# Mutational signatures

Code

Hide

```
library(readr)
```

```
Warning: package 'readr' was built under R version 4.1.2
```

Load VCF samples

Hide

```
# Load InDels
indel_grl <- read_vcfs_as_granges(indel_vcf_files, group_names, ref_genome, type = "i
ndel")
indel_grl
```

```
GRangesList object of length 84:
$`WT-day0_chr1`
GRanges object with 449 ranges and 5 metadata columns:
                       seqnames              ranges strand | paramRangeID
REF                 ALT      QUAL
                          <Rle>           <IRanges>  <Rle> |     <factor> <DNAStringS
et> <DNAStringSetList> <numeric>
    chr1:3555815_GTC/G     chr1     3555815-3555817      * |           NA
GTC                  G     55.12
     chr1:4009853_G/GA     chr1             4009853      * |           NA
G                   GA     66.12
     chr1:4624693_CA/C     chr1     4624693-4624694      * |           NA
CA                   C     47.12
     chr1:4851624_TA/T     chr1     4851624-4851625      * |           NA
TA                   T     47.12
    chr1:4894076_TTG/T     chr1     4894076-4894078      * |           NA
TTG                  T     61.12
                   ...      ...                 ...    ... .          ...
...                ...       ...
  chr1:194095071_GGT/G     chr1 194095071-194095073      * |           NA
GGT                  G     48.12
  chr1:194301892_TTC/T     chr1 194301892-194301894      * |           NA
TTC                  T     44.12
   chr1:194365483_CT/C     chr1 194365483-194365484      * |           NA
CT                   C     43.12
   chr1:194846624_GC/G     chr1 194846624-194846625      * |           NA
GC                   G     41.02
```

```
  chr1:194852291_AAC/A     chr1 194852291-194852293      * |           NA
AAC                  A     50.02
                            FILTER
                        <character>
    chr1:3555815_GTC/G        PASS
     chr1:4009853_G/GA        PASS
     chr1:4624693_CA/C        PASS
     chr1:4851624_TA/T        PASS
    chr1:4894076_TTG/T        PASS
                    ...         ...
  chr1:194095071_GGT/G        PASS
  chr1:194301892_TTC/T        PASS
   chr1:194365483_CT/C        PASS
   chr1:194846624_GC/G        PASS
  chr1:194852291_AAC/A        PASS
  -------
  seqinfo: 21 sequences from mm10 genome

...
<83 more elements>
```

Hide

```
sample_names <-c("WT-day0", "WT-day77", "Pms1-day0", "Pms1-day77")


snp_vcf_files <- c("./cov_x5/WT-day0_tk39_filtered_snps_x5.vcf","./cov_x5/WT-day77_tk
39_filtered_snps_x5.vcf","./cov_x5/Pms1-day0_tk39_filtered_snps_x5.vcf","./cov_x5/Pms
1-day77_tk39_filtered_snps_x5.vcf")
indel_vcf_files <- c("./cov_x5/WT-day0_tk39_filtered_indels_x5.vcf","./cov_x5/WT-day7
7_tk39_filtered_indels_x5.vcf","./cov_x5/Pms1-day0_tk39_filtered_indels_x5.vcf","./co
v_x5/Pms1-day77_tk39_filtered_indels_x5.vcf")
group_names <- c("WT-day0", "WT-day77", "Pms1-day0", "Pms1-day77")


ref_genome <- "BSgenome.Mmusculus.UCSC.mm10"
library(ref_genome, character.only = TRUE)

groups <- c("WT-day0", "WT-day77", "Pms1-day0", "Pms1-day77")
genotype <- c(rep("WT", 2),rep("Pms1", 2))
time <- c(rep("day0", 1),rep("day77", 1),rep("day0", 1),rep("day77", 1))

# Load SNPs
snp_grl <- read_vcfs_as_granges(snp_vcf_files, group_names, ref_genome)
```

```
Any neighbouring SNVs will be merged into DBS/MBS variants.
Set the 'predefined_dbs_mbs' to 'TRUE' if you don't want this.
```

Hide

```
snp_grl
```

```
GRangesList object of length 4:
$`WT-day0`
GRanges object with 537 ranges and 5 metadata columns:
                    seqnames    ranges strand | paramRangeID            REF
ALT      QUAL
                       <Rle> <IRanges>  <Rle> |     <factor> <DNAStringSet> <DNAStrin
gSetList> <numeric>
  chr1:19143048_T/A     chr1  19143048      * |           NA              T
A     44.75
  chr1:26687461_G/A     chr1  26687461      * |           NA              G
A     83.03
  chr1:33853419_A/G     chr1  33853419      * |           NA              A
G     41.75
  chr1:33853432_A/T     chr1  33853432      * |           NA              A
T     47.82
  chr1:53208190_A/C     chr1  53208190      * |           NA              A
C     41.75
                ...      ...       ...    ... .          ...            ...
...       ...
   chrY:1364719_C/A     chrY   1364719      * |           NA              C
A     50.75
  chrY:90744547_G/C     chrY  90744547      * |           NA              G
C    446.10
  chrY:90744549_G/C     chrY  90744549      * |           NA              G
C    446.10
  chrY:90744585_G/A     chrY  90744585      * |           NA              G
A    206.08
  chrY:90744601_G/A     chrY  90744601      * |           NA              G
A    119.18
                      FILTER
                    <character>
  chr1:19143048_T/A        PASS
  chr1:26687461_G/A       QDlow
  chr1:33853419_A/G        PASS
  chr1:33853432_A/T        PASS
  chr1:53208190_A/C        PASS
                ...         ...
   chrY:1364719_C/A        PASS
  chrY:90744547_G/C        PASS
```

```
  chrY:90744549_G/C        PASS
  chrY:90744585_G/A        PASS
  chrY:90744601_G/A        PASS
  -------
  seqinfo: 21 sequences from mm10 genome

...
<3 more elements>
```

