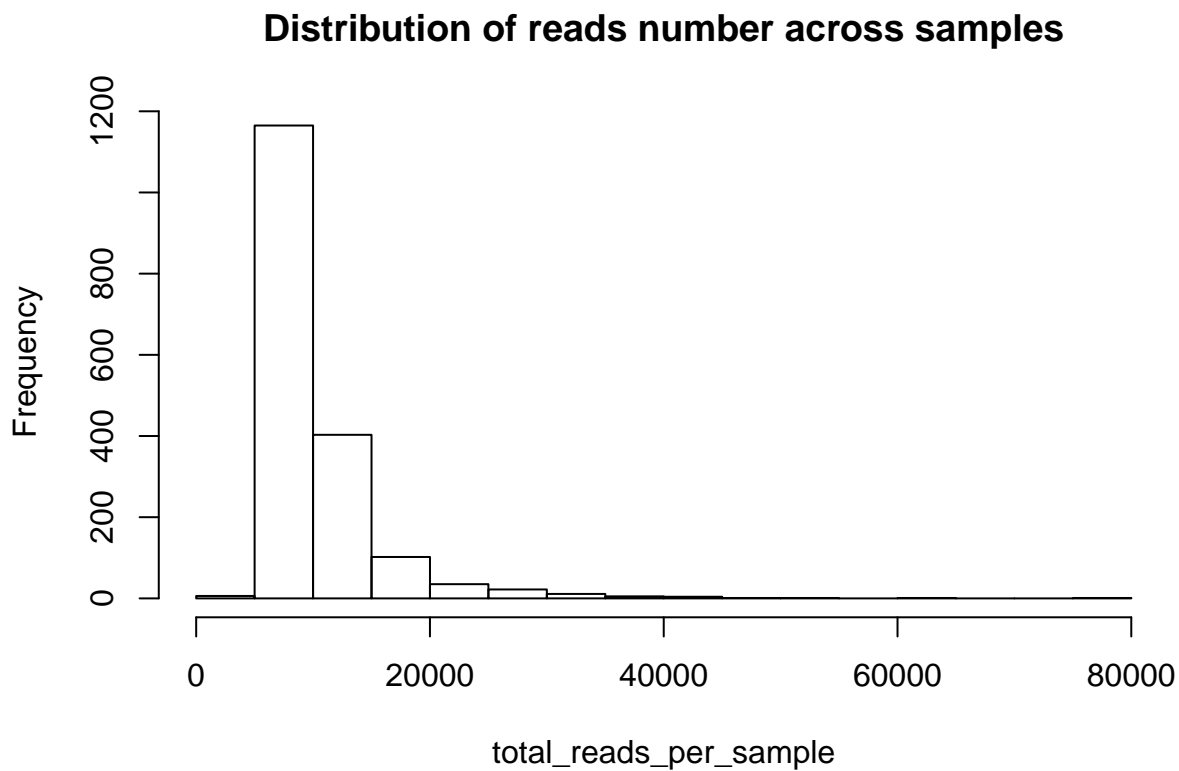
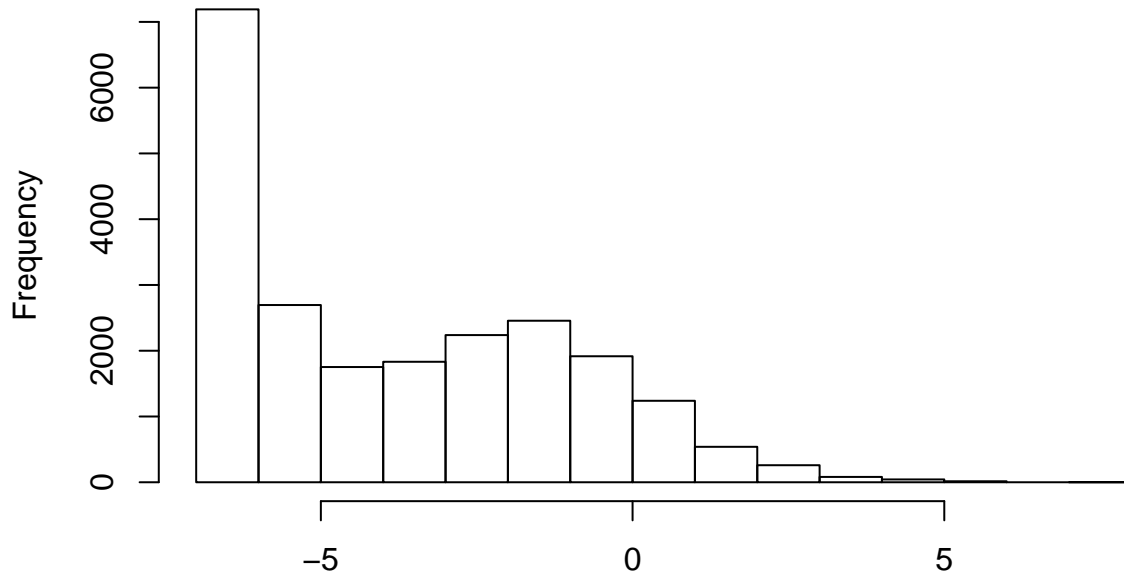


