nature research

https://doi.org/10.1038/s41596-020-00437-6

Supplementary information

Reliable detection of somatic mutations in solid tissues by laser-capture microdissection and low-input DNA sequencing

In the format provided by the authors and unedited

Sonication vs Enzymatic Fragmentation Coverage Comparisons

Low-input, LCM-derived whole genome libraries were prepared from 5 colonic crypts fragmented via sonication, and 7 colonic crypts fragmented via enzymatic fragmentation.

Each was sequenced to a median depth whole-genome depth of 20X or 21X, and found to have a median VAF >= 0.40.

Load Libraries

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(data.table)
```

```
##
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
library(ggplot2)
library(readxl)
library(VennDiagram)
```

```
## Loading required package: grid

## Loading required package: futile.logger
```

Input Files

```
mosdeth.dir <- "~/nfs_tb14/scratch_117/LCM_Methods_Paper/coverage_comparisons/"
threshold.files <- list.files(mosdeth.dir, pattern = "*thresholds.bed.gz$", full.names = TRUE)
quantile.file <- list.files(mosdeth.dir, pattern = "quantized.bed.gz", full.names = TRUE)

# load sample metadata
sample_metadata <- read_excel("~/nfs_tb14/Projects/LCM_Methods_Paper/revisions_analysis/coverage_compari
```

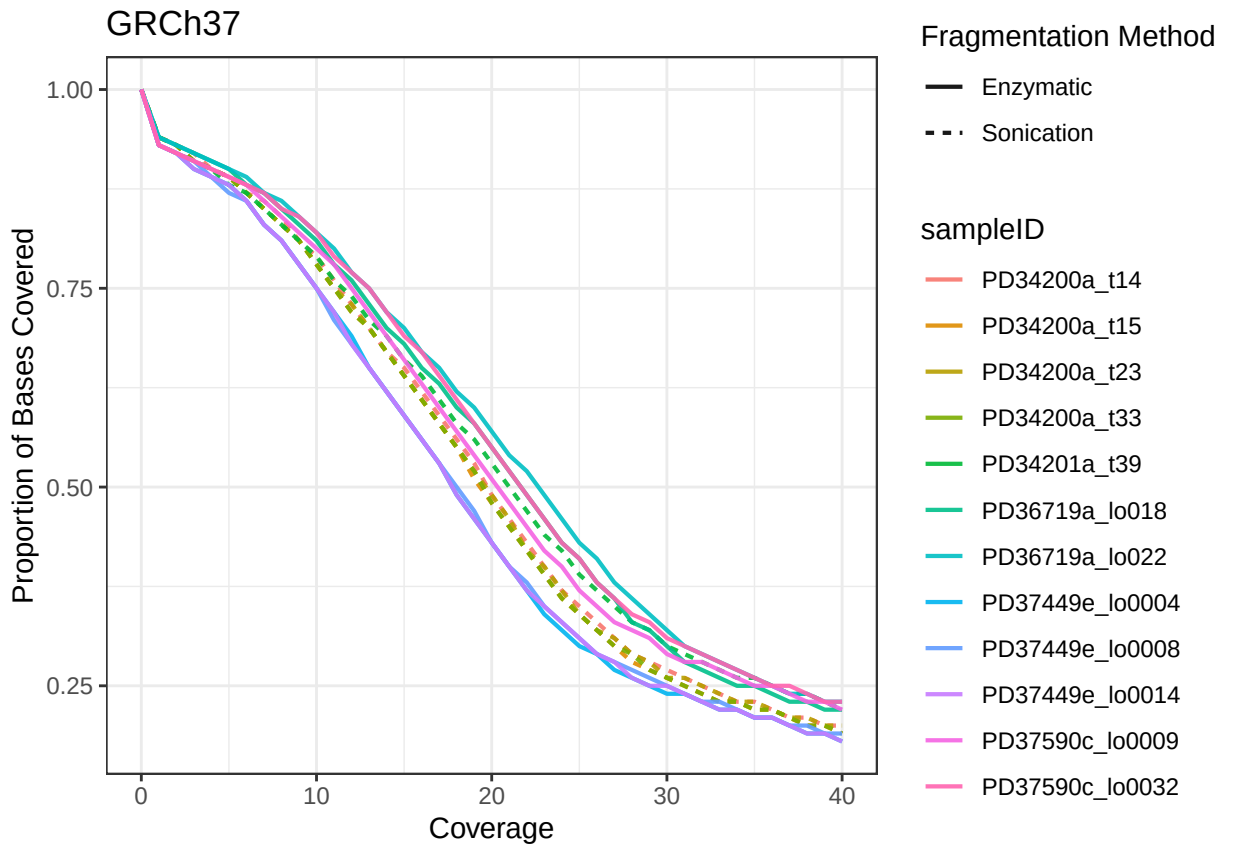
Functions

```
# Function for calculating fraction of genome covered at X depth (via mosdepth output)
mosdepth_thresh <- function(x) {
  genome_size = 3095693981
  chr_list <- seq(1,22,1) # ignoring "X" and "Y" intentionally in order to compare men and women
  this_sample = basename(x)
  this_sample_name = gsub('.mosdepth.thresholds.bed.gz', '', this_sample)
  no_analysis_size = scan(gsub('.mosdepth.thresholds.bed.gz', '.caveman_analyzed_size.txt', x))
  mosdepth_tbl = fread(x, header = F, skip = 1)
  mosdepth_tbl_filtered <- mosdepth_tbl %>%
    dplyr::filter(V1 %in% chr_list)
  cumulative_total_tbl <- data.frame(no_analysis_size,  t(colSums(mosdepth_tbl_filtered[,5:12]))) # get
  colnames(cumulative_total_tbl) <- c("no_analysis_frac", "cov_2X_frac", "cov_4X_frac", "cov_6X_frac",
  cumulative_frac_tbl <- cumulative_total_tbl/genome_size
  cbind(sampleID = this_sample_name, cumulative_frac_tbl)
}

# function to get cumulative coverage stats
myCumulative_Files <- function(x) {
  chr_list <- seq(1,22,1)
  chr_list[23] <- "X"
  chr_list[24] <- "Y"
  chr_list[35] <- "hs37d5" # gets overall summary for whole genome
  sample = gsub(".mosdepth.mosdepth.global.dist.txt", "", basename(x))
  cov_tbl <- fread(x)
  cov_tbl_filtered <- cov_tbl %>%
    dplyr::filter(V1 %in% chr_list)
  cbind(sampleID = sample, cov_tbl_filtered)
}
```

Cumulative coverage Plots - Plotting the output of mosdepth

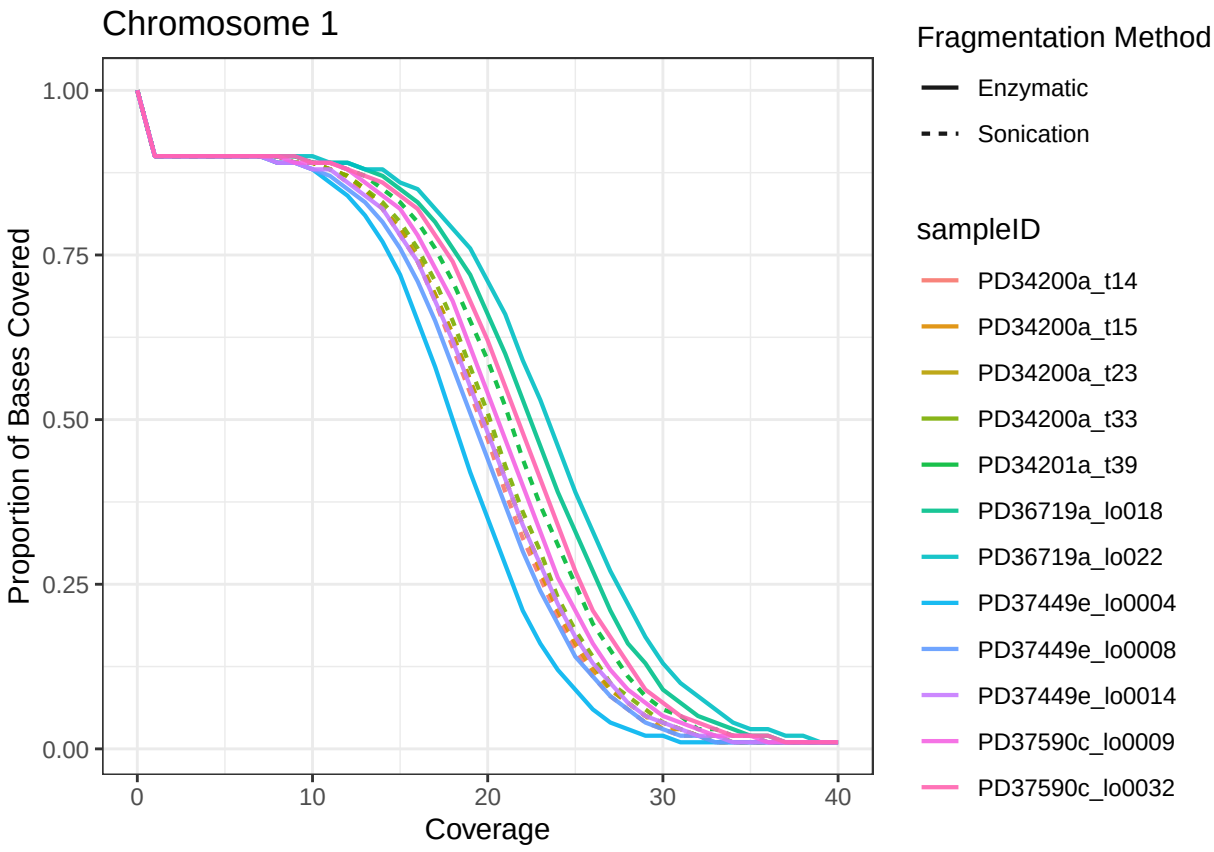
```
cumulative_cov_files <- list.files("~/nfs_tb14/scratch_117/LCM_Methods_Paper/coverage_comparisons/", pa

cumulative_cov_tbl <- lapply(cumulative_cov_files, myCumulative_Files) %>% bind_rows()
colnames(cumulative_cov_tbl) <- c("sampleID", "chr", "coverage", "proportion_of_bases")

cumulative_cov_meta <- left_join(cumulative_cov_tbl, sample_metadata[,c("sampleID", "fragmentation_meth