Hide

```
# Load InDels
indel_grl <- read_vcfs_as_granges(indel_vcf_files, group_names, ref_genome, type = "i
ndel")
indel_grl
```

```
GRangesList object of length 4:
$`WT-day0`
GRanges object with 686 ranges and 5 metadata columns:
                            seqnames              ranges strand | paramRangeID
REF
                               <Rle>           <IRanges>  <Rle> |     <factor> <DNAS
tringSet>
           chr1:6923386_GA/G     chr1     6923386-6923387      * |           NA
GA
           chr1:6923627_T/TA     chr1             6923627      * |           NA
T
  chr1:16437642_T/TGAGGAGGAG     chr1            16437642      * |           NA
T
        chr1:16496611_TTGC/T     chr1   16496611-16496614      * |           NA
TTGC
        chr1:16648011_GAAA/G     chr1   16648011-16648014      * |           NA
GAAA
                         ...      ...                 ...    ... .          ...
...
        chrX:113064791_GGA/G     chrX 113064791-113064793      * |           NA
GGA
         chrX:139823014_TA/T     chrX 139823014-139823015      * |           NA
TA
         chrX:157341920_G/GT     chrX           157341920      * |           NA
G
          chrY:90744554_TC/T     chrY   90744554-90744555      * |           NA
TC
      chrY:90744588_CCCTAG/C     chrY   90744588-90744593      * |           NA
CCCTAG
                                          ALT      QUAL      FILTER
```

```
                           <DNAStringSetList> <numeric> <character>
          chr1:6923386_GA/G                  G     44.84        PASS
          chr1:6923627_T/TA                 TA     61.72        PASS
  chr1:16437642_T/TGAGGAGGAG         TGAGGAGGAG    214.28        PASS
       chr1:16496611_TTGC/T                  T     40.95        PASS
       chr1:16648011_GAAA/G                  G     35.73        PASS
                        ...                ...       ...         ...
       chrX:113064791_GGA/G                  G    159.71        PASS
        chrX:139823014_TA/T                  T     45.28        PASS
        chrX:157341920_G/GT                 GT     59.75        PASS
         chrY:90744554_TC/T                  T    443.85        PASS
     chrY:90744588_CCCTAG/C                  C    204.85        PASS
  -------
  seqinfo: 21 sequences from mm10 genome

...
<3 more elements>
```

Hide

```
sample_names <-c("WT-day0", "WT-day77", "Pms1-day0", "Pms1-day77")


snp_vcf_files <- c("./cov_x2/WT-day0_tk39_filtered_snps.vcf","./cov_x2/WT-day77_tk39_
filtered_snps.vcf","./cov_x2/Pms1-day0_tk39_filtered_snps.vcf","./cov_x2/Pms1-day77_t
k39_filtered_snps.vcf")
indel_vcf_files <- c("./cov_x2/WT-day0_tk39_filtered_indels.vcf","./cov_x2/WT-day77_t
k39_filtered_indels.vcf","./cov_x2/Pms1-day0_tk39_filtered_indels.vcf","./cov_x2/Pms1
-day77_tk39_filtered_indels.vcf")
group_names <- c("WT-day0", "WT-day77", "Pms1-day0", "Pms1-day77")


ref_genome <- "BSgenome.Mmusculus.UCSC.mm10"
library(ref_genome, character.only = TRUE)

groups <- c("WT-day0", "WT-day77", "Pms1-day0", "Pms1-day77")
genotype <- c(rep("WT", 2),rep("Pms1", 2))
time <- c(rep("day0", 1),rep("day77", 1),rep("day0", 1),rep("day77", 1))

# Load SNPs
snp_grl <- read_vcfs_as_granges(snp_vcf_files, group_names, ref_genome)
```

```
Any neighbouring SNVs will be merged into DBS/MBS variants.
Set the 'predefined_dbs_mbs' to 'TRUE' if you don't want this.
```

Hide

Hide

```
snp_grl
```

```
GRangesList object of length 4:
$`WT-day0`
GRanges object with 28481 ranges and 5 metadata columns:
                    seqnames    ranges strand | paramRangeID            REF
ALT      QUAL
                       <Rle> <IRanges>  <Rle> |     <factor> <DNAStringSet> <DNAStrin
gSetList> <numeric>
   chr1:3425962_C/A     chr1   3425962      * |           NA              C
A     40.18
   chr1:3512366_T/C     chr1   3512366      * |           NA              T
C     40.18
   chr1:3559604_T/C     chr1   3559604      * |           NA              T
C     40.18
   chr1:3932140_G/T     chr1   3932140      * |           NA              G
T     40.18
   chr1:4108374_C/T     chr1   4108374      * |           NA              C
T     40.18
                ...      ...       ...    ... .          ...            ...
...       ...
  chrY:90744547_G/C     chrY  90744547      * |           NA              G
C    446.10
  chrY:90744549_G/C     chrY  90744549      * |           NA              G
C    446.10
  chrY:90744585_G/A     chrY  90744585      * |           NA              G
A    206.08
  chrY:90744601_G/A     chrY  90744601      * |           NA              G
A    119.18
  chrY:90805307_G/C     chrY  90805307      * |           NA              G
C     37.07
                      FILTER
                    <character>
   chr1:3425962_C/A        PASS
   chr1:3512366_T/C        PASS
   chr1:3559604_T/C        PASS
   chr1:3932140_G/T        PASS
   chr1:4108374_C/T        PASS
                ...         ...
  chrY:90744547_G/C        PASS
  chrY:90744549_G/C        PASS
  chrY:90744585_G/A        PASS
  chrY:90744601_G/A        PASS
  chrY:90805307_G/C        PASS
  -------
```

```
  seqinfo: 21 sequences from mm10 genome

...
<3 more elements>
```

