# Variance All the Way Down: Exploring the Impact of RNA-Seq Pipeline Choices on Differential Expression Variance

Hunter Schuler and Art Tay

## Methods

Assume there are $n$ samples of $G$ gene counts. Let $B_{gi}$ denote the count for gene $g$ in sample $i$ reported to the NIH database, and let $C_{giX}$ denote the count obtained from pipeline with choices $X$. Similar let $D_g$ and $E_{gX}$ denote the p-values obtained from `edgeR`. Now,

$$Y_{1X}^2 = \frac{1}{nG} \sum_{i=1}^{n} \sum_{g=1}^{G} (C_{giX} - B_{gi})^2 \tag{1}$$

and

$$Y_{2X}^2 = \frac{1}{G} \sum_{g=1}^{G} (E_{gX} - D_g)^2 \tag{2}$$

Our primary analysis will focus on the two following regression models:

$$Y_{1X} = \beta_0 + \sum_{i=1}^{p} \beta_i X_i + \sum_{1 \leq i < j \leq p} \beta_{ij}(X_i \times X_j) + \epsilon \tag{3}$$

and

$$Y_{2X} = \beta_0 + \sum_{i=1}^{p} \beta_i X_i + \sum_{1 \leq i < j \leq p} \beta_{ij}(X_i \times X_j) + \epsilon \tag{4}$$

where $p$ is the number of pipeline choices from tbl-1. The first model studies the effect of each pipeline choice, include all pairwise interactions, on the average square deviation from the official NIH count matrix. The second model does the same, but for the p-values from a differential expression analysis.

## Code

### EDA

```r
sample_names <- c(
    "gene",
    "SRR31476642",
    "SRR31476643",
    "SRR31476644",
    "SRR31476645",
    "SRR31476646",
    "SRR31476647",
    "SRR31476648",
    "SRR31476649",
    "SRR31476650"
)
```

```r
treatments <- c(
    "DMSO",
    "DMSO",
    "DMSO",
    "DMSO",
    "EPZ015666",
    "EPZ015666",
    "EPZ015666",
    "DMSO",
    "DMSO"
)

factors <- c("aligner", "trim_poly_g", "trim_poly_x", "norm_method")
```

```r
count_sd_df_salmon_kallisto <- read.csv("./data/gen_samples/count_sd_df.csv")
count_sd_df_STAR_HISAT2 <- read.csv("./STAR_HISAT2_combined_results.csv")

count_sd_df <- rbind(count_sd_df_salmon_kallisto, count_sd_df_STAR_HISAT2)
count_sd_df <- count_sd_df |>
    mutate(across(any_of(factors), ~ as.factor(.)))

DE_sd_df_salmon_kallisto <- read.csv("./data/gen_samples/DE_sd_df.csv")
DE_sd_df <- DE_sd_df_salmon_kallisto
DE_sd_df <- DE_sd_df |>
    mutate(across(any_of(factors), ~ as.factor(.)))
```

```r
count_sd_df |>
    group_by(aligner) |>
    summarize(mean_count_sd = mean(count_sd))
```

```
# A tibble: 4 x 2
  aligner   mean_count_sd
  <fct>            <dbl>
1 HISAT2           2986.
2 kallisto         2230.
3 salmon           2312.
4 STAR             3404.
```

```r
count_sd_df |>
    group_by(trim_poly_g) |>
    summarize(mean_count_sd = mean(count_sd))
```

```
# A tibble: 2 x 2
  trim_poly_g mean_count_sd
  <fct>              <dbl>
1 0                  2897.
2 1                  2873.
```

```r
count_sd_df |>
    group_by(trim_poly_x) |>
    summarize(mean_count_sd = mean(count_sd))
```

```
# A tibble: 2 x 2
  trim_poly_x mean_count_sd
  <fct>              <dbl>
```

```
1 0                  2928.
2 1                  2842.
```

```r
count_sd_df |>
    group_by(aligner) |>
    summarize(mean_runtime = mean(runtime_sec))
```

```
# A tibble: 4 x 2
  aligner   mean_runtime
  <fct>           <dbl>
1 HISAT2             NA
2 kallisto          181.
3 salmon           1269.
4 STAR               NA
```

```r
DE_sd_df |>
    group_by(aligner, norm_method, trim_poly_x) |>
    summarize(mean_p_value_sd = mean(p_value_sd))
```

```
# A tibble: 24 x 4
# Groups:    aligner, norm_method [12]
   aligner   norm_method trim_poly_x mean_p_value_sd
   <fct>     <fct>       <fct>                 <dbl>
 1 kallisto ALDEx2       0                     0.374
 2 kallisto ALDEx2       1                     0.374
 3 kallisto none         0                     0.553
 4 kallisto none         1                     0.553
 5 kallisto RLE          0                     0.553
 6 kallisto RLE          1                     0.553
 7 kallisto TMM          0                     0.553
 8 kallisto TMM          1                     0.553
 9 kallisto TMMwsp       0                     0.553
10 kallisto TMMwsp       1                     0.553
# i 14 more rows
```

```r
qqnorm(log(DE_sd_df$p_value_sd), main = "p-value SD Q-Q Plot")
qqline(log(DE_sd_df$p_value_sd))
```

## p–value SD Q–Q Plot
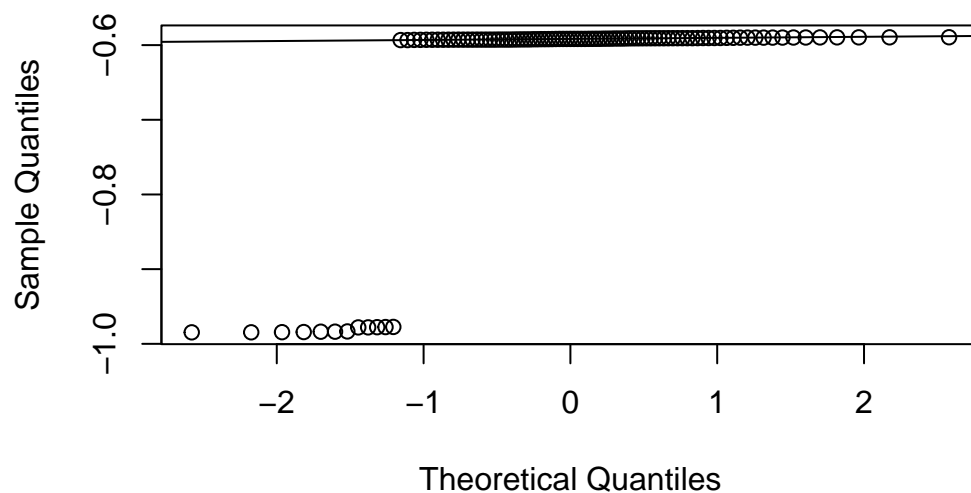


```
qqnorm(DE_sd_df$effect_size_sd^2, main = "Effect Size SD Q-Q Plot")
qqline(DE_sd_df$effect_size_sd^2)
```

## Effect Size SD Q–Q Plot
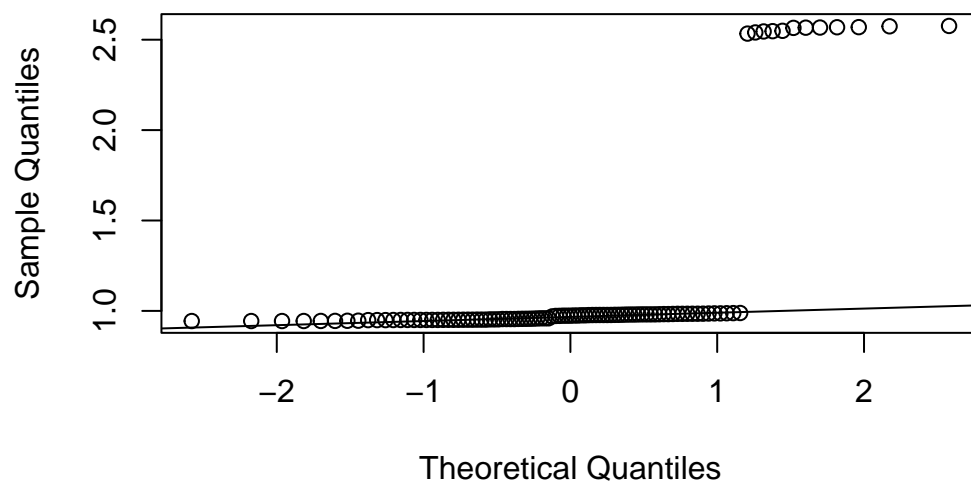


```
count_sd_df |> ggplot(
    aes(x = count_sd)
) + geom_histogram(binwidth = 250)
```

```
DE_sd_df |> ggplot(
    aes(x = p_value_sd)
) + geom_histogram(binwidth = 0.01)
```



