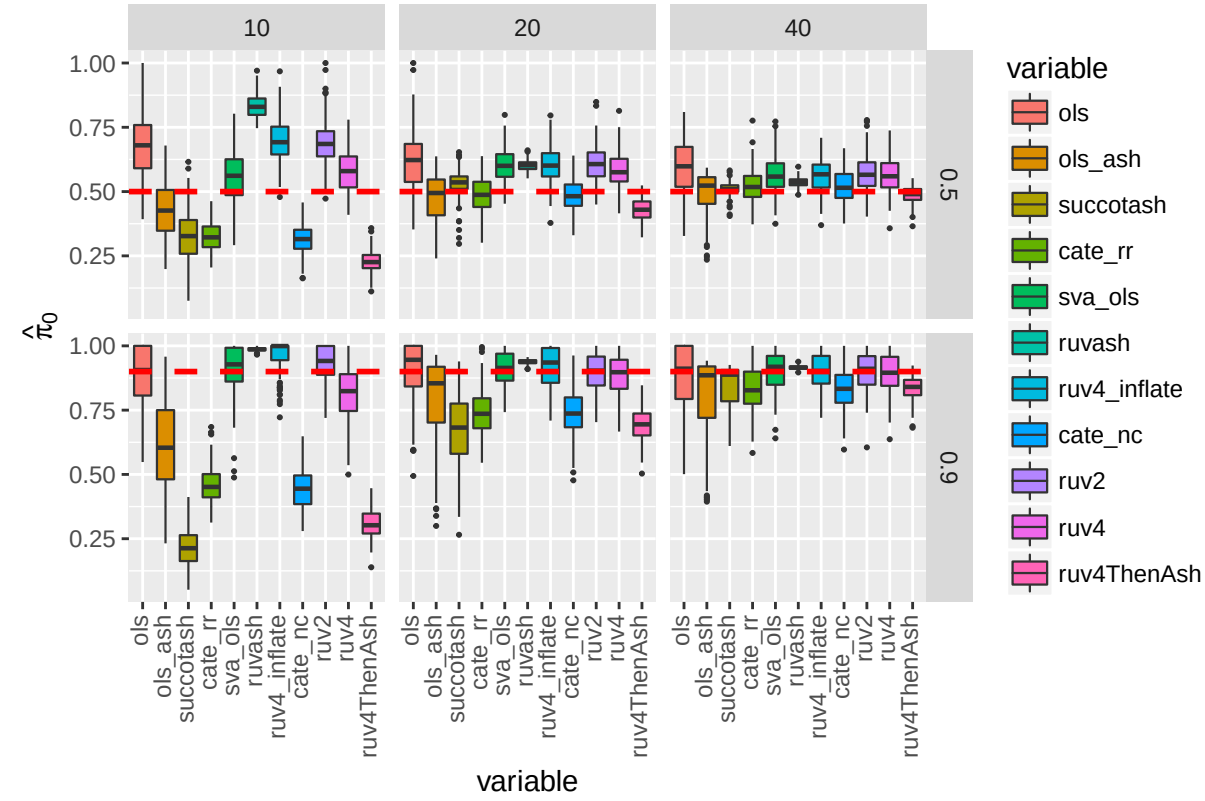
AUC performance.

- RUVASH is the clear winner in all ways except some weird behavior is going on when n = 10 and the alternative is the flattop scenario. This might just be because of using the limma-shrunk variances.
- ruv4_inflate works well except when n = 10 and $\pi_0 = 0.9$. I don't know whey it doesn't work here.

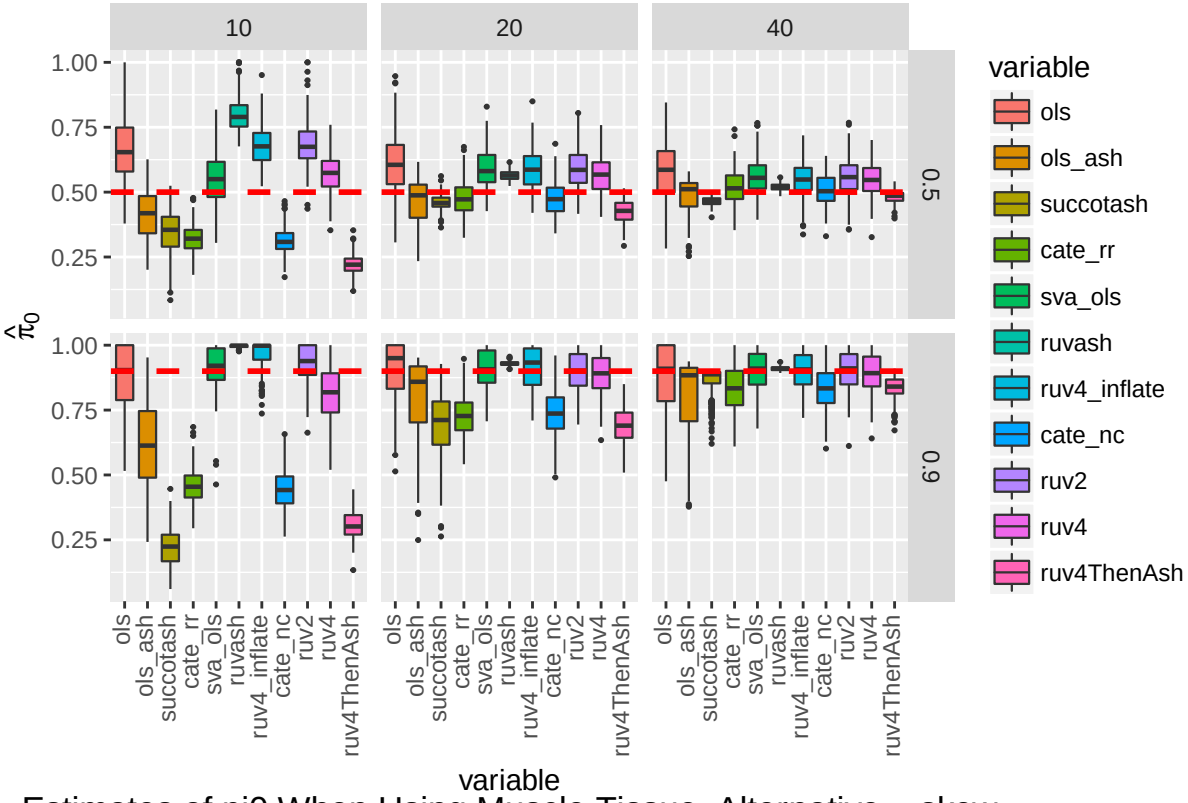
MSE

- All of the ASH methods work well in terms of MSE.