ggplot(subset(cumulative_cov_meta, chr == "hs37d5" & coverage <=40)) +
  theme_bw() +
```

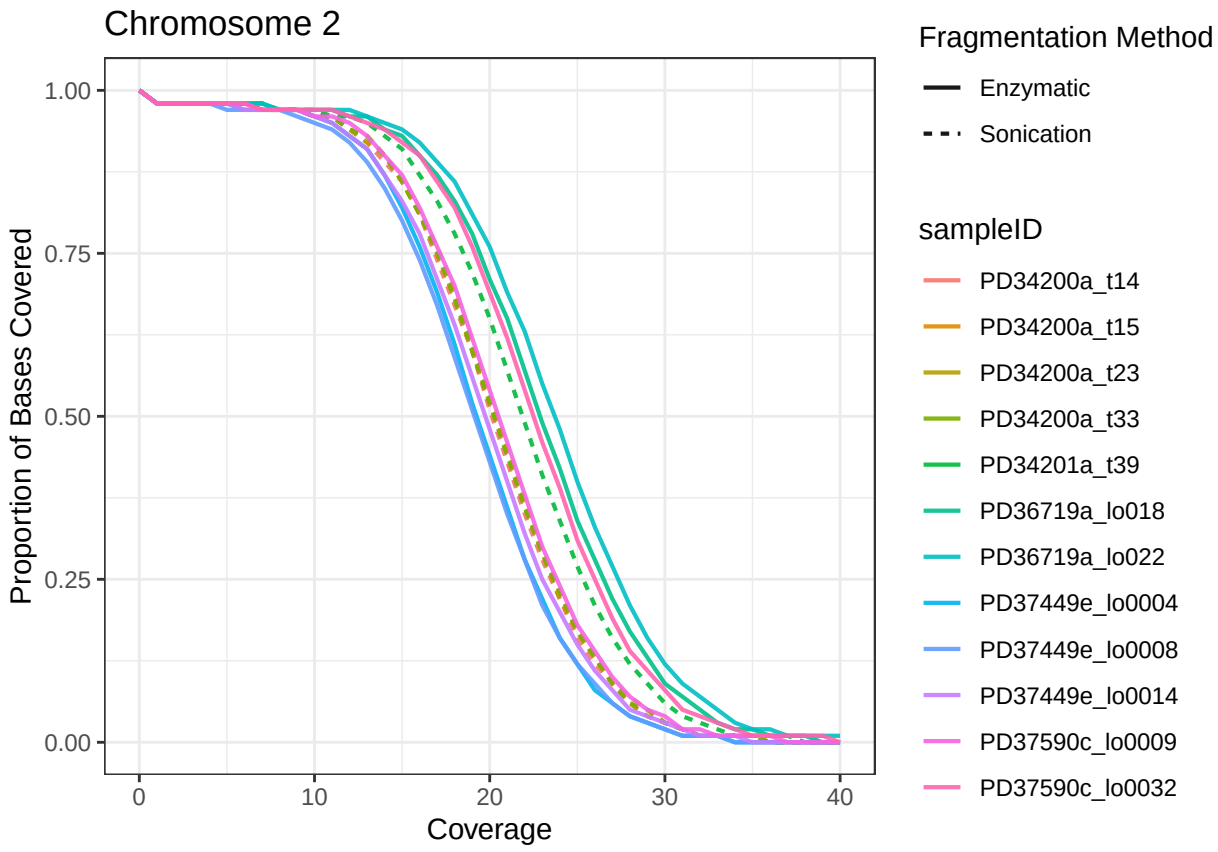
```
  aes(x=coverage, y=proportion_of_bases, color = sampleID, linetype = fragmentation_method) +
  geom_line(size=0.75, alpha=0.9) +
  ylab("Proportion of Bases Covered") +
  xlab("Coverage")+
  ggtitle("GRCh37") +
  scale_linetype_discrete(name = "Fragmentation Method")
```

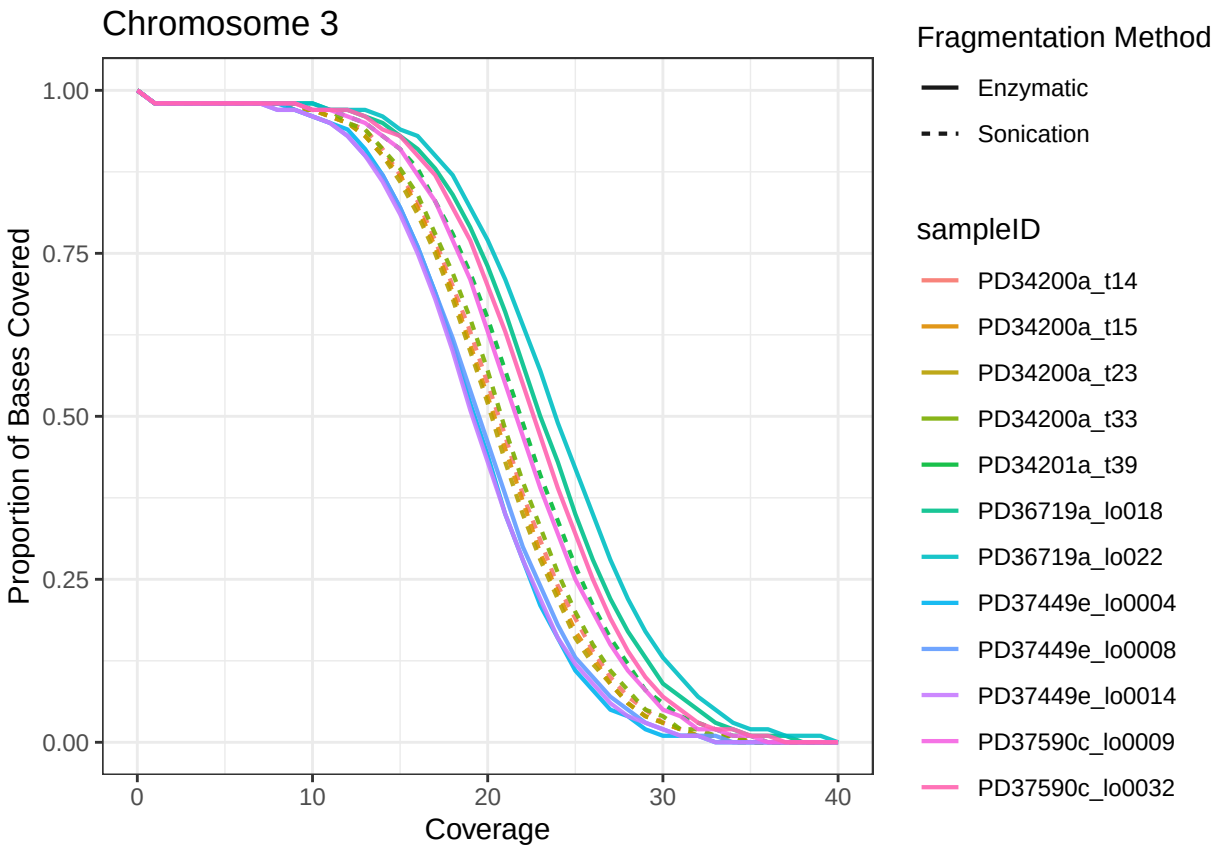


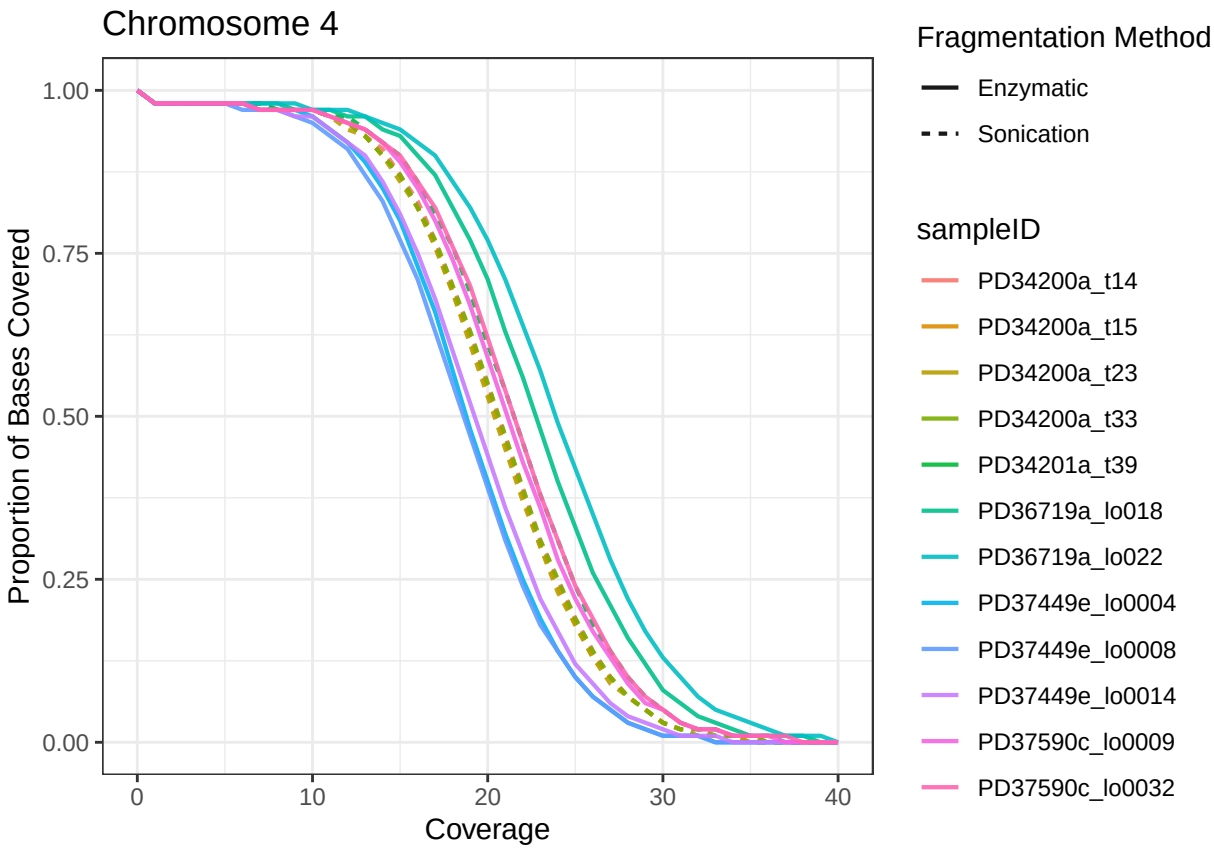
Per chromosome cumulative coverage

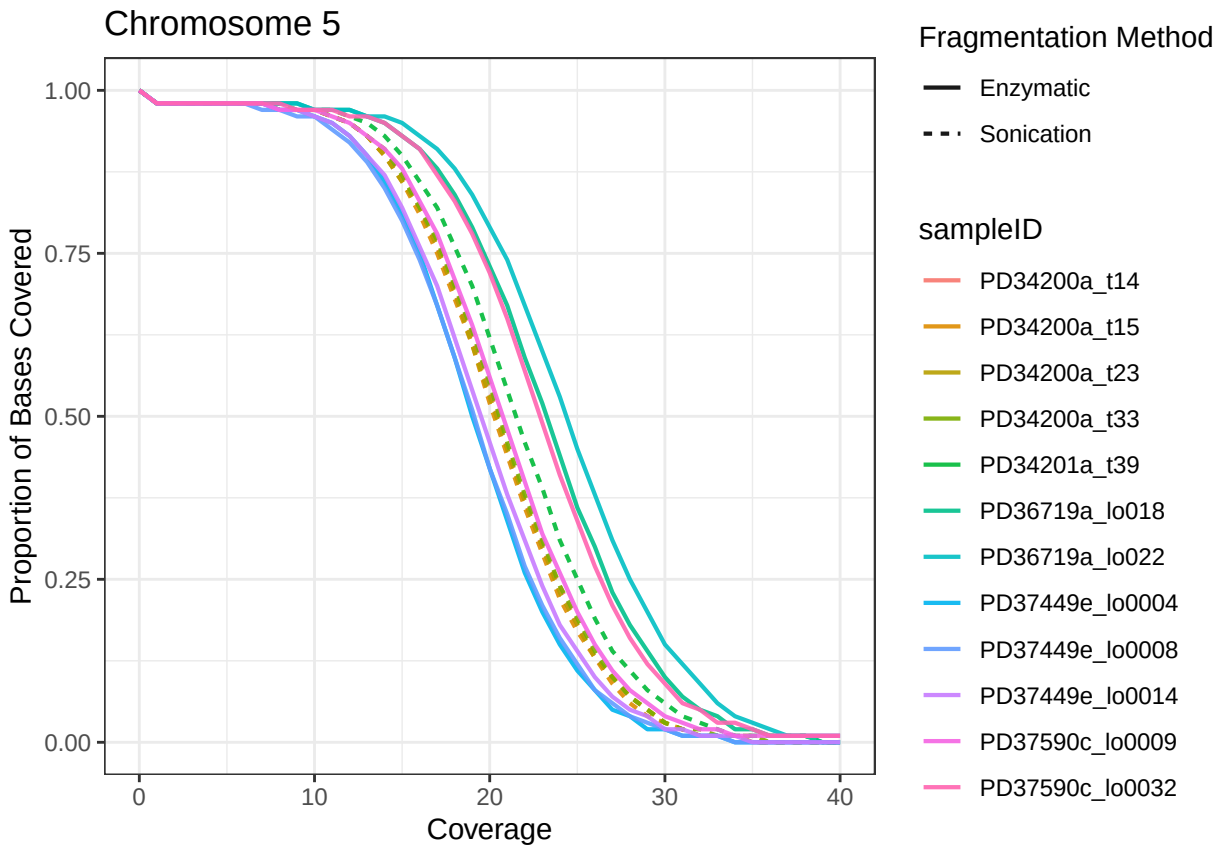
```
per_chr <- subset(cumulative_cov_meta, chr != "hs37d5" & coverage <=40)
chr_list <- unique(per_chr$chr)