Hide

```
# Load InDels
indel_grl <- read_vcfs_as_granges(indel_vcf_files, group_names, ref_genome, type = "i
ndel")
indel_grl
```

```
GRangesList object of length 4:
$`WT-day0`
GRanges object with 6029 ranges and 5 metadata columns:
                         seqnames            ranges strand | paramRangeID
REF
                            <Rle>         <IRanges>  <Rle> |     <factor> <DNAStringS
et>
      chr1:3555815_GTC/G     chr1   3555815-3555817      * |           NA
GTC
       chr1:4009853_G/GA     chr1           4009853      * |           NA
G
       chr1:4624693_CA/C     chr1   4624693-4624694      * |           NA
CA
       chr1:4851624_TA/T     chr1   4851624-4851625      * |           NA
TA
      chr1:4894076_TTG/T     chr1   4894076-4894078      * |           NA
TTG
                     ...      ...               ...    ... .          ...
...
     chrY:39960747_AAC/A     chrY 39960747-39960749      * |           NA
AAC
     chrY:76250767_CTG/C     chrY 76250767-76250769      * |           NA
CTG
     chrY:86118528_GGA/G     chrY 86118528-86118530      * |           NA
GGA
      chrY:90744554_TC/T     chrY 90744554-90744555      * |           NA
TC
  chrY:90744588_CCCTAG/C     chrY 90744588-90744593      * |           NA         CCC
TAG
                                      ALT      QUAL      FILTER
                         <DNAStringSetList> <numeric> <character>
      chr1:3555815_GTC/G                  G     55.12        PASS
       chr1:4009853_G/GA                 GA     66.12        PASS
       chr1:4624693_CA/C                  C     47.12        PASS
```

```
       chr1:4851624_TA/T                  T     47.12        PASS
      chr1:4894076_TTG/T                  T     61.12        PASS
                     ...                ...       ...         ...
     chrY:39960747_AAC/A                  A     60.98        PASS
     chrY:76250767_CTG/C                  C     44.12        PASS
     chrY:86118528_GGA/G                  G     58.12        PASS
      chrY:90744554_TC/T                  T    443.85        PASS
  chrY:90744588_CCCTAG/C                  C    204.85        PASS
  -------
  seqinfo: 21 sequences from mm10 genome

...
<3 more elements>
```

Remove filtered sites

Hide

```
snp_grl_pass <- list()
indel_grl_pass <- list()
for (i in names(snp_grl)){
  snp_grl_pass[[i]] <- subset(snp_grl[[i]], FILTER == "PASS")
  indel_grl_pass[[i]] <- subset(indel_grl[[i]], FILTER == "PASS")
}
```

SNVs

Base substitution types

You can retrieve base substitution types from the VCF GRanges object as “REF>ALT” using mutations_from_vcf:

Hide

```
muts <- mutations_from_vcf(snp_grl_pass[[1]])
head(muts, 12)
```

```
 [1] "C>A" "T>C" "T>C" "G>T" "C>T" "T>C" "C>A" "G>T" "C>A" "C>A" "C>A" "C>T"
```

Hide

```
types <- mut_type(snp_grl_pass[[1]])
head(types, 12)
```

```
 [1] "C>A" "T>C" "T>C" "C>A" "C>T" "T>C" "C>A" "C>A" "C>A" "C>A" "C>A" "C>T"
```

Hide

```
context <- mut_context(snp_grl_pass[[1]], ref_genome)
head(context, 12)
```

```
 chr1  chr1  chr1  chr1  chr1  chr1  chr1  chr1  chr1  chr1  chr1  chr1
"GCT" "CTG" "ATT" "GGA" "TCA" "TTC" "ACA" "TGA" "GCG" "GCT" "ACA" "TCA"
```

Hide

```
type_context <- type_context(snp_grl_pass[[1]], ref_genome)
lapply(type_context, head, 12)
```

```
$types
 [1] "C>A" "T>C" "T>C" "C>A" "C>T" "T>C" "C>A" "C>A" "C>A" "C>A" "C>A" "C>T"

$context
 chr1  chr1  chr1  chr1  chr1  chr1  chr1  chr1  chr1  chr1  chr1  chr1
"GCT" "CTG" "ATT" "TCC" "TCA" "TTC" "ACA" "TCA" "GCG" "GCT" "ACA" "TCA"
```

Hide

```
type_occurrences <- mut_type_occurrences(snp_grl_pass, ref_genome)
type_occurrences
```

| | C>A | C>G | C>T | T>A | T>C | T>G | C>T at CpG | C>T other |
|---|---|---|---|---|---|---|---|---|
| | <int> | <int> | <int> | <int> | <int> | <int> | <int> | <int> |
| WT-day0 | 8279 | 1299 | 5691 | 2030 | 9821 | 697 | 666 | 5025 |
| WT-day77 | 9463 | 1357 | 5496 | 1745 | 9431 | 752 | 628 | 4868 |
| Pms1-day0 | 6040 | 1113 | 5037 | 1830 | 9214 | 642 | 601 | 4436 |
| Pms1-day77 | 9843 | 1475 | 5994 | 2094 | 10428 | 810 | 749 | 5245 |