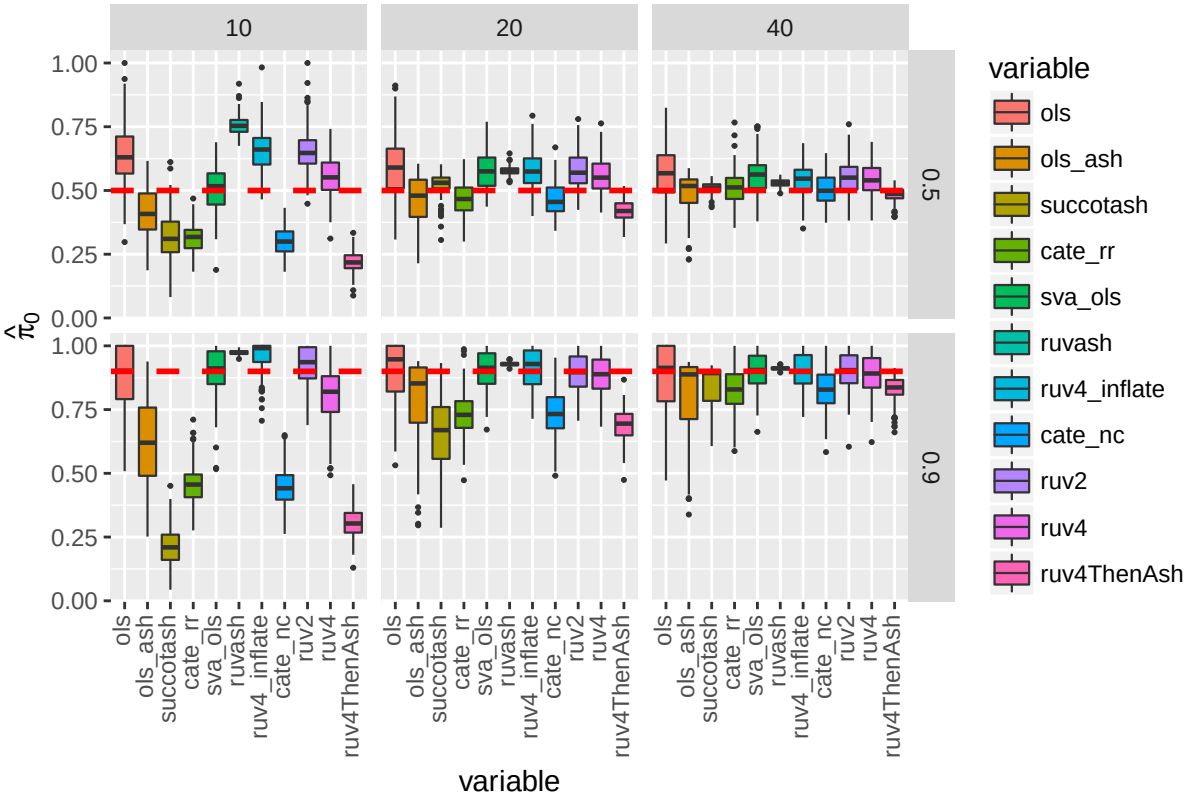

Estimates of pi0 When Using Muscle Tissue, Alternative = spiky
10
20
40
0.5
0.9
1.00
0.75
0.50
0.25
$\hat{\pi}_0$
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh
variable
variable


stimates of pi0 When Using Muscle Tissue, Alternative = near_normal
10
20
40
0.5
0.9
1.00
0.75
0.50
0.25
$\hat{\pi}_0$
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh
variable
variable

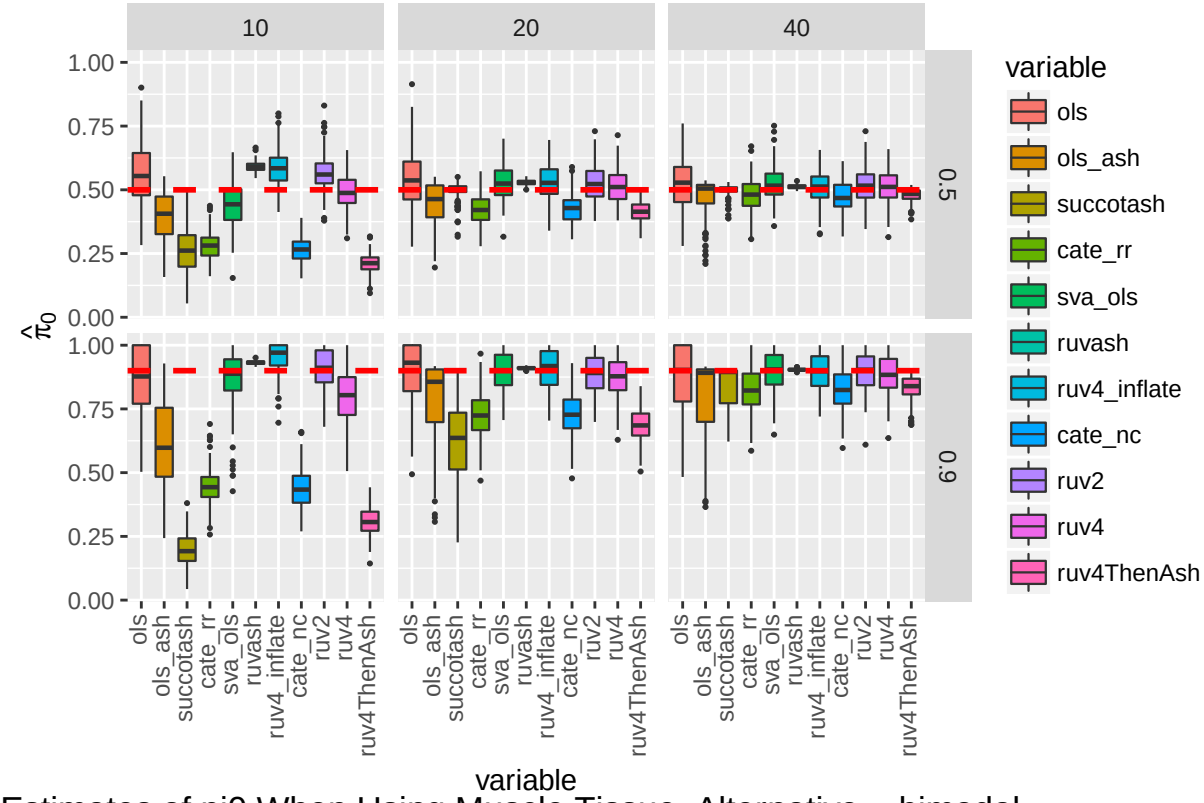
Estimates of pi0 When Using Muscle Tissue, Alternative = flattop



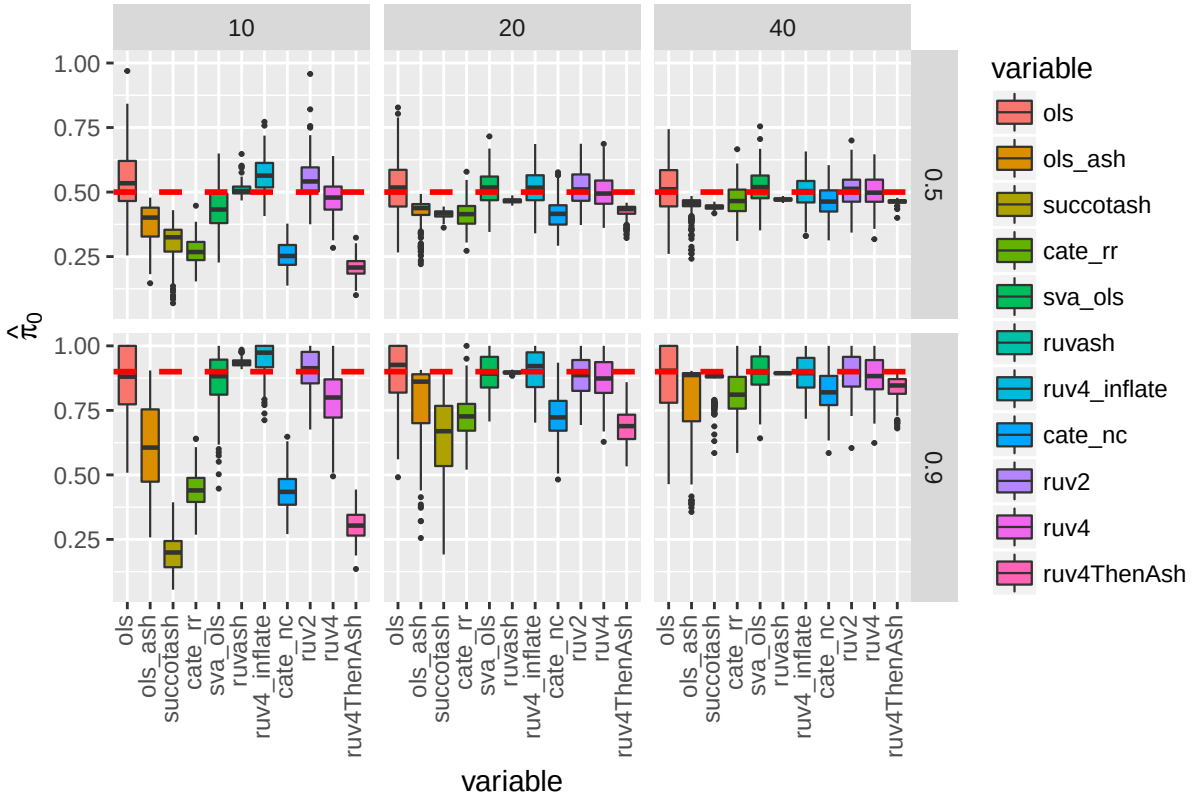
Estimates of pi0 When Using Muscle Tissue, Alternative = skew