Power calculation

```
knitr::opts_chunk$set(comment = NA, echo=FALSE, message = FALSE, warnings = FALSE)
library(data.table)
library(MASS)
library(ggplot2)
```



Distribution of mean reads per gene across samples



$\log(\text{total_reads_per_gene_across_samples} + 0.01, 2)$

0% 25% 50% 75% 100%
5.691520e-04 2.276608e-03 4.097894e-02 3.164485e-01 1.530324e+02

Need to come up a function to random sample 20000 reads for each cell with more than 20000 reads.

| 0% | 1% | 2% | 3% | 4% |
|------------|------------|------------|------------|------------|
| 0.00000000 | 0.00000000 | 0.00000000 | 0.00000000 | 0.00000000 |
| 5% | 6% | 7% | 8% | 9% |
| 0.00000000 | 0.00000000 | 0.00000000 | 0.00000000 | 0.00000000 |
| 10% | 11% | 12% | 13% | 14% |
| 0.00000000 | 0.00000000 | 0.00000000 | 0.00000000 | 0.00000000 |
| 15% | 16% | 17% | 18% | 19% |
| 0.00000000 | 0.00000000 | 0.00000000 | 0.00000000 | 0.00000000 |
| 20% | 21% | 22% | 23% | 24% |
| 0.00000000 | 0.00000000 | 0.00000000 | 0.00000000 | 0.00000000 |
| 25% | 26% | 27% | 28% | 29% |
| 0.00000000 | 0.00000000 | 0.00000000 | 0.01234568 | 0.01234568 |
| 30% | 31% | 32% | 33% | 34% |
| 0.01234568 | 0.01234568 | 0.01234568 | 0.01234568 | 0.01234568 |
| 35% | 36% | 37% | 38% | 39% |
| 0.01234568 | 0.01234568 | 0.01234568 | 0.02469136 | 0.02469136 |
| 40% | 41% | 42% | 43% | 44% |
| 0.02469136 | 0.02469136 | 0.02469136 | 0.03703704 | 0.03703704 |
| 45% | 46% | 47% | 48% | 49% |
| 0.03703704 | 0.04938272 | 0.04938272 | 0.06172840 | 0.06172840 |
| 50% | 51% | 52% | 53% | 54% |
| 0.07407407 | 0.08641975 | 0.09876543 | 0.09876543 | 0.11111111 |
| 55% | 56% | 57% | 58% | 59% |
| 0.12345679 | 0.13580247 | 0.14814815 | 0.17283951 | 0.18518519 |
| 60% | 61% | 62% | 63% | 64% |

| | | | | |
|------------|------------|------------|------------|-------------|
| 0.19753086 | 0.22222222 | 0.23456790 | 0.25925926 | 0.28395062 |
| 65% | 66% | 67% | 68% | 69% |
| 0.30864198 | 0.32098765 | 0.34567901 | 0.37037037 | 0.39506173 |
| 70% | 71% | 72% | 73% | 74% |
| 0.43209877 | 0.45679012 | 0.49382716 | 0.51851852 | 0.55555556 |
| 75% | 76% | 77% | 78% | 79% |
| 0.60493827 | 0.64197531 | 0.69135802 | 0.72839506 | 0.79012346 |
| 80% | 81% | 82% | 83% | 84% |
| 0.85185185 | 0.91358025 | 0.97530864 | 1.04938272 | 1.13580247 |
| 85% | 86% | 87% | 88% | 89% |
| 1.22222222 | 1.32098765 | 1.44444444 | 1.58024691 | 1.72839506 |
| 90% | 91% | 92% | 93% | 94% |
| 1.88888889 | 2.11111111 | 2.33333333 | 2.62962963 | 2.96296296 |
| 95% | 96% | 97% | 98% | 99% |
| 3.40740741 | 4.08641975 | 5.18913580 | 7.28246914 | 12.22222222 |
| 100% | | | | |