## Frequentist

### Counts

```
# Classic LM
lm_fit <- count_sd_df |>
```

```
    select(-c(runtime_sec, gene_overlap_percent)) |>
    (\(x) glm(count_sd ~ (.)^2, family = gaussian(), data = x))()
summary(lm_fit)
```
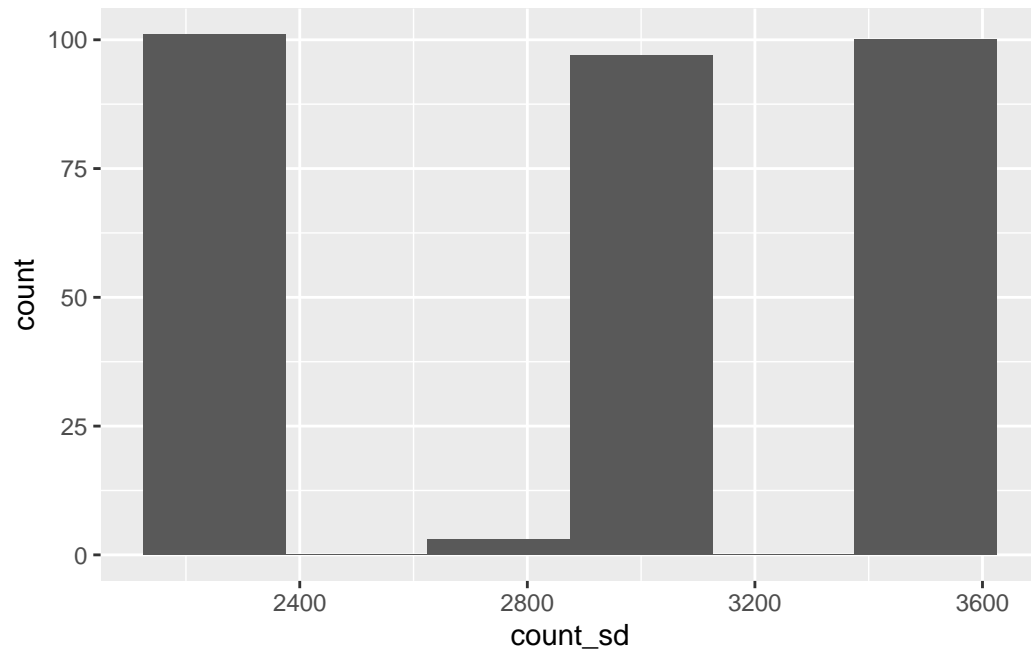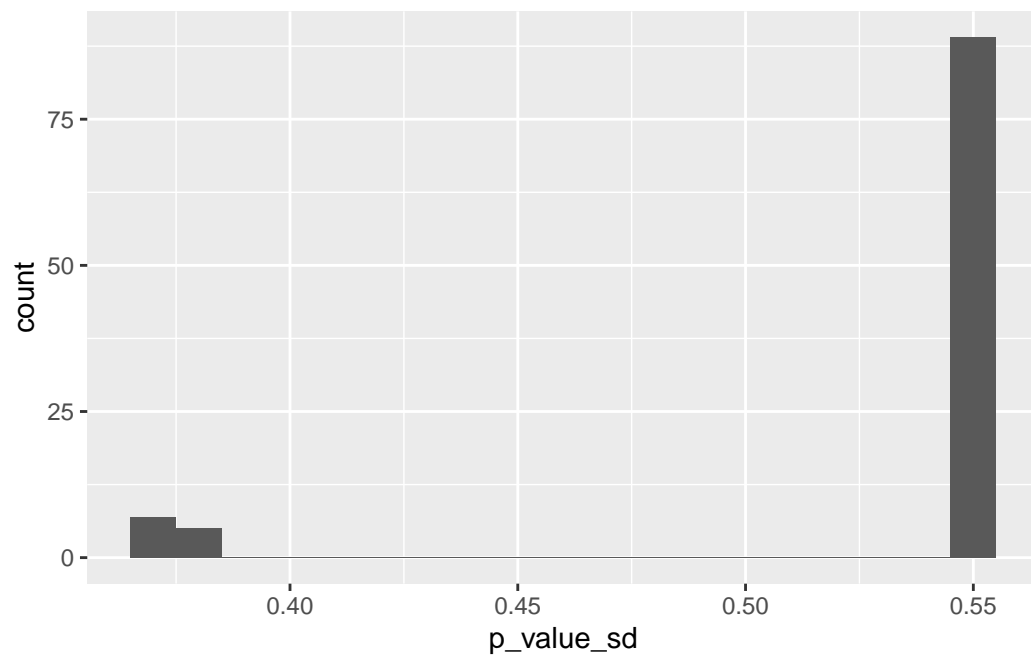
```
Call:
glm(formula = count_sd ~ (.)^2, family = gaussian(), data = x)

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                 2993.46484   56.11471  53.345   <2e-16 ***
alignerkallisto             -742.62396   30.47180 -24.371   <2e-16 ***
alignersalmon               -662.39020   29.62698 -22.358   <2e-16 ***
alignerSTAR                  455.13567   24.28028  18.745   <2e-16 ***
min_phred                     -0.21463    2.14742  -0.100    0.920
min_length                     0.25311    1.32070   0.192    0.848
trim_poly_g1                 -17.25640   20.02630  -0.862    0.390
trim_poly_x1                  23.54365   20.18331   1.166    0.244
alignerkallisto:min_phred     -0.84130    0.89165  -0.944    0.346
alignersalmon:min_phred       -0.74020    0.93781  -0.789    0.431
alignerSTAR:min_phred         -0.86916    0.68790  -1.263    0.207
alignerkallisto:min_length     0.23846    0.48461   0.492    0.623
alignersalmon:min_length       0.17077    0.44391   0.385    0.701
alignerSTAR:min_length        -0.33702    0.38003  -0.887    0.376
alignerkallisto:trim_poly_g1   1.38542    5.62166   0.246    0.806
alignersalmon:trim_poly_g1     4.19567    5.45869   0.769    0.443
alignerSTAR:trim_poly_g1       2.35752    4.34865   0.542    0.588
alignerkallisto:trim_poly_x1  -5.13513    5.39388  -0.952    0.342
alignersalmon:trim_poly_x1    -2.44843    5.65208  -0.433    0.665
alignerSTAR:trim_poly_x1      -4.33325    4.38889  -0.987    0.324
min_phred:min_length          -0.01147    0.05065  -0.226    0.821
min_phred:trim_poly_g1         0.69233    0.58472   1.184    0.237
min_phred:trim_poly_x1         0.10339    0.58526   0.177    0.860
min_length:trim_poly_g1       -0.08697    0.31120  -0.279    0.780
min_length:trim_poly_x1       -0.59818    0.31548  -1.896    0.059 .
trim_poly_g1:trim_poly_x1     -0.07453    3.63010  -0.021    0.984
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 234.6174)

    Null deviance: 66457950  on 300  degrees of freedom
Residual deviance:    64520  on 275  degrees of freedom
AIC: 2523.9

Number of Fisher Scoring iterations: 2
```

```
# Log-normal GLM
glm_log_fit <- count_sd_df |>
    select(-c(runtime_sec, gene_overlap_percent)) |>
    (\(x) glm(count_sd ~ (.)^2, family = gaussian(link = "log"), data = x))()
summary(glm_log_fit)
```

```
Call:
glm(formula = count_sd ~ (.)^2, family = gaussian(link = "log"),
    data = x)

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                8.004e+00  1.911e-02 418.841   <2e-16 ***
alignerkallisto           -2.833e-01  1.259e-02 -22.494   <2e-16 ***
alignersalmon             -2.481e-01  1.188e-02 -20.887   <2e-16 ***
alignerSTAR                1.409e-01  7.645e-03  18.430   <2e-16 ***
min_phred                 -5.907e-05  7.297e-04  -0.081   0.9355
min_length                 9.293e-05  4.500e-04   0.207   0.8366
trim_poly_g1              -5.819e-03  6.828e-03  -0.852   0.3948
trim_poly_x1               8.095e-03  6.866e-03   1.179   0.2394
alignerkallisto:min_phred -3.900e-04  3.688e-04  -1.057   0.2913
alignersalmon:min_phred   -3.549e-04  3.817e-04  -0.930   0.3533
alignerSTAR:min_phred     -2.435e-04  2.167e-04  -1.124   0.2620
alignerkallisto:min_length 4.786e-05  2.008e-04   0.238   0.8117
alignersalmon:min_length   3.077e-05  1.768e-04   0.174   0.8620
alignerSTAR:min_length    -8.188e-05  1.197e-04  -0.684   0.4944
alignerkallisto:trim_poly_g1 3.273e-04 2.328e-03  0.141   0.8883
alignersalmon:trim_poly_g1 1.454e-03  2.199e-03   0.661   0.5090
alignerSTAR:trim_poly_g1   8.419e-04  1.370e-03   0.615   0.5393
alignerkallisto:trim_poly_x1 -1.971e-03 2.224e-03 -0.886  0.3764
alignersalmon:trim_poly_x1 -9.356e-04  2.285e-03  -0.409   0.6826
alignerSTAR:trim_poly_x1   -1.348e-03  1.382e-03  -0.975   0.3304
min_phred:min_length      -4.128e-06  1.724e-05  -0.240   0.8109
min_phred:trim_poly_g1     2.284e-04  1.968e-04   1.160   0.2469
min_phred:trim_poly_x1     3.511e-05  1.978e-04   0.178   0.8592
min_length:trim_poly_g1   -2.545e-05  1.063e-04  -0.239   0.8109
min_length:trim_poly_x1   -2.053e-04  1.081e-04  -1.899   0.0586 .
trim_poly_g1:trim_poly_x1 -7.099e-05  1.235e-03  -0.057   0.9542
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 234.5776)

    Null deviance: 66457950  on 300  degrees of freedom
Residual deviance:    64509  on 275  degrees of freedom
AIC: 2523.8

Number of Fisher Scoring iterations: 3
```

```
# Quasi GLM
quasi_fit <- count_sd_df |>
    select(-c(runtime_sec, gene_overlap_percent)) |>
    (\(x) glm(count_sd ~ (.)^2, family = quasi(), data = x))()
summary(quasi_fit)
```

```
Call:
glm(formula = count_sd ~ (.)^2, family = quasi(), data = x)

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)                        2993.46484    56.11471  53.345    <2e-16 ***
alignerkallisto                    -742.62396     30.47180 -24.371    <2e-16 ***
alignersalmon                      -662.39020     29.62698 -22.358    <2e-16 ***
alignerSTAR                         455.13567     24.28028  18.745    <2e-16 ***
min_phred                            -0.21463      2.14742  -0.100     0.920
min_length                            0.25311      1.32070   0.192     0.848
trim_poly_g1                        -17.25640     20.02630  -0.862     0.390
trim_poly_x1                         23.54365     20.18331   1.166     0.244
alignerkallisto:min_phred            -0.84130      0.89165  -0.944     0.346
alignersalmon:min_phred              -0.74020      0.93781  -0.789     0.431
alignerSTAR:min_phred                -0.86916      0.68790  -1.263     0.207
alignerkallisto:min_length            0.23846      0.48461   0.492     0.623
alignersalmon:min_length              0.17077      0.44391   0.385     0.701
alignerSTAR:min_length               -0.33702      0.38003  -0.887     0.376
alignerkallisto:trim_poly_g1          1.38542      5.62166   0.246     0.806
alignersalmon:trim_poly_g1            4.19567      5.45869   0.769     0.443
alignerSTAR:trim_poly_g1              2.35752      4.34865   0.542     0.588
alignerkallisto:trim_poly_x1         -5.13513      5.39388  -0.952     0.342
alignersalmon:trim_poly_x1           -2.44843      5.65208  -0.433     0.665
alignerSTAR:trim_poly_x1             -4.33325      4.38889  -0.987     0.324
min_phred:min_length                 -0.01147      0.05065  -0.226     0.821
min_phred:trim_poly_g1                0.69233      0.58472   1.184     0.237
min_phred:trim_poly_x1                0.10339      0.58526   0.177     0.860
min_length:trim_poly_g1              -0.08697      0.31120  -0.279     0.780
min_length:trim_poly_x1              -0.59818      0.31548  -1.896     0.059 .
trim_poly_g1:trim_poly_x1            -0.07453      3.63010  -0.021     0.984
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


(Dispersion parameter for quasi family taken to be 234.6174)

    Null deviance: 66457950  on 300  degrees of freedom
Residual deviance:    64520  on 275  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 2
```