4 rows

Hide

```
NA
```

# Fisher’s exact test of mutation types

Hide

```
type_occurrences_norm_t <- t(type_occurrences_norm)
my_fisher_p <-c()
my_conf_int_dn <-c()
my_conf_int_up <-c()
my_odds <- c()
my_fdr <- c()

for (mut in 1:nrow(type_occurrences_norm_t)){
  my_matrix <- t(matrix(type_occurrences_norm_t[mut,], nrow = 2, dimnames = list( Tim
epoint = c("d0", "d77"),
                        Genotype = c("WT", "Pms1"))))
  my_mut_name <- rownames(type_occurrences_norm_t)[mut]
  print(my_mut_name)
  print(my_matrix)
  my_test <- fisher.test(my_matrix)
  print(my_test)
  my_fisher_p <- c(my_fisher_p, my_test$p.value)
  my_conf_int_dn <- c(my_conf_int_dn, my_test$conf.int[1])
  my_conf_int_up <- c(my_conf_int_up, my_test$conf.int[2])
  my_odds <- c(my_odds, my_test$estimate)
  my_fdr <- c(my_fdr, p.adjust(p = my_test$p.value, method = "BH", n = nrow(type_occu
rrences_norm_t)))

}
```

```
[1] "C>A"
        Timepoint
Genotype   d0  d77
    WT   2471 2805
    Pms1 2089 2687

    Fisher's Exact Test for Count Data

data:  my_matrix
p-value = 0.001877
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 1.046534 1.226850
sample estimates:
odds ratio
  1.133086

[1] "C>G"
        Timepoint
Genotype  d0 d77
```

```
    WT   388 402
    Pms1 385 403

    Fisher's Exact Test for Count Data

data:  my_matrix
p-value = 0.9199
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.8252233 1.2368935
sample estimates:
odds ratio
  1.010294

[1] "C>T"
        Timepoint
Genotype   d0  d77
    WT   1698 1629
    Pms1 1742 1636

    Fisher's Exact Test for Count Data

data:  my_matrix
p-value = 0.6779
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.8884766 1.0786029
sample estimates:
odds ratio
 0.9789266

[1] "T>A"
        Timepoint
Genotype  d0 d77
    WT   606 517
    Pms1 633 572

    Fisher's Exact Test for Count Data

data:  my_matrix
p-value = 0.5061
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.8968763 1.2508951
sample estimates:
odds ratio
  1.059143
```

```
[1] "T>C"
        Timepoint
Genotype   d0  d77
    WT   2931 2795
    Pms1 3187 2846

    Fisher's Exact Test for Count Data

data:  my_matrix
p-value = 0.07631
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.8704766 1.0074501
sample estimates:
odds ratio
 0.9364641

[1] "T>G"
        Timepoint
Genotype  d0 d77
    WT   208 223
    Pms1 222 221

    Fisher's Exact Test for Count Data

data:  my_matrix
p-value = 0.589
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.7058685 1.2214320
sample estimates:
odds ratio
 0.9286132

[1] "C>T at CpG"
        Timepoint
Genotype  d0 d77
    WT   199 186
    Pms1 208 204

    Fisher's Exact Test for Count Data

data:  my_matrix
p-value = 0.7768
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
```

```
 0.7869958 1.3991067
sample estimates:
odds ratio
  1.049221

[1] "C>T other"
        Timepoint
Genotype   d0  d77
    WT   1500 1443
    Pms1 1534 1432

    Fisher's Exact Test for Count Data

data:  my_matrix
p-value = 0.567
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.8750866 1.0760213
sample estimates:
odds ratio
 0.9703735
```

Hide

```
type_occurrence_report <- tibble(as.data.frame(cbind(mutation = rownames(type_occurre
nces_norm_t),type_occurrences_norm_t, p.value = round(my_fisher_p, digits = 4), p.adj
= round(my_fdr, digits = 4), conf.int.down = round(my_conf_int_dn,digits = 2), conf.i
nt.up = round(my_conf_int_up, digits = 2), odds.ratio = round(my_odds, digits = 2))))

filter(type_occurrence_report, type_occurrence_report$p.value <= 0.1)
```

| mutation <chr> | WT-day0 <chr> | WT-day77 <chr> | Pms1-day0 <chr> | Pms1-day77 <chr> | p.value <chr> | p.adj <chr> | conf.int.down <chr> | conf.int.u <chr> |
|---|---|---|---|---|---|---|---|---|
| C>A | 2471 | 2805 | 2089 | 2687 | 0.0019 | 0.015 | 1.05 | 1.23 |
| T>C | 2931 | 2795 | 3187 | 2846 | 0.0763 | 0.6105 | 0.87 | 1.01 |