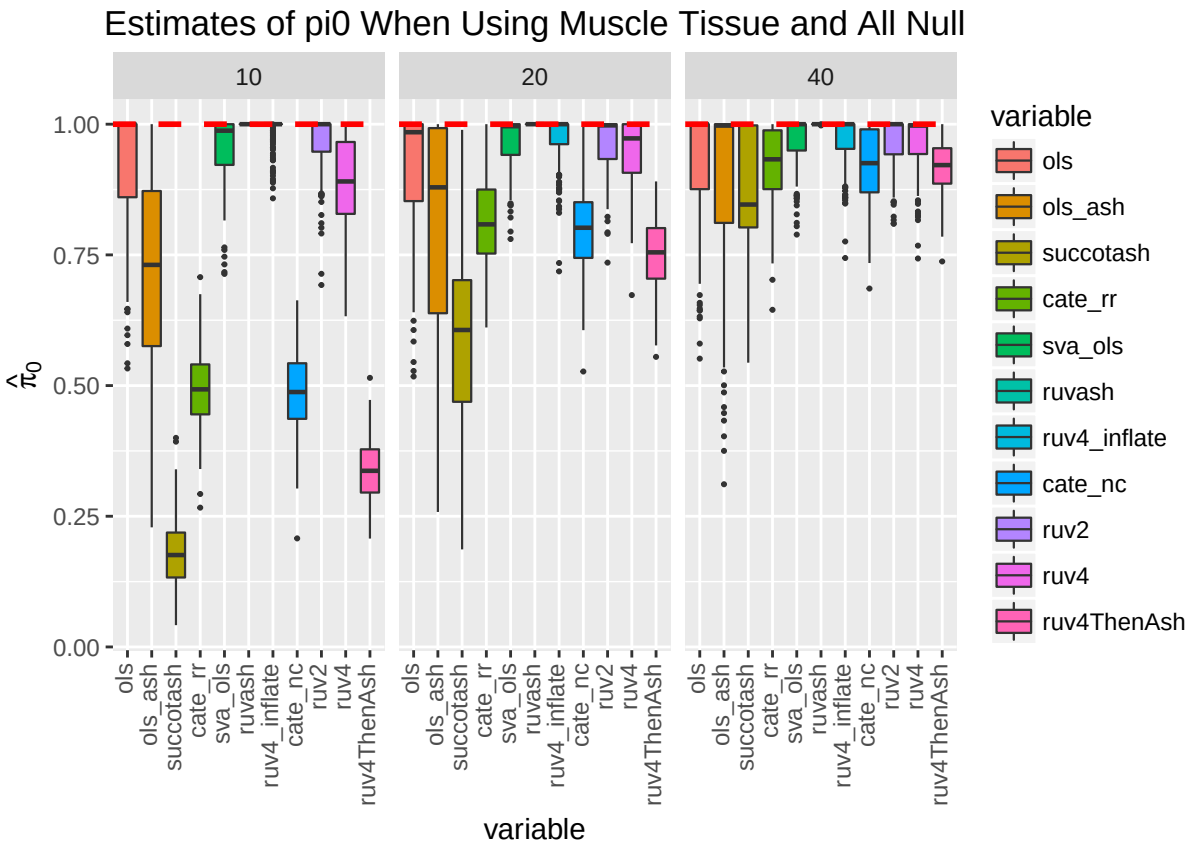


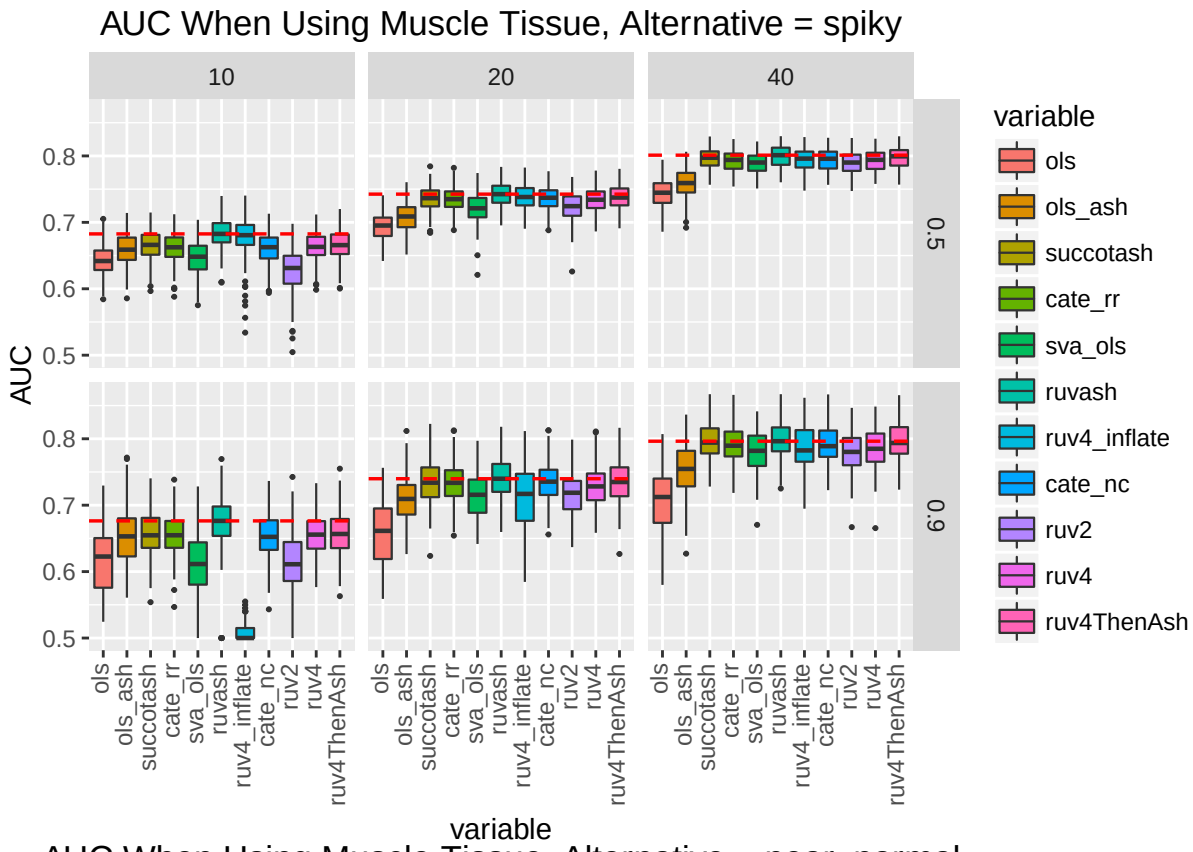
Estimates of pi0 When Using Muscle Tissue, Alternative = big_normal