for (i in 1:length(chr_list)) {
  this_chr = chr_list[i]
  this_tbl <- subset(per_chr, chr == this_chr)
  chr.p <- ggplot(this_tbl) +
    theme_bw() +
    aes(x=coverage, y=proportion_of_bases, color = sampleID, linetype = fragmentation_method) +
    geom_line(size=0.75, alpha=0.9) +
    ylab("Proportion of Bases Covered") +
    xlab("Coverage") +
    ggtitle(paste("Chromosome", this_chr)) +
    scale_linetype_discrete(name = "Fragmentation Method")
  print(chr.p)
}
```

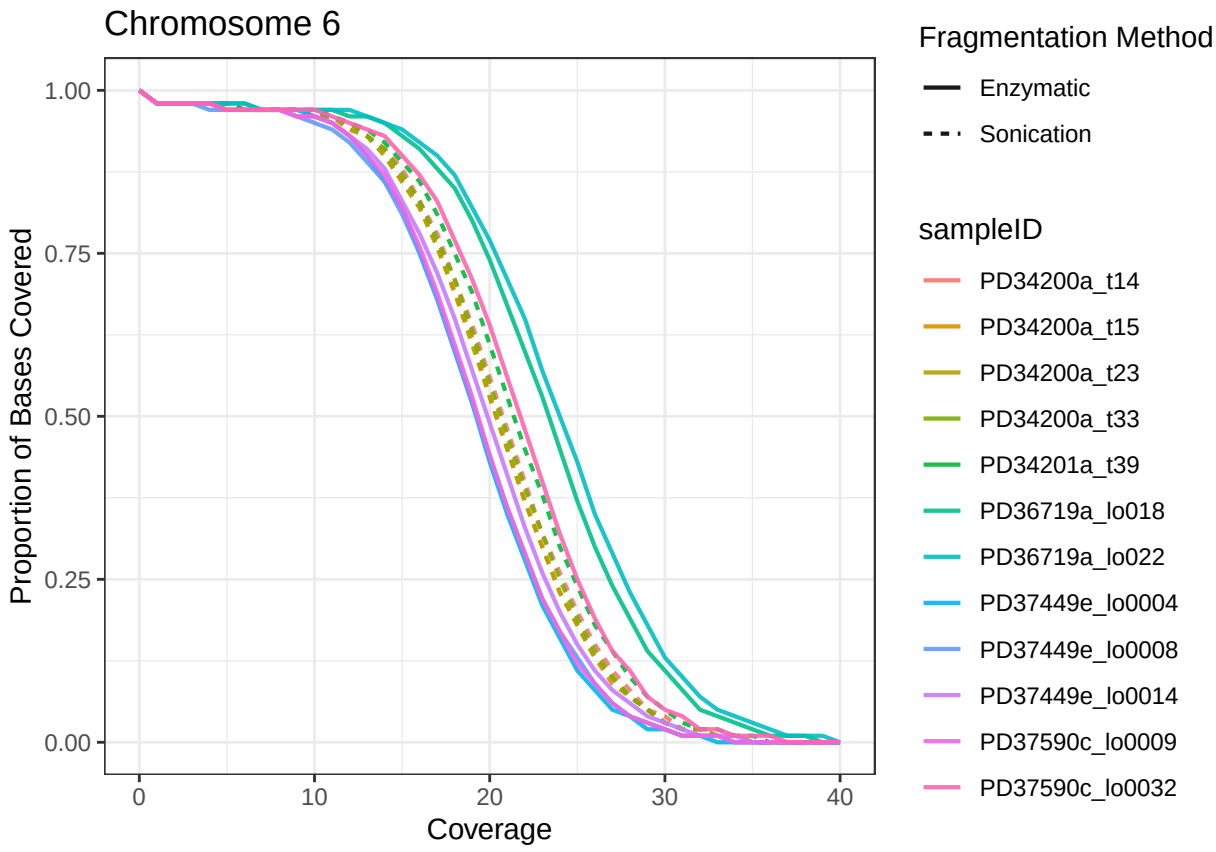

Chromosome 1
Proportion of Bases Covered
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Coverage
Fragmentation Method
Enzymatic
Sonication
sampleID
PD34200a_t14
PD34200a_t15
PD34200a_t23
PD34200a_t33
PD34201a_t39
PD36719a_lo018
PD36719a_lo022
PD37449e_lo0004
PD37449e_lo0008
PD37449e_lo0014
PD37590c_lo0009
PD37590c_lo0032

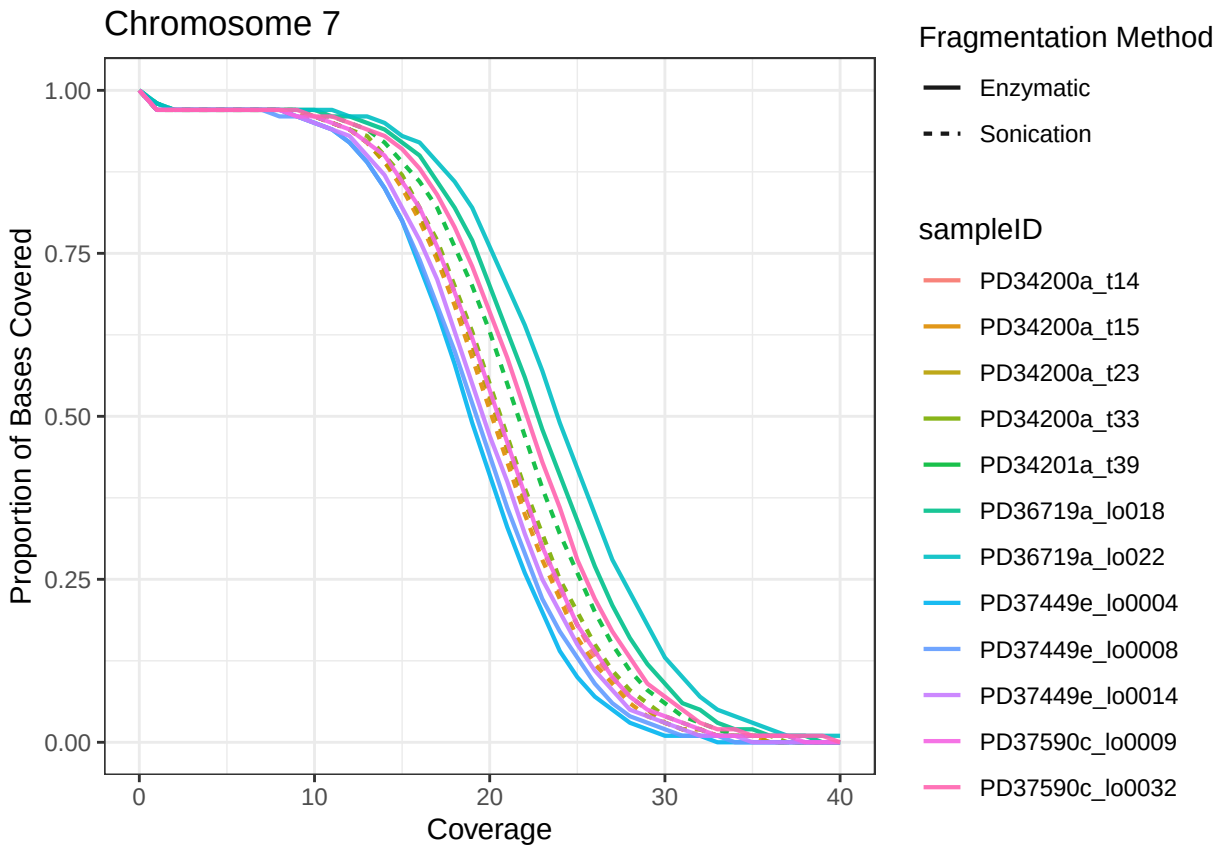

Chromosome 2
Proportion of Bases Covered
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Coverage
Fragmentation Method
Enzymatic
Sonication
sampleID
PD34200a_t14
PD34200a_t15
PD34200a_t23
PD34200a_t33
PD34201a_t39
PD36719a_lo018
PD36719a_lo022
PD37449e_lo0004
PD37449e_lo0008
PD37449e_lo0014
PD37590c_lo0009
PD37590c_lo0032

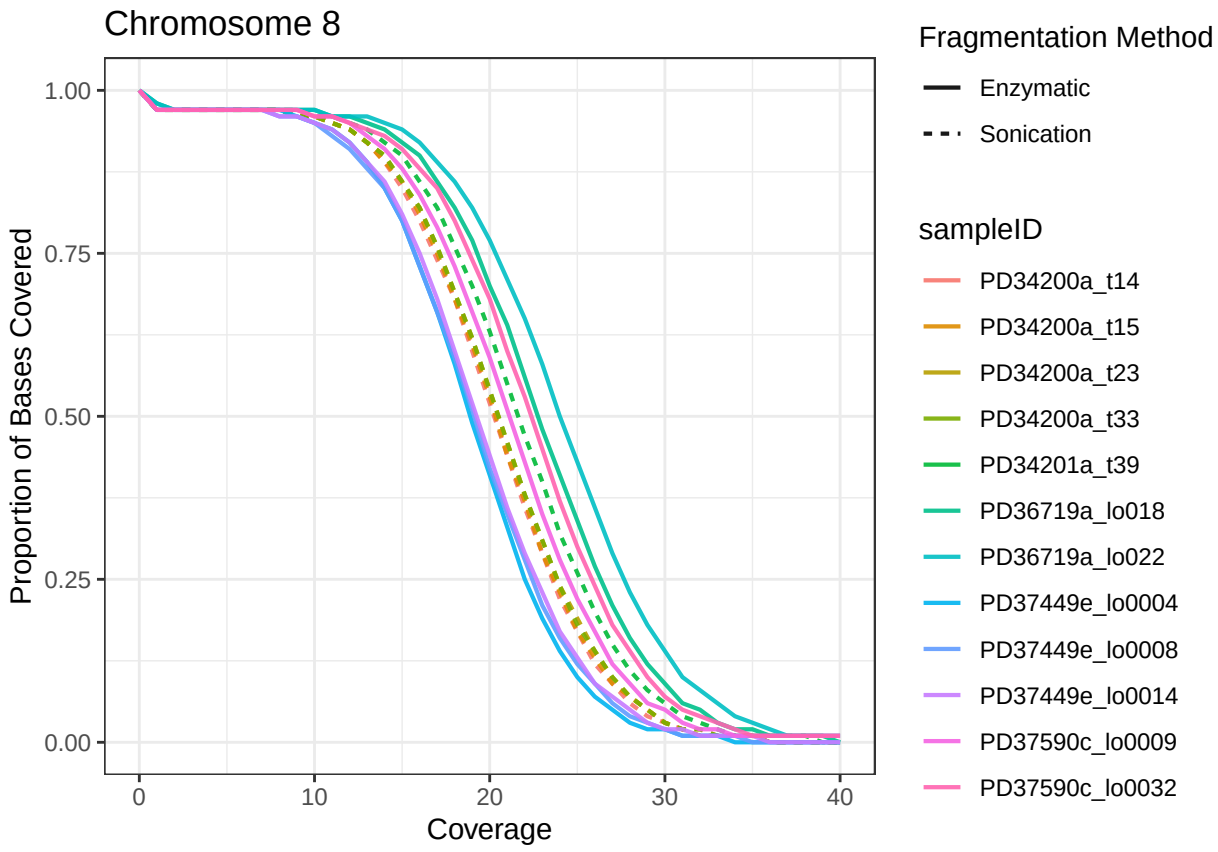

Chromosome 3
Proportion of Bases Covered
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Coverage
Fragmentation Method
Enzymatic
Sonication
sampleID
PD34200a_t14
PD34200a_t15
PD34200a_t23
PD34200a_t33
PD34201a_t39
PD36719a_lo018
PD36719a_lo022
PD37449e_lo0004
PD37449e_lo0008
PD37449e_lo0014
PD37590c_lo0009
PD37590c_lo0032

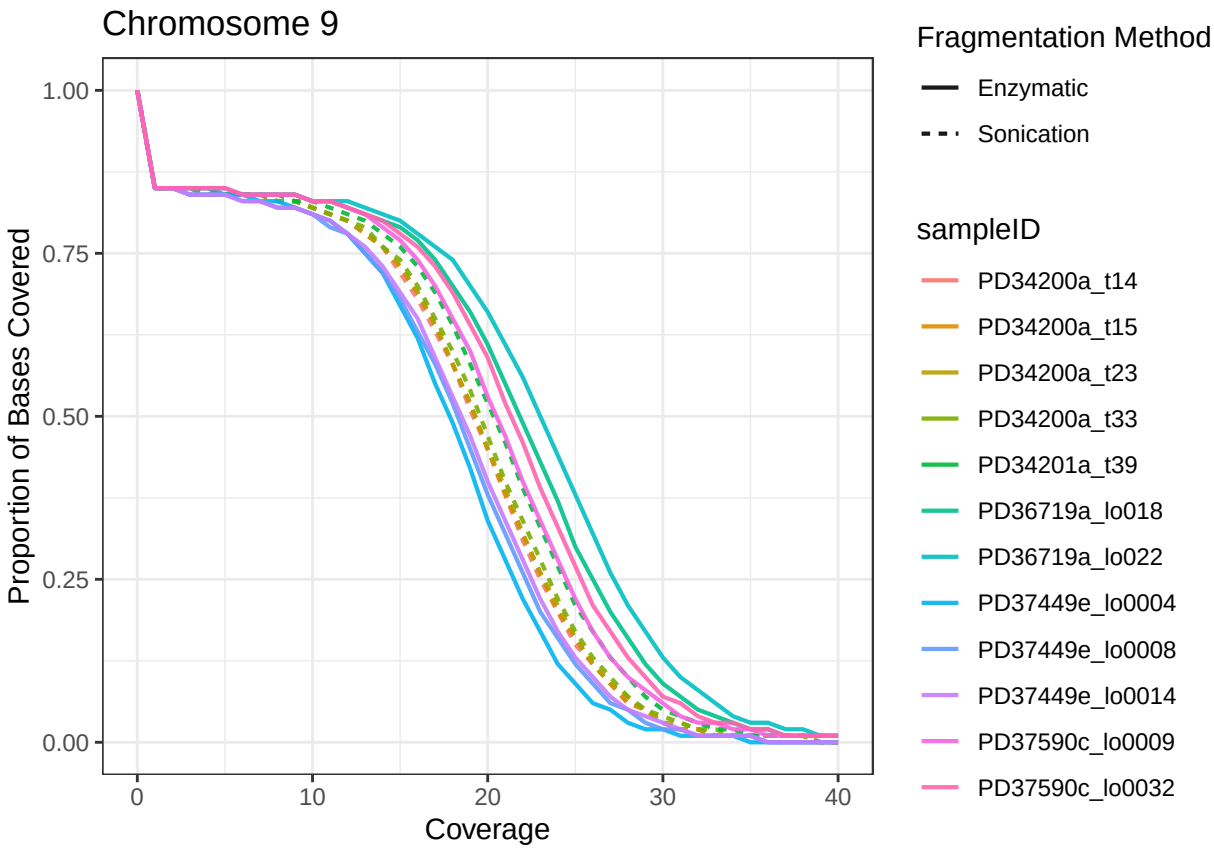

Chromosome 4
Proportion of Bases Covered
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Coverage
Fragmentation Method
Enzymatic
Sonication
sampleID