**P-Values**

```
# Classic LM
lm_fit <- DE_sd_df |>
    select(-c(runtime_sec, gene_overlap_percent, effect_size_sd)) |>
    (\(x) glm(p_value_sd ~ (.)^2, family = gaussian(), data = x))()
summary(lm_fit)
```

```
Call:
glm(formula = p_value_sd ~ (.)^2, family = gaussian(), data = x)

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                    3.728e-01  5.737e-04 649.784   < 2e-16
alignersalmon                  2.832e-03  2.051e-04  13.804   < 2e-16
min_phred                      2.546e-05  2.089e-05   1.219 0.228202
```

8

```
min_length                                -2.342e-07  1.495e-05   -0.016 0.987561
trim_poly_g1                              -2.496e-04  1.982e-04   -1.259 0.213197
trim_poly_x1                               4.324e-04  2.299e-04    1.881 0.065314
norm_methodnone                            1.806e-01  3.431e-04  526.491   < 2e-16
norm_methodRLE                             1.804e-01  4.391e-04  410.781   < 2e-16
norm_methodTMM                             1.807e-01  4.215e-04  428.799   < 2e-16
norm_methodTMMwsp                          1.804e-01  3.435e-04  525.295   < 2e-16
norm_methodupperquartile                   1.801e-01  3.680e-04  489.437   < 2e-16
alignersalmon:min_phred                   -1.620e-06  6.428e-06   -0.252 0.801897
alignersalmon:min_length                  -4.985e-06  3.110e-06   -1.603 0.114682
alignersalmon:trim_poly_g1                -3.041e-05  3.802e-05   -0.800 0.427282
alignersalmon:trim_poly_x1                -2.208e-04  3.905e-05   -5.654 5.78e-07
alignersalmon:norm_methodnone             -1.190e-03  6.087e-05  -19.542   < 2e-16
alignersalmon:norm_methodRLE              -1.434e-03  7.271e-05  -19.720   < 2e-16
alignersalmon:norm_methodTMM              -1.492e-03  7.149e-05  -20.867   < 2e-16
alignersalmon:norm_methodTMMwsp           -1.544e-03  6.941e-05  -22.238   < 2e-16
alignersalmon:norm_methodupperquartile    -1.685e-03  6.575e-05  -25.623   < 2e-16
min_phred:min_length                       2.280e-07  5.347e-07    0.426 0.671532
min_phred:trim_poly_g1                     1.059e-05  6.173e-06    1.715 0.091940
min_phred:trim_poly_x1                    -2.048e-05  6.812e-06   -3.007 0.003972
min_phred:norm_methodnone                 -4.257e-05  1.206e-05   -3.529 0.000851
min_phred:norm_methodRLE                  -4.931e-05  1.497e-05   -3.294 0.001734
min_phred:norm_methodTMM                  -5.246e-05  1.346e-05   -3.899 0.000265
min_phred:norm_methodTMMwsp               -4.681e-05  1.201e-05   -3.899 0.000265
min_phred:norm_methodupperquartile        -4.395e-05  1.435e-05   -3.063 0.003393
min_length:trim_poly_g1                    -1.495e-06  3.157e-06   -0.474 0.637611
min_length:trim_poly_x1                     3.833e-06  3.662e-06    1.047 0.299701
min_length:norm_methodnone                 -4.954e-06  4.591e-06   -1.079 0.285301
min_length:norm_methodRLE                  -1.246e-06  6.416e-06   -0.194 0.846778
min_length:norm_methodTMM                  -3.915e-06  6.378e-06   -0.614 0.541856
min_length:norm_methodTMMwsp               -1.416e-06  5.439e-06   -0.260 0.795651
min_length:norm_methodupperquartile        -1.731e-06  5.391e-06   -0.321 0.749331
trim_poly_g1:trim_poly_x1                   1.084e-04  3.804e-05    2.851 0.006128
trim_poly_g1:norm_methodnone               -3.855e-05  6.028e-05   -0.639 0.525164
trim_poly_g1:norm_methodRLE                 6.003e-05  8.167e-05    0.735 0.465442
trim_poly_g1:norm_methodTMM                -4.483e-05  7.366e-05   -0.609 0.545314
trim_poly_g1:norm_methodTMMwsp              3.669e-05  6.876e-05    0.534 0.595730
trim_poly_g1:norm_methodupperquartile      -6.271e-06  7.049e-05   -0.089 0.929436
trim_poly_x1:norm_methodnone                2.539e-04  7.195e-05    3.528 0.000853
trim_poly_x1:norm_methodRLE                 2.788e-04  8.077e-05    3.452 0.001078
trim_poly_x1:norm_methodTMM                 2.960e-04  8.771e-05    3.374 0.001362
trim_poly_x1:norm_methodTMMwsp              2.435e-04  7.191e-05    3.386 0.001316
trim_poly_x1:norm_methodupperquartile       2.670e-04  7.899e-05    3.381 0.001337

(Intercept)                               ***
alignersalmon                             ***
min_phred
min_length
trim_poly_g1
trim_poly_x1                              .
norm_methodnone                           ***
norm_methodRLE                            ***
norm_methodTMM                            ***
norm_methodTMMwsp                         ***
```

```
norm_methodupperquartile                    ***
alignersalmon:min_phred
alignersalmon:min_length
alignersalmon:trim_poly_g1
alignersalmon:trim_poly_x1                   ***
alignersalmon:norm_methodnone                ***
alignersalmon:norm_methodRLE                 ***
alignersalmon:norm_methodTMM                 ***
alignersalmon:norm_methodTMMwsp              ***
alignersalmon:norm_methodupperquartile ***
min_phred:min_length
min_phred:trim_poly_g1                         .