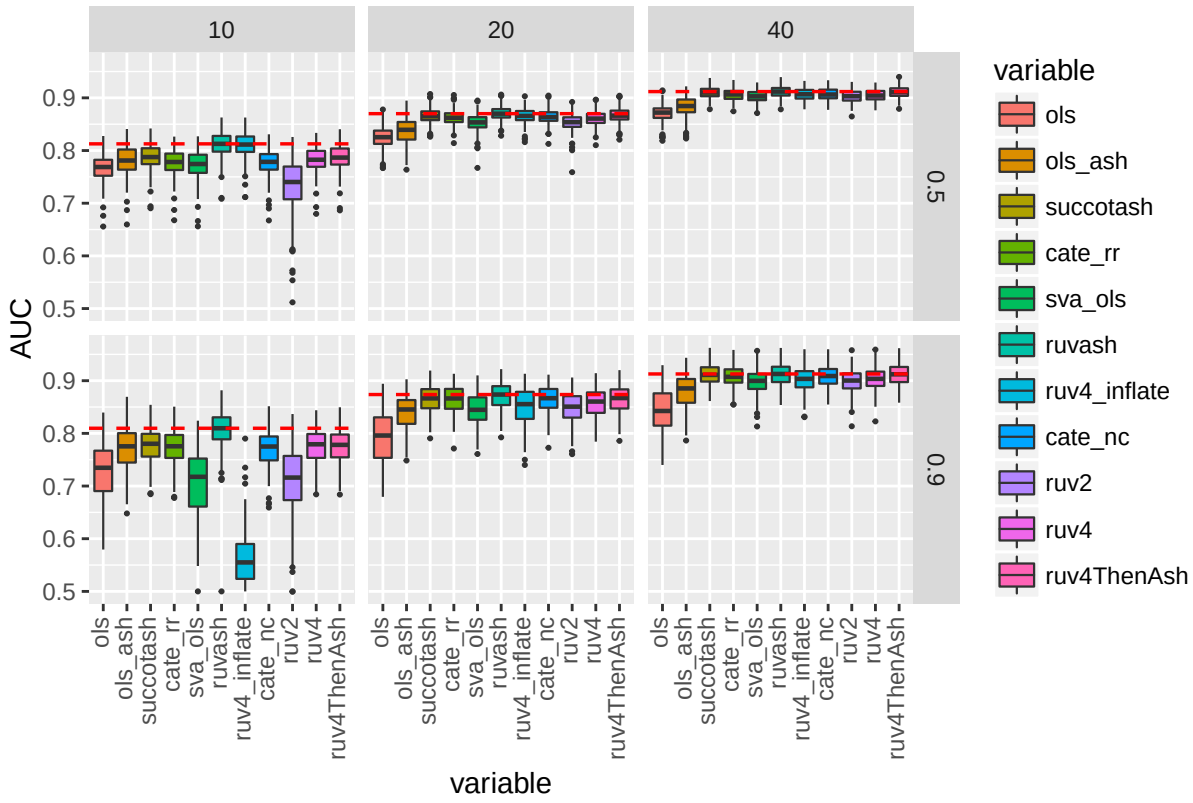


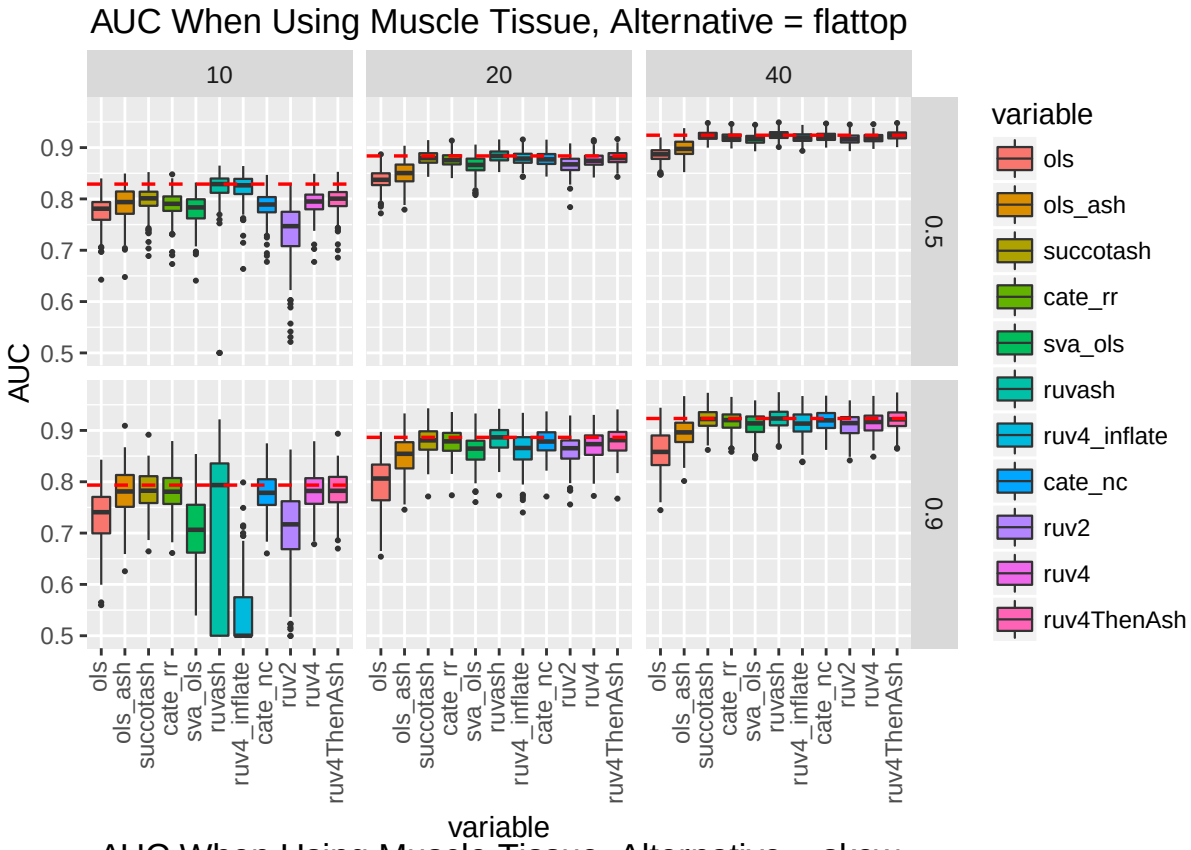
Estimates of pi0 When Using Muscle Tissue, Alternative = bimodal

