

# 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,

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

and

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

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

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

and

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

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

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

```

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

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

# Load all relevant summary data
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_star_hisat2 <- read.csv("count_sd_df_star_hisat2.csv")

# Combine ALL count summary data
count_sd_df <- bind_rows(
  count_sd_df_salmon_kallisto,
  count_sd_df_STAR_HISAT2,
  count_sd_df_star_hisat2
)

# Convert factor columns
factors <- c("aligner", "trim_poly_g", "trim_poly_x", "norm_method")
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_star_hisat2 <- read.csv("DE_sd_df_star_hisat2.csv")
DE_sd_df <- bind_rows(DE_sd_df_salmon_kallisto, DE_sd_df_star_hisat2)
factors <- c("aligner", "trim_poly_g", "trim_poly_x", "norm_method")
DE_sd_df <- DE_sd_df |> mutate(across(any_of(factors), ~ as.factor(.)))

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      3002.
2 kallisto    2230.
3 salmon      2312.
4 STAR       3423.

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          3031.
2 1          3013.
```

```
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           3053.
2 1           2990.
```

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

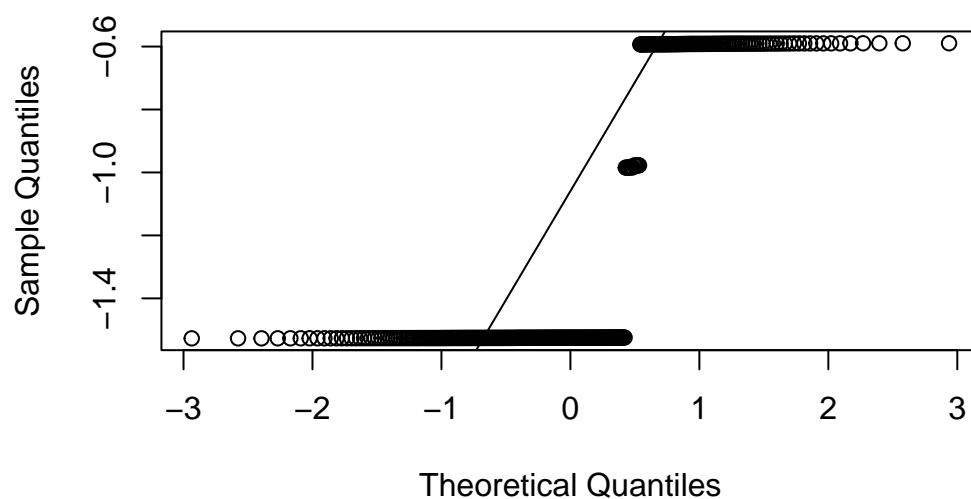
```
# A tibble: 4 x 2
  aligner mean_runtime
  <fct>         <dbl>
1 HISAT2         NA
2 kallisto       181.
3 salmon       1269.
4 STAR           NA
```

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

```
# A tibble: 44 x 4
# Groups:   aligner, norm_method [22]
  aligner norm_method trim_poly_x mean_p_value_sd
  <fct>   <fct>         <fct>         <dbl>
1 HISAT2 none          0           0.217
2 HISAT2 none          1           0.217
3 HISAT2 RLE           0           0.218
4 HISAT2 RLE           1           0.218
5 HISAT2 TMM           0           0.217
6 HISAT2 TMM           1           0.217
7 HISAT2 TMMwsp        0           0.217
8 HISAT2 TMMwsp        1           0.217
9 HISAT2 upperquartile 0           0.217
10 HISAT2 upperquartile 1           0.217
# i 34 more rows
```

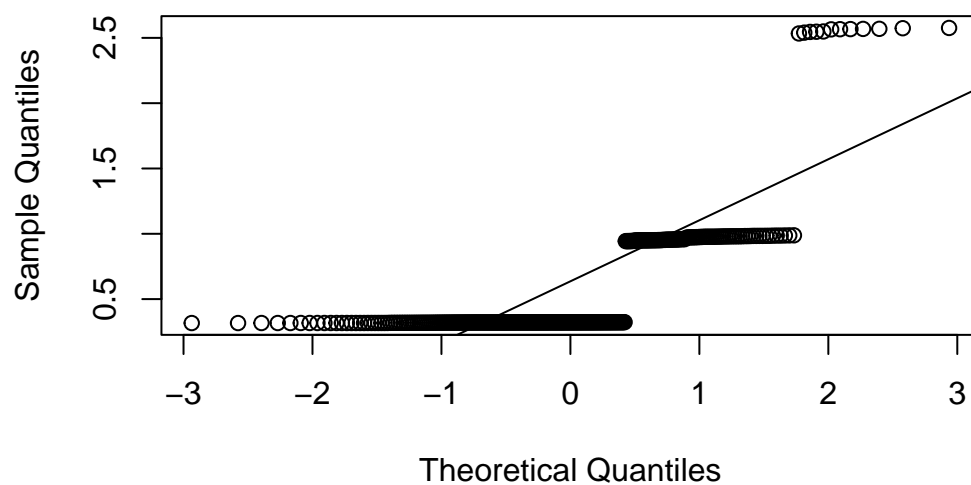
```
qqnorm(log(DE_sd_df$p_value_sd), main = "p-value SD Q-Q Plot")
qqline(log(DE_sd_df$p_value_sd))
```

### p-value SD Q-Q Plot

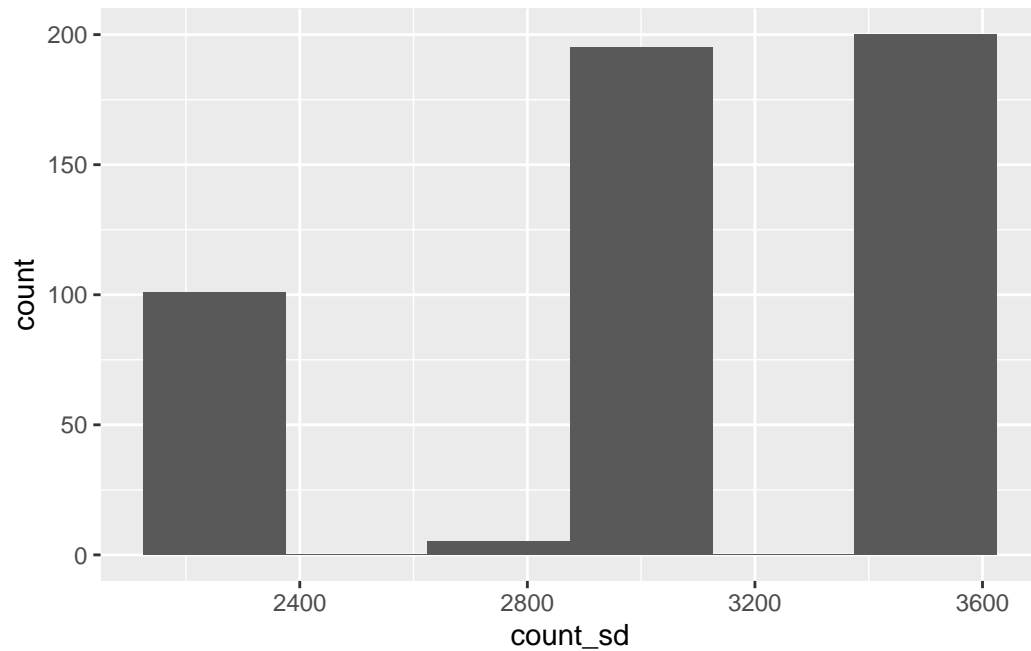


```
qqnorm(DE_sd_df$effect_size_sd^2, main = "Effect Size SD Q-Q Plot")
qqline(DE_sd_df$effect_size_sd^2)
```