192.74074074

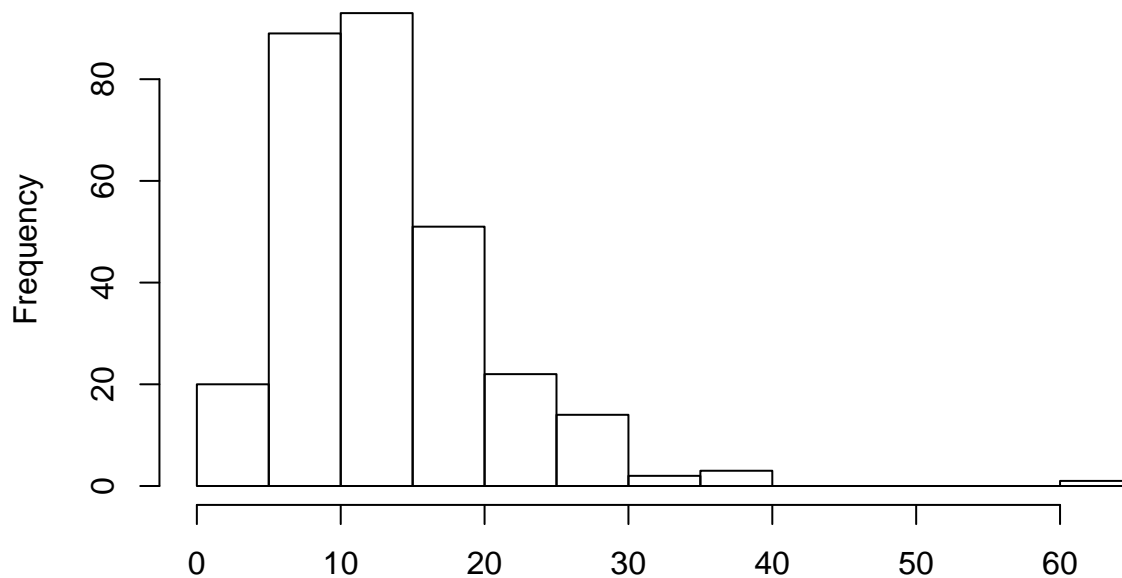
[1] NA

[1] 3.550555

[1] 10.16409

[1] 7.194391

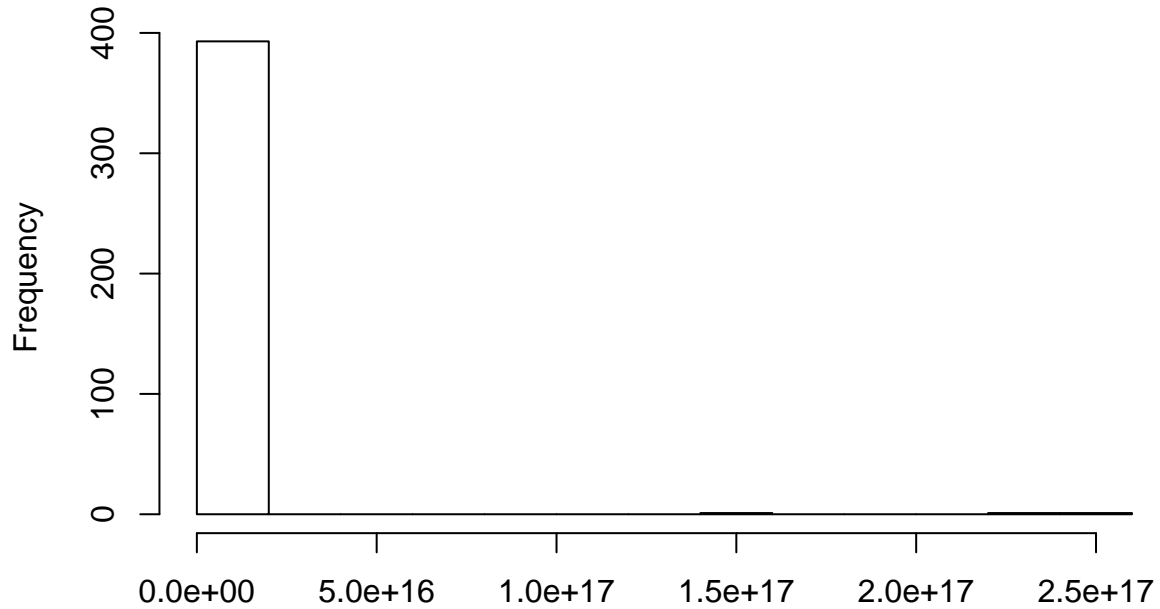
Genes with mean reads greater than 10



```
apply(dispersion_calculation, genes_above10, MoreArgs = list(count_matrix = t, genes =
```