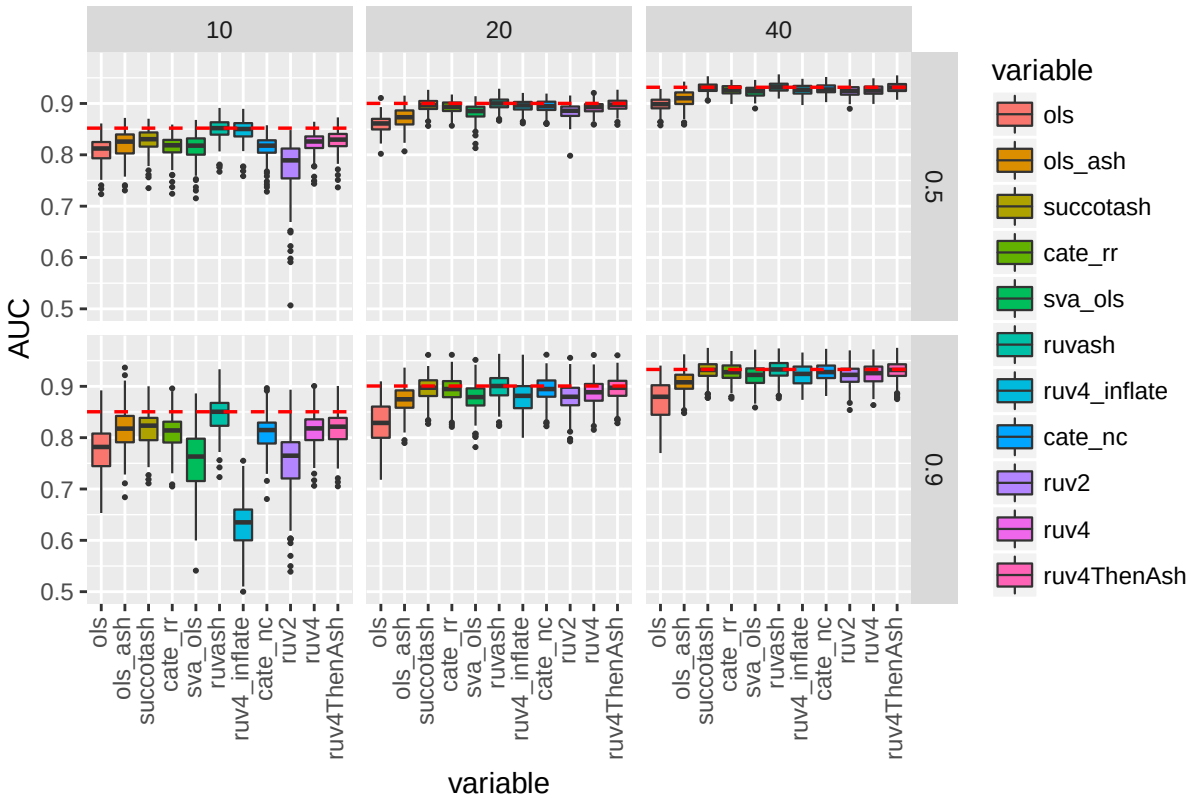


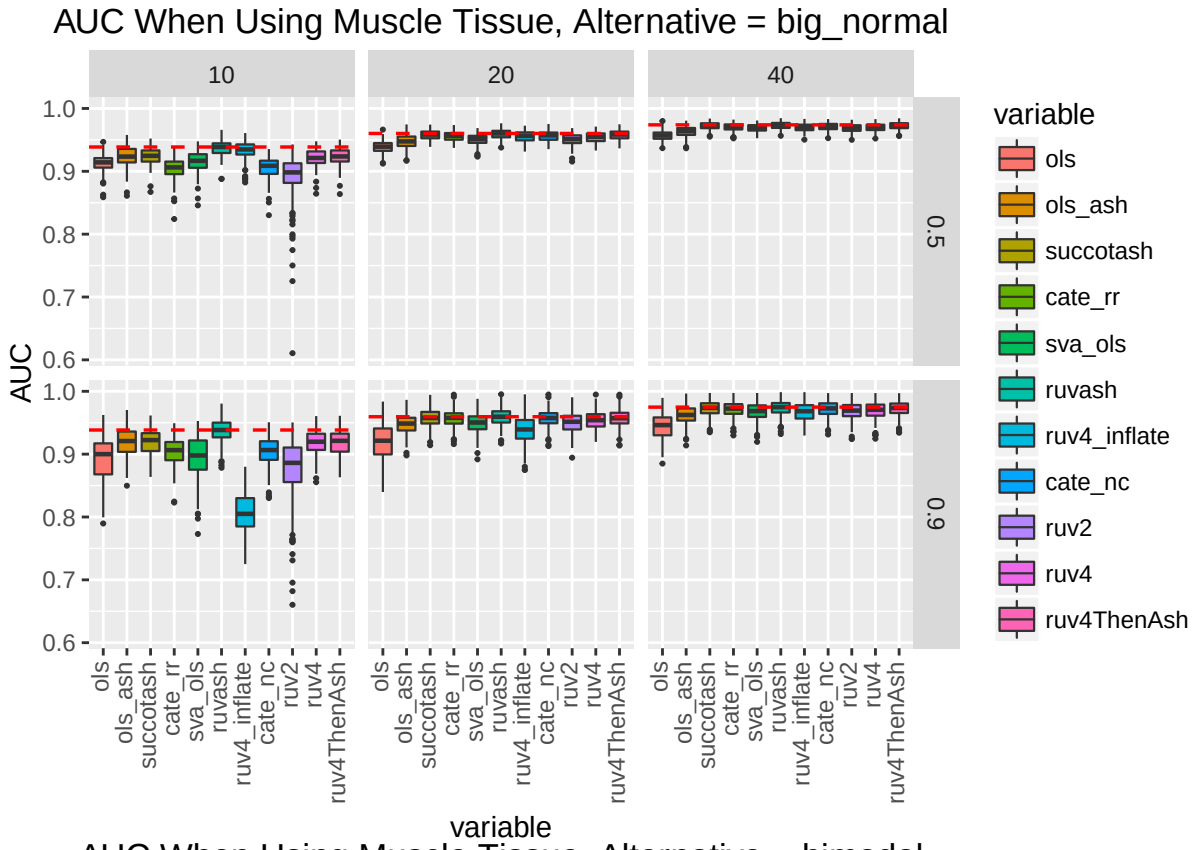

Estimates of pi0 When Using Muscle Tissue and All Null
10
20
40
1.00
0.75
0.50
0.25
0.00
$\hat{\pi}_0$
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh
variable
variable
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh

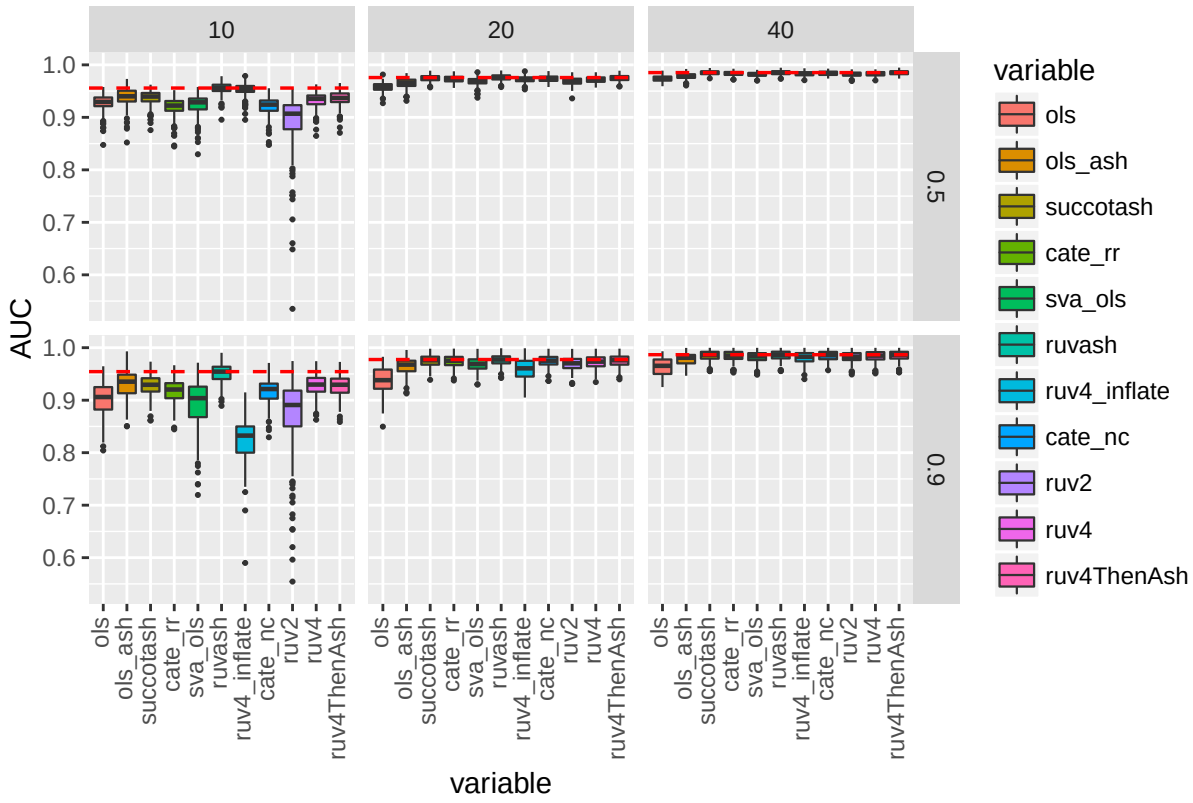

AUC When Using Muscle Tissue, Alternative = spiky
10
20
40
0.5
0.9
AUC
0.8
0.7
0.6
0.5
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh
variable
variable