min_phred:trim_poly_x1                       **
min_phred:norm_methodnone                    ***
min_phred:norm_methodRLE                     **
min_phred:norm_methodTMM                     ***
min_phred:norm_methodTMMwsp                  ***
min_phred:norm_methodupperquartile          **
min_length:trim_poly_g1
min_length:trim_poly_x1
min_length:norm_methodnone
min_length:norm_methodRLE
min_length:norm_methodTMM
min_length:norm_methodTMMwsp
min_length:norm_methodupperquartile
trim_poly_g1:trim_poly_x1                    **
trim_poly_g1:norm_methodnone
trim_poly_g1:norm_methodRLE
trim_poly_g1:norm_methodTMM
trim_poly_g1:norm_methodTMMwsp
trim_poly_g1:norm_methodupperquartile
trim_poly_x1:norm_methodnone                 ***
trim_poly_x1:norm_methodRLE                  **
trim_poly_x1:norm_methodTMM                  **
trim_poly_x1:norm_methodTMMwsp               **
trim_poly_x1:norm_methodupperquartile        **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


(Dispersion parameter for gaussian family taken to be 5.143455e-09)

    Null deviance: 3.3790e-01  on 100  degrees of freedom
Residual deviance: 2.8289e-07  on  55  degrees of freedom
AIC: -1608.4

Number of Fisher Scoring iterations: 2
```

```
# Log-normal GLM
glm_log_fit <- DE_sd_df |>
    select(-c(runtime_sec, gene_overlap_percent, effect_size_sd)) |>
    (\(x) glm(p_value_sd ~ (.)^2, family = gaussian(link = "log"), data = x))()
summary(glm_log_fit)
```

```
Call:
```

```
glm(formula = p_value_sd ~ (.)^2, family = gaussian(link = "log"),
    data = x)
```

Coefficients:

|  | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -9.867e-01 | 1.193e-03 | -826.915 | < 2e-16 |
| alignersalmon | 7.156e-03 | 4.128e-04 | 17.336 | < 2e-16 |
| min_phred | 7.427e-05 | 4.395e-05 | 1.690 | 0.096701 |
| min_length | 4.462e-06 | 2.985e-05 | 0.149 | 0.881718 |
| trim_poly_g1 | -3.872e-04 | 3.897e-04 | -0.993 | 0.324854 |
| trim_poly_x1 | 7.470e-04 | 4.377e-04 | 1.707 | 0.093481 |
| norm_methodnone | 3.950e-01 | 7.848e-04 | 503.292 | < 2e-16 |
| norm_methodRLE | 3.945e-01 | 9.342e-04 | 422.271 | < 2e-16 |
| norm_methodTMM | 3.951e-01 | 9.039e-04 | 437.126 | < 2e-16 |
| norm_methodTMMwsp | 3.946e-01 | 7.855e-04 | 502.304 | < 2e-16 |
| norm_methodupperquartile | 3.941e-01 | 8.227e-04 | 479.029 | < 2e-16 |
| alignersalmon:min_phred | 4.339e-08 | 1.235e-05 | 0.004 | 0.997210 |
| alignersalmon:min_length | -9.052e-06 | 5.906e-06 | -1.533 | 0.131098 |
| alignersalmon:trim_poly_g1 | -3.813e-05 | 7.201e-05 | -0.530 | 0.598555 |
| alignersalmon:trim_poly_x1 | -4.204e-04 | 7.335e-05 | -5.731 | 4.35e-07 |
| alignersalmon:norm_methodnone | -4.257e-03 | 1.498e-04 | -28.429 | < 2e-16 |
| alignersalmon:norm_methodRLE | -4.704e-03 | 1.665e-04 | -28.249 | < 2e-16 |
| alignersalmon:norm_methodTMM | -4.811e-03 | 1.649e-04 | -29.185 | < 2e-16 |
| alignersalmon:norm_methodTMMwsp | -4.904e-03 | 1.620e-04 | -30.272 | < 2e-16 |
| alignersalmon:norm_methodupperquartile | -5.142e-03 | 1.569e-04 | -32.778 | < 2e-16 |
| min_phred:min_length | 2.982e-07 | 1.040e-06 | 0.287 | 0.775353 |
| min_phred:trim_poly_g1 | 1.715e-05 | 1.169e-05 | 1.467 | 0.148062 |
| min_phred:trim_poly_x1 | -3.617e-05 | 1.251e-05 | -2.892 | 0.005480 |
| min_phred:norm_methodnone | -1.007e-04 | 2.864e-05 | -3.518 | 0.000881 |
| min_phred:norm_methodRLE | -1.116e-04 | 3.296e-05 | -3.388 | 0.001309 |
| min_phred:norm_methodTMM | -1.185e-04 | 3.066e-05 | -3.866 | 0.000295 |
| min_phred:norm_methodTMMwsp | -1.082e-04 | 2.860e-05 | -3.785 | 0.000382 |
| min_phred:norm_methodupperquartile | -1.048e-04 | 3.203e-05 | -3.273 | 0.001844 |
| min_length:trim_poly_g1 | -2.797e-06 | 6.044e-06 | -0.463 | 0.645370 |
| min_length:trim_poly_x1 | 7.540e-06 | 6.726e-06 | 1.121 | 0.267192 |
| min_length:norm_methodnone | -1.109e-05 | 1.107e-05 | -1.001 | 0.320977 |
| min_length:norm_methodRLE | -4.645e-06 | 1.386e-05 | -0.335 | 0.738841 |
| min_length:norm_methodTMM | -9.089e-06 | 1.374e-05 | -0.661 | 0.511157 |
| min_length:norm_methodTMMwsp | -4.437e-06 | 1.237e-05 | -0.359 | 0.721164 |
| min_length:norm_methodupperquartile | -5.708e-06 | 1.227e-05 | -0.465 | 0.643584 |
| trim_poly_g1:trim_poly_x1 | 2.016e-04 | 7.187e-05 | 2.805 | 0.006942 |
| trim_poly_g1:norm_methodnone | -9.327e-05 | 1.482e-04 | -0.629 | 0.531782 |
| trim_poly_g1:norm_methodRLE | 9.026e-05 | 1.799e-04 | 0.502 | 0.617935 |
| trim_poly_g1:norm_methodTMM | -1.078e-04 | 1.673e-04 | -0.644 | 0.522067 |
| trim_poly_g1:norm_methodTMMwsp | 4.047e-05 | 1.601e-04 | 0.253 | 0.801403 |
| trim_poly_g1:norm_methodupperquartile | -3.709e-05 | 1.626e-04 | -0.228 | 0.820448 |
| trim_poly_x1:norm_methodnone | 4.523e-04 | 1.771e-04 | 2.554 | 0.013444 |
| trim_poly_x1:norm_methodRLE | 4.961e-04 | 1.892e-04 | 2.622 | 0.011287 |
| trim_poly_x1:norm_methodTMM | 5.301e-04 | 1.997e-04 | 2.655 | 0.010351 |
| trim_poly_x1:norm_methodTMMwsp | 4.366e-04 | 1.770e-04 | 2.467 | 0.016762 |
| trim_poly_x1:norm_methodupperquartile | 4.805e-04 | 1.869e-04 | 2.571 | 0.012883 |