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|--------|---------|---------|---------|---------|---------|
| 0.8323 | 8.7606 | 12.2035 | 13.2920 | 16.2635 | 64.5893 |

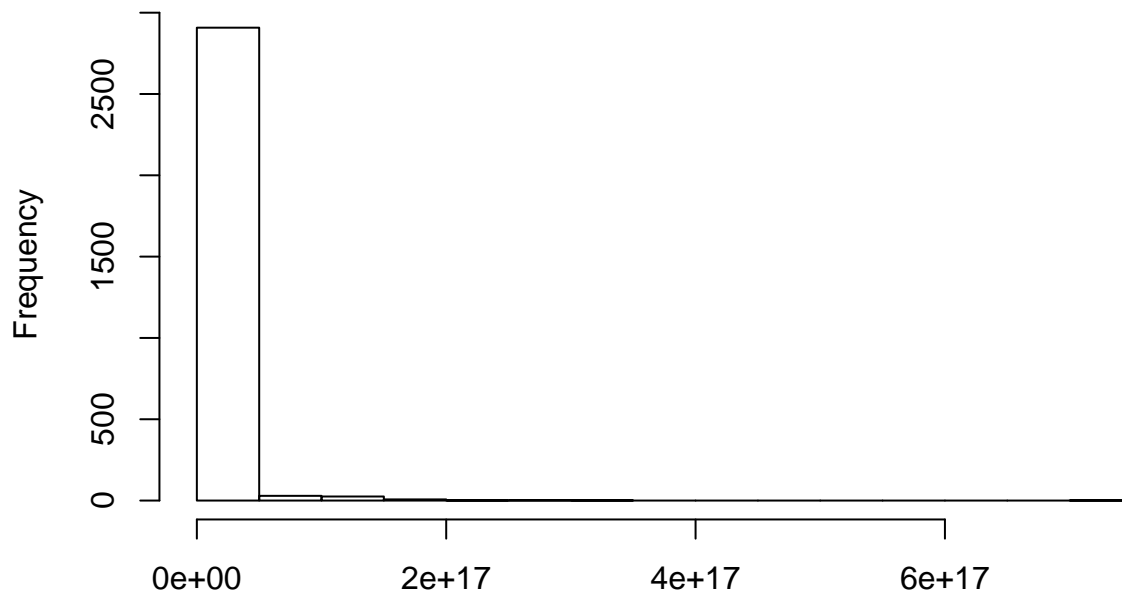
Genes with mean reads between 5 to 10



```
apply(dispersion_calculation, genes_5_to_10, MoreArgs = list(count_matrix = t, genes =
```

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | NA's |
|-----------|-----------|-----------|-----------|-----------|-----------|------|
| 0.000e+00 | 7.000e+00 | 1.000e+01 | 1.552e+15 | 1.500e+01 | 2.443e+17 | 1 |