2 rows

Hide

```
type_occurrence_report
```

| WT- | WT- | Pms1- | Pms1- |
|---|---|---|---|

| mutation <chr> | day0 <chr> | day77 <chr> | day0 <chr> | day77 <chr> | p.value <chr> | p.adj <chr> | conf.int.down <chr> | conf.in <chr> |
|---|---|---|---|---|---|---|---|---|
| C>A | 2471 | 2805 | 2089 | 2687 | 0.0019 | 0.015 | 1.05 | 1.23 |
| C>G | 388 | 402 | 385 | 403 | 0.9199 | 1 | 0.83 | 1.24 |
| C>T | 1698 | 1629 | 1742 | 1636 | 0.6779 | 1 | 0.89 | 1.08 |
| T>A | 606 | 517 | 633 | 572 | 0.5061 | 1 | 0.9 | 1.25 |
| T>C | 2931 | 2795 | 3187 | 2846 | 0.0763 | 0.6105 | 0.87 | 1.01 |
| T>G | 208 | 223 | 222 | 221 | 0.589 | 1 | 0.71 | 1.22 |
| C>T at CpG | 199 | 186 | 208 | 204 | 0.7768 | 1 | 0.79 | 1.4 |
| C>T other | 1500 | 1443 | 1534 | 1432 | 0.567 | 1 | 0.88 | 1.08 |

8 rows | 1-9 of 10 columns

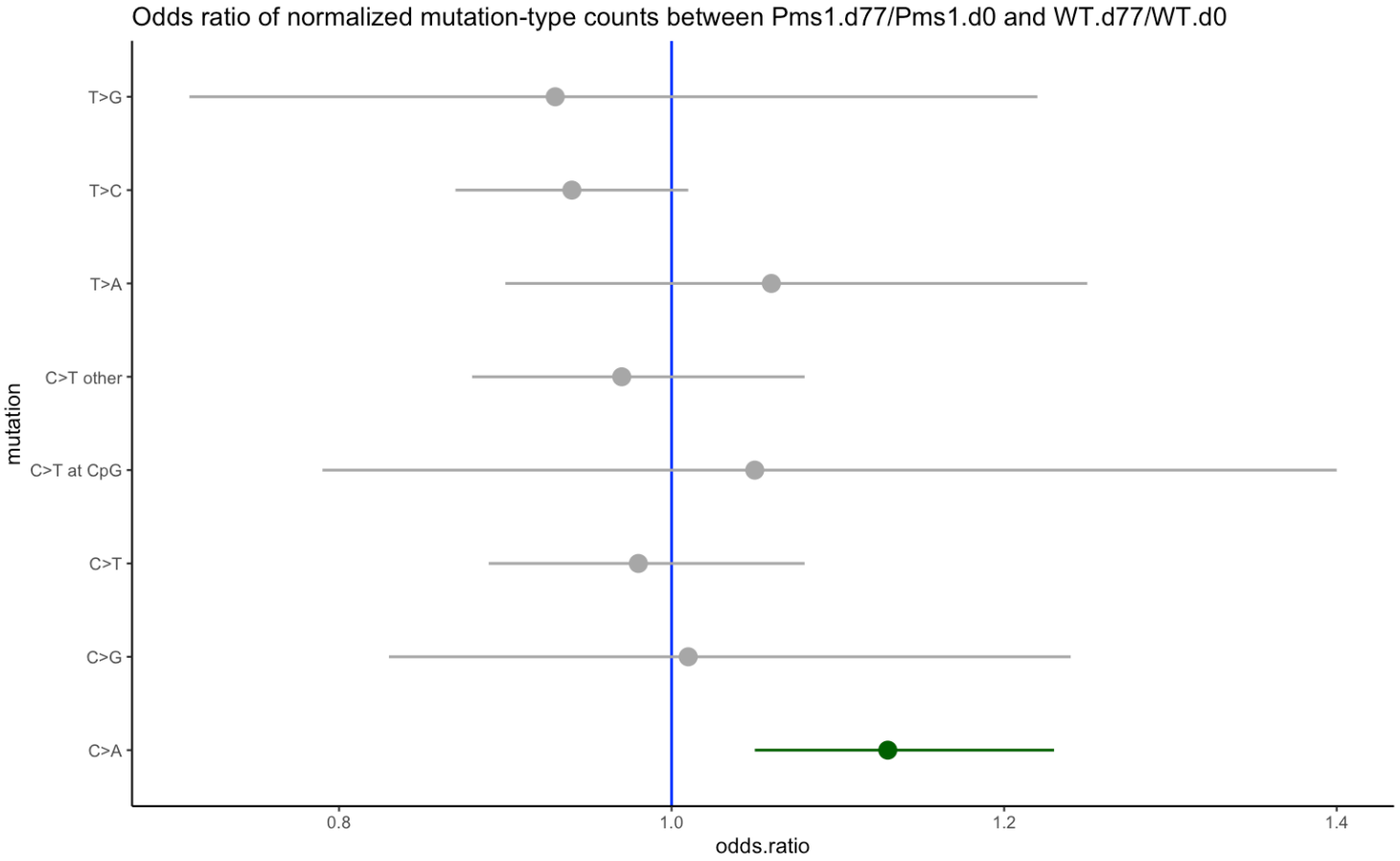
Hide

```
library(ggplot2)
library(cowplot)
library(extrafont)

type_occurrence_report$odds.ratio <- as.numeric(type_occurrence_report$odds.ratio)
type_occurrence_report$conf.int.down <- as.numeric(type_occurrence_report$conf.int.do
wn)
type_occurrence_report$conf.int.up <- as.numeric(type_occurrence_report$conf.int.up)


p.occurrence.coi <- ggplot(data = type_occurrence_report, aes(x = mutation, y = odds.
ratio, ymin = conf.int.down, ymax = conf.int.up)) +
  geom_hline(yintercept = 1, color = "blue" ) +
  geom_pointrange(colour=ifelse(type_occurrence_report$p.adj <= 0.05,"darkgreen",ifel
se(type_occurrence_report$p.value <= 0.05,"orange","darkgrey"))) +
  ggtitle("Odds ratio of normalized mutation-type counts between Pms1.d77/Pms1.d0 and
WT.d77/WT.d0") +
  theme_classic(base_size = 8) +
  coord_flip()

ggsave2(filename = "mut_ocurrence_report.pdf", plot = p.occurrence.coi)
```

```
Saving 7 x 7 in image
```