| | |
|---|---|
| (Intercept) | *** |
| alignersalmon | *** |

```
min_phred                                   .
min_length
trim_poly_g1
trim_poly_x1                                .
norm_methodnone                         ***
norm_methodRLE                          ***
norm_methodTMM                          ***
norm_methodTMMwsp                       ***
norm_methodupperquartile                ***
alignersalmon:min_phred
alignersalmon:min_length
alignersalmon:trim_poly_g1
alignersalmon:trim_poly_x1              ***
alignersalmon:norm_methodnone           ***
alignersalmon:norm_methodRLE            ***
alignersalmon:norm_methodTMM            ***
alignersalmon:norm_methodTMMwsp         ***
alignersalmon:norm_methodupperquartile ***
min_phred:min_length
min_phred:trim_poly_g1
min_phred:trim_poly_x1                  **
min_phred:norm_methodnone               ***
min_phred:norm_methodRLE                **
min_phred:norm_methodTMM                ***
min_phred:norm_methodTMMwsp             ***
min_phred:norm_methodupperquartile      **
min_length:trim_poly_g1
min_length:trim_poly_x1
min_length:norm_methodnone
min_length:norm_methodRLE
min_length:norm_methodTMM
min_length:norm_methodTMMwsp
min_length:norm_methodupperquartile
trim_poly_g1:trim_poly_x1               **
trim_poly_g1:norm_methodnone
trim_poly_g1:norm_methodRLE
trim_poly_g1:norm_methodTMM
trim_poly_g1:norm_methodTMMwsp
trim_poly_g1:norm_methodupperquartile
trim_poly_x1:norm_methodnone            *
trim_poly_x1:norm_methodRLE             *
trim_poly_x1:norm_methodTMM             *
trim_poly_x1:norm_methodTMMwsp          *
trim_poly_x1:norm_methodupperquartile   *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 5.261523e-09)

    Null deviance: 3.3790e-01  on 100  degrees of freedom
Residual deviance: 2.8938e-07  on  55  degrees of freedom
AIC: -1606.1

Number of Fisher Scoring iterations: 2
```

```
# Quasi GLM
quasi_fit <- DE_sd_df |>
    select(-c(runtime_sec, gene_overlap_percent, effect_size_sd)) |>
    (\(x) glm(p_value_sd ~ (.)^2, family = quasi(), data = x))()
summary(quasi_fit)
```

Call:
glm(formula = p_value_sd ~ (.)^2, family = quasi(), data = x)

Coefficients:

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 3.728e-01 | 5.737e-04 | 649.784 | < 2e-16 |
| alignersalmon | 2.832e-03 | 2.051e-04 | 13.804 | < 2e-16 |
| min_phred | 2.546e-05 | 2.089e-05 | 1.219 | 0.228202 |
| min_length | -2.342e-07 | 1.495e-05 | -0.016 | 0.987561 |
| trim_poly_g1 | -2.496e-04 | 1.982e-04 | -1.259 | 0.213197 |
| trim_poly_x1 | 4.324e-04 | 2.299e-04 | 1.881 | 0.065314 |
| norm_methodnone | 1.806e-01 | 3.431e-04 | 526.491 | < 2e-16 |
| norm_methodRLE | 1.804e-01 | 4.391e-04 | 410.781 | < 2e-16 |
| norm_methodTMM | 1.807e-01 | 4.215e-04 | 428.799 | < 2e-16 |
| norm_methodTMMwsp | 1.804e-01 | 3.435e-04 | 525.295 | < 2e-16 |
| norm_methodupperquartile | 1.801e-01 | 3.680e-04 | 489.437 | < 2e-16 |
| alignersalmon:min_phred | -1.620e-06 | 6.428e-06 | -0.252 | 0.801897 |
| alignersalmon:min_length | -4.985e-06 | 3.110e-06 | -1.603 | 0.114682 |
| alignersalmon:trim_poly_g1 | -3.041e-05 | 3.802e-05 | -0.800 | 0.427282 |
| alignersalmon:trim_poly_x1 | -2.208e-04 | 3.905e-05 | -5.654 | 5.78e-07 |
| alignersalmon:norm_methodnone | -1.190e-03 | 6.087e-05 | -19.542 | < 2e-16 |
| alignersalmon:norm_methodRLE | -1.434e-03 | 7.271e-05 | -19.720 | < 2e-16 |
| alignersalmon:norm_methodTMM | -1.492e-03 | 7.149e-05 | -20.867 | < 2e-16 |
| alignersalmon:norm_methodTMMwsp | -1.544e-03 | 6.941e-05 | -22.238 | < 2e-16 |
| alignersalmon:norm_methodupperquartile | -1.685e-03 | 6.575e-05 | -25.623 | < 2e-16 |
| min_phred:min_length | 2.280e-07 | 5.347e-07 | 0.426 | 0.671532 |
| min_phred:trim_poly_g1 | 1.059e-05 | 6.173e-06 | 1.715 | 0.091940 |
| min_phred:trim_poly_x1 | -2.048e-05 | 6.812e-06 | -3.007 | 0.003972 |
| min_phred:norm_methodnone | -4.257e-05 | 1.206e-05 | -3.529 | 0.000851 |
| min_phred:norm_methodRLE | -4.931e-05 | 1.497e-05 | -3.294 | 0.001734 |
| min_phred:norm_methodTMM | -5.246e-05 | 1.346e-05 | -3.899 | 0.000265 |
| min_phred:norm_methodTMMwsp | -4.681e-05 | 1.201e-05 | -3.899 | 0.000265 |
| min_phred:norm_methodupperquartile | -4.395e-05 | 1.435e-05 | -3.063 | 0.003393 |
| min_length:trim_poly_g1 | -1.495e-06 | 3.157e-06 | -0.474 | 0.637611 |
| min_length:trim_poly_x1 | 3.833e-06 | 3.662e-06 | 1.047 | 0.299701 |
| min_length:norm_methodnone | -4.954e-06 | 4.591e-06 | -1.079 | 0.285301 |
| min_length:norm_methodRLE | -1.246e-06 | 6.416e-06 | -0.194 | 0.846778 |
| min_length:norm_methodTMM | -3.915e-06 | 6.378e-06 | -0.614 | 0.541856 |
| min_length:norm_methodTMMwsp | -1.416e-06 | 5.439e-06 | -0.260 | 0.795651 |
| min_length:norm_methodupperquartile | -1.731e-06 | 5.391e-06 | -0.321 | 0.749331 |
| trim_poly_g1:trim_poly_x1 | 1.084e-04 | 3.804e-05 | 2.851 | 0.006128 |
| trim_poly_g1:norm_methodnone | -3.855e-05 | 6.028e-05 | -0.639 | 0.525164 |
| trim_poly_g1:norm_methodRLE | 6.003e-05 | 8.167e-05 | 0.735 | 0.465442 |
| trim_poly_g1:norm_methodTMM | -4.483e-05 | 7.366e-05 | -0.609 | 0.545314 |
| trim_poly_g1:norm_methodTMMwsp | 3.669e-05 | 6.876e-05 | 0.534 | 0.595730 |
| trim_poly_g1:norm_methodupperquartile | -6.271e-06 | 7.049e-05 | -0.089 | 0.929436 |
| trim_poly_x1:norm_methodnone | 2.539e-04 | 7.195e-05 | 3.528 | 0.000853 |

```
trim_poly_x1:norm_methodRLE              2.788e-04  8.077e-05   3.452 0.001078
trim_poly_x1:norm_methodTMM              2.960e-04  8.771e-05   3.374 0.001362
trim_poly_x1:norm_methodTMMwsp           2.435e-04  7.191e-05   3.386 0.001316
trim_poly_x1:norm_methodupperquartile    2.670e-04  7.899e-05   3.381 0.001337

(Intercept)                              ***
alignersalmon                            ***
min_phred
min_length
trim_poly_g1
trim_poly_x1                               .
norm_methodnone                          ***
norm_methodRLE                           ***
norm_methodTMM                           ***
norm_methodTMMwsp                        ***
norm_methodupperquartile                 ***
alignersalmon:min_phred
alignersalmon:min_length
alignersalmon:trim_poly_g1
alignersalmon:trim_poly_x1               ***
alignersalmon:norm_methodnone            ***
alignersalmon:norm_methodRLE             ***
alignersalmon:norm_methodTMM             ***
alignersalmon:norm_methodTMMwsp          ***
alignersalmon:norm_methodupperquartile   ***
min_phred:min_length
min_phred:trim_poly_g1                      .
min_phred:trim_poly_x1                    **
min_phred:norm_methodnone                ***
min_phred:norm_methodRLE                  **
min_phred:norm_methodTMM                 ***
min_phred:norm_methodTMMwsp              ***
min_phred:norm_methodupperquartile        **
min_length:trim_poly_g1
min_length:trim_poly_x1
min_length:norm_methodnone
min_length:norm_methodRLE
min_length:norm_methodTMM
min_length:norm_methodTMMwsp
min_length:norm_methodupperquartile
trim_poly_g1:trim_poly_x1                 **
trim_poly_g1:norm_methodnone
trim_poly_g1:norm_methodRLE
trim_poly_g1:norm_methodTMM
trim_poly_g1:norm_methodTMMwsp
trim_poly_g1:norm_methodupperquartile
trim_poly_x1:norm_methodnone             ***
trim_poly_x1:norm_methodRLE               **
trim_poly_x1:norm_methodTMM               **
trim_poly_x1:norm_methodTMMwsp            **
trim_poly_x1:norm_methodupperquartile     **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

14