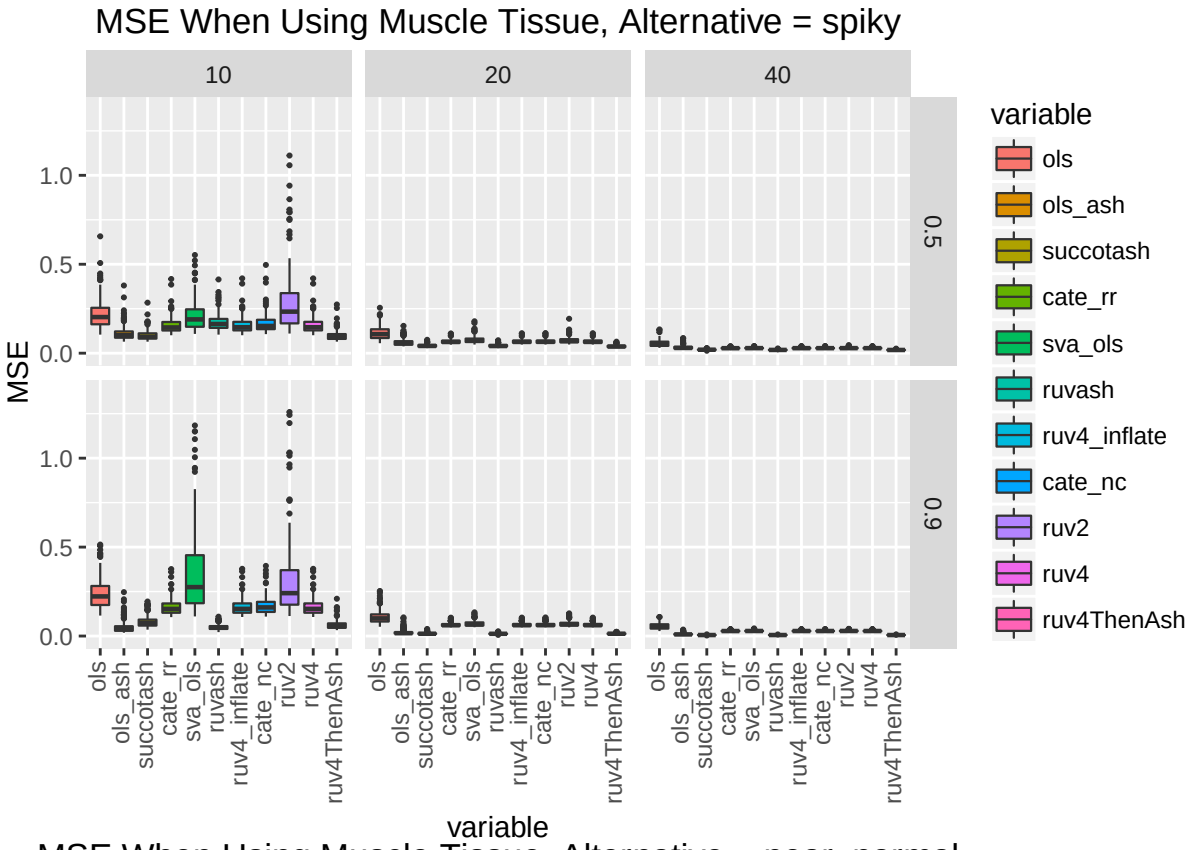

AUC When Using Muscle Tissue, Alternative = near_normal
10
20
40
0.5
0.9
AUC
0.9
0.8
0.7
0.6
0.5
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh
variable
variable

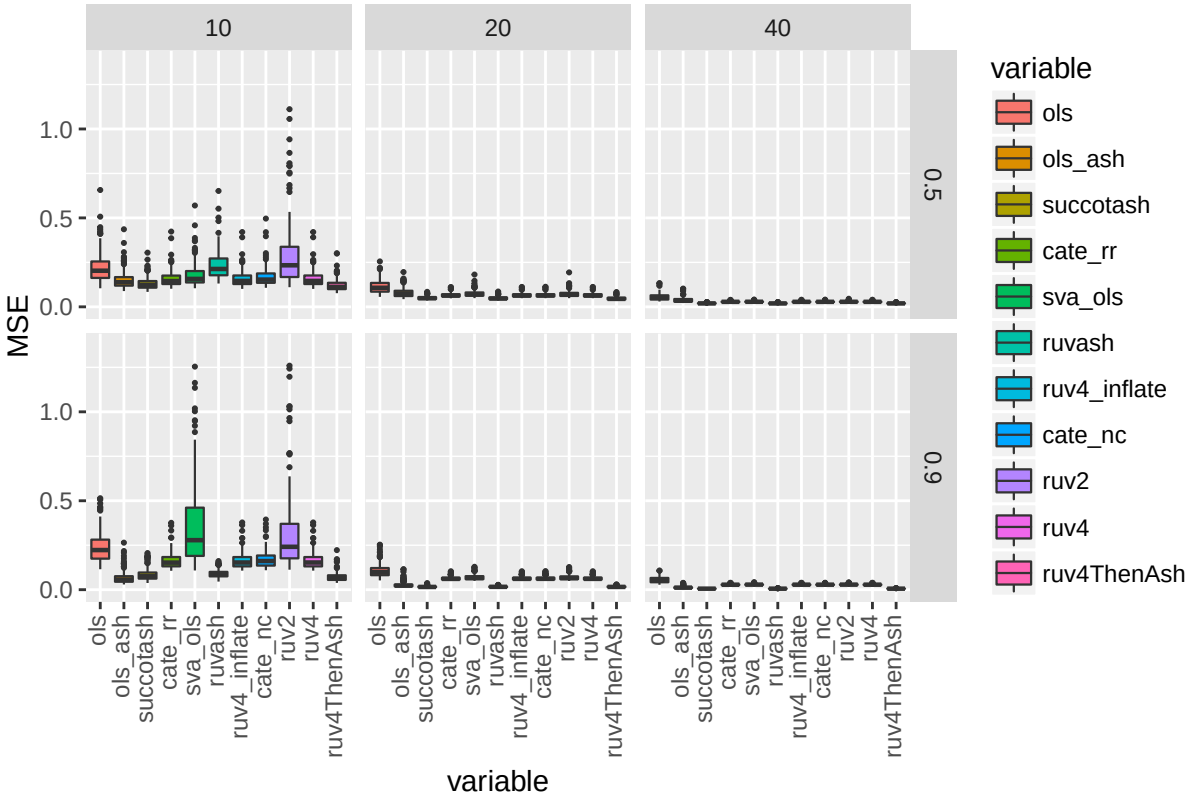

AUC When Using Muscle Tissue, Alternative = flattop
10
20
40
0.5
0.9
AUC
0.9
0.8
0.7
0.6
0.5
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh
variable
variable
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh

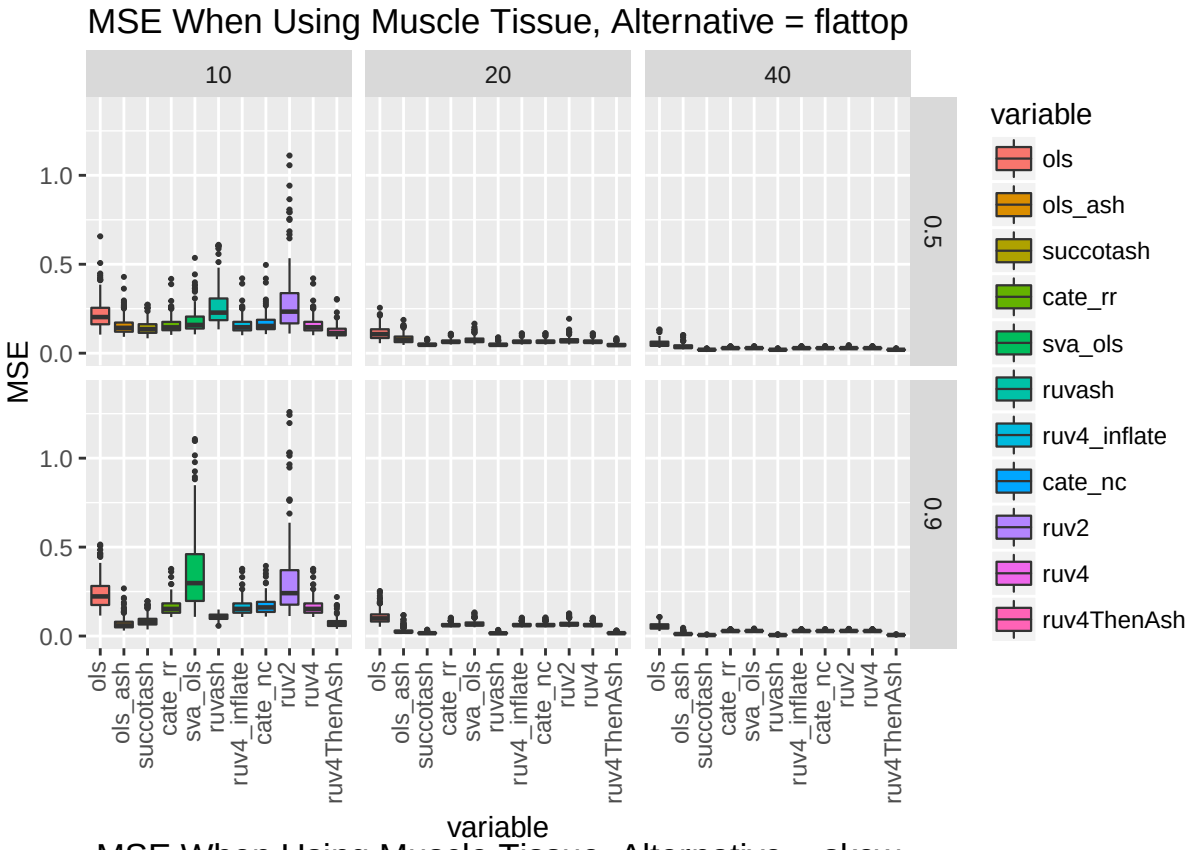

AUC When Using Muscle Tissue, Alternative = skew
10
20
40
0.5
0.9
AUC
0.9
0.8
0.7
0.6
0.5
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh
variable
variable
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh