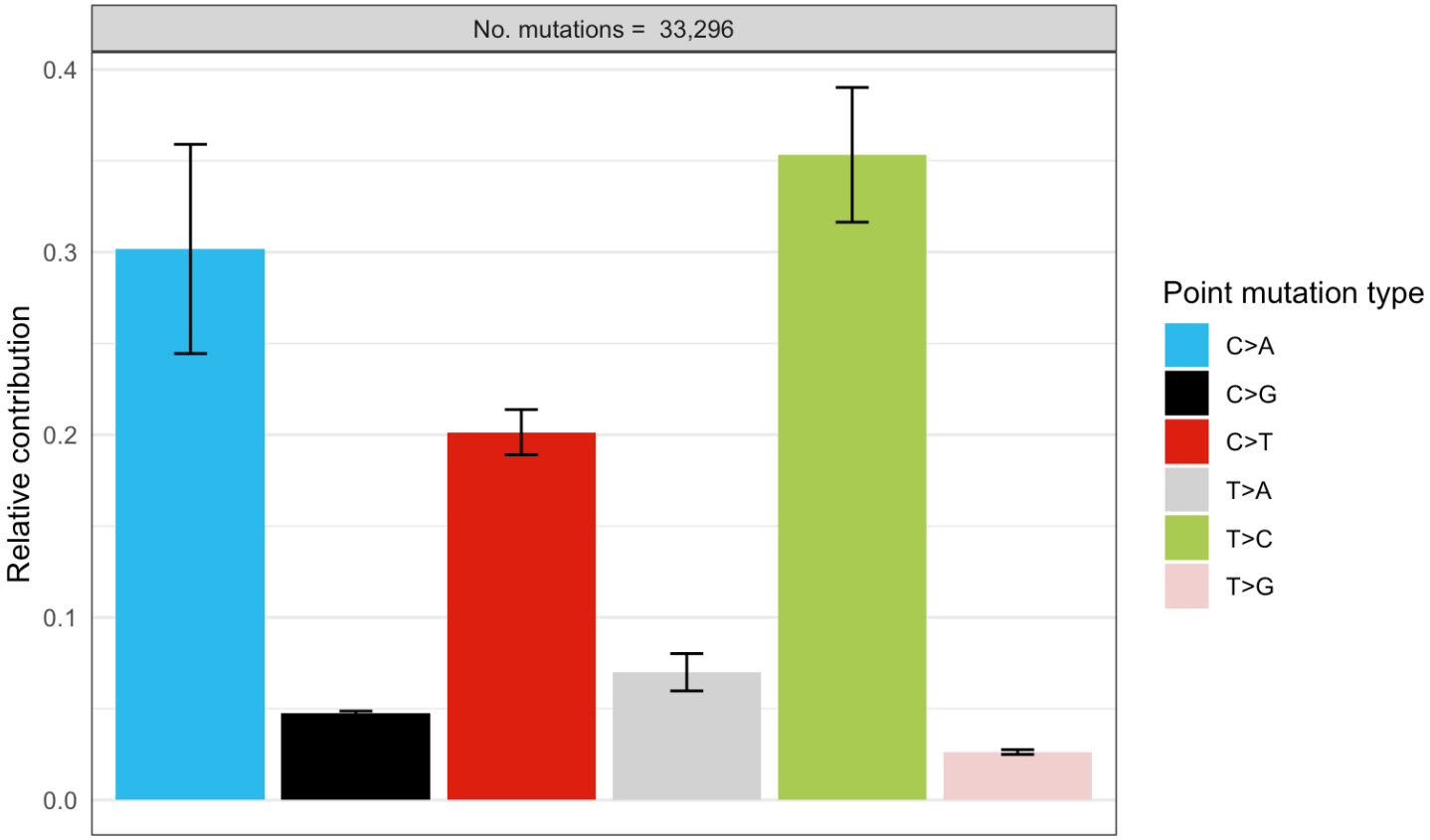
Hide

```
p.occurrence.coi
```



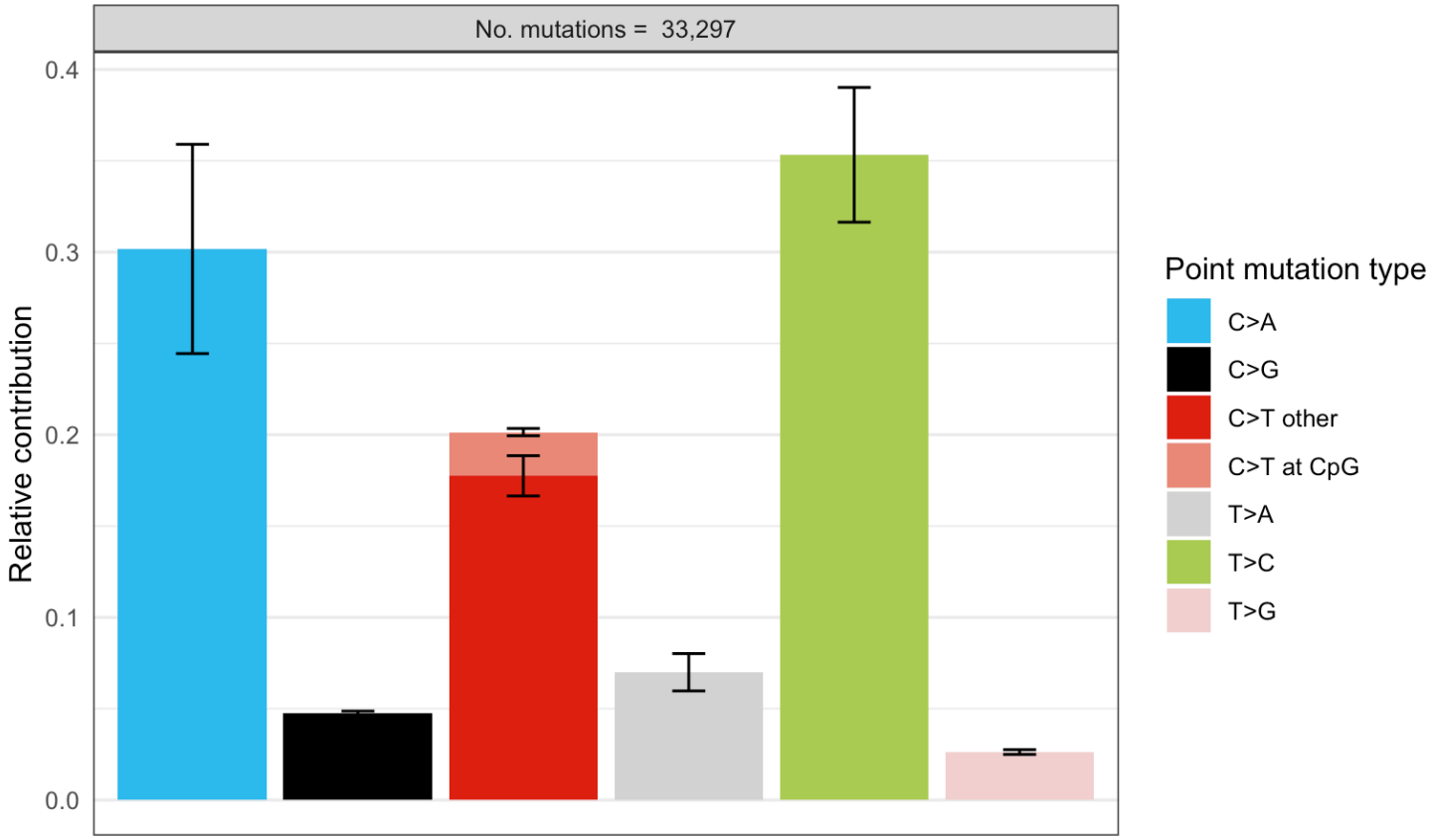

Hide

```
p1 <- plot_spectrum(type_occurrences_norm)