PD34200a_t14
PD34200a_t15
PD34200a_t23
PD34200a_t33
PD34201a_t39
PD36719a_lo018
PD36719a_lo022
PD37449e_lo0004
PD37449e_lo0008
PD37449e_lo0014
PD37590c_lo0009
PD37590c_lo0032

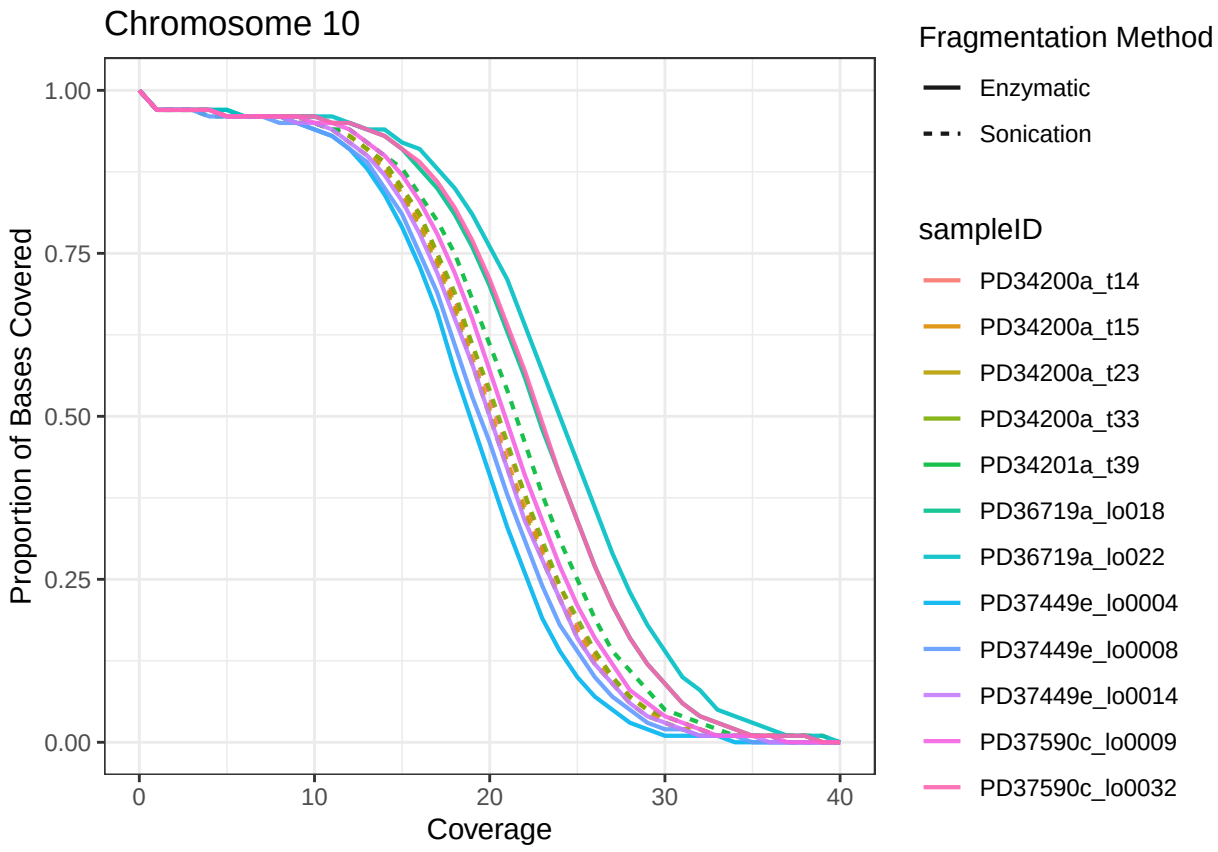

Chromosome 5
Proportion of Bases Covered
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Coverage
Fragmentation Method
Enzymatic
Sonication
sampleID
PD34200a_t14
PD34200a_t15
PD34200a_t23
PD34200a_t33
PD34201a_t39
PD36719a_lo018
PD36719a_lo022
PD37449e_lo0004
PD37449e_lo0008
PD37449e_lo0014
PD37590c_lo0009
PD37590c_lo0032

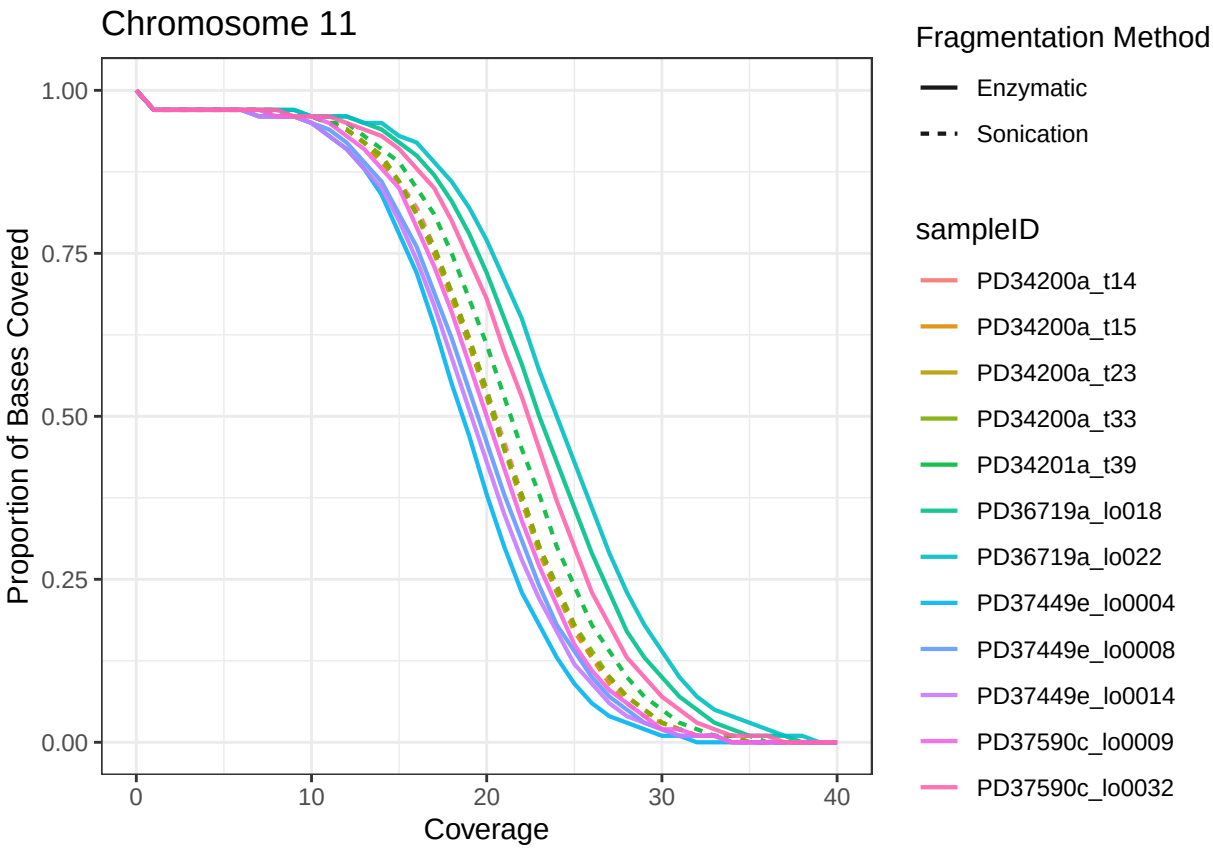

Chromosome 6
Proportion of Bases Covered
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Coverage
Fragmentation Method
Enzymatic
Sonication
sampleID
PD34200a_t14
PD34200a_t15
PD34200a_t23
PD34200a_t33
PD34201a_t39
PD36719a_lo018
PD36719a_lo022
PD37449e_lo0004
PD37449e_lo0008
PD37449e_lo0014
PD37590c_lo0009
PD37590c_lo0032

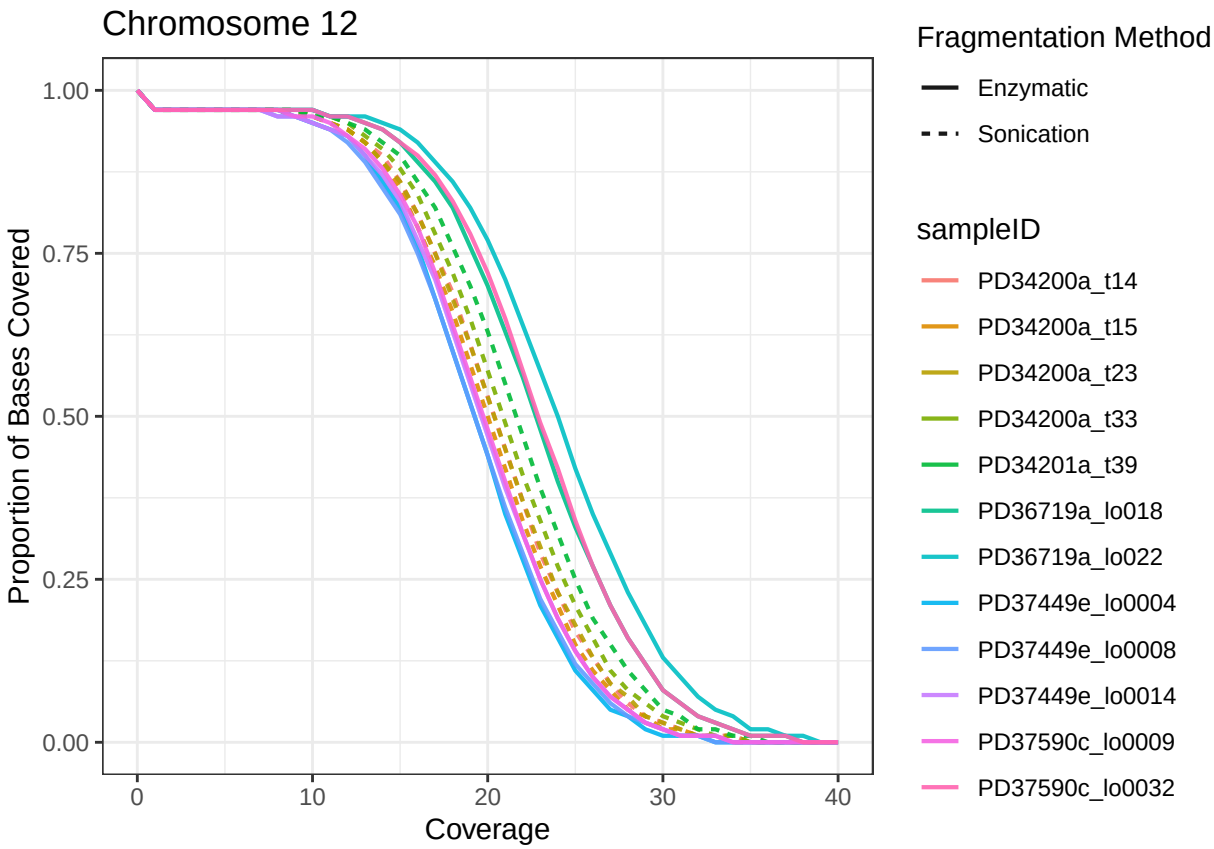

Chromosome 7
Proportion of Bases Covered
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Coverage
Fragmentation Method
Enzymatic
Sonication
sampleID
PD34200a_t14
PD34200a_t15
PD34200a_t23
PD34200a_t33
PD34201a_t39
PD36719a_lo018
PD36719a_lo022
PD37449e_lo0004
PD37449e_lo0008
PD37449e_lo0014
PD37590c_lo0009
PD37590c_lo0032

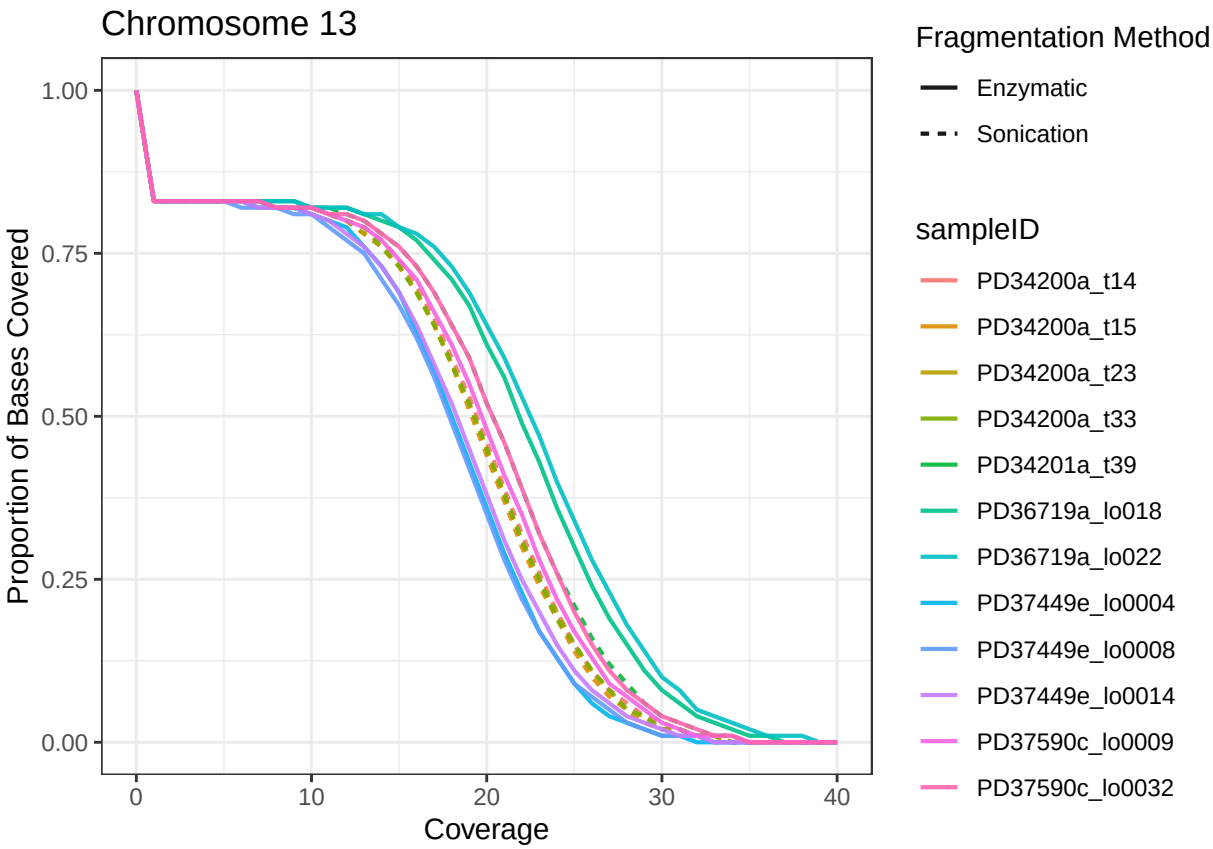

Chromosome 8
Proportion of Bases Covered
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Coverage
Fragmentation Method
Enzymatic
Sonication
sampleID
PD34200a_t14
PD34200a_t15
PD34200a_t23
PD34200a_t33
PD34201a_t39
PD36719a_lo018
PD36719a_lo022
PD37449e_lo0004
PD37449e_lo0008
PD37449e_lo0014
PD37590c_lo0009
PD37590c_lo0032

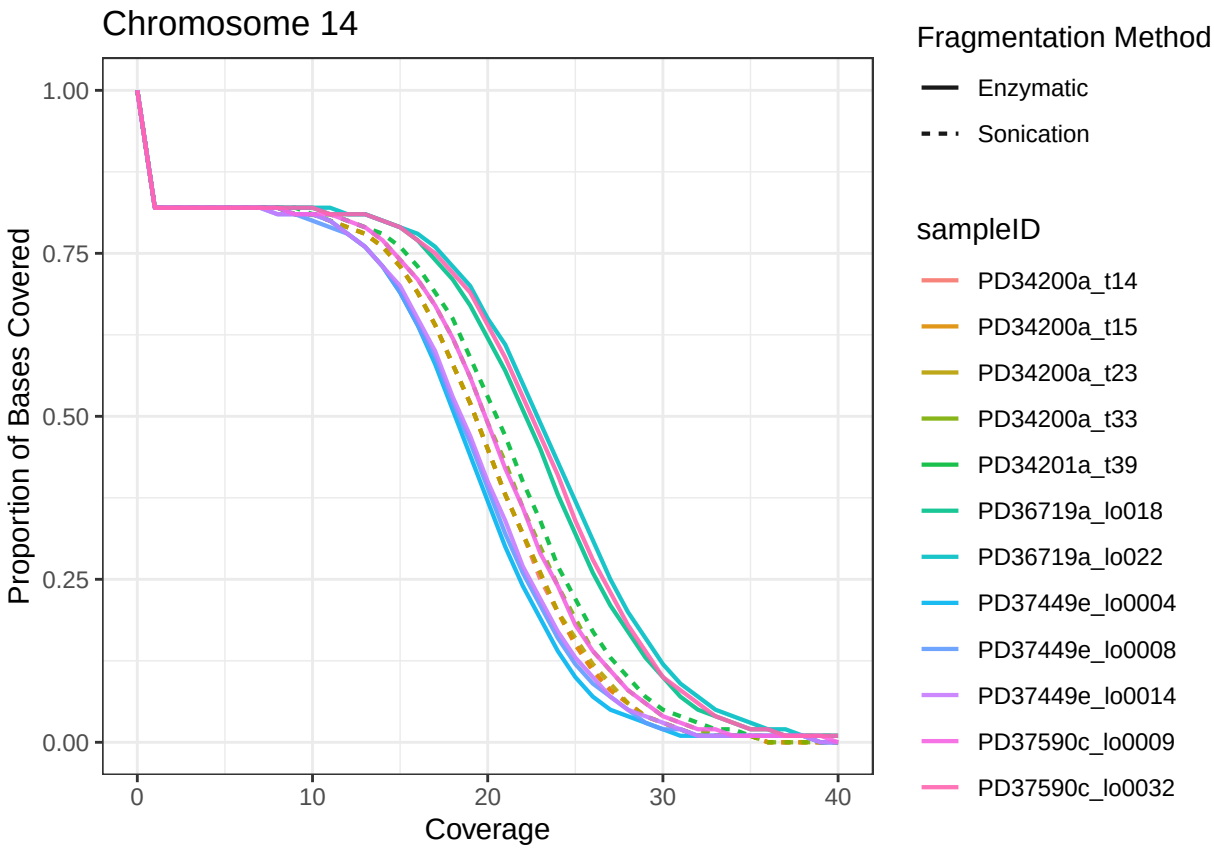

Chromosome 9
Proportion of Bases Covered
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Coverage