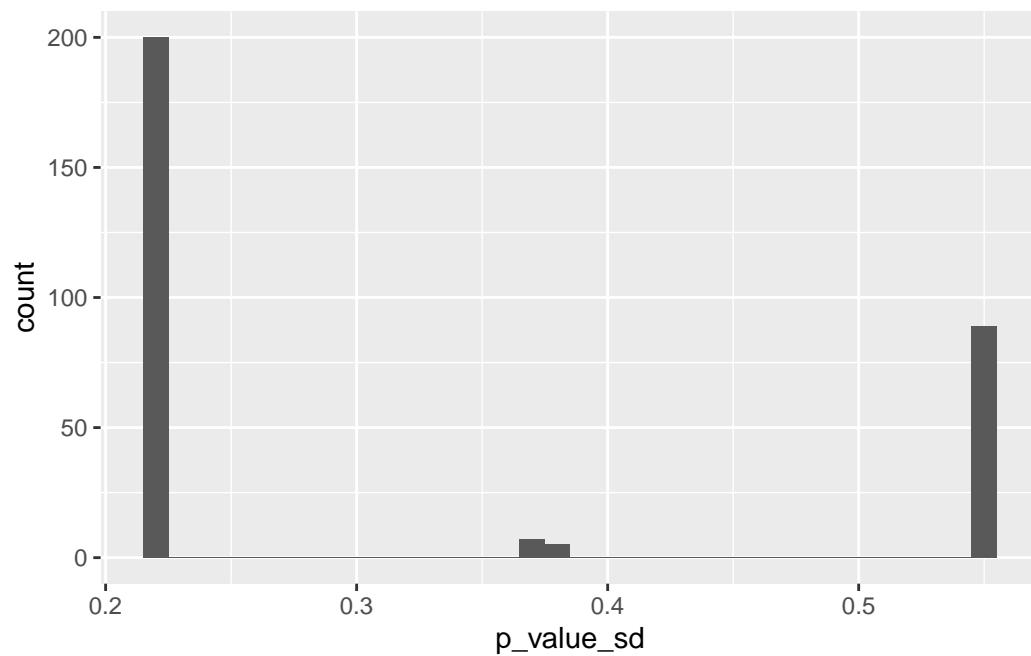
### Effect Size SD Q-Q Plot



```
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)	3008.39972	64.14166	46.902	<2e-16 ***
alignerkallisto	-758.29185	41.39051	-18.320	<2e-16 ***
alignersalmon	-678.82916	40.10585	-16.926	<2e-16 ***
alignerSTAR	457.75110	25.56446	17.906	<2e-16 ***
min_phred	-0.24431	2.45522	-0.100	0.9208
min_length	0.32039	1.51564	0.211	0.8327
trim_poly_g1	-20.11668	23.01991	-0.874	0.3826
trim_poly_x1	26.13987	23.13600	1.130	0.2591
alignerkallisto:min_phred	-0.86556	1.21289	-0.714	0.4758
alignersalmon:min_phred	-0.73292	1.29424	-0.566	0.5715
alignerSTAR:min_phred	-0.87433	0.72429	-1.207	0.2280
alignerkallisto:min_length	0.25140	0.65977	0.381	0.7033
alignersalmon:min_length	0.17783	0.59584	0.298	0.7655
alignerSTAR:min_length	-0.33909	0.40013	-0.847	0.3972
alignerkallisto:trim_poly_g1	1.34140	7.65664	0.175	0.8610
alignersalmon:trim_poly_g1	4.27302	7.42879	0.575	0.5654
alignerSTAR:trim_poly_g1	2.37221	4.57865	0.518	0.6046
alignerkallisto:trim_poly_x1	-5.27661	7.30506	-0.722	0.4705
alignersalmon:trim_poly_x1	-2.37450	7.73406	-0.307	0.7590
alignerSTAR:trim_poly_x1	-4.35868	4.62101	-0.943	0.3460
min_phred:min_length	-0.01253	0.05819	-0.215	0.8296
min_phred:trim_poly_g1	0.81609	0.66411	1.229	0.2197
min_phred:trim_poly_x1	0.12361	0.66712	0.185	0.8531
min_length:trim_poly_g1	-0.09030	0.35893	-0.252	0.8015
min_length:trim_poly_x1	-0.67106	0.36485	-1.839	0.0665 .
trim_poly_g1:trim_poly_x1	-0.33214	4.16675	-0.080	0.9365

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 520.1829)

Null deviance: 89766478 on 500 degrees of freedom  
 Residual deviance: 247087 on 475 degrees of freedom  
 AIC: 4582.4

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)	8.009e+00	2.102e-02	380.963	<2e-16 ***
alignerkallisto	-2.886e-01	1.768e-02	-16.322	<2e-16 ***
alignersalmon	-2.535e-01	1.657e-02	-15.297	<2e-16 ***
alignerSTAR	1.410e-01	8.009e-03	17.599	<2e-16 ***
min_phred	-6.944e-05	8.026e-04	-0.087	0.9311
min_length	1.014e-04	4.962e-04	0.204	0.8381
trim_poly_g1	-6.350e-03	7.538e-03	-0.842	0.4000
trim_poly_x1	8.471e-03	7.563e-03	1.120	0.2632
alignerkallisto:min_phred	-3.947e-04	5.181e-04	-0.762	0.4466
alignersalmon:min_phred	-3.546e-04	5.403e-04	-0.656	0.5120
alignerSTAR:min_phred	-2.437e-04	2.270e-04	-1.074	0.2836
alignerkallisto:min_length	4.956e-05	2.823e-04	0.176	0.8607
alignersalmon:min_length	3.154e-05	2.451e-04	0.129	0.8976
alignerSTAR:min_length	-8.194e-05	1.254e-04	-0.654	0.5137
alignerkallisto:trim_poly_g1	3.213e-04	3.275e-03	0.098	0.9219
alignersalmon:trim_poly_g1	1.467e-03	3.080e-03	0.476	0.6341
alignerSTAR:trim_poly_g1	8.427e-04	1.435e-03	0.587	0.5573
alignerkallisto:trim_poly_x1	-1.991e-03	3.116e-03	-0.639	0.5231
alignersalmon:trim_poly_x1	-9.208e-04	3.214e-03	-0.286	0.7747
alignerSTAR:trim_poly_x1	-1.349e-03	1.448e-03	-0.931	0.3522
min_phred:min_length	-4.204e-06	1.901e-05	-0.221	0.8251
min_phred:trim_poly_g1	2.515e-04	2.155e-04	1.167	0.2437
min_phred:trim_poly_x1	3.947e-05	2.169e-04	0.182	0.8557
min_length:trim_poly_g1	-2.596e-05	1.176e-04	-0.221	0.8254
min_length:trim_poly_x1	-2.165e-04	1.199e-04	-1.806	0.0716
trim_poly_g1:trim_poly_x1	-1.290e-04	1.362e-03	-0.095	0.9246

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 520.4323)

Null deviance: 89766478 on 500 degrees of freedom  
Residual deviance: 247205 on 475 degrees of freedom  
AIC: 4582.7

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)	3008.39972	64.14166	46.902	<2e-16 ***
alignerkallisto	-758.29185	41.39051	-18.320	<2e-16 ***
alignersalmon	-678.82916	40.10585	-16.926	<2e-16 ***
alignerSTAR	457.75110	25.56446	17.906	<2e-16 ***
min_phred	-0.24431	2.45522	-0.100	0.9208
min_length	0.32039	1.51564	0.211	0.8327
trim_poly_g1	-20.11668	23.01991	-0.874	0.3826
trim_poly_x1	26.13987	23.13600	1.130	0.2591
alignerkallisto:min_phred	-0.86556	1.21289	-0.714	0.4758
alignersalmon:min_phred	-0.73292	1.29424	-0.566	0.5715
alignerSTAR:min_phred	-0.87433	0.72429	-1.207	0.2280
alignerkallisto:min_length	0.25140	0.65977	0.381	0.7033
alignersalmon:min_length	0.17783	0.59584	0.298	0.7655
alignerSTAR:min_length	-0.33909	0.40013	-0.847	0.3972
alignerkallisto:trim_poly_g1	1.34140	7.65664	0.175	0.8610
alignersalmon:trim_poly_g1	4.27302	7.42879	0.575	0.5654
alignerSTAR:trim_poly_g1	2.37221	4.57865	0.518	0.6046
alignerkallisto:trim_poly_x1	-5.27661	7.30506	-0.722	0.4705
alignersalmon:trim_poly_x1	-2.37450	7.73406	-0.307	0.7590
alignerSTAR:trim_poly_x1	-4.35868	4.62101	-0.943	0.3460
min_phred:min_length	-0.01253	0.05819	-0.215	0.8296
min_phred:trim_poly_g1	0.81609	0.66411	1.229	0.2197
min_phred:trim_poly_x1	0.12361	0.66712	0.185	0.8531
min_length:trim_poly_g1	-0.09030	0.35893	-0.252	0.8015
min_length:trim_poly_x1	-0.67106	0.36485	-1.839	0.0665 .
trim_poly_g1:trim_poly_x1	-0.33214	4.16675	-0.080	0.9365

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasi family taken to be 520.1829)

Null deviance: 89766478 on 500 degrees of freedom

Residual deviance: 247087 on 475 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: (2 not defined because of singularities)

	Estimate	Std. Error	t value
(Intercept)	3.918e-02	2.577e-04	152.067
alignerkallisto	3.337e-01	1.230e-04	2712.604
alignersalmon	3.366e-01	1.061e-04	3173.283



alignerSTAR	2.967e-04	8.204e-05	3.616
min_phred	2.714e-05	9.541e-06	2.845
min_length	4.527e-06	5.481e-06	0.826
trim_poly_g1	1.798e-05	8.249e-05	0.218
trim_poly_x1	-2.673e-04	8.583e-05	-3.114
norm_methodnone	1.782e-01	1.913e-04	931.629
norm_methodRLE	1.784e-01	2.012e-04	886.599
norm_methodTMM	1.783e-01	1.945e-04	916.609
norm_methodTMMwsp	1.783e-01	1.946e-04	916.425
norm_methodupperquartile	1.782e-01	2.027e-04	879.217
alignerkallisto:min_phred	-6.902e-06	3.346e-06	-2.063
alignersalmon:min_phred	-9.897e-06	3.321e-06	-2.980
alignerSTAR:min_phred	3.343e-06	2.255e-06	1.483
alignerkallisto:min_length	-1.810e-06	1.779e-06	-1.017
alignersalmon:min_length	-7.153e-06	1.568e-06	-4.561
alignerSTAR:min_length	-5.327e-06	1.231e-06	-4.328
alignerkallisto:trim_poly_g1	-6.916e-06	2.069e-05	-0.334
alignersalmon:trim_poly_g1	-9.928e-06	1.925e-05	-0.516
alignerSTAR:trim_poly_g1	5.839e-08	1.429e-05	0.004
alignerkallisto:trim_poly_x1	4.341e-04	2.045e-05	21.232
alignersalmon:trim_poly_x1	2.119e-04	2.031e-05	10.432
alignerSTAR:trim_poly_x1	-1.485e-05	1.435e-05	-1.035
alignerkallisto:norm_methodnone	2.299e-03	4.536e-05	50.676
alignersalmon:norm_methodnone	1.067e-03	2.862e-05	37.275
alignerSTAR:norm_methodnone	5.449e-05	2.146e-05	2.539
alignerkallisto:norm_methodRLE	1.784e-03	4.864e-05	36.676
alignersalmon:norm_methodRLE	3.079e-04	3.251e-05	9.471
alignerSTAR:norm_methodRLE	-4.501e-05	2.142e-05	-2.102
alignerkallisto:norm_methodTMM	2.029e-03	4.660e-05	43.528
alignersalmon:norm_methodTMM	4.510e-04	3.181e-05	14.176
alignerSTAR:norm_methodTMM	9.326e-05	2.059e-05	4.529
alignerkallisto:norm_methodTMMwsp	1.948e-03	4.735e-05	41.141
alignersalmon:norm_methodTMMwsp	3.949e-04	3.248e-05	12.158
alignerSTAR:norm_methodTMMwsp	-4.031e-05	2.567e-05	-1.570
alignerkallisto:norm_methodupperquartile	1.685e-03	4.345e-05	38.777
alignersalmon:norm_methodupperquartile	NA	NA	NA
alignerSTAR:norm_methodupperquartile	NA	NA	NA
min_phred:min_length	1.613e-07	1.794e-07	0.900
min_phred:trim_poly_g1	2.240e-06	2.085e-06	1.074
min_phred:trim_poly_x1	-6.087e-06	2.114e-06	-2.879
min_phred:norm_methodnone	-3.743e-05	7.284e-06	-5.139
min_phred:norm_methodRLE	-3.966e-05	7.530e-06	-5.268
min_phred:norm_methodTMM	-4.071e-05	7.374e-06	-5.522
min_phred:norm_methodTMMwsp	-3.983e-05	7.330e-06	-5.434
min_phred:norm_methodupperquartile	-3.682e-05	7.485e-06	-4.919
min_length:trim_poly_g1	-1.180e-06	1.178e-06	-1.002
min_length:trim_poly_x1	2.143e-06	1.163e-06	1.842
min_length:norm_methodnone	-4.891e-06	2.852e-06	-1.715
min_length:norm_methodRLE	-1.737e-06	3.049e-06	-0.570
min_length:norm_methodTMM	-1.442e-06	2.974e-06	-0.485
min_length:norm_methodTMMwsp	-1.840e-06	2.999e-06	-0.614
min_length:norm_methodupperquartile	-1.010e-06	3.010e-06	-0.335
trim_poly_g1:trim_poly_x1	2.939e-05	1.280e-05	2.296
trim_poly_g1:norm_methodnone	-2.816e-05	3.875e-05	-0.727

trim_poly_g1:norm_methodRLE	-2.792e-05	4.034e-05	-0.692
trim_poly_g1:norm_methodTMM	-2.844e-05	4.007e-05	-0.710
trim_poly_g1:norm_methodTMMwsp	-2.639e-05	4.074e-05	-0.648
trim_poly_g1:norm_methodupperquartile	-5.982e-05	3.953e-05	-1.513
trim_poly_x1:norm_methodnone	2.684e-04	4.587e-05	5.851
trim_poly_x1:norm_methodRLE	2.924e-04	4.602e-05	6.355
trim_poly_x1:norm_methodTMM	2.917e-04	4.628e-05	6.303
trim_poly_x1:norm_methodTMMwsp	2.724e-04	4.548e-05	5.989
trim_poly_x1:norm_methodupperquartile	3.120e-04	4.559e-05	6.845
	Pr(> t )		
(Intercept)	< 2e-16	***	
alignerkallisto	< 2e-16	***	
alignersalmon	< 2e-16	***	
alignerSTAR	0.000365	***	
min_phred	0.004835	**	
min_length	0.409628		
trim_poly_g1	0.827624		
trim_poly_x1	0.002074	**	
norm_methodnone	< 2e-16	***	
norm_methodRLE	< 2e-16	***	
norm_methodTMM	< 2e-16	***	
norm_methodTMMwsp	< 2e-16	***	
norm_methodupperquartile	< 2e-16	***	
alignerkallisto:min_phred	0.040236	*	
alignersalmon:min_phred	0.003178	**	
alignerSTAR:min_phred	0.139428		
alignerkallisto:min_length	0.309955		
alignersalmon:min_length	8.15e-06	***	
alignerSTAR:min_length	2.22e-05	***	
alignerkallisto:trim_poly_g1	0.738446		
alignersalmon:trim_poly_g1	0.606461		
alignerSTAR:trim_poly_g1	0.996742		
alignerkallisto:trim_poly_x1	< 2e-16	***	
alignersalmon:trim_poly_x1	< 2e-16	***	
alignerSTAR:trim_poly_x1	0.301738		
alignerkallisto:norm_methodnone	< 2e-16	***	
alignersalmon:norm_methodnone	< 2e-16	***	
alignerSTAR:norm_methodnone	0.011746	*	
alignerkallisto:norm_methodRLE	< 2e-16	***	
alignersalmon:norm_methodRLE	< 2e-16	***	
alignerSTAR:norm_methodRLE	0.036641	*	
alignerkallisto:norm_methodTMM	< 2e-16	***	
alignersalmon:norm_methodTMM	< 2e-16	***	
alignerSTAR:norm_methodTMM	9.38e-06	***	
alignerkallisto:norm_methodTMMwsp	< 2e-16	***	
alignersalmon:norm_methodTMMwsp	< 2e-16	***	
alignerSTAR:norm_methodTMMwsp	0.117677		
alignerkallisto:norm_methodupperquartile	< 2e-16	***	
alignersalmon:norm_methodupperquartile	NA		
alignerSTAR:norm_methodupperquartile	NA		
min_phred:min_length	0.369291		
min_phred:trim_poly_g1	0.283748		
min_phred:trim_poly_x1	0.004356	**	
min_phred:norm_methodnone	5.78e-07	***	