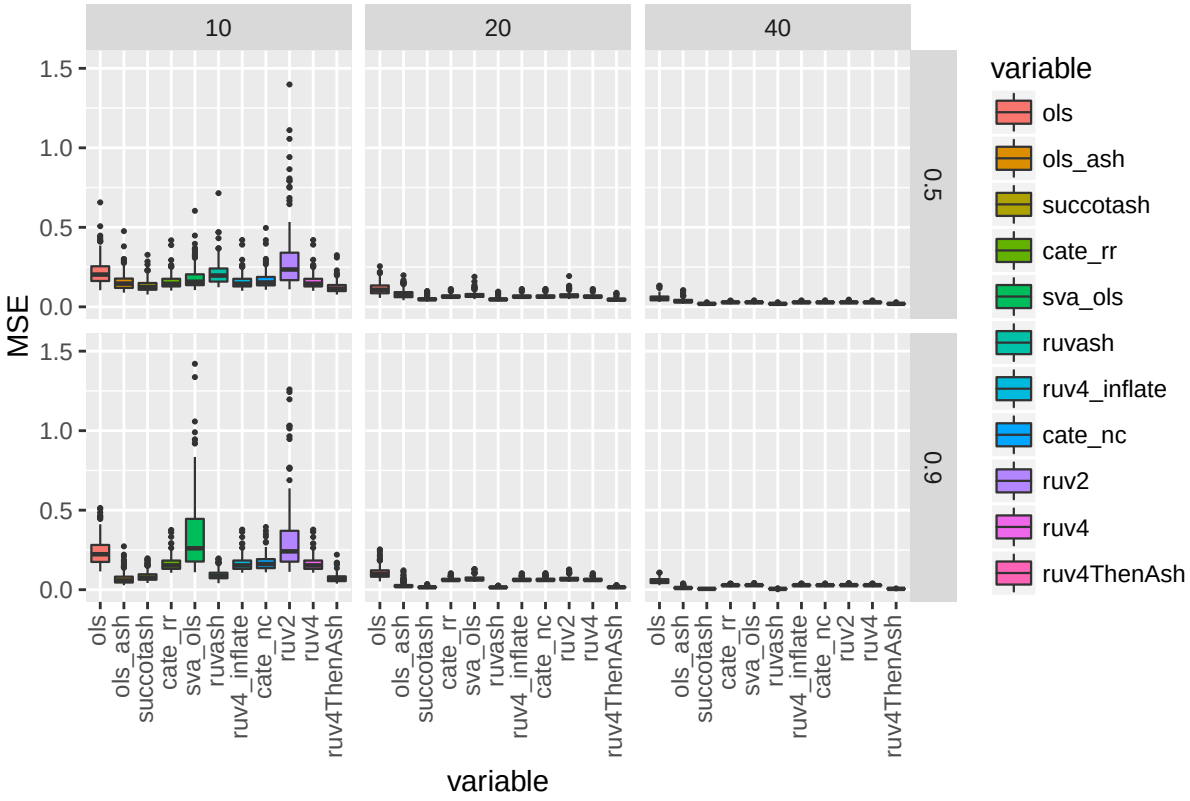

AUC When Using Muscle Tissue, Alternative = big_normal
10
20
40
0.5
0.9
AUC
1.0
0.9
0.8
0.7
0.6
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh
variable
variable
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh

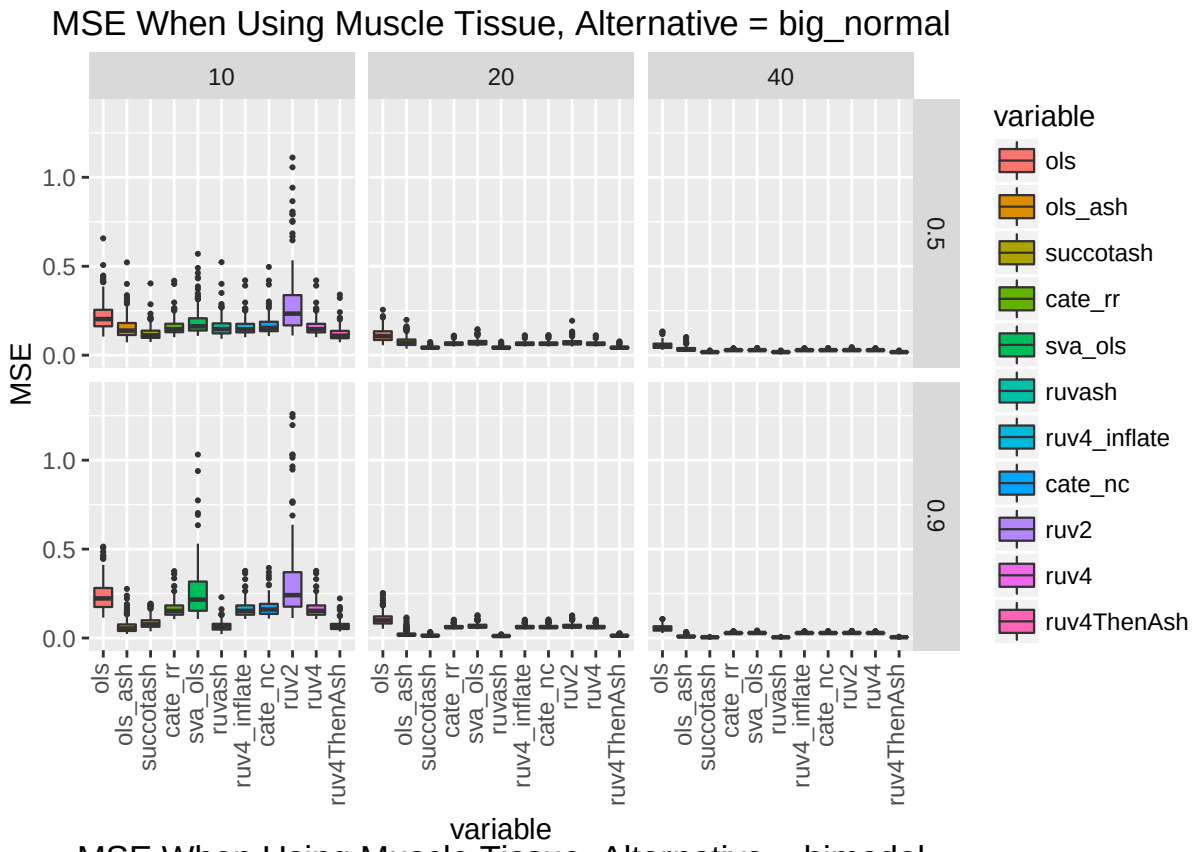

AUC When Using Muscle Tissue, Alternative = bimodal
10
20
40
0.5
0.9
AUC
1.0
0.9
0.8
0.7
0.6
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh
variable
variable
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh