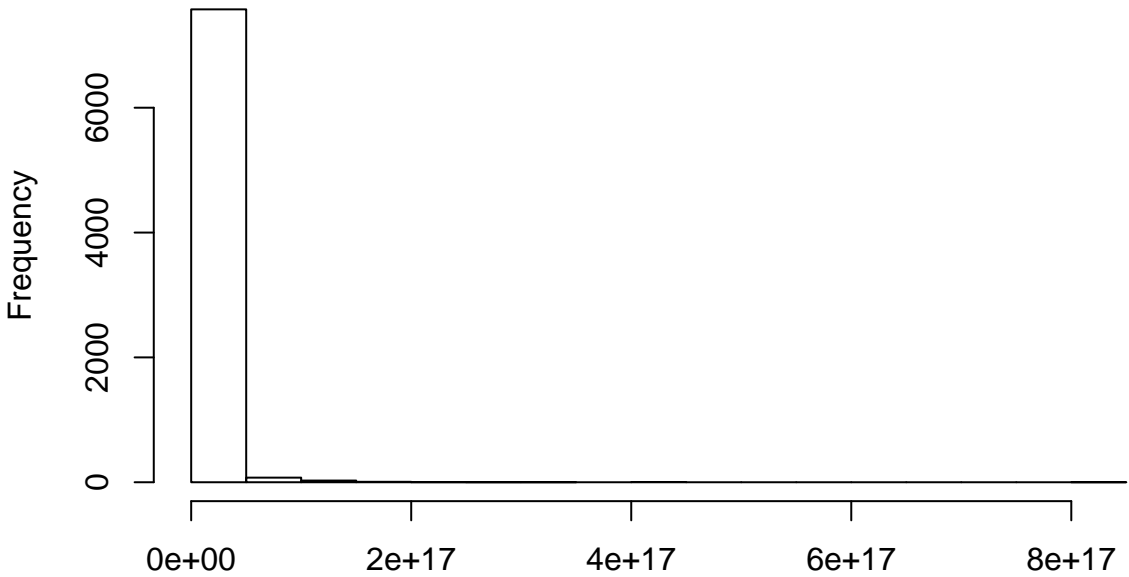
Genes with mean reads between 1 to 5



```
apply(dispersion_calculation, genes_1_to_5, MoreArgs = list(count_matrix = t, genes = c
```

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | NA's |
|-----------|-----------|-----------|-----------|-----------|-----------|------|
| 0.000e+00 | 4.000e+00 | 7.000e+00 | 2.868e+15 | 1.400e+01 | 7.113e+17 | 285 |