p1
```



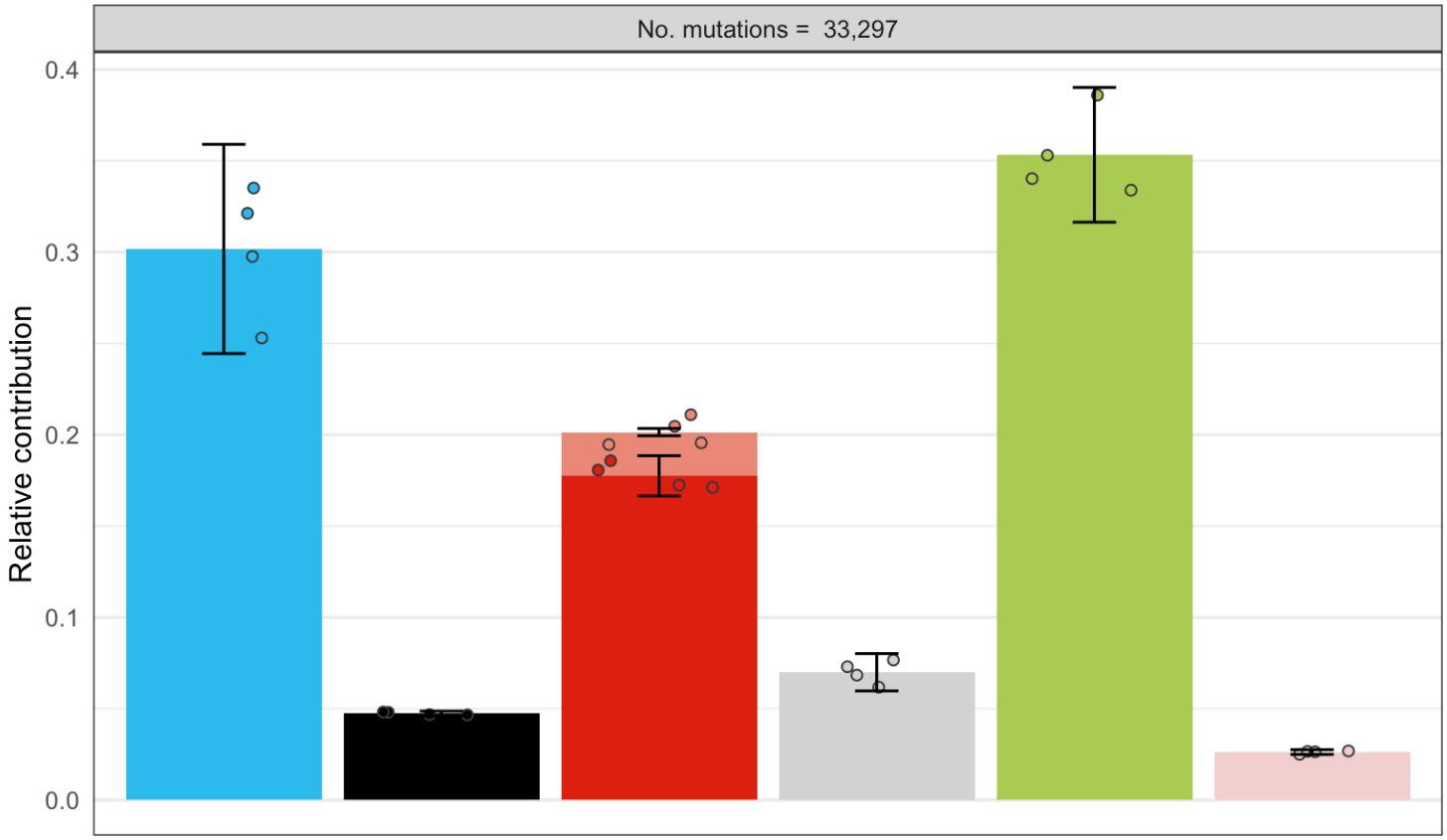

Hide

```
p2 <- plot_spectrum(type_occurrences_norm, CT = TRUE)
p2
```