```

min_phred:norm_methodRLE          3.10e-07 ***
min_phred:norm_methodTMM          8.81e-08 ***
min_phred:norm_methodTMMwsp       1.37e-07 ***
min_phred:norm_methodupperquartile 1.63e-06 ***
min_length:trim_poly_g1          0.317576
min_length:trim_poly_x1          0.066714 .
min_length:norm_methodnone        0.087639 .
min_length:norm_methodRLE         0.569404
min_length:norm_methodTMM         0.628281
min_length:norm_methodTMMwsp      0.540020
min_length:norm_methodupperquartile 0.737584
trim_poly_g1:trim_poly_x1        0.022568 *
trim_poly_g1:norm_methodnone      0.468152
trim_poly_g1:norm_methodRLE       0.489497
trim_poly_g1:norm_methodTMM       0.478482
trim_poly_g1:norm_methodTMMwsp    0.517737
trim_poly_g1:norm_methodupperquartile 0.131489
trim_poly_x1:norm_methodnone      1.62e-08 ***
trim_poly_x1:norm_methodRLE       1.06e-09 ***
trim_poly_x1:norm_methodTMM       1.41e-09 ***
trim_poly_x1:norm_methodTMMwsp    7.78e-09 ***
trim_poly_x1:norm_methodupperquartile 6.51e-11 ***

```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 2.43689e-09)