```
(Dispersion parameter for quasi family taken to be 5.143455e-09)

    Null deviance: 3.3790e-01  on 100  degrees of freedom
Residual deviance: 2.8289e-07  on  55  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 2
```

**Effect Size**

```
# Classic LM
lm_fit <- DE_sd_df |>
    select(-c(runtime_sec, gene_overlap_percent, p_value_sd)) |>
    (\(x) glm(effect_size_sd ~ (.)^2, family = gaussian(), data = x))()
summary(lm_fit)
```

```
Call:
glm(formula = effect_size_sd ~ (.)^2, family = gaussian(), data = x)

Coefficients:
                                      Estimate Std. Error  t value Pr(>|t|)
(Intercept)                          1.602e+00  7.069e-03  226.583  < 2e-16
alignersalmon                       -6.532e-03  2.527e-03   -2.585  0.01243
min_phred                            1.802e-04  2.574e-04    0.700  0.48675
min_length                           1.674e-04  1.842e-04    0.909  0.36751
trim_poly_g1                        -7.183e-03  2.441e-03   -2.942  0.00476
trim_poly_x1                        -1.815e-03  2.833e-03   -0.641  0.52447
norm_methodnone                     -6.321e-01  4.227e-03 -149.546  < 2e-16
norm_methodRLE                      -6.356e-01  5.410e-03 -117.491  < 2e-16
norm_methodTMM                      -6.369e-01  5.193e-03 -122.645  < 2e-16
norm_methodTMMwsp                   -6.375e-01  4.232e-03 -150.651  < 2e-16
norm_methodupperquartile            -6.398e-01  4.534e-03 -141.116  < 2e-16
alignersalmon:min_phred             -1.978e-05  7.920e-05   -0.250  0.80374
alignersalmon:min_length            -2.752e-05  3.832e-05   -0.718  0.47563
alignersalmon:trim_poly_g1          -4.399e-04  4.685e-04   -0.939  0.35184
alignersalmon:trim_poly_x1          -5.199e-04  4.811e-04   -1.081  0.28461
alignersalmon:norm_methodnone        2.334e-02  7.500e-04   31.117  < 2e-16
alignersalmon:norm_methodRLE         2.251e-02  8.959e-04   25.120  < 2e-16
alignersalmon:norm_methodTMM         2.263e-02  8.808e-04   25.695  < 2e-16
alignersalmon:norm_methodTMMwsp      2.270e-02  8.552e-04   26.544  < 2e-16
alignersalmon:norm_methodupperquartile 2.328e-02 8.100e-04  28.740  < 2e-16
min_phred:min_length                -1.027e-05  6.588e-06   -1.559  0.12464
min_phred:trim_poly_g1               1.321e-04  7.606e-05    1.736  0.08809
min_phred:trim_poly_x1               3.718e-05  8.393e-05    0.443  0.65954
min_phred:norm_methodnone            1.798e-04  1.486e-04    1.210  0.23161
min_phred:norm_methodRLE             1.637e-04  1.845e-04    0.888  0.37863
min_phred:norm_methodTMM             2.937e-04  1.658e-04    1.772  0.08200
min_phred:norm_methodTMMwsp          1.606e-04  1.479e-04    1.085  0.28249
min_phred:norm_methodupperquartile   3.119e-04  1.768e-04    1.764  0.08331
min_length:trim_poly_g1              9.227e-05  3.890e-05    2.372  0.02122
min_length:trim_poly_x1              4.609e-05  4.511e-05    1.022  0.31148
min_length:norm_methodnone           6.692e-05  5.657e-05    1.183  0.24189
min_length:norm_methodRLE            7.430e-05  7.905e-05    0.940  0.35138
```

```
min_length:norm_methodTMM               5.787e-05  7.858e-05   0.736  0.46457
min_length:norm_methodTMMwsp            1.479e-04  6.702e-05   2.207  0.03147
min_length:norm_methodupperquartile     1.110e-05  6.643e-05   0.167  0.86788
trim_poly_g1:trim_poly_x1              -2.227e-04  4.686e-04  -0.475  0.63654
trim_poly_g1:norm_methodnone            7.887e-04  7.427e-04   1.062  0.29287
trim_poly_g1:norm_methodRLE             1.008e-03  1.006e-03   1.001  0.32105
trim_poly_g1:norm_methodTMM             1.301e-04  9.076e-04   0.143  0.88653
trim_poly_g1:norm_methodTMMwsp          1.861e-04  8.472e-04   0.220  0.82694
trim_poly_g1:norm_methodupperquartile   1.550e-03  8.684e-04   1.785  0.07974
trim_poly_x1:norm_methodnone           -6.796e-04  8.865e-04  -0.767  0.44655
trim_poly_x1:norm_methodRLE            -2.663e-04  9.952e-04  -0.268  0.79005
trim_poly_x1:norm_methodTMM            -4.644e-05  1.081e-03  -0.043  0.96588
trim_poly_x1:norm_methodTMMwsp         -4.481e-05  8.860e-04  -0.051  0.95985
trim_poly_x1:norm_methodupperquartile  -2.910e-04  9.732e-04  -0.299  0.76607

(Intercept)                             ***
alignersalmon                           *
min_phred
min_length
trim_poly_g1                            **
trim_poly_x1
norm_methodnone                         ***
norm_methodRLE                          ***
norm_methodTMM                          ***
norm_methodTMMwsp                       ***
norm_methodupperquartile                ***
alignersalmon:min_phred
alignersalmon:min_length
alignersalmon:trim_poly_g1
alignersalmon:trim_poly_x1
alignersalmon:norm_methodnone           ***
alignersalmon:norm_methodRLE            ***
alignersalmon:norm_methodTMM            ***
alignersalmon:norm_methodTMMwsp         ***
alignersalmon:norm_methodupperquartile  ***
min_phred:min_length
min_phred:trim_poly_g1                  .
min_phred:trim_poly_x1
min_phred:norm_methodnone
min_phred:norm_methodRLE
min_phred:norm_methodTMM                .
min_phred:norm_methodTMMwsp
min_phred:norm_methodupperquartile      .
min_length:trim_poly_g1                 *
min_length:trim_poly_x1
min_length:norm_methodnone
min_length:norm_methodRLE
min_length:norm_methodTMM
min_length:norm_methodTMMwsp            *
min_length:norm_methodupperquartile
trim_poly_g1:trim_poly_x1
trim_poly_g1:norm_methodnone
trim_poly_g1:norm_methodRLE
trim_poly_g1:norm_methodTMM
```