Genes with mean reads smaller than 1

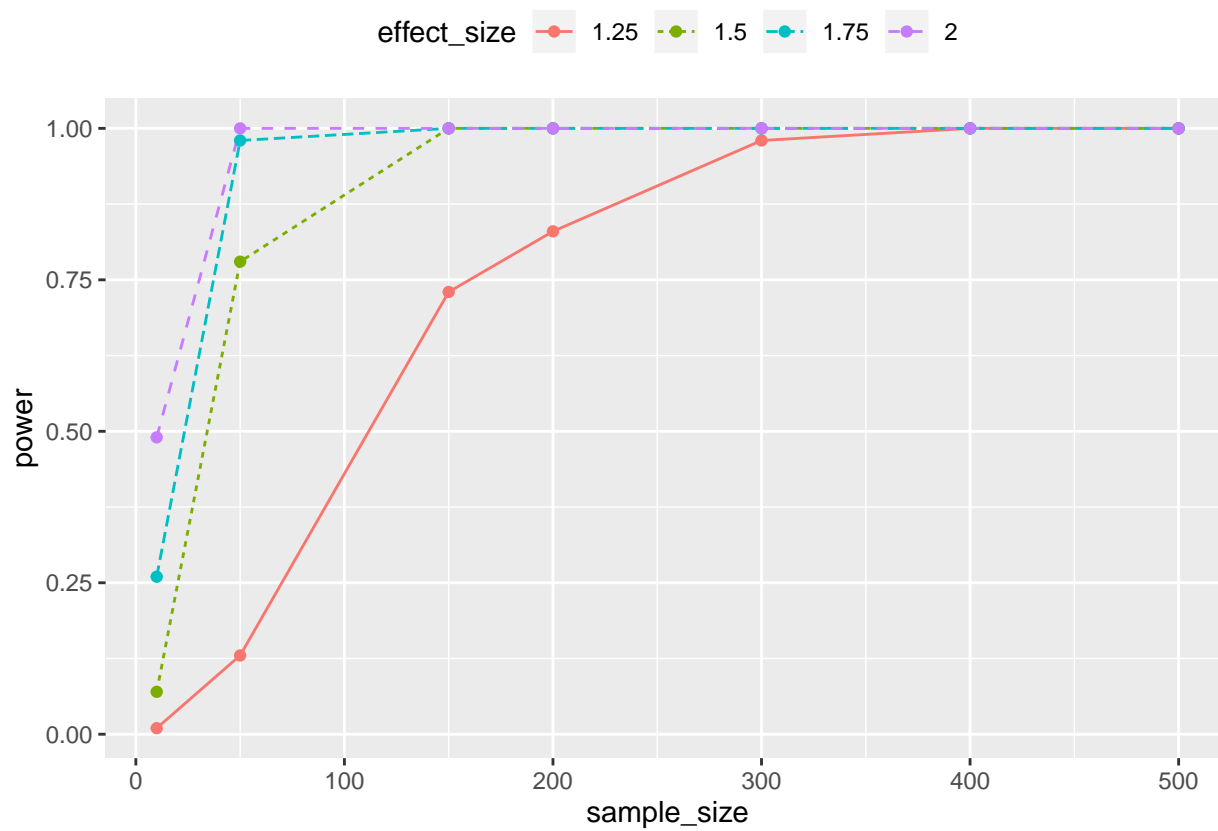


`apply(dispersion_calculation, genes_below1, MoreArgs = list(count_matrix = t, genes = c`

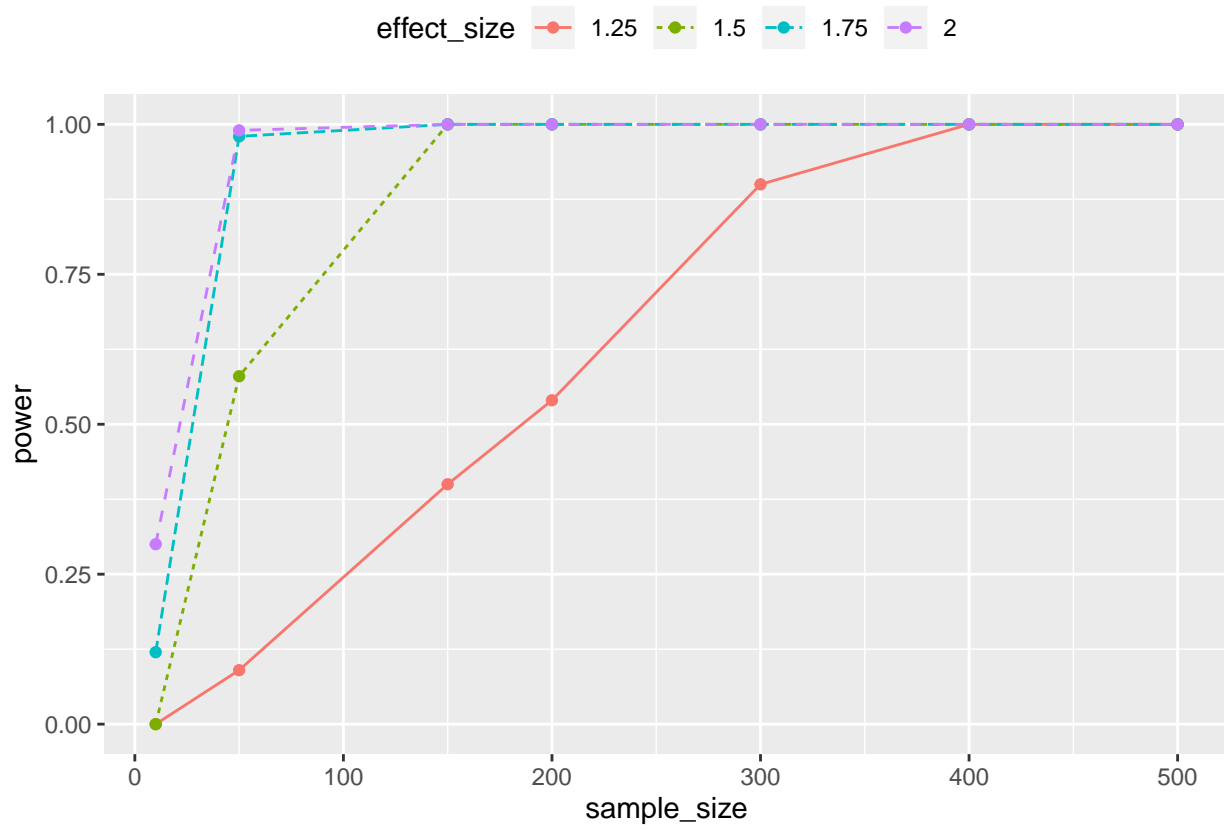
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. | NA's |
|-----------|-----------|-----------|-----------|-----------|-----------|-------|
| 0.000e+00 | 1.000e+00 | 6.000e+00 | 8.457e+15 | 9.909e+15 | 8.454e+17 | 10613 |