```

Null deviance: 6.9874e+00 on 300 degrees of freedom
Residual deviance: 5.7754e-07 on 237 degrees of freedom
AIC: -5057.3

```

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: (2 not defined because of singularities)

	Estimate	Std. Error	t value
(Intercept)	-1.915e+00	7.362e-04	-2601.443
alignerkallisto	9.284e-01	3.392e-04	2736.762
alignersalmon	9.356e-01	3.141e-04	2978.583
alignerSTAR	1.364e-03	3.680e-04	3.707
min_phred	6.706e-05	2.696e-05	2.488
min_length	2.788e-05	1.596e-05	1.747
trim_poly_g1	-8.422e-05	2.228e-04	-0.378
trim_poly_x1	-3.133e-04	2.412e-04	-1.299
norm_methodnone	3.882e-01	4.802e-04	808.307

norm_methodRLE	3.894e-01	5.319e-04	732.181
norm_methodTMM	3.891e-01	5.133e-04	758.088
norm_methodTMMwsp	3.890e-01	4.904e-04	793.102
norm_methodupperquartile	3.887e-01	5.016e-04	775.011
alignerkallisto:min_phred	1.007e-06	9.185e-06	0.110
alignersalmon:min_phred	-7.508e-09	9.102e-06	-0.001
alignerSTAR:min_phred	1.539e-05	1.011e-05	1.522
alignerkallisto:min_length	-2.294e-05	4.934e-06	-4.650
alignersalmon:min_length	-3.238e-05	4.679e-06	-6.920
alignerSTAR:min_length	-2.451e-05	5.520e-06	-4.440
alignerkallisto:trim_poly_g1	-2.717e-05	5.745e-05	-0.473
alignersalmon:trim_poly_g1	-4.176e-05	5.535e-05	-0.755
alignerSTAR:trim_poly_g1	2.314e-07	6.408e-05	0.004
alignerkallisto:trim_poly_x1	8.545e-04	5.754e-05	14.852
alignersalmon:trim_poly_x1	4.454e-04	5.674e-05	7.849
alignerSTAR:trim_poly_x1	-6.817e-05	6.437e-05	-1.059
alignerkallisto:norm_methodnone	6.724e-03	1.242e-04	54.127
alignersalmon:norm_methodnone	2.429e-03	8.399e-05	28.917
alignerSTAR:norm_methodnone	2.512e-04	9.628e-05	2.609
alignerkallisto:norm_methodRLE	4.947e-03	1.309e-04	37.784
alignersalmon:norm_methodRLE	2.049e-04	8.942e-05	2.292
alignerSTAR:norm_methodRLE	-2.072e-04	9.604e-05	-2.157
alignerkallisto:norm_methodTMM	5.687e-03	1.242e-04	45.781
alignersalmon:norm_methodTMM	8.023e-04	8.629e-05	9.298
alignerSTAR:norm_methodTMM	4.285e-04	9.234e-05	4.640
alignerkallisto:norm_methodTMMwsp	5.545e-03	1.329e-04	41.713
alignersalmon:norm_methodTMMwsp	6.353e-04	9.755e-05	6.513
alignerSTAR:norm_methodTMMwsp	-1.854e-04	1.151e-04	-1.610
alignerkallisto:norm_methodupperquartile	5.152e-03	1.028e-04	50.101
alignersalmon:norm_methodupperquartile	NA	NA	NA
alignerSTAR:norm_methodupperquartile	NA	NA	NA
min_phred:min_length	3.553e-07	5.320e-07	0.668
min_phred:trim_poly_g1	9.703e-06	6.031e-06	1.609
min_phred:trim_poly_x1	-2.526e-05	6.225e-06	-4.058
min_phred:norm_methodnone	-9.742e-05	1.841e-05	-5.290
min_phred:norm_methodRLE	-1.070e-04	1.969e-05	-5.433
min_phred:norm_methodTMM	-1.119e-04	1.904e-05	-5.879
min_phred:norm_methodTMMwsp	-1.048e-04	1.841e-05	-5.693
min_phred:norm_methodupperquartile	-1.023e-04	1.939e-05	-5.277
min_length:trim_poly_g1	-4.077e-06	3.190e-06	-1.278
min_length:trim_poly_x1	6.178e-06	3.359e-06	1.839
min_length:norm_methodnone	-1.165e-05	7.177e-06	-1.624
min_length:norm_methodRLE	-3.280e-06	8.020e-06	-0.409
min_length:norm_methodTMM	-5.468e-06	7.886e-06	-0.693
min_length:norm_methodTMMwsp	-5.077e-06	7.681e-06	-0.661
min_length:norm_methodupperquartile	-2.956e-06	7.624e-06	-0.388
trim_poly_g1:trim_poly_x1	1.480e-04	3.749e-05	3.949
trim_poly_g1:norm_methodnone	-7.235e-05	9.705e-05	-0.746
trim_poly_g1:norm_methodRLE	8.188e-06	1.052e-04	0.078
trim_poly_g1:norm_methodTMM	-5.327e-05	1.037e-04	-0.514
trim_poly_g1:norm_methodTMMwsp	1.376e-06	1.027e-04	0.013
trim_poly_g1:norm_methodupperquartile	-1.026e-04	1.013e-04	-1.012
trim_poly_x1:norm_methodnone	4.526e-04	1.156e-04	3.914
trim_poly_x1:norm_methodRLE	5.125e-04	1.186e-04	4.320

trim_poly_x1:norm_methodTMM	5.093e-04	1.211e-04	4.204
trim_poly_x1:norm_methodTMMwsp	4.501e-04	1.151e-04	3.911
trim_poly_x1:norm_methodupperquartile	5.381e-04	1.174e-04	4.584
	Pr(> t )		
(Intercept)	< 2e-16	***	
alignerkallisto	< 2e-16	***	
alignersalmon	< 2e-16	***	
alignerSTAR	0.000261	***	
min_phred	0.013547	*	
min_length	0.081946	.	
trim_poly_g1	0.705804		
trim_poly_x1	0.195273		
norm_methodnone	< 2e-16	***	
norm_methodRLE	< 2e-16	***	
norm_methodTMM	< 2e-16	***	
norm_methodTMMwsp	< 2e-16	***	
norm_methodupperquartile	< 2e-16	***	
alignerkallisto:min_phred	0.912830		
alignersalmon:min_phred	0.999343		
alignerSTAR:min_phred	0.129248		
alignerkallisto:min_length	5.51e-06	***	
alignersalmon:min_length	4.19e-11	***	
alignerSTAR:min_length	1.38e-05	***	
alignerkallisto:trim_poly_g1	0.636736		
alignersalmon:trim_poly_g1	0.451259		
alignerSTAR:trim_poly_g1	0.997121		
alignerkallisto:trim_poly_x1	< 2e-16	***	
alignersalmon:trim_poly_x1	1.44e-13	***	
alignerSTAR:trim_poly_x1	0.290671		
alignerkallisto:norm_methodnone	< 2e-16	***	
alignersalmon:norm_methodnone	< 2e-16	***	
alignerSTAR:norm_methodnone	0.009646	**	
alignerkallisto:norm_methodRLE	< 2e-16	***	
alignersalmon:norm_methodRLE	0.022799	*	
alignerSTAR:norm_methodRLE	0.031983	*	
alignerkallisto:norm_methodTMM	< 2e-16	***	
alignersalmon:norm_methodTMM	< 2e-16	***	
alignerSTAR:norm_methodTMM	5.76e-06	***	
alignerkallisto:norm_methodTMMwsp	< 2e-16	***	
alignersalmon:norm_methodTMMwsp	4.37e-10	***	
alignerSTAR:norm_methodTMMwsp	0.108770		
alignerkallisto:norm_methodupperquartile	< 2e-16	***	
alignersalmon:norm_methodupperquartile	NA		
alignerSTAR:norm_methodupperquartile	NA		
min_phred:min_length	0.504848		
min_phred:trim_poly_g1	0.108954		
min_phred:trim_poly_x1	6.73e-05	***	
min_phred:norm_methodnone	2.77e-07	***	
min_phred:norm_methodRLE	1.37e-07	***	
min_phred:norm_methodTMM	1.39e-08	***	
min_phred:norm_methodTMMwsp	3.68e-08	***	
min_phred:norm_methodupperquartile	2.97e-07	***	
min_length:trim_poly_g1	0.202546		
min_length:trim_poly_x1	0.067163	.	

```

min_length:norm_methodnone          0.105740
min_length:norm_methodRLE            0.682941
min_length:norm_methodTMM            0.488755
min_length:norm_methodTMMwsp         0.509210
min_length:norm_methodupperquartile  0.698614
trim_poly_g1:trim_poly_x1            0.000103 ***
trim_poly_g1:norm_methodnone          0.456680
trim_poly_g1:norm_methodRLE          0.938045
trim_poly_g1:norm_methodTMM          0.607914
trim_poly_g1:norm_methodTMMwsp       0.989325
trim_poly_g1:norm_methodupperquartile 0.312466
trim_poly_x1:norm_methodnone          0.000119 ***
trim_poly_x1:norm_methodRLE          2.30e-05 ***
trim_poly_x1:norm_methodTMM          3.71e-05 ***
trim_poly_x1:norm_methodTMMwsp       0.000120 ***
trim_poly_x1:norm_methodupperquartile 7.40e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for gaussian family taken to be 2.318872e-09)

```

Null deviance: 6.9874e+00 on 300 degrees of freedom
Residual deviance: 5.4957e-07 on 237 degrees of freedom
AIC: -5072.3

```

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: (2 not defined because of singularities)