Fragmentation Method
Enzymatic
Sonication
sampleID
PD34200a_t14
PD34200a_t15
PD34200a_t23
PD34200a_t33
PD34201a_t39
PD36719a_lo018
PD36719a_lo022
PD37449e_lo0004
PD37449e_lo0008
PD37449e_lo0014
PD37590c_lo0009
PD37590c_lo0032

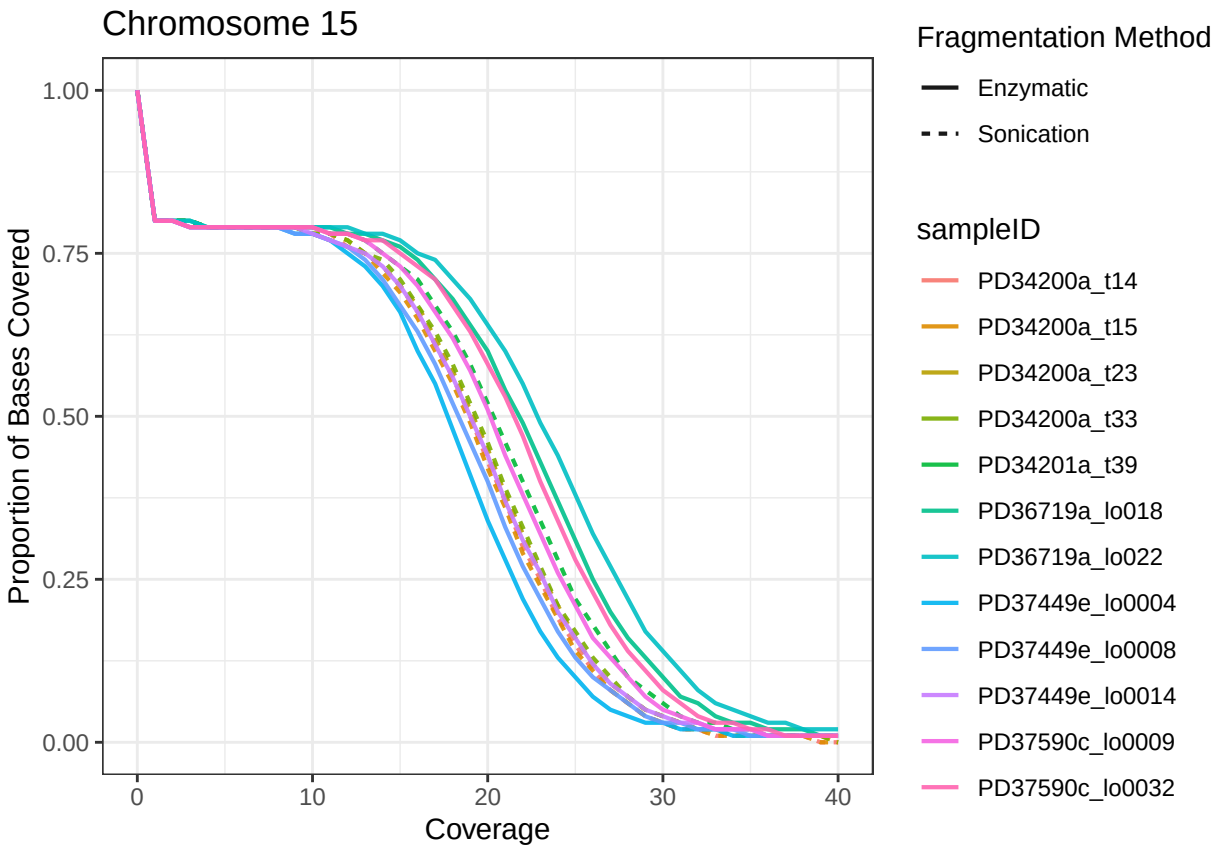

Chromosome 10
Proportion of Bases Covered
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Coverage
Fragmentation Method
Enzymatic
Sonication
sampleID
PD34200a_t14
PD34200a_t15
PD34200a_t23
PD34200a_t33
PD34201a_t39
PD36719a_lo018
PD36719a_lo022
PD37449e_lo0004
PD37449e_lo0008
PD37449e_lo0014
PD37590c_lo0009
PD37590c_lo0032

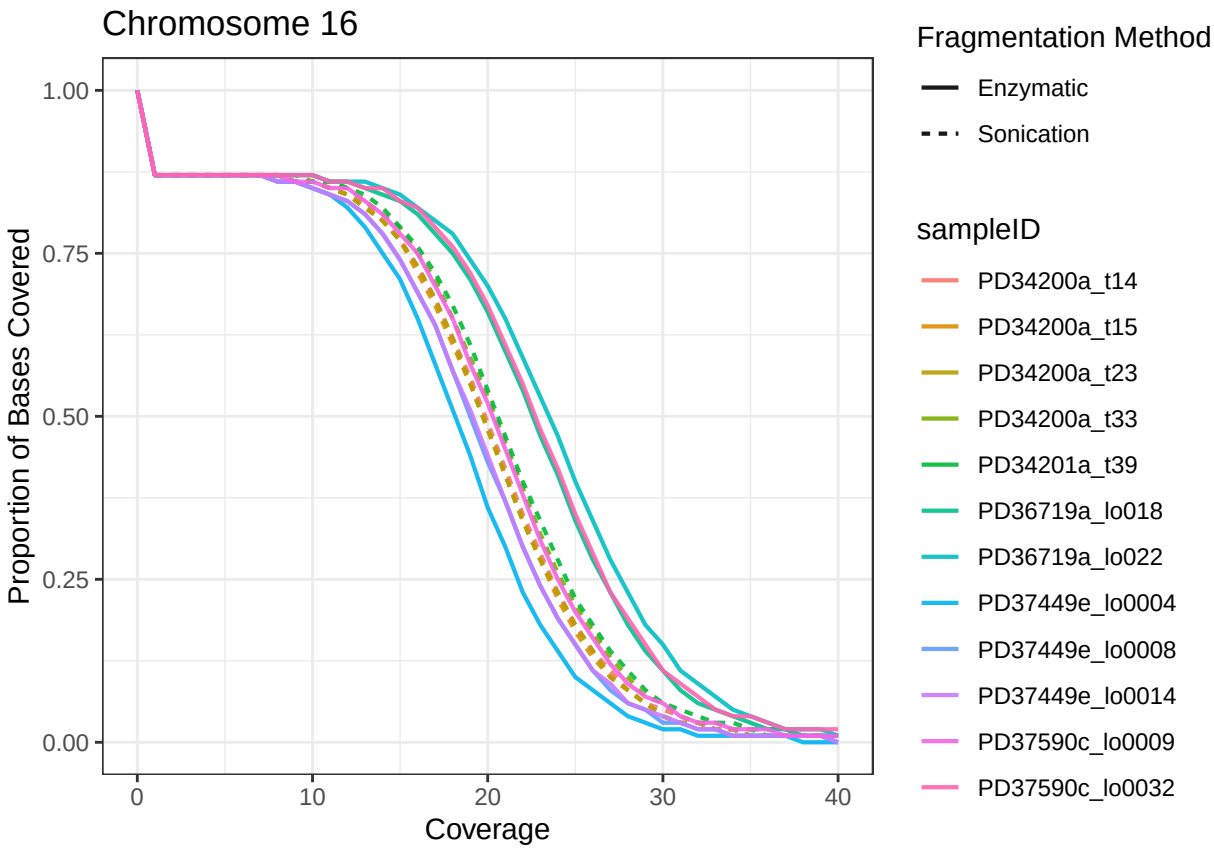

Chromosome 11
Proportion of Bases Covered
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Coverage
Fragmentation Method
Enzymatic
Sonication
sampleID
PD34200a_t14
PD34200a_t15
PD34200a_t23
PD34200a_t33
PD34201a_t39
PD36719a_lo018
PD36719a_lo022
PD37449e_lo0004
PD37449e_lo0008
PD37449e_lo0014
PD37590c_lo0009
PD37590c_lo0032

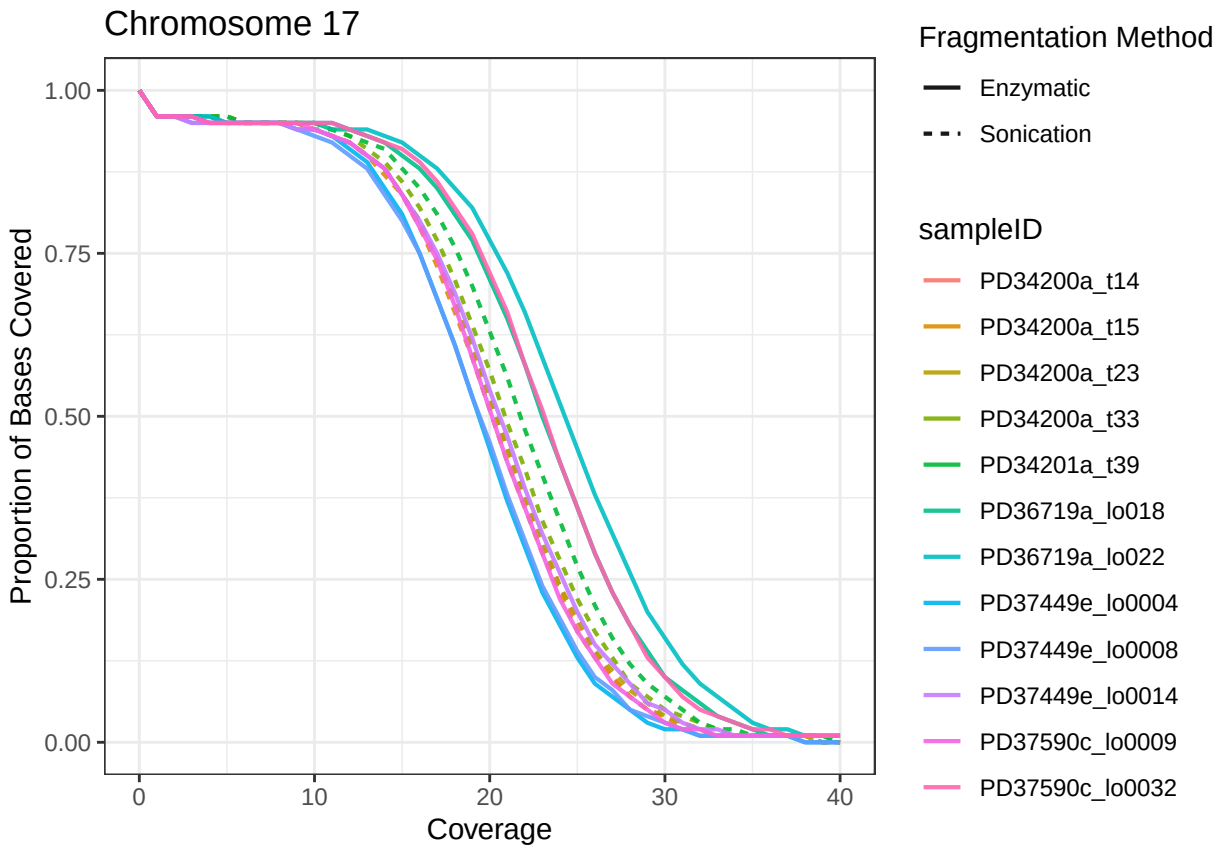

Chromosome 12
Proportion of Bases Covered
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Coverage
Fragmentation Method
Enzymatic
Sonication
sampleID
PD34200a_t14
PD34200a_t15
PD34200a_t23
PD34200a_t33
PD34201a_t39
PD36719a_lo018
PD36719a_lo022
PD37449e_lo0004
PD37449e_lo0008
PD37449e_lo0014
PD37590c_lo0009
PD37590c_lo0032

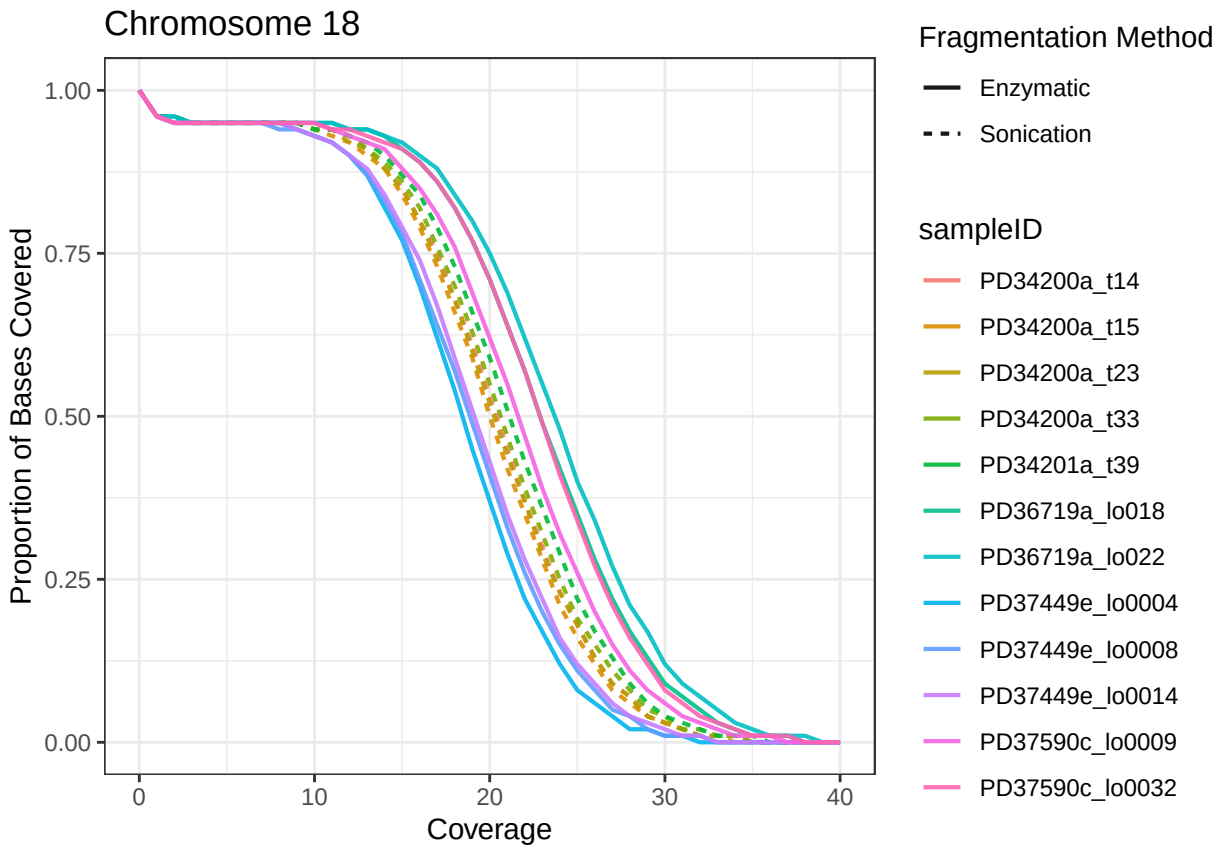

Chromosome 13
Proportion of Bases Covered
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Coverage
Fragmentation Method
Enzymatic
Sonication
sampleID
PD34200a_t14
PD34200a_t15
PD34200a_t23
PD34200a_t33
PD34201a_t39
PD36719a_lo018
PD36719a_lo022
PD37449e_lo0004
PD37449e_lo0008
PD37449e_lo0014
PD37590c_lo0009
PD37590c_lo0032

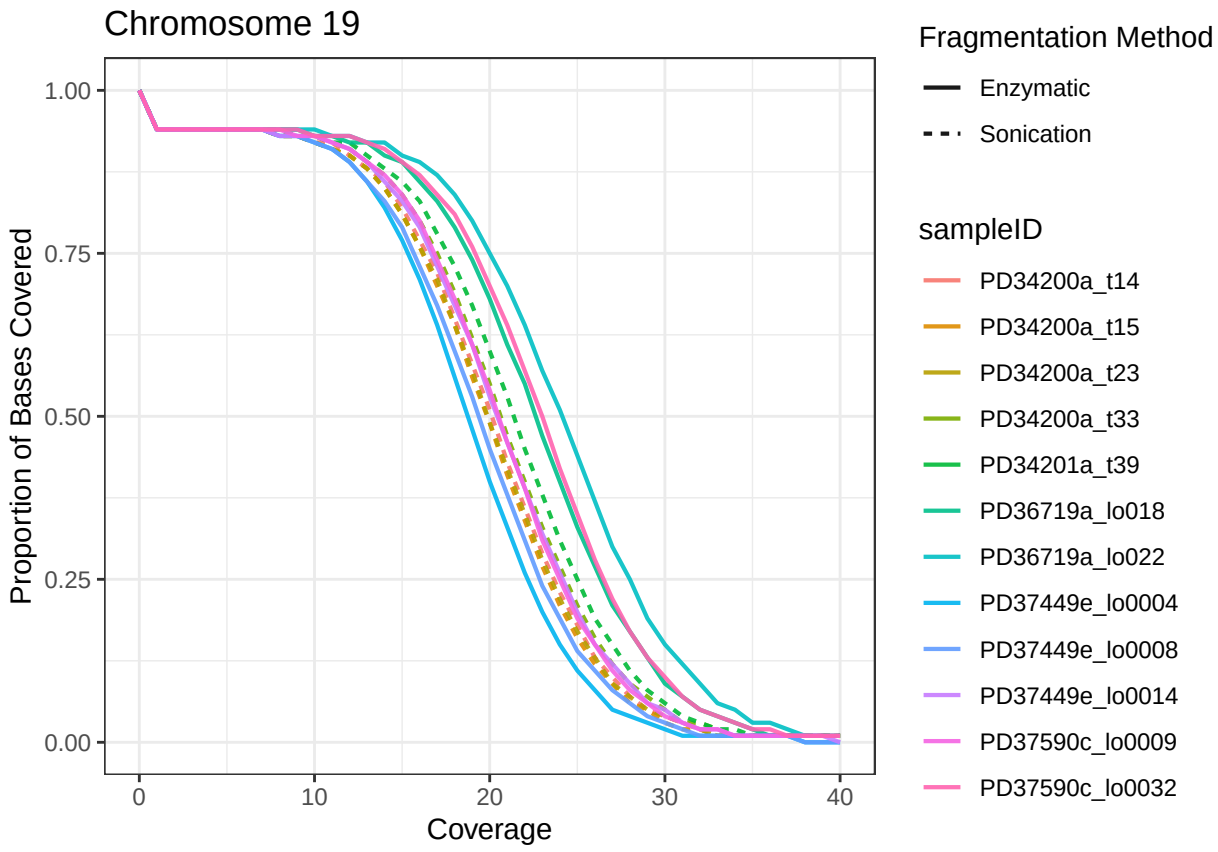