Simulation will be performed assume there are 50K reads for each cell.

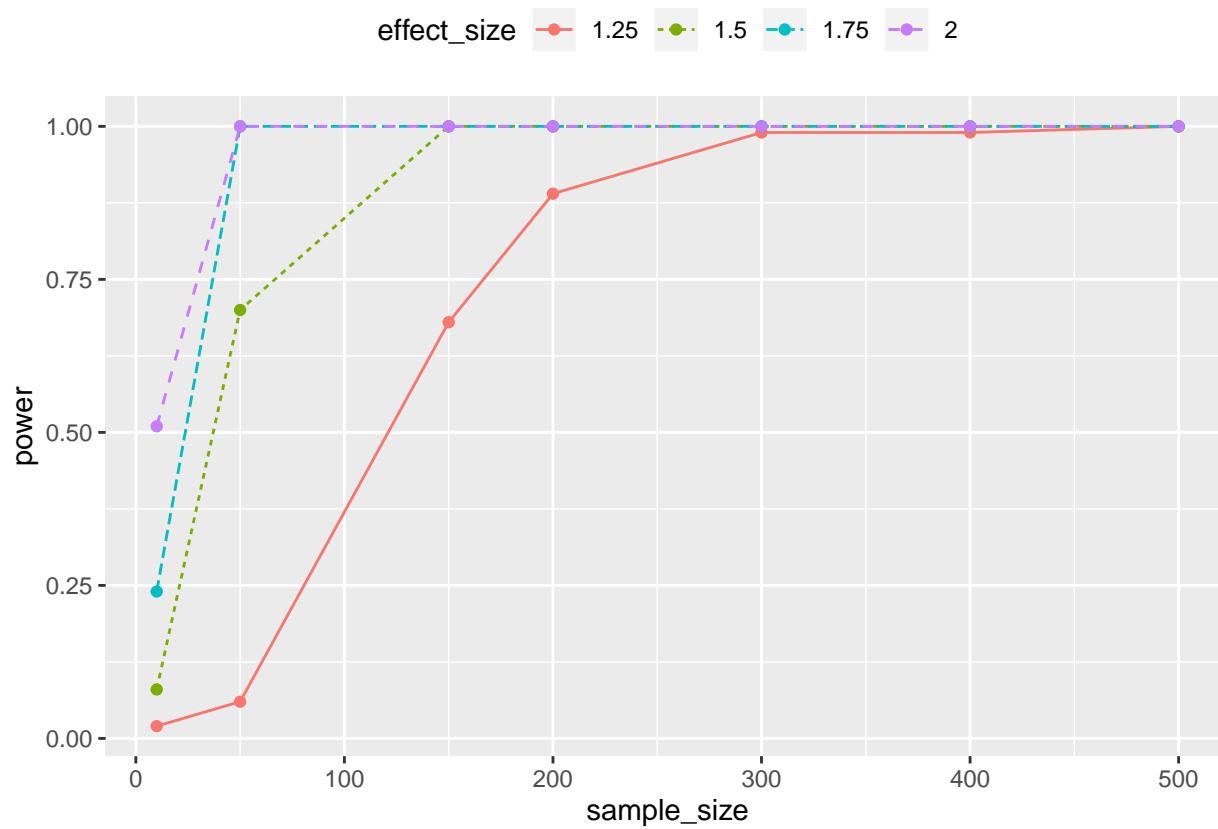
10 reads on average in control, 500 SNP-gene



5 reads on average in control, 500 SNP-gene



10 reads on average in control, 1000 SNP-gene



5 reads on average in control, 1000 SNP-gene