	Estimate	Std. Error	t value
(Intercept)	3.918e-02	2.577e-04	152.067
alignerkallisto	3.337e-01	1.230e-04	2712.604
alignersalmon	3.366e-01	1.061e-04	3173.283
alignerSTAR	2.967e-04	8.204e-05	3.616
min_phred	2.714e-05	9.541e-06	2.845
min_length	4.527e-06	5.481e-06	0.826
trim_poly_g1	1.798e-05	8.249e-05	0.218
trim_poly_x1	-2.673e-04	8.583e-05	-3.114
norm_methodnone	1.782e-01	1.913e-04	931.629
norm_methodRLE	1.784e-01	2.012e-04	886.599
norm_methodTMM	1.783e-01	1.945e-04	916.609
norm_methodTMMwsp	1.783e-01	1.946e-04	916.425
norm_methodupperquartile	1.782e-01	2.027e-04	879.217
alignerkallisto:min_phred	-6.902e-06	3.346e-06	-2.063
alignersalmon:min_phred	-9.897e-06	3.321e-06	-2.980
alignerSTAR:min_phred	3.343e-06	2.255e-06	1.483

alignerkallisto:min_length	-1.810e-06	1.779e-06	-1.017
alignersalmon:min_length	-7.153e-06	1.568e-06	-4.561
alignerSTAR:min_length	-5.327e-06	1.231e-06	-4.328
alignerkallisto:trim_poly_g1	-6.916e-06	2.069e-05	-0.334
alignersalmon:trim_poly_g1	-9.928e-06	1.925e-05	-0.516
alignerSTAR:trim_poly_g1	5.839e-08	1.429e-05	0.004
alignerkallisto:trim_poly_x1	4.341e-04	2.045e-05	21.232
alignersalmon:trim_poly_x1	2.119e-04	2.031e-05	10.432
alignerSTAR:trim_poly_x1	-1.485e-05	1.435e-05	-1.035
alignerkallisto:norm_methodnone	2.299e-03	4.536e-05	50.676
alignersalmon:norm_methodnone	1.067e-03	2.862e-05	37.275
alignerSTAR:norm_methodnone	5.449e-05	2.146e-05	2.539
alignerkallisto:norm_methodRLE	1.784e-03	4.864e-05	36.676
alignersalmon:norm_methodRLE	3.079e-04	3.251e-05	9.471
alignerSTAR:norm_methodRLE	-4.501e-05	2.142e-05	-2.102
alignerkallisto:norm_methodTMM	2.029e-03	4.660e-05	43.528
alignersalmon:norm_methodTMM	4.510e-04	3.181e-05	14.176
alignerSTAR:norm_methodTMM	9.326e-05	2.059e-05	4.529
alignerkallisto:norm_methodTMMwsp	1.948e-03	4.735e-05	41.141
alignersalmon:norm_methodTMMwsp	3.949e-04	3.248e-05	12.158
alignerSTAR:norm_methodTMMwsp	-4.031e-05	2.567e-05	-1.570
alignerkallisto:norm_methodupperquartile	1.685e-03	4.345e-05	38.777
alignersalmon:norm_methodupperquartile	NA	NA	NA
alignerSTAR:norm_methodupperquartile	NA	NA	NA
min_phred:min_length	1.613e-07	1.794e-07	0.900
min_phred:trim_poly_g1	2.240e-06	2.085e-06	1.074
min_phred:trim_poly_x1	-6.087e-06	2.114e-06	-2.879
min_phred:norm_methodnone	-3.743e-05	7.284e-06	-5.139
min_phred:norm_methodRLE	-3.966e-05	7.530e-06	-5.268
min_phred:norm_methodTMM	-4.071e-05	7.374e-06	-5.522
min_phred:norm_methodTMMwsp	-3.983e-05	7.330e-06	-5.434
min_phred:norm_methodupperquartile	-3.682e-05	7.485e-06	-4.919
min_length:trim_poly_g1	-1.180e-06	1.178e-06	-1.002
min_length:trim_poly_x1	2.143e-06	1.163e-06	1.842
min_length:norm_methodnone	-4.891e-06	2.852e-06	-1.715
min_length:norm_methodRLE	-1.737e-06	3.049e-06	-0.570
min_length:norm_methodTMM	-1.442e-06	2.974e-06	-0.485
min_length:norm_methodTMMwsp	-1.840e-06	2.999e-06	-0.614
min_length:norm_methodupperquartile	-1.010e-06	3.010e-06	-0.335
trim_poly_g1:trim_poly_x1	2.939e-05	1.280e-05	2.296
trim_poly_g1:norm_methodnone	-2.816e-05	3.875e-05	-0.727
trim_poly_g1:norm_methodRLE	-2.792e-05	4.034e-05	-0.692
trim_poly_g1:norm_methodTMM	-2.844e-05	4.007e-05	-0.710
trim_poly_g1:norm_methodTMMwsp	-2.639e-05	4.074e-05	-0.648
trim_poly_g1:norm_methodupperquartile	-5.982e-05	3.953e-05	-1.513
trim_poly_x1:norm_methodnone	2.684e-04	4.587e-05	5.851
trim_poly_x1:norm_methodRLE	2.924e-04	4.602e-05	6.355
trim_poly_x1:norm_methodTMM	2.917e-04	4.628e-05	6.303
trim_poly_x1:norm_methodTMMwsp	2.724e-04	4.548e-05	5.989
trim_poly_x1:norm_methodupperquartile	3.120e-04	4.559e-05	6.845
Pr(> t )			
(Intercept)	< 2e-16 ***		
alignerkallisto	< 2e-16 ***		
alignersalmon	< 2e-16 ***		

alignerSTAR	0.000365	***
min_phred	0.004835	**
min_length	0.409628	
trim_poly_g1	0.827624	
trim_poly_x1	0.002074	**
norm_methodnone	< 2e-16	***
norm_methodRLE	< 2e-16	***
norm_methodTMM	< 2e-16	***
norm_methodTMMwsp	< 2e-16	***
norm_methodupperquartile	< 2e-16	***
alignerkallisto:min_phred	0.040236	*
alignersalmon:min_phred	0.003178	**
alignerSTAR:min_phred	0.139428	
alignerkallisto:min_length	0.309955	
alignersalmon:min_length	8.15e-06	***
alignerSTAR:min_length	2.22e-05	***
alignerkallisto:trim_poly_g1	0.738446	
alignersalmon:trim_poly_g1	0.606461	
alignerSTAR:trim_poly_g1	0.996742	
alignerkallisto:trim_poly_x1	< 2e-16	***
alignersalmon:trim_poly_x1	< 2e-16	***
alignerSTAR:trim_poly_x1	0.301738	
alignerkallisto:norm_methodnone	< 2e-16	***
alignersalmon:norm_methodnone	< 2e-16	***
alignerSTAR:norm_methodnone	0.011746	*
alignerkallisto:norm_methodRLE	< 2e-16	***
alignersalmon:norm_methodRLE	< 2e-16	***
alignerSTAR:norm_methodRLE	0.036641	*
alignerkallisto:norm_methodTMM	< 2e-16	***
alignersalmon:norm_methodTMM	< 2e-16	***
alignerSTAR:norm_methodTMM	9.38e-06	***
alignerkallisto:norm_methodTMMwsp	< 2e-16	***
alignersalmon:norm_methodTMMwsp	< 2e-16	***
alignerSTAR:norm_methodTMMwsp	0.117677	
alignerkallisto:norm_methodupperquartile	< 2e-16	***
alignersalmon:norm_methodupperquartile	NA	
alignerSTAR:norm_methodupperquartile	NA	
min_phred:min_length	0.369291	
min_phred:trim_poly_g1	0.283748	
min_phred:trim_poly_x1	0.004356	**
min_phred:norm_methodnone	5.78e-07	***
min_phred:norm_methodRLE	3.10e-07	***
min_phred:norm_methodTMM	8.81e-08	***
min_phred:norm_methodTMMwsp	1.37e-07	***
min_phred:norm_methodupperquartile	1.63e-06	***
min_length:trim_poly_g1	0.317576	
min_length:trim_poly_x1	0.066714	.
min_length:norm_methodnone	0.087639	.
min_length:norm_methodRLE	0.569404	
min_length:norm_methodTMM	0.628281	
min_length:norm_methodTMMwsp	0.540020	
min_length:norm_methodupperquartile	0.737584	
trim_poly_g1:trim_poly_x1	0.022568	*
trim_poly_g1:norm_methodnone	0.468152	



```

trim_poly_g1:norm_methodRLE          0.489497
trim_poly_g1:norm_methodTMM          0.478482
trim_poly_g1:norm_methodTMMwsp       0.517737
trim_poly_g1:norm_methodupperquartile 0.131489
trim_poly_x1:norm_methodnone          1.62e-08 ***
trim_poly_x1:norm_methodRLE          1.06e-09 ***
trim_poly_x1:norm_methodTMM          1.41e-09 ***
trim_poly_x1:norm_methodTMMwsp       7.78e-09 ***
trim_poly_x1:norm_methodupperquartile 6.51e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for quasi family taken to be 2.43689e-09)

```

Null deviance: 6.9874e+00 on 300 degrees of freedom
Residual deviance: 5.7754e-07 on 237 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: (2 not defined because of singularities)