Chromosome 14
Proportion of Bases Covered
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Coverage
Fragmentation Method
Enzymatic
Sonication
sampleID
PD34200a_t14
PD34200a_t15
PD34200a_t23
PD34200a_t33
PD34201a_t39
PD36719a_lo018
PD36719a_lo022
PD37449e_lo0004
PD37449e_lo0008
PD37449e_lo0014
PD37590c_lo0009
PD37590c_lo0032

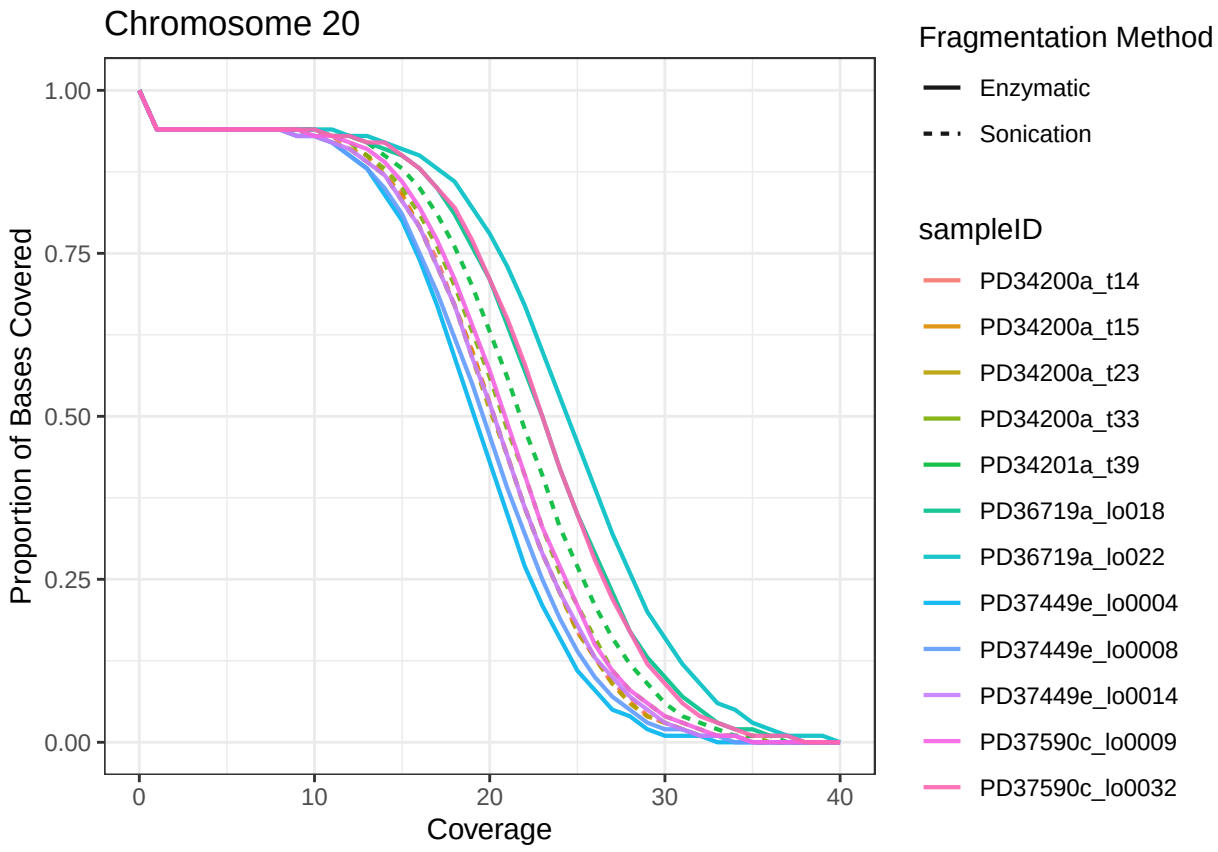

Chromosome 15
Proportion of Bases Covered
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Coverage
Fragmentation Method
Enzymatic
Sonication
sampleID
PD34200a_t14
PD34200a_t15
PD34200a_t23
PD34200a_t33
PD34201a_t39
PD36719a_lo018
PD36719a_lo022
PD37449e_lo0004
PD37449e_lo0008
PD37449e_lo0014
PD37590c_lo0009
PD37590c_lo0032

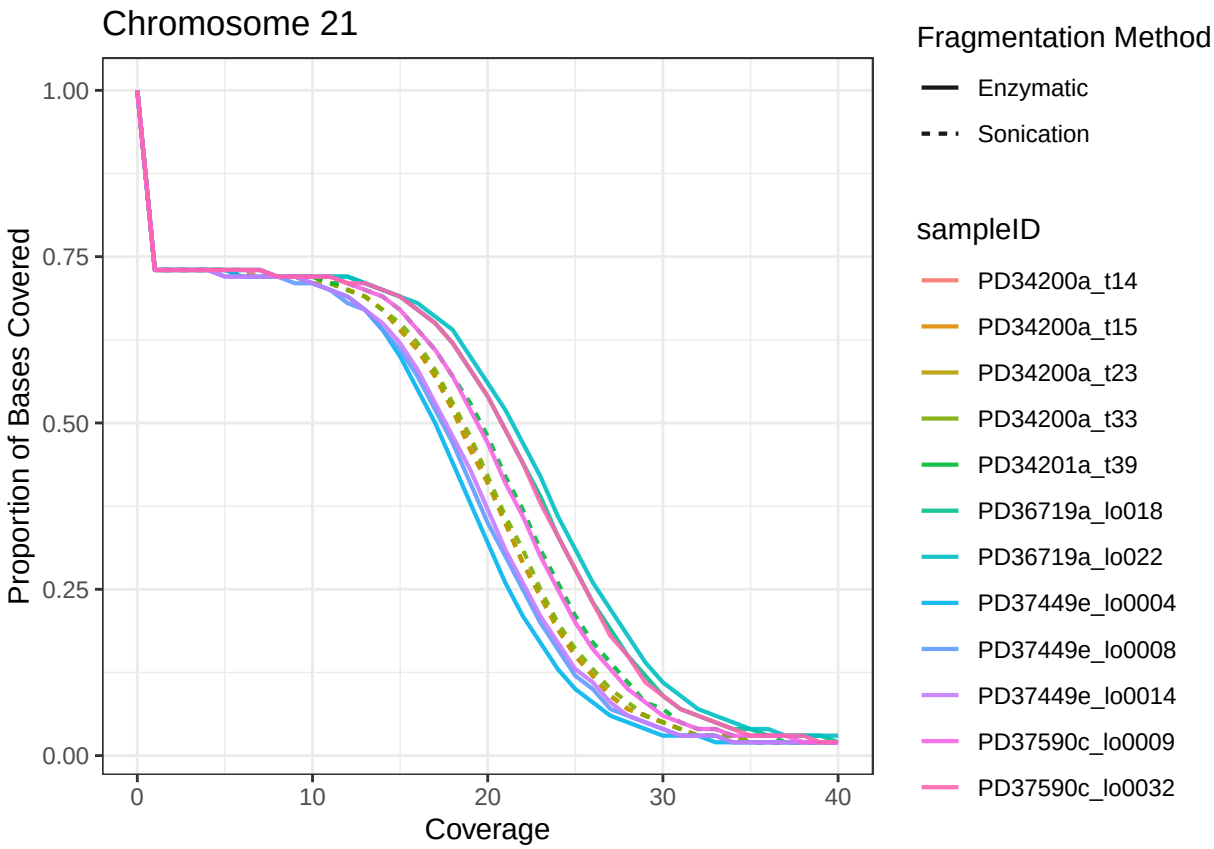

Chromosome 16
Proportion of Bases Covered
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Coverage
Fragmentation Method
Enzymatic
Sonication
sampleID
PD34200a_t14
PD34200a_t15
PD34200a_t23
PD34200a_t33
PD34201a_t39
PD36719a_lo018
PD36719a_lo022
PD37449e_lo0004
PD37449e_lo0008
PD37449e_lo0014
PD37590c_lo0009
PD37590c_lo0032

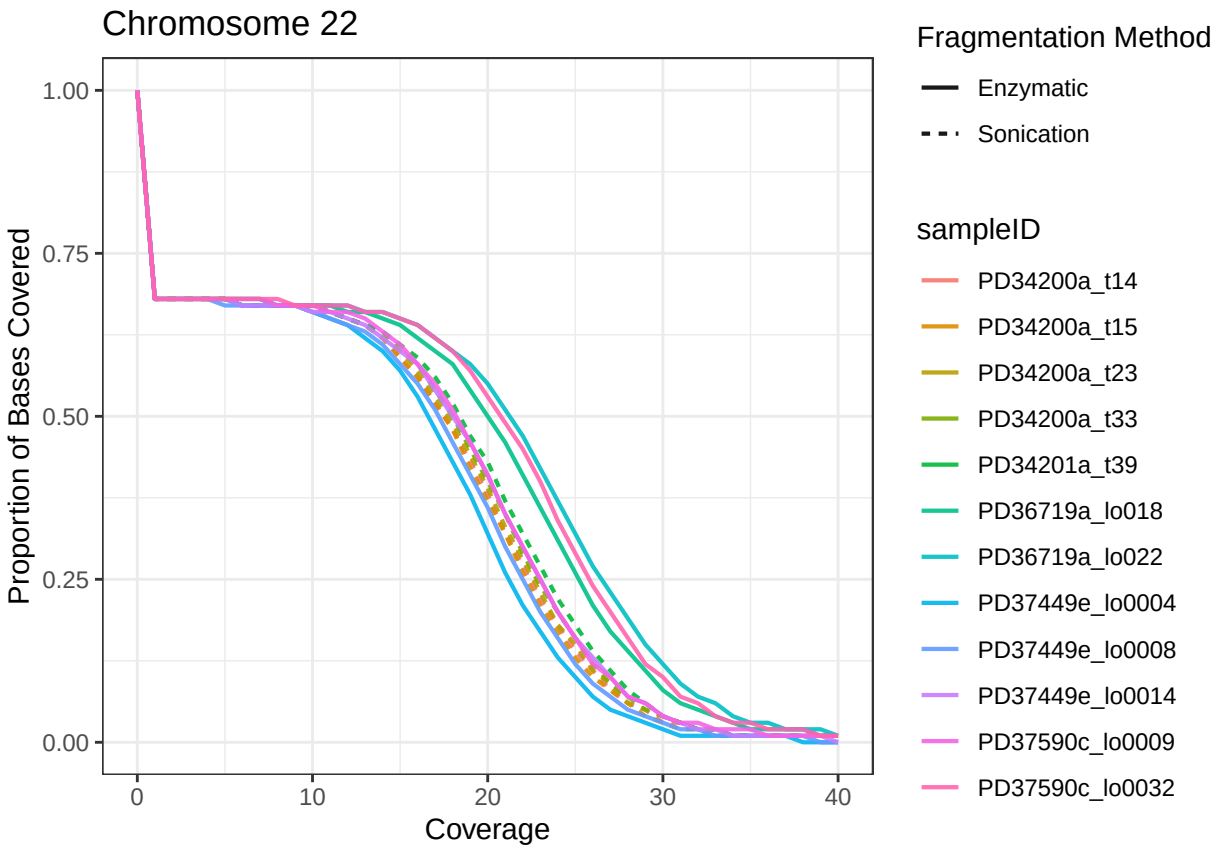

Chromosome 17
Proportion of Bases Covered
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Coverage
Fragmentation Method
Enzymatic
Sonication
sampleID
PD34200a_t14
PD34200a_t15
PD34200a_t23
PD34200a_t33
PD34201a_t39
PD36719a_lo018
PD36719a_lo022
PD37449e_lo0004
PD37449e_lo0008
PD37449e_lo0014
PD37590c_lo0009
PD37590c_lo0032

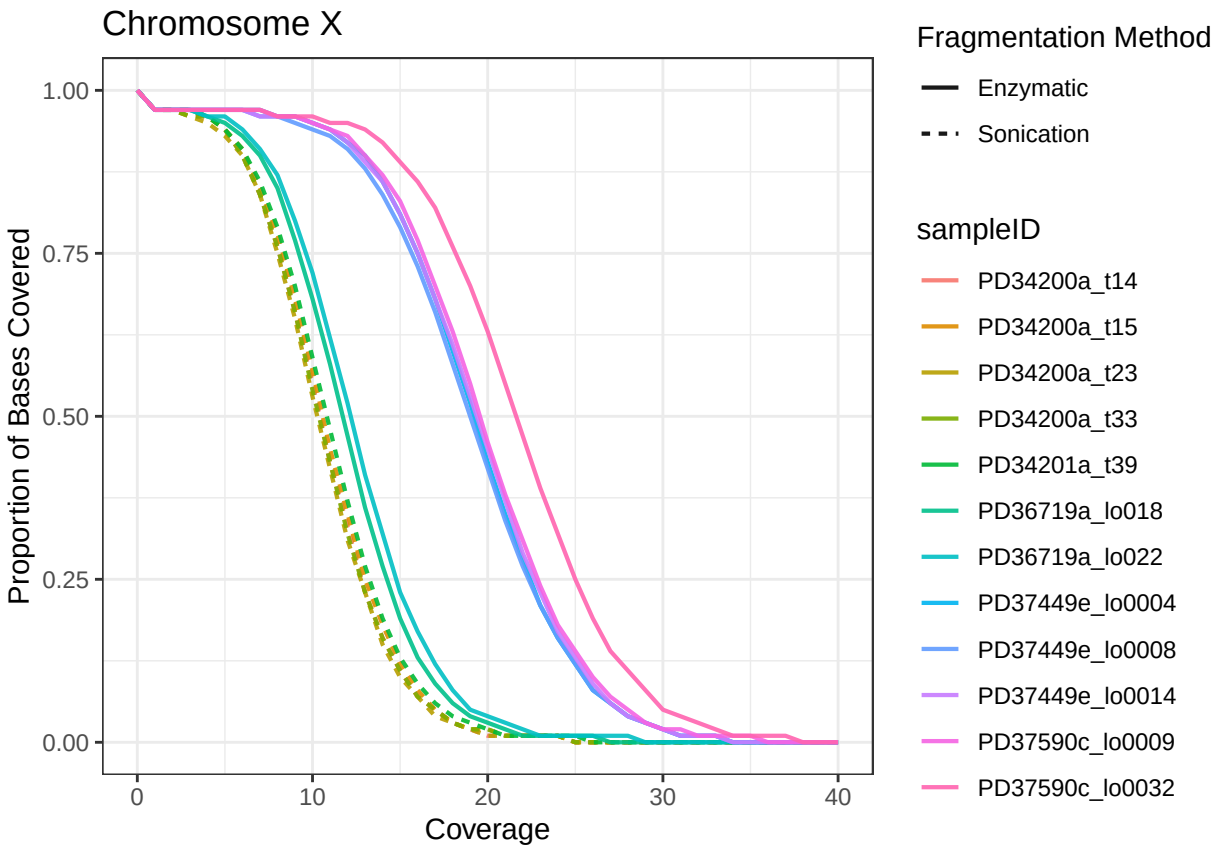

Chromosome 18
Proportion of Bases Covered
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Coverage
Fragmentation Method
Enzymatic
Sonication
sampleID
PD34200a_t14
PD34200a_t15
PD34200a_t23
PD34200a_t33
PD34201a_t39
PD36719a_lo018
PD36719a_lo022
PD37449e_lo0004
PD37449e_lo0008
PD37449e_lo0014