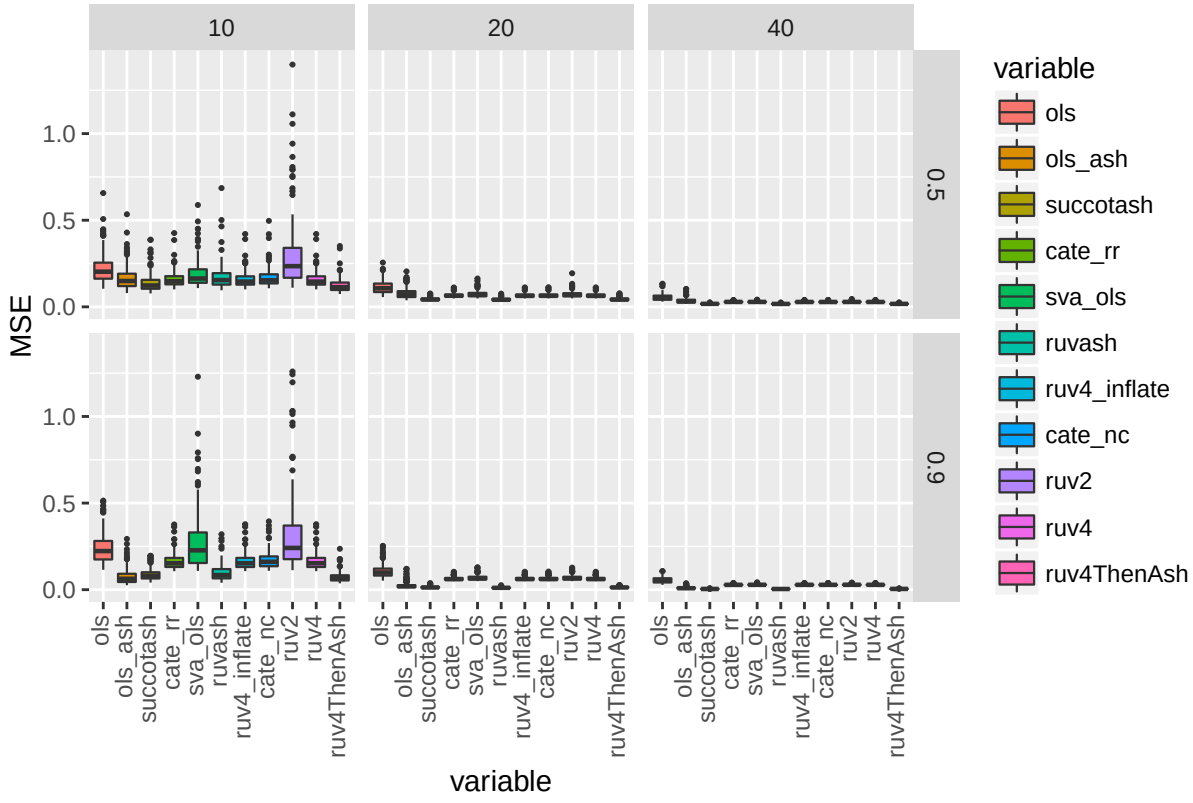

MSE When Using Muscle Tissue, Alternative = spiky
10
20
40
0.5
0.9
MSE
1.0
0.5
0.0
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh
variable
variable
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh

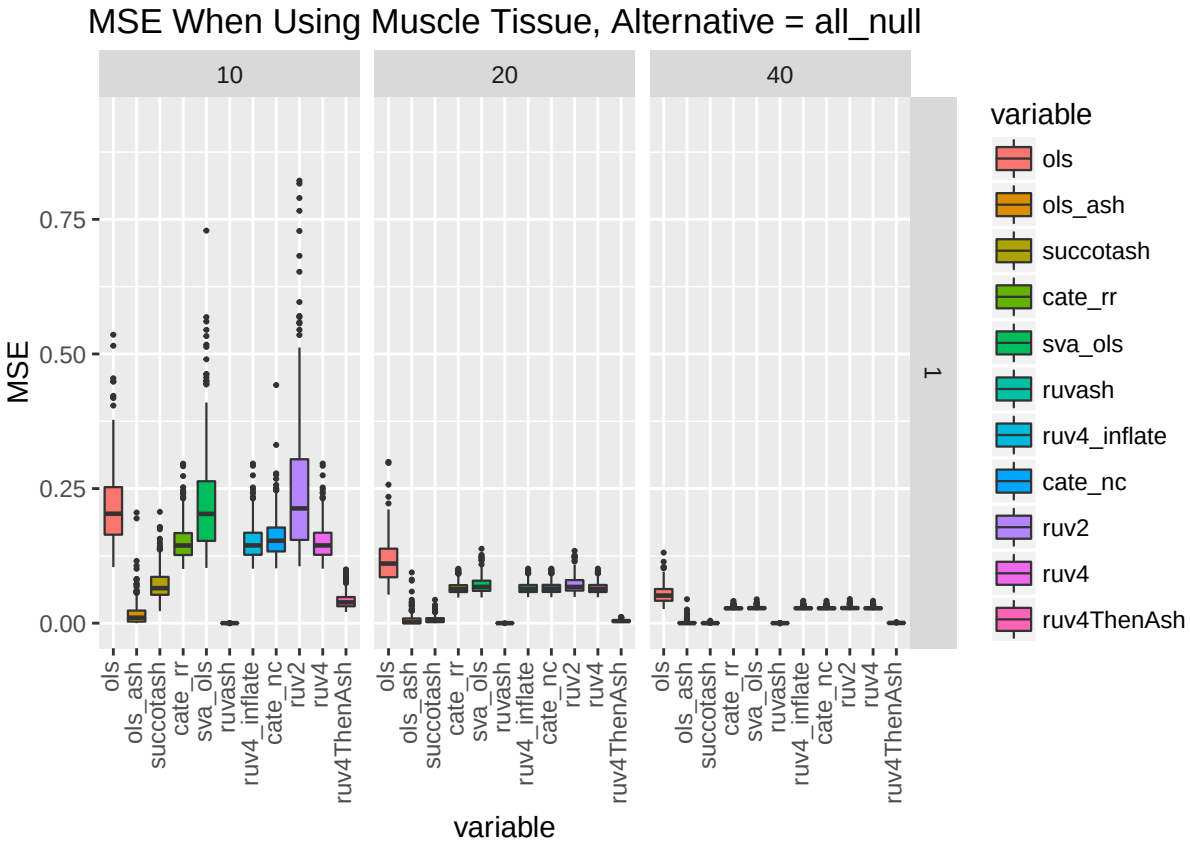

MSE When Using Muscle Tissue, Alternative = near_normal
10
20
40
0.5
0.9
MSE
1.0
0.5
0.0
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh
variable
variable
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh


MSE When Using Muscle Tissue, Alternative = flattop
10
20
40
0.5
0.9
MSE
1.0
0.5
0.0
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh
variable
variable
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh


MSE When Using Muscle Tissue, Alternative = skew
10
20
40
0.5
0.9
MSE
1.5
1.0
0.5
0.0
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh
variable
variable
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh


MSE When Using Muscle Tissue, Alternative = big_normal
10
20
40
0.5
0.9
MSE
1.0
0.5
0.0
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh
variable
variable
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh


MSE When Using Muscle Tissue, Alternative = bimodal
10
20
40
0.5
0.9
MSE
1.0
0.5
0.0
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh
variable
variable
ols
ols_ash
succotash
cate_rr
sva_ols
ruvash
ruv4_inflate
cate_nc
ruv2
ruv4
ruv4ThenAsh




```
sessionInfo()
```

```
## R version 3.3.0 (2016-05-03)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 14.04.4 LTS
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] pROC_1.8       dplyr_0.4.3    reshape2_1.4.1 ggplot2_2.1.0
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_0.12.5      knitr_1.12.28    magrittr_1.5     munsell_0.4.3
##  [5] colorspace_1.2-6 R6_2.1.2         stringr_1.0.0    plyr_1.8.3
##  [9] tools_3.3.0      parallel_3.3.0   grid_3.3.0       gtable_0.2.0
## [13] DBI_0.4          htmltools_0.3.5  yaml_2.1.13      lazyeval_0.1.10
## [17] assertthat_0.1   digest_0.6.9     formatR_1.3      codetools_0.2-14
## [21] evaluate_0.9     rmarkdown_0.9.6  labeling_0.3     stringi_1.0-1
```

```
## [25] compiler_3.3.0   scales_0.4.0
```

Stephens, Matthew. 2016. "False Discovery Rates: A New Deal." *BioRxiv*. Cold Spring Harbor Labs Journals, 038216.