	Estimate	Std. Error	t value
(Intercept)	1.177e+00	3.174e-03	370.900
alignerkallisto	4.235e-01	1.515e-03	279.494
alignersalmon	4.177e-01	1.306e-03	319.734
alignerSTAR	2.147e-04	1.011e-03	0.212
min_phred	5.000e-05	1.175e-04	0.425
min_length	3.519e-05	6.751e-05	0.521
trim_poly_g1	-3.726e-03	1.016e-03	-3.667
trim_poly_x1	1.156e-03	1.057e-03	1.093
norm_methodnone	-6.174e-01	2.356e-03	-262.031
norm_methodRLE	-6.148e-01	2.479e-03	-247.997
norm_methodTMM	-6.142e-01	2.396e-03	-256.317
norm_methodTMMwsp	-6.147e-01	2.397e-03	-256.465
norm_methodupperquartile	-6.140e-01	2.497e-03	-245.932
alignerkallisto:min_phred	-1.619e-07	4.122e-05	-0.004
alignersalmon:min_phred	-1.989e-05	4.090e-05	-0.486
alignerSTAR:min_phred	3.258e-05	2.777e-05	1.173
alignerkallisto:min_length	8.449e-05	2.191e-05	3.855
alignersalmon:min_length	3.359e-05	1.932e-05	1.739
alignerSTAR:min_length	-1.831e-05	1.516e-05	-1.208
alignerkallisto:trim_poly_g1	2.356e-04	2.548e-04	0.925

alignersalmon:trim_poly_g1	9.016e-05	2.371e-04	0.380
alignerSTAR:trim_poly_g1	-1.000e-04	1.760e-04	-0.568
alignerkallisto:trim_poly_x1	9.388e-05	2.519e-04	0.373
alignersalmon:trim_poly_x1	-2.600e-05	2.502e-04	-0.104
alignerSTAR:trim_poly_x1	-7.516e-04	1.768e-04	-4.251
alignerkallisto:norm_methodnone	-1.207e-02	5.587e-04	-21.594
alignersalmon:norm_methodnone	1.097e-02	3.525e-04	31.106
alignerSTAR:norm_methodnone	1.222e-03	2.643e-04	4.624
alignerkallisto:norm_methodRLE	-1.910e-02	5.992e-04	-31.876
alignersalmon:norm_methodRLE	2.970e-03	4.004e-04	7.416
alignerSTAR:norm_methodRLE	-1.518e-04	2.638e-04	-0.575
alignerkallisto:norm_methodTMM	-1.803e-02	5.741e-04	-31.400
alignersalmon:norm_methodTMM	4.514e-03	3.919e-04	11.519
alignerSTAR:norm_methodTMM	7.109e-04	2.536e-04	2.803
alignerkallisto:norm_methodTMMwsp	-1.788e-02	5.833e-04	-30.649
alignersalmon:norm_methodTMMwsp	5.164e-03	4.001e-04	12.906
alignerSTAR:norm_methodTMMwsp	6.243e-04	3.162e-04	1.974
alignerkallisto:norm_methodupperquartile	-2.288e-02	5.352e-04	-42.754
alignersalmon:norm_methodupperquartile	NA	NA	NA
alignerSTAR:norm_methodupperquartile	NA	NA	NA
min_phred:min_length	-3.753e-06	2.210e-06	-1.699
min_phred:trim_poly_g1	6.923e-05	2.568e-05	2.696
min_phred:trim_poly_x1	2.049e-05	2.605e-05	0.787
min_phred:norm_methodnone	8.511e-05	8.973e-05	0.949
min_phred:norm_methodRLE	1.161e-04	9.275e-05	1.251
min_phred:norm_methodTMM	1.238e-04	9.083e-05	1.363
min_phred:norm_methodTMMwsp	7.548e-05	9.030e-05	0.836
min_phred:norm_methodupperquartile	1.369e-04	9.220e-05	1.485
min_length:trim_poly_g1	2.743e-05	1.451e-05	1.890
min_length:trim_poly_x1	-2.866e-05	1.433e-05	-2.000
min_length:norm_methodnone	5.675e-05	3.513e-05	1.615
min_length:norm_methodRLE	6.846e-05	3.755e-05	1.823
min_length:norm_methodTMM	4.433e-05	3.664e-05	1.210
min_length:norm_methodTMMwsp	6.641e-05	3.694e-05	1.798
min_length:norm_methodupperquartile	5.390e-05	3.708e-05	1.453
trim_poly_g1:trim_poly_x1	2.004e-05	1.577e-04	0.127
trim_poly_g1:norm_methodnone	8.316e-04	4.774e-04	1.742
trim_poly_g1:norm_methodRLE	8.856e-04	4.969e-04	1.782
trim_poly_g1:norm_methodTMM	7.182e-04	4.936e-04	1.455
trim_poly_g1:norm_methodTMMwsp	8.575e-04	5.019e-04	1.709
trim_poly_g1:norm_methodupperquartile	1.013e-03	4.869e-04	2.081
trim_poly_x1:norm_methodnone	-5.442e-04	5.651e-04	-0.963
trim_poly_x1:norm_methodRLE	-3.722e-04	5.669e-04	-0.657
trim_poly_x1:norm_methodTMM	-3.921e-04	5.701e-04	-0.688
trim_poly_x1:norm_methodTMMwsp	-9.051e-05	5.602e-04	-0.162
trim_poly_x1:norm_methodupperquartile	-2.997e-04	5.616e-04	-0.534
Pr(> t )			
(Intercept)	< 2e-16 ***		
alignerkallisto	< 2e-16 ***		
alignersalmon	< 2e-16 ***		
alignerSTAR	0.831920		
min_phred	0.670921		
min_length	0.602697		
trim_poly_g1	0.000303 ***		

trim_poly_x1	0.275405
norm_methodnone	< 2e-16 ***
norm_methodRLE	< 2e-16 ***
norm_methodTMM	< 2e-16 ***
norm_methodTMMwsp	< 2e-16 ***
norm_methodupperquartile	< 2e-16 ***
alignerkallisto:min_phred	0.996868
alignersalmon:min_phred	0.627266
alignerSTAR:min_phred	0.241959
alignerkallisto:min_length	0.000149 ***
alignersalmon:min_length	0.083354 .
alignerSTAR:min_length	0.228228
alignerkallisto:trim_poly_g1	0.356074
alignersalmon:trim_poly_g1	0.704084
alignerSTAR:trim_poly_g1	0.570294
alignerkallisto:trim_poly_x1	0.709654
alignersalmon:trim_poly_x1	0.917304
alignerSTAR:trim_poly_x1	3.06e-05 ***
alignerkallisto:norm_methodnone	< 2e-16 ***
alignersalmon:norm_methodnone	< 2e-16 ***
alignerSTAR:norm_methodnone	6.19e-06 ***
alignerkallisto:norm_methodRLE	< 2e-16 ***
alignersalmon:norm_methodRLE	2.14e-12 ***
alignerSTAR:norm_methodRLE	0.565618
alignerkallisto:norm_methodTMM	< 2e-16 ***
alignersalmon:norm_methodTMM	< 2e-16 ***
alignerSTAR:norm_methodTMM	0.005487 **
alignerkallisto:norm_methodTMMwsp	< 2e-16 ***
alignersalmon:norm_methodTMMwsp	< 2e-16 ***
alignerSTAR:norm_methodTMMwsp	0.049526 *
alignerkallisto:norm_methodupperquartile	< 2e-16 ***
alignersalmon:norm_methodupperquartile	NA
alignerSTAR:norm_methodupperquartile	NA
min_phred:min_length	0.090724 .
min_phred:trim_poly_g1	0.007516 **
min_phred:trim_poly_x1	0.432200
min_phred:norm_methodnone	0.343821
min_phred:norm_methodRLE	0.212028
min_phred:norm_methodTMM	0.174287
min_phred:norm_methodTMMwsp	0.404034
min_phred:norm_methodupperquartile	0.138840
min_length:trim_poly_g1	0.059959 .
min_length:trim_poly_x1	0.046595 *
min_length:norm_methodnone	0.107584
min_length:norm_methodRLE	0.069550 .
min_length:norm_methodTMM	0.227475
min_length:norm_methodTMMwsp	0.073518 .
min_length:norm_methodupperquartile	0.147411
trim_poly_g1:trim_poly_x1	0.898980
trim_poly_g1:norm_methodnone	0.082792 .
trim_poly_g1:norm_methodRLE	0.075991 .
trim_poly_g1:norm_methodTMM	0.146954
trim_poly_g1:norm_methodTMMwsp	0.088834 .
trim_poly_g1:norm_methodupperquartile	0.038472 *

```

trim_poly_x1:norm_methodnone          0.336507
trim_poly_x1:norm_methodRLE           0.512069
trim_poly_x1:norm_methodTMM           0.492192
trim_poly_x1:norm_methodTMMwsp        0.871795
trim_poly_x1:norm_methodupperquartile 0.594029

```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 3.697764e-07)

```

Null deviance: 2.0106e+01 on 300 degrees of freedom
Residual deviance: 8.7637e-05 on 237 degrees of freedom
AIC: -3545.7