PD37590c_lo0009
PD37590c_lo0032

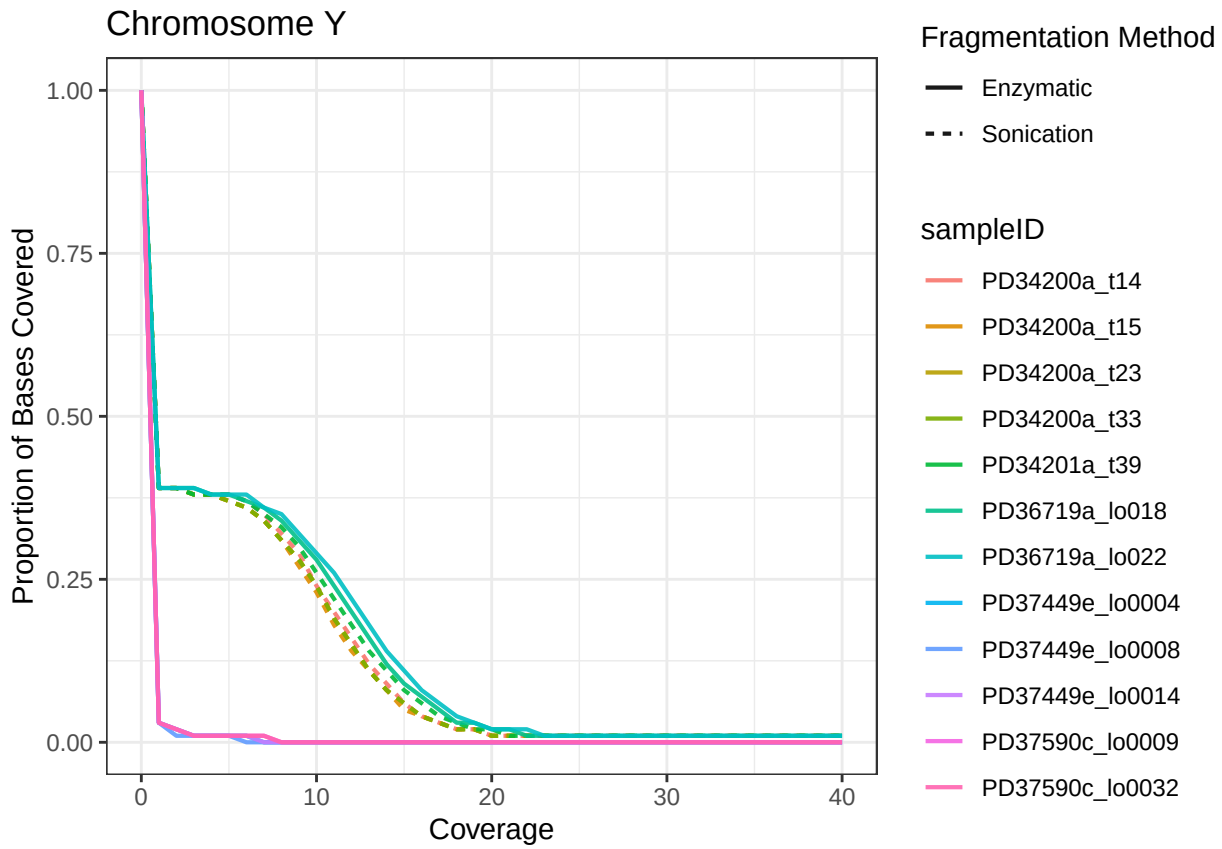

Chromosome 19
Proportion of Bases Covered
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Coverage
Fragmentation Method
Enzymatic
Sonication
sampleID
PD34200a_t14
PD34200a_t15
PD34200a_t23
PD34200a_t33
PD34201a_t39
PD36719a_lo018
PD36719a_lo022
PD37449e_lo0004
PD37449e_lo0008
PD37449e_lo0014
PD37590c_lo0009
PD37590c_lo0032


Chromosome 20
Proportion of Bases Covered
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Coverage
Fragmentation Method
Enzymatic
Sonication
sampleID
PD34200a_t14
PD34200a_t15
PD34200a_t23
PD34200a_t33
PD34201a_t39
PD36719a_lo018
PD36719a_lo022
PD37449e_lo0004
PD37449e_lo0008
PD37449e_lo0014
PD37590c_lo0009
PD37590c_lo0032


Chromosome 21
Proportion of Bases Covered
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Coverage
Fragmentation Method
Enzymatic
Sonication
sampleID
PD34200a_t14
PD34200a_t15
PD34200a_t23
PD34200a_t33
PD34201a_t39
PD36719a_lo018
PD36719a_lo022
PD37449e_lo0004
PD37449e_lo0008
PD37449e_lo0014
PD37590c_lo0009
PD37590c_lo0032


Chromosome 22
Proportion of Bases Covered
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Coverage
Fragmentation Method
Enzymatic
Sonication
sampleID
PD34200a_t14
PD34200a_t15
PD34200a_t23
PD34200a_t33
PD34201a_t39
PD36719a_lo018
PD36719a_lo022
PD37449e_lo0004
PD37449e_lo0008
PD37449e_lo0014
PD37590c_lo0009
PD37590c_lo0032


Chromosome X
Proportion of Bases Covered
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
Coverage
Fragmentation Method
Enzymatic
Sonication
sampleID
PD34200a_t14
PD34200a_t15
PD34200a_t23
PD34200a_t33
PD34201a_t39
PD36719a_lo018
PD36719a_lo022
PD37449e_lo0004
PD37449e_lo0008
PD37449e_lo0014
PD37590c_lo0009
PD37590c_lo0032



Grab stats of interest from mosdepth files