```
trim_poly_g1:norm_methodTMMwsp
trim_poly_g1:norm_methodupperquartile   .
trim_poly_x1:norm_methodnone
trim_poly_x1:norm_methodRLE
trim_poly_x1:norm_methodTMM
trim_poly_x1:norm_methodTMMwsp
trim_poly_x1:norm_methodupperquartile
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 7.807996e-07)

    Null deviance: 4.0307e+00  on 100  degrees of freedom
Residual deviance: 4.2944e-05  on  55  degrees of freedom
AIC: -1101.1

Number of Fisher Scoring iterations: 2
```

```r
# Log-normal GLM
glm_log_fit <- DE_sd_df |>
    select(-c(runtime_sec, gene_overlap_percent, p_value_sd)) |>
    (\(x) glm(effect_size_sd ~ (.)^2, family = gaussian(link = "log"), data = x))()
summary(glm_log_fit)
```

```
Call:
glm(formula = effect_size_sd ~ (.)^2, family = gaussian(link = "log"),
    data = x)

Coefficients:
                                         Estimate Std. Error  t value Pr(>|t|)
(Intercept)                             4.700e-01  6.209e-03   75.699  < 2e-16
alignersalmon                          -3.787e-03  2.250e-03   -1.683  0.09808
min_phred                               2.197e-04  2.238e-04    0.981  0.33069
min_length                              1.687e-04  1.646e-04    1.024  0.31016
trim_poly_g1                           -6.523e-03  2.254e-03   -2.894  0.00545
trim_poly_x1                           -2.301e-03  2.767e-03   -0.831  0.40932
norm_methodnone                        -5.011e-01  3.536e-03 -141.702  < 2e-16
norm_methodRLE                         -5.053e-01  4.878e-03 -103.591  < 2e-16
norm_methodTMM                         -5.056e-01  4.660e-03 -108.495  < 2e-16
norm_methodTMMwsp                      -5.066e-01  3.544e-03 -142.968  < 2e-16
norm_methodupperquartile               -5.085e-01  3.884e-03 -130.948  < 2e-16
alignersalmon:min_phred                -3.432e-05  7.361e-05   -0.466  0.64283
alignersalmon:min_length               -6.788e-06  3.591e-05   -0.189  0.85076
alignersalmon:trim_poly_g1             -5.265e-04  4.396e-04   -1.198  0.23618
alignersalmon:trim_poly_x1             -4.660e-04  4.557e-04   -1.023  0.31097
alignersalmon:norm_methodnone           2.034e-02  5.677e-04   35.839  < 2e-16
alignersalmon:norm_methodRLE            1.958e-02  7.528e-04   26.010  < 2e-16
alignersalmon:norm_methodTMM            1.976e-02  7.322e-04   26.986  < 2e-16
alignersalmon:norm_methodTMMwsp         1.977e-02  6.962e-04   28.403  < 2e-16
alignersalmon:norm_methodupperquartile  2.039e-02  6.423e-04   31.746  < 2e-16
min_phred:min_length                   -1.074e-05  5.997e-06   -1.790  0.07892
min_phred:trim_poly_g1                  9.871e-05  7.110e-05    1.388  0.17065
min_phred:trim_poly_x1                  5.454e-05  8.360e-05    0.652  0.51682
min_phred:norm_methodnone               1.712e-04  1.185e-04    1.445  0.15410
```

```
min_phred:norm_methodRLE                   1.668e-04  1.611e-04   1.035  0.30498
min_phred:norm_methodTMM                    2.991e-04  1.394e-04   2.145  0.03639
min_phred:norm_methodTMMwsp                 1.591e-04  1.176e-04   1.352  0.18187
min_phred:norm_methodupperquartile          2.828e-04  1.520e-04   1.861  0.06812
min_length:trim_poly_g1                      1.049e-04  3.578e-05   2.932  0.00490
min_length:trim_poly_x1                      3.703e-05  4.481e-05   0.826  0.41215
min_length:norm_methodnone                   6.718e-05  4.427e-05   1.518  0.13483
min_length:norm_methodRLE                    8.358e-05  7.029e-05   1.189  0.23954
min_length:norm_methodTMM                    3.889e-05  6.997e-05   0.556  0.58061
min_length:norm_methodTMMwsp                 1.460e-04  5.593e-05   2.611  0.01162
min_length:norm_methodupperquartile          1.124e-05  5.585e-05   0.201  0.84129
trim_poly_g1:trim_poly_x1                   -1.288e-04  4.390e-04  -0.293  0.77029
trim_poly_g1:norm_methodnone                 5.002e-04  5.614e-04   0.891  0.37681
trim_poly_g1:norm_methodRLE                  6.819e-04  8.758e-04   0.779  0.43957
trim_poly_g1:norm_methodTMM                 -1.516e-04  7.649e-04  -0.198  0.84358
trim_poly_g1:norm_methodTMMwsp              -1.659e-04  6.947e-04  -0.239  0.81215
trim_poly_g1:norm_methodupperquartile        1.166e-03  7.200e-04   1.619  0.11113
trim_poly_x1:norm_methodnone                -3.782e-04  6.669e-04  -0.567  0.57300
trim_poly_x1:norm_methodRLE                  8.369e-05  8.103e-04   0.103  0.91811
trim_poly_x1:norm_methodTMM                  3.671e-04  9.047e-04   0.406  0.68648
trim_poly_x1:norm_methodTMMwsp               3.189e-04  6.674e-04   0.478  0.63469
trim_poly_x1:norm_methodupperquartile        1.510e-04  7.803e-04   0.193  0.84732

(Intercept)                              ***
alignersalmon                            .
min_phred
min_length
trim_poly_g1                             **
trim_poly_x1
norm_methodnone                          ***
norm_methodRLE                           ***
norm_methodTMM                           ***
norm_methodTMMwsp                        ***
norm_methodupperquartile                 ***
alignersalmon:min_phred
alignersalmon:min_length
alignersalmon:trim_poly_g1
alignersalmon:trim_poly_x1
alignersalmon:norm_methodnone            ***
alignersalmon:norm_methodRLE             ***
alignersalmon:norm_methodTMM             ***
alignersalmon:norm_methodTMMwsp          ***
alignersalmon:norm_methodupperquartile   ***
min_phred:min_length                     .
min_phred:trim_poly_g1
min_phred:trim_poly_x1
min_phred:norm_methodnone
min_phred:norm_methodRLE
min_phred:norm_methodTMM                 *
min_phred:norm_methodTMMwsp
min_phred:norm_methodupperquartile       .
min_length:trim_poly_g1                  **
min_length:trim_poly_x1
min_length:norm_methodnone
```

```
min_length:norm_methodRLE
min_length:norm_methodTMM
min_length:norm_methodTMMwsp                 *
min_length:norm_methodupperquartile
trim_poly_g1:trim_poly_x1
trim_poly_g1:norm_methodnone
trim_poly_g1:norm_methodRLE
trim_poly_g1:norm_methodTMM
trim_poly_g1:norm_methodTMMwsp
trim_poly_g1:norm_methodupperquartile
trim_poly_x1:norm_methodnone
trim_poly_x1:norm_methodRLE
trim_poly_x1:norm_methodTMM
trim_poly_x1:norm_methodTMMwsp
trim_poly_x1:norm_methodupperquartile
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


(Dispersion parameter for gaussian family taken to be 7.627892e-07)


    Null deviance: 4.0307e+00  on 100  degrees of freedom
Residual deviance: 4.1953e-05  on  55  degrees of freedom
AIC: -1103.5


Number of Fisher Scoring iterations: 2
```

```
# Quasi GLM
quasi_fit <- DE_sd_df |>
    select(-c(runtime_sec, gene_overlap_percent, p_value_sd)) |>
    (\(x) glm(effect_size_sd ~ (.)^2, family = quasi(), data = x))()
summary(quasi_fit)
```

```
Call:
glm(formula = effect_size_sd ~ (.)^2, family = quasi(), data = x)