```

Number of Fisher Scoring iterations: 2

```

# 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: (2 not defined because of singularities)

	Estimate	Std. Error	t value
(Intercept)	-8.570e-02	3.504e-03	-24.454
alignerkallisto	5.551e-01	1.795e-03	309.279
alignersalmon	5.518e-01	1.658e-03	332.889
alignerSTAR	3.797e-04	1.753e-03	0.217
min_phred	1.285e-04	1.247e-04	1.031
min_length	7.446e-05	8.124e-05	0.917
trim_poly_g1	-4.972e-03	1.196e-03	-4.157
trim_poly_x1	3.609e-04	1.306e-03	0.276
norm_methodnone	-4.936e-01	2.025e-03	-243.722
norm_methodRLE	-4.883e-01	2.369e-03	-206.116
norm_methodTMM	-4.881e-01	2.221e-03	-219.744
norm_methodTMMwsp	-4.891e-01	2.112e-03	-231.565
norm_methodupperquartile	-4.869e-01	2.233e-03	-218.009
alignerkallisto:min_phred	-2.968e-05	5.060e-05	-0.587
alignersalmon:min_phred	-5.799e-05	4.931e-05	-1.176
alignerSTAR:min_phred	5.745e-05	4.819e-05	1.192
alignerkallisto:min_length	7.431e-05	2.677e-05	2.776
alignersalmon:min_length	4.569e-05	2.473e-05	1.847
alignerSTAR:min_length	-3.233e-05	2.631e-05	-1.229
alignerkallisto:trim_poly_g1	3.549e-04	3.120e-04	1.138
alignersalmon:trim_poly_g1	7.568e-05	2.975e-04	0.254
alignerSTAR:trim_poly_g1	-1.753e-04	3.052e-04	-0.574
alignerkallisto:trim_poly_x1	3.948e-05	3.162e-04	0.125
alignersalmon:trim_poly_x1	-1.191e-04	3.052e-04	-0.390
alignerSTAR:trim_poly_x1	-1.326e-03	3.066e-04	-4.326
alignerkallisto:norm_methodnone	-5.320e-03	5.326e-04	-9.988

alignersalmon:norm_methodnone	1.481e-02	4.551e-04	32.543
alignerSTAR:norm_methodnone	2.167e-03	4.591e-04	4.720
alignerkallisto:norm_methodRLE	-1.561e-02	5.925e-04	-26.346
alignersalmon:norm_methodRLE	3.654e-03	4.969e-04	7.353
alignerSTAR:norm_methodRLE	-2.678e-04	4.568e-04	-0.586
alignerkallisto:norm_methodTMM	-1.414e-02	5.437e-04	-26.002
alignersalmon:norm_methodTMM	5.546e-03	4.787e-04	11.586
alignerSTAR:norm_methodTMM	1.254e-03	4.392e-04	2.855
alignerkallisto:norm_methodTMMwsp	-1.356e-02	5.863e-04	-23.134
alignersalmon:norm_methodTMMwsp	6.496e-03	5.234e-04	12.410
alignerSTAR:norm_methodTMMwsp	1.102e-03	5.480e-04	2.011
alignerkallisto:norm_methodupperquartile	-2.009e-02	4.226e-04	-47.542
alignersalmon:norm_methodupperquartile	NA	NA	NA
alignerSTAR:norm_methodupperquartile	NA	NA	NA
min_phred:min_length	-5.961e-06	2.929e-06	-2.035
min_phred:trim_poly_g1	8.891e-05	3.413e-05	2.605
min_phred:trim_poly_x1	4.203e-05	3.637e-05	1.156
min_phred:norm_methodnone	1.044e-04	7.098e-05	1.471
min_phred:norm_methodRLE	1.375e-04	8.028e-05	1.713
min_phred:norm_methodTMM	1.770e-04	7.551e-05	2.345
min_phred:norm_methodTMMwsp	8.679e-05	7.155e-05	1.213
min_phred:norm_methodupperquartile	1.798e-04	7.807e-05	2.303
min_length:trim_poly_g1	5.466e-05	1.796e-05	3.044
min_length:trim_poly_x1	-2.946e-05	1.968e-05	-1.497
min_length:norm_methodnone	5.500e-05	2.714e-05	2.026
min_length:norm_methodRLE	7.388e-05	3.363e-05	2.197
min_length:norm_methodTMM	3.164e-05	3.205e-05	0.987
min_length:norm_methodTMMwsp	8.125e-05	3.135e-05	2.592
min_length:norm_methodupperquartile	4.361e-05	3.096e-05	1.409
trim_poly_g1:trim_poly_x1	4.004e-05	2.111e-04	0.190
trim_poly_g1:norm_methodnone	5.309e-04	3.588e-04	1.480
trim_poly_g1:norm_methodRLE	5.187e-04	4.201e-04	1.235
trim_poly_g1:norm_methodTMM	2.137e-04	4.109e-04	0.520
trim_poly_g1:norm_methodTMMwsp	3.349e-04	4.141e-04	0.809
trim_poly_g1:norm_methodupperquartile	7.949e-04	3.916e-04	2.030
trim_poly_x1:norm_methodnone	-2.541e-04	4.223e-04	-0.602
trim_poly_x1:norm_methodRLE	-3.109e-06	4.435e-04	-0.007
trim_poly_x1:norm_methodTMM	7.559e-05	4.582e-04	0.165
trim_poly_x1:norm_methodTMMwsp	3.563e-04	4.181e-04	0.852
trim_poly_x1:norm_methodupperquartile	1.248e-04	4.310e-04	0.289
Pr(> t )			
(Intercept)	< 2e-16 ***		
alignerkallisto	< 2e-16 ***		
alignersalmon	< 2e-16 ***		
alignerSTAR	0.82877		
min_phred	0.30378		
min_length	0.36032		
trim_poly_g1	4.50e-05 ***		
trim_poly_x1	0.78260		
norm_methodnone	< 2e-16 ***		
norm_methodRLE	< 2e-16 ***		
norm_methodTMM	< 2e-16 ***		
norm_methodTMMwsp	< 2e-16 ***		
norm_methodupperquartile	< 2e-16 ***		

alignerkallisto:min_phred	0.55806
alignersalmon:min_phred	0.24072
alignerSTAR:min_phred	0.23437
alignerkallisto:min_length	0.00595 **
alignersalmon:min_length	0.06596 .
alignerSTAR:min_length	0.22038
alignerkallisto:trim_poly_g1	0.25641
alignersalmon:trim_poly_g1	0.79937
alignerSTAR:trim_poly_g1	0.56623
alignerkallisto:trim_poly_x1	0.90075
alignersalmon:trim_poly_x1	0.69672
alignerSTAR:trim_poly_x1	2.24e-05 ***
alignerkallisto:norm_methodnone	< 2e-16 ***
alignersalmon:norm_methodnone	< 2e-16 ***
alignerSTAR:norm_methodnone	4.03e-06 ***
alignerkallisto:norm_methodRLE	< 2e-16 ***
alignersalmon:norm_methodRLE	3.14e-12 ***
alignerSTAR:norm_methodRLE	0.55824
alignerkallisto:norm_methodTMM	< 2e-16 ***
alignersalmon:norm_methodTMM	< 2e-16 ***
alignerSTAR:norm_methodTMM	0.00469 **
alignerkallisto:norm_methodTMMwsp	< 2e-16 ***
alignersalmon:norm_methodTMMwsp	< 2e-16 ***
alignerSTAR:norm_methodTMMwsp	0.04547 *
alignerkallisto:norm_methodupperquartile	< 2e-16 ***
alignersalmon:norm_methodupperquartile	NA
alignerSTAR:norm_methodupperquartile	NA
min_phred:min_length	0.04295 *
min_phred:trim_poly_g1	0.00977 **
min_phred:trim_poly_x1	0.24898
min_phred:norm_methodnone	0.14265
min_phred:norm_methodRLE	0.08802 .
min_phred:norm_methodTMM	0.01987 *
min_phred:norm_methodTMMwsp	0.22634
min_phred:norm_methodupperquartile	0.02214 *
min_length:trim_poly_g1	0.00260 **
min_length:trim_poly_x1	0.13571
min_length:norm_methodnone	0.04385 *
min_length:norm_methodRLE	0.02899 *
min_length:norm_methodTMM	0.32462
min_length:norm_methodTMMwsp	0.01014 *
min_length:norm_methodupperquartile	0.16028
trim_poly_g1:trim_poly_x1	0.84970
trim_poly_g1:norm_methodnone	0.14028
trim_poly_g1:norm_methodRLE	0.21813
trim_poly_g1:norm_methodTMM	0.60355
trim_poly_g1:norm_methodTMMwsp	0.41952
trim_poly_g1:norm_methodupperquartile	0.04350 *
trim_poly_x1:norm_methodnone	0.54792
trim_poly_x1:norm_methodRLE	0.99441
trim_poly_x1:norm_methodTMM	0.86910
trim_poly_x1:norm_methodTMMwsp	0.39501
trim_poly_x1:norm_methodupperquartile	0.77248

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 3.568389e-07)

Null deviance: 2.0106e+01 on 300 degrees of freedom  
Residual deviance: 8.4571e-05 on 237 degrees of freedom  
AIC: -3556.4

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: (2 not defined because of singularities)

	Estimate	Std. Error	t value
(Intercept)	1.177e+00	3.174e-03	370.900
alignerkallisto	4.235e-01	1.515e-03	279.494
alignersalmon	4.177e-01	1.306e-03	319.734
alignerSTAR	2.147e-04	1.011e-03	0.212
min_phred	5.000e-05	1.175e-04	0.425