```
# calculate fraction of genome covered at various depths
thresholds_table <- lapply(threshold.files,mosdepth_thresh) %>% bind_rows()

head(thresholds_table)
```

```
##         sampleID no_analysis_frac cov_2X_frac cov_4X_frac cov_6X_frac
## 1   PD34200a_t14        0.9127584   0.8590431   0.8583380   0.8571117
## 2   PD34200a_t15        0.9127319   0.8589954   0.8582524   0.8569633
## 3   PD34200a_t23        0.9127791   0.8590275   0.8582471   0.8569012
## 4   PD34200a_t33        0.9126637   0.8589543   0.8582383   0.8570301
## 5   PD34201a_t39        0.9130080   0.8592566   0.8585469   0.8574264
## 6 PD36719a_lo018        0.9128503   0.8591792   0.8586292   0.8578011
##   cov_8X_frac cov_10X_frac cov_15X_frac cov_20X_frac cov_30X_frac
## 1   0.8544807    0.8477960    0.7595776    0.4701017   0.02971666
## 2   0.8541956    0.8471880    0.7558631    0.4623420   0.02828662
## 3   0.8540539    0.8470080    0.7571492    0.4674295   0.02974251
## 4   0.8545302    0.8483351    0.7662509    0.4869252   0.03405921
## 5   0.8553329    0.8506772    0.7906483    0.5522727   0.05055420
## 6   0.8563765    0.8534800    0.8161422    0.6341772   0.08580012
```

combine with sample metadata

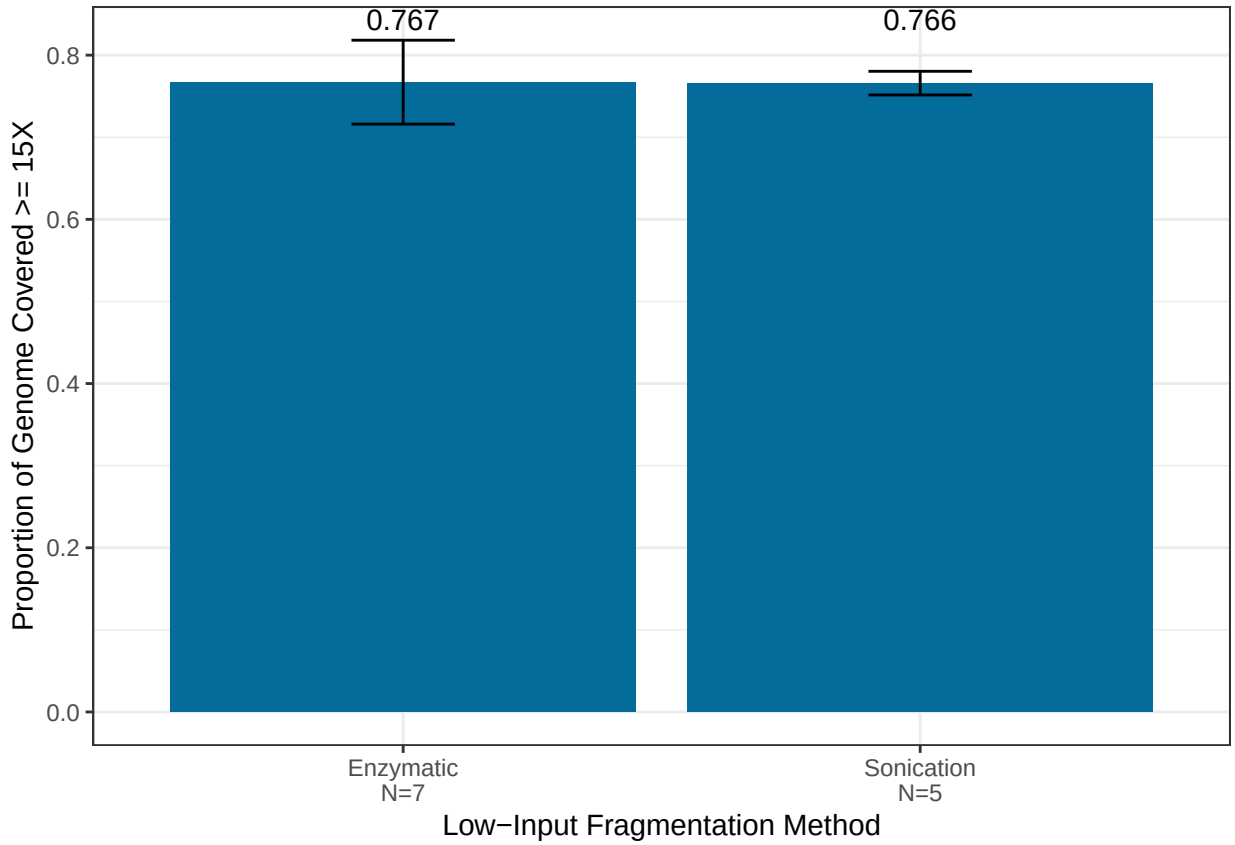
```
thresholds_table_meta <- left_join(sample_metadata, thresholds_table, by = "sampleID")
```

plot coverage comparison

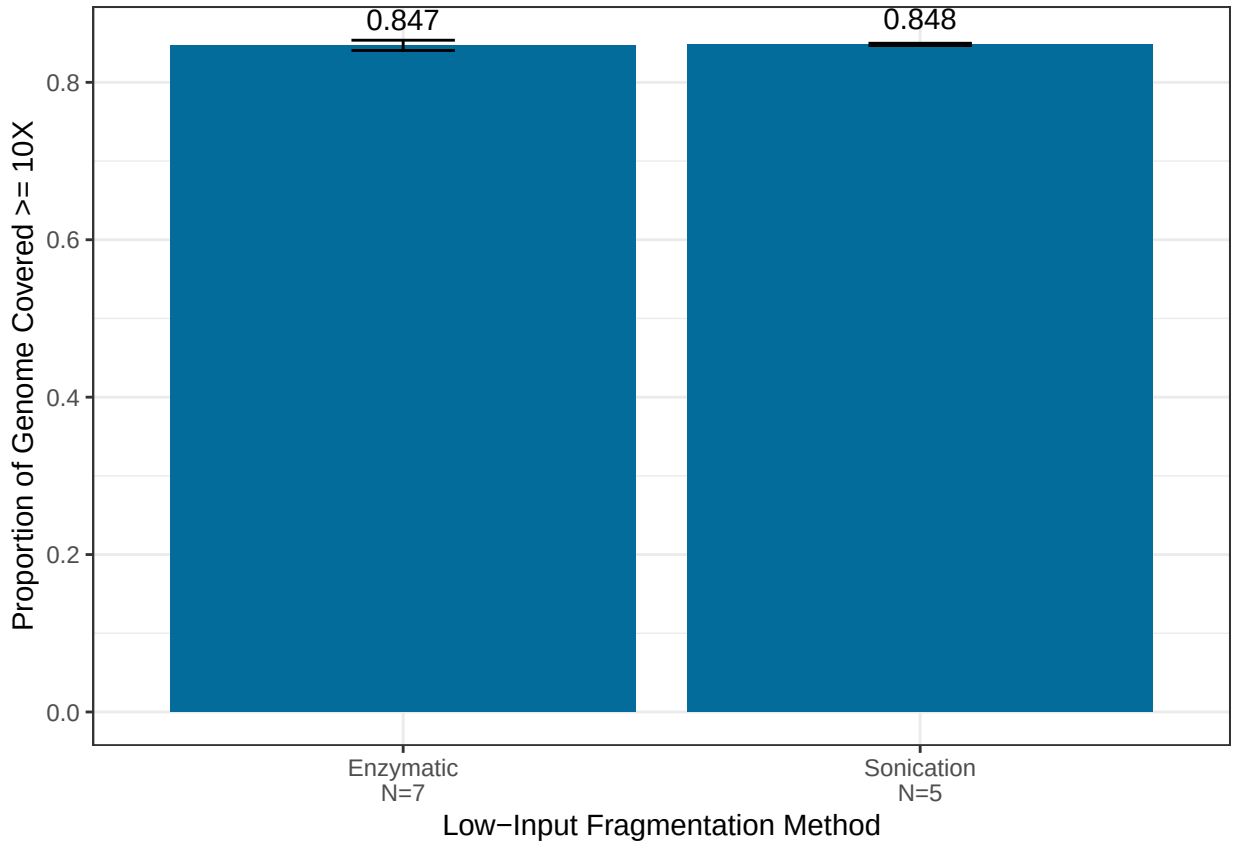
```
threshold_sum <- thresholds_table_meta %>%
  group_by(fragmentation_method) %>%
  dplyr::summarise(cov_15X = mean(cov_15X_frac),
                   sd_15X = sd(cov_15X_frac),
                   cov_10X = mean(cov_10X_frac),
                   sd_10X = sd(cov_10X_frac))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
my_xlab <- c("Enzymatic\nN=7", "Sonication\nN=5")

ggplot(threshold_sum) +
  theme_bw() +
  aes(x=fragmentation_method, y = cov_15X) +
  geom_col(fill = "#046C9A") +
  geom_errorbar(aes(ymin=cov_15X-sd_15X, ymax=cov_15X+sd_15X), width =0.2) +
  xlab("Low-Input Fragmentation Method") +
  ylab("Proportion of Genome Covered >= 15X") +
  scale_x_discrete(labels = my_xlab) +
  geom_text(aes(label=round(cov_15X, digits = 3)), position=position_dodge(width = 0.9), vjust =-2.5)
```




```
ggplot(threshold_sum) +
  theme_bw() +
  aes(x=fragmentation_method, y = cov_10X) +
  geom_col(fill = "#046C9A") +
  geom_errorbar(aes(ymin=cov_10X-sd_10X, ymax=cov_10X+sd_10X), width =0.2) +
  xlab("Low-Input Fragmentation Method") +
  ylab("Proportion of Genome Covered >= 10X") +
  scale_x_discrete(labels = my_xlab) +
  geom_text(aes(label=round(cov_10X, digits = 3)), position=position_dodge(width = 0.9), vjust =-0.75)
```



Overlapping Regions

Bedfiles of regions covered >10X were generated for each sample.

Consensus bedfiles for each fragmentation method were created, requiring consensus in >= 80% of samples

```
genome_size = 3095693981
shared_bases <- 2644281393
frag_unique_bases <- 4754732
sonic_unique_bases <- 11761529

shared_frac = shared_bases/genome_size
frag_frac = frag_unique_bases/genome_size
sonic_frac =  sonic_unique_bases/genome_size

overlap_tbl <- fread("~/nfs_tb14/Projects/LCM_Methods_Paper/revisions_analysis/coverage_comparisons/10x

my_xlab <- c("Common", "Unique to Sonication", "Unique to Enzymatic")

ggplot(overlap_tbl) +
```

```
  theme_bw() +
  aes(x=reorder(Overlap, -genomic_fraction), y=genomic_fraction) +
  geom_col() +
  geom_text(aes(label=round(genomic_fraction, digits = 3)), position=position_dodge(width = 0.9), vjust
  xlab("") +
  ylab("Proportion of Genome >= 10X\n >= 80% of Samples") +
  scale_x_discrete(labels = my_xlab)
```

Bedtools intersect was used to identify regions covered by both methods and unique to each