Coefficients:
                                  Estimate Std. Error  t value Pr(>|t|)
(Intercept)                      1.602e+00  7.069e-03  226.583  < 2e-16
alignersalmon                   -6.532e-03  2.527e-03   -2.585  0.01243
min_phred                        1.802e-04  2.574e-04    0.700  0.48675
min_length                       1.674e-04  1.842e-04    0.909  0.36751
trim_poly_g1                    -7.183e-03  2.441e-03   -2.942  0.00476
trim_poly_x1                    -1.815e-03  2.833e-03   -0.641  0.52447
norm_methodnone                 -6.321e-01  4.227e-03 -149.546  < 2e-16
norm_methodRLE                  -6.356e-01  5.410e-03 -117.491  < 2e-16
norm_methodTMM                  -6.369e-01  5.193e-03 -122.645  < 2e-16
norm_methodTMMwsp               -6.375e-01  4.232e-03 -150.651  < 2e-16
norm_methodupperquartile        -6.398e-01  4.534e-03 -141.116  < 2e-16
alignersalmon:min_phred         -1.978e-05  7.920e-05   -0.250  0.80374
alignersalmon:min_length        -2.752e-05  3.832e-05   -0.718  0.47563
alignersalmon:trim_poly_g1      -4.399e-04  4.685e-04   -0.939  0.35184
alignersalmon:trim_poly_x1      -5.199e-04  4.811e-04   -1.081  0.28461
alignersalmon:norm_methodnone    2.334e-02  7.500e-04   31.117  < 2e-16
alignersalmon:norm_methodRLE     2.251e-02  8.959e-04   25.120  < 2e-16
```

```
alignersalmon:norm_methodTMM              2.263e-02  8.808e-04   25.695  < 2e-16
alignersalmon:norm_methodTMMwsp           2.270e-02  8.552e-04   26.544  < 2e-16
alignersalmon:norm_methodupperquartile    2.328e-02  8.100e-04   28.740  < 2e-16
min_phred:min_length                     -1.027e-05  6.588e-06   -1.559  0.12464
min_phred:trim_poly_g1                     1.321e-04  7.606e-05    1.736  0.08809
min_phred:trim_poly_x1                     3.718e-05  8.393e-05    0.443  0.65954
min_phred:norm_methodnone                  1.798e-04  1.486e-04    1.210  0.23161
min_phred:norm_methodRLE                   1.637e-04  1.845e-04    0.888  0.37863
min_phred:norm_methodTMM                   2.937e-04  1.658e-04    1.772  0.08200
min_phred:norm_methodTMMwsp                1.606e-04  1.479e-04    1.085  0.28249
min_phred:norm_methodupperquartile         3.119e-04  1.768e-04    1.764  0.08331
min_length:trim_poly_g1                    9.227e-05  3.890e-05    2.372  0.02122
min_length:trim_poly_x1                     4.609e-05  4.511e-05    1.022  0.31148
min_length:norm_methodnone                 6.692e-05  5.657e-05    1.183  0.24189
min_length:norm_methodRLE                  7.430e-05  7.905e-05    0.940  0.35138
min_length:norm_methodTMM                  5.787e-05  7.858e-05    0.736  0.46457
min_length:norm_methodTMMwsp               1.479e-04  6.702e-05    2.207  0.03147
min_length:norm_methodupperquartile        1.110e-05  6.643e-05    0.167  0.86788
trim_poly_g1:trim_poly_x1                 -2.227e-04  4.686e-04   -0.475  0.63654
trim_poly_g1:norm_methodnone               7.887e-04  7.427e-04    1.062  0.29287
trim_poly_g1:norm_methodRLE                1.008e-03  1.006e-03    1.001  0.32105
trim_poly_g1:norm_methodTMM                1.301e-04  9.076e-04    0.143  0.88653
trim_poly_g1:norm_methodTMMwsp             1.861e-04  8.472e-04    0.220  0.82694
trim_poly_g1:norm_methodupperquartile      1.550e-03  8.684e-04    1.785  0.07974
trim_poly_x1:norm_methodnone              -6.796e-04  8.865e-04   -0.767  0.44655
trim_poly_x1:norm_methodRLE               -2.663e-04  9.952e-04   -0.268  0.79005
trim_poly_x1:norm_methodTMM               -4.644e-05  1.081e-03   -0.043  0.96588
trim_poly_x1:norm_methodTMMwsp            -4.481e-05  8.860e-04   -0.051  0.95985
trim_poly_x1:norm_methodupperquartile     -2.910e-04  9.732e-04   -0.299  0.76607

(Intercept)                               ***
alignersalmon                             *
min_phred
min_length
trim_poly_g1                              **
trim_poly_x1
norm_methodnone                           ***
norm_methodRLE                            ***
norm_methodTMM                            ***
norm_methodTMMwsp                         ***
norm_methodupperquartile                  ***
alignersalmon:min_phred
alignersalmon:min_length
alignersalmon:trim_poly_g1
alignersalmon:trim_poly_x1
alignersalmon:norm_methodnone             ***
alignersalmon:norm_methodRLE              ***
alignersalmon:norm_methodTMM              ***
alignersalmon:norm_methodTMMwsp           ***
alignersalmon:norm_methodupperquartile    ***
min_phred:min_length
min_phred:trim_poly_g1                     .
min_phred:trim_poly_x1
min_phred:norm_methodnone
```

```
min_phred:norm_methodRLE
min_phred:norm_methodTMM                     .
min_phred:norm_methodTMMwsp
min_phred:norm_methodupperquartile        .
min_length:trim_poly_g1                   *
min_length:trim_poly_x1
min_length:norm_methodnone
min_length:norm_methodRLE
min_length:norm_methodTMM
min_length:norm_methodTMMwsp              *
min_length:norm_methodupperquartile
trim_poly_g1:trim_poly_x1
trim_poly_g1:norm_methodnone
trim_poly_g1:norm_methodRLE
trim_poly_g1:norm_methodTMM
trim_poly_g1:norm_methodTMMwsp
trim_poly_g1:norm_methodupperquartile     .
trim_poly_x1:norm_methodnone
trim_poly_x1:norm_methodRLE
trim_poly_x1:norm_methodTMM
trim_poly_x1:norm_methodTMMwsp
trim_poly_x1:norm_methodupperquartile
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasi family taken to be 7.807996e-07)

    Null deviance: 4.0307e+00  on 100  degrees of freedom
Residual deviance: 4.2944e-05  on  55  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 2
```

## Bayesian

We know that

$$Y^2 \xrightarrow{d} \mathcal{N}(\mu, \sigma^2) \tag{5}$$

by the central limit theorem since $Y^2$ is an average. This is not completely accurate because $Y^2 > 0$, but if $\mu >> 0$, then the truncation is inconsequential. Using the 1-1 transformation formula we can derive that the distribution of $Y$ must be:

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(y^2 - \mu)^2}{2\sigma^2}} \cdot 2y \tag{6}$$

Unfortunately this doesn't have a close form expectation, which makes it difficult to model $\mathbb{E}\, Y = X\beta$. Since a mean and variance function can be derived, it is possible to fit a model with something like general estimating equation, but there a two key problems. First, the mean function is an integral which most likely needs to be approximated. Second, the necessary link function results in a non-linear relationship between the $\hat{\beta}$s and $Y$ making interpretation difficult.

Instead, we will build from the fact that $Y \geq 0$. There are several common likelihoods that have support $[0, \infty)$ such as the log-normal, gamma, weibull, etc. Since we are looking to model $\mathbb{E}\, Y = X\beta$, the log-normal is the simplest choice since the default parameterization is a location-scale family.

Consider the following Bayesian Hierarchical Model:

$$
\begin{aligned}
Y_i &\sim \log-\mathcal{N}(\mu_i, \sigma_i^2) \\
\mu_i &= X_i\beta \\
\sigma_i &= a \cdot \mu_i^b \\
\beta &\sim \mathcal{N}(0, 100) \\
a &\sim \text{Gamma}(c, d) \\
b &\sim \mathcal{N}(0, 10)
\end{aligned} \tag{7}
$$

This set up has a couple of key advantages.

1. The interpretation is still linear on the $Y$ scale since we are modeling $\mathbb{E}\, Y_i = \mu_i = X_i\beta$.

2. Natural parameter shrinkage via the prior on $\beta$. Handles multicollinearity and high dimensionality of $X$.

3. Does not assume constant variance. Specifically, we are applying the variance-power law from the Tweedie family of distributions, which the log-Normal is a member.

$$
\text{Var}\, Y \propto (\mathbb{E}\, Y)^p \tag{8}
$$

   $a > 0$ and represents a common variance scale ie if $b = 0$ we recover the classical log-Normal regression model. $b \in \mathbb{R}$ where $b > 0$ indicates over-dispersion and $b < 0$ indicates under-dispersion.

4. We can use the posterior predictive distribution to check whether the model is consistent with the fact that $Y^2 \sim \mathcal{N}$.