min_length	3.519e-05	6.751e-05	0.521
trim_poly_g1	-3.726e-03	1.016e-03	-3.667
trim_poly_x1	1.156e-03	1.057e-03	1.093
norm_methodnone	-6.174e-01	2.356e-03	-262.031
norm_methodRLE	-6.148e-01	2.479e-03	-247.997
norm_methodTMM	-6.142e-01	2.396e-03	-256.317
norm_methodTMMwsp	-6.147e-01	2.397e-03	-256.465
norm_methodupperquartile	-6.140e-01	2.497e-03	-245.932
alignerkallisto:min_phred	-1.619e-07	4.122e-05	-0.004
alignersalmon:min_phred	-1.989e-05	4.090e-05	-0.486
alignerSTAR:min_phred	3.258e-05	2.777e-05	1.173
alignerkallisto:min_length	8.449e-05	2.191e-05	3.855
alignersalmon:min_length	3.359e-05	1.932e-05	1.739
alignerSTAR:min_length	-1.831e-05	1.516e-05	-1.208
alignerkallisto:trim_poly_g1	2.356e-04	2.548e-04	0.925
alignersalmon:trim_poly_g1	9.016e-05	2.371e-04	0.380
alignerSTAR:trim_poly_g1	-1.000e-04	1.760e-04	-0.568
alignerkallisto:trim_poly_x1	9.388e-05	2.519e-04	0.373
alignersalmon:trim_poly_x1	-2.600e-05	2.502e-04	-0.104
alignerSTAR:trim_poly_x1	-7.516e-04	1.768e-04	-4.251
alignerkallisto:norm_methodnone	-1.207e-02	5.587e-04	-21.594
alignersalmon:norm_methodnone	1.097e-02	3.525e-04	31.106
alignerSTAR:norm_methodnone	1.222e-03	2.643e-04	4.624
alignerkallisto:norm_methodRLE	-1.910e-02	5.992e-04	-31.876
alignersalmon:norm_methodRLE	2.970e-03	4.004e-04	7.416
alignerSTAR:norm_methodRLE	-1.518e-04	2.638e-04	-0.575
alignerkallisto:norm_methodTMM	-1.803e-02	5.741e-04	-31.400
alignersalmon:norm_methodTMM	4.514e-03	3.919e-04	11.519

alignerSTAR:norm_methodTMM	7.109e-04	2.536e-04	2.803
alignerkallisto:norm_methodTMMwsp	-1.788e-02	5.833e-04	-30.649
alignersalmon:norm_methodTMMwsp	5.164e-03	4.001e-04	12.906
alignerSTAR:norm_methodTMMwsp	6.243e-04	3.162e-04	1.974
alignerkallisto:norm_methodupperquartile	-2.288e-02	5.352e-04	-42.754
alignersalmon:norm_methodupperquartile	NA	NA	NA
alignerSTAR:norm_methodupperquartile	NA	NA	NA
min_phred:min_length	-3.753e-06	2.210e-06	-1.699
min_phred:trim_poly_g1	6.923e-05	2.568e-05	2.696
min_phred:trim_poly_x1	2.049e-05	2.605e-05	0.787
min_phred:norm_methodnone	8.511e-05	8.973e-05	0.949
min_phred:norm_methodRLE	1.161e-04	9.275e-05	1.251
min_phred:norm_methodTMM	1.238e-04	9.083e-05	1.363
min_phred:norm_methodTMMwsp	7.548e-05	9.030e-05	0.836
min_phred:norm_methodupperquartile	1.369e-04	9.220e-05	1.485
min_length:trim_poly_g1	2.743e-05	1.451e-05	1.890
min_length:trim_poly_x1	-2.866e-05	1.433e-05	-2.000
min_length:norm_methodnone	5.675e-05	3.513e-05	1.615
min_length:norm_methodRLE	6.846e-05	3.755e-05	1.823
min_length:norm_methodTMM	4.433e-05	3.664e-05	1.210
min_length:norm_methodTMMwsp	6.641e-05	3.694e-05	1.798
min_length:norm_methodupperquartile	5.390e-05	3.708e-05	1.453
trim_poly_g1:trim_poly_x1	2.004e-05	1.577e-04	0.127
trim_poly_g1:norm_methodnone	8.316e-04	4.774e-04	1.742
trim_poly_g1:norm_methodRLE	8.856e-04	4.969e-04	1.782
trim_poly_g1:norm_methodTMM	7.182e-04	4.936e-04	1.455
trim_poly_g1:norm_methodTMMwsp	8.575e-04	5.019e-04	1.709
trim_poly_g1:norm_methodupperquartile	1.013e-03	4.869e-04	2.081
trim_poly_x1:norm_methodnone	-5.442e-04	5.651e-04	-0.963
trim_poly_x1:norm_methodRLE	-3.722e-04	5.669e-04	-0.657
trim_poly_x1:norm_methodTMM	-3.921e-04	5.701e-04	-0.688
trim_poly_x1:norm_methodTMMwsp	-9.051e-05	5.602e-04	-0.162
trim_poly_x1:norm_methodupperquartile	-2.997e-04	5.616e-04	-0.534
Pr(> t )			
(Intercept)	< 2e-16 ***		
alignerkallisto	< 2e-16 ***		
alignersalmon	< 2e-16 ***		
alignerSTAR	0.831920		
min_phred	0.670921		
min_length	0.602697		
trim_poly_g1	0.000303 ***		
trim_poly_x1	0.275405		
norm_methodnone	< 2e-16 ***		
norm_methodRLE	< 2e-16 ***		
norm_methodTMM	< 2e-16 ***		
norm_methodTMMwsp	< 2e-16 ***		
norm_methodupperquartile	< 2e-16 ***		
alignerkallisto:min_phred	0.996868		
alignersalmon:min_phred	0.627266		
alignerSTAR:min_phred	0.241959		
alignerkallisto:min_length	0.000149 ***		
alignersalmon:min_length	0.083354 .		
alignerSTAR:min_length	0.228228		
alignerkallisto:trim_poly_g1	0.356074		



alignersalmon:trim_poly_g1	0.704084
alignerSTAR:trim_poly_g1	0.570294
alignerkallisto:trim_poly_x1	0.709654
alignersalmon:trim_poly_x1	0.917304
alignerSTAR:trim_poly_x1	3.06e-05 ***
alignerkallisto:norm_methodnone	< 2e-16 ***
alignersalmon:norm_methodnone	< 2e-16 ***
alignerSTAR:norm_methodnone	6.19e-06 ***
alignerkallisto:norm_methodRLE	< 2e-16 ***
alignersalmon:norm_methodRLE	2.14e-12 ***
alignerSTAR:norm_methodRLE	0.565618
alignerkallisto:norm_methodTMM	< 2e-16 ***
alignersalmon:norm_methodTMM	< 2e-16 ***
alignerSTAR:norm_methodTMM	0.005487 **
alignerkallisto:norm_methodTMMwsp	< 2e-16 ***
alignersalmon:norm_methodTMMwsp	< 2e-16 ***
alignerSTAR:norm_methodTMMwsp	0.049526 *
alignerkallisto:norm_methodupperquartile	< 2e-16 ***
alignersalmon:norm_methodupperquartile	NA
alignerSTAR:norm_methodupperquartile	NA
min_phred:min_length	0.090724 .
min_phred:trim_poly_g1	0.007516 **
min_phred:trim_poly_x1	0.432200
min_phred:norm_methodnone	0.343821
min_phred:norm_methodRLE	0.212028
min_phred:norm_methodTMM	0.174287
min_phred:norm_methodTMMwsp	0.404034
min_phred:norm_methodupperquartile	0.138840
min_length:trim_poly_g1	0.059959 .
min_length:trim_poly_x1	0.046595 *
min_length:norm_methodnone	0.107584
min_length:norm_methodRLE	0.069550 .
min_length:norm_methodTMM	0.227475
min_length:norm_methodTMMwsp	0.073518 .
min_length:norm_methodupperquartile	0.147411
trim_poly_g1:trim_poly_x1	0.898980
trim_poly_g1:norm_methodnone	0.082792 .
trim_poly_g1:norm_methodRLE	0.075991 .
trim_poly_g1:norm_methodTMM	0.146954
trim_poly_g1:norm_methodTMMwsp	0.088834 .
trim_poly_g1:norm_methodupperquartile	0.038472 *
trim_poly_x1:norm_methodnone	0.336507
trim_poly_x1:norm_methodRLE	0.512069
trim_poly_x1:norm_methodTMM	0.492192
trim_poly_x1:norm_methodTMMwsp	0.871795
trim_poly_x1:norm_methodupperquartile	0.594029

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasi family taken to be 3.697764e-07)

Null deviance: 2.0106e+01 on 300 degrees of freedom  
Residual deviance: 8.7637e-05 on 237 degrees of freedom  
AIC: NA

Number of Fisher Scoring iterations: 2

## Bayesian

We know that

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

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

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

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 are 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:

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

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.

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

$a > 0$  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  $Y^2 \sim \mathcal{N}$ .