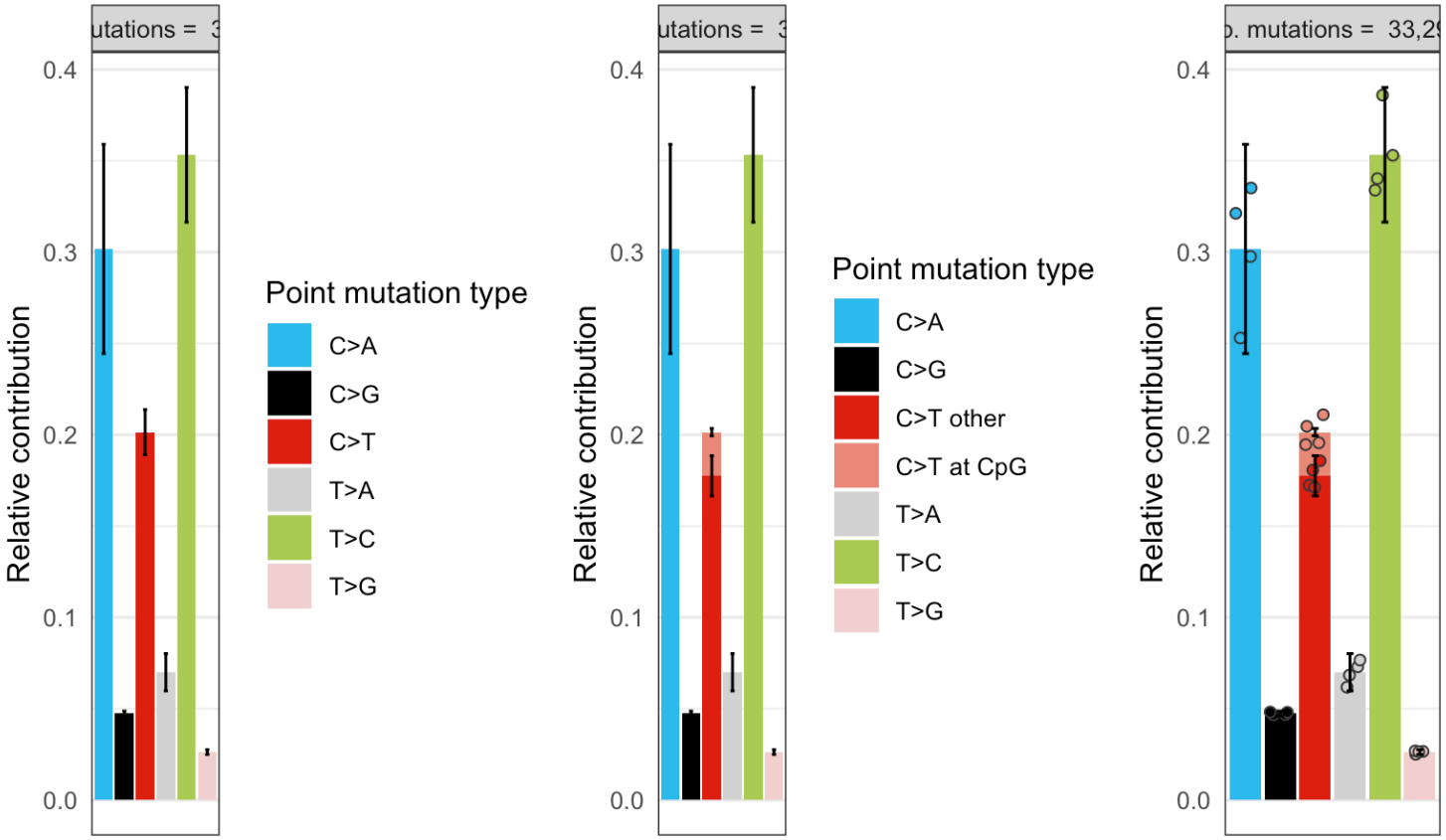

Hide

```
p3 <- plot_spectrum(type_occurrences_norm, CT = TRUE,
                    indv_points = TRUE, legend = FALSE)
p3
```




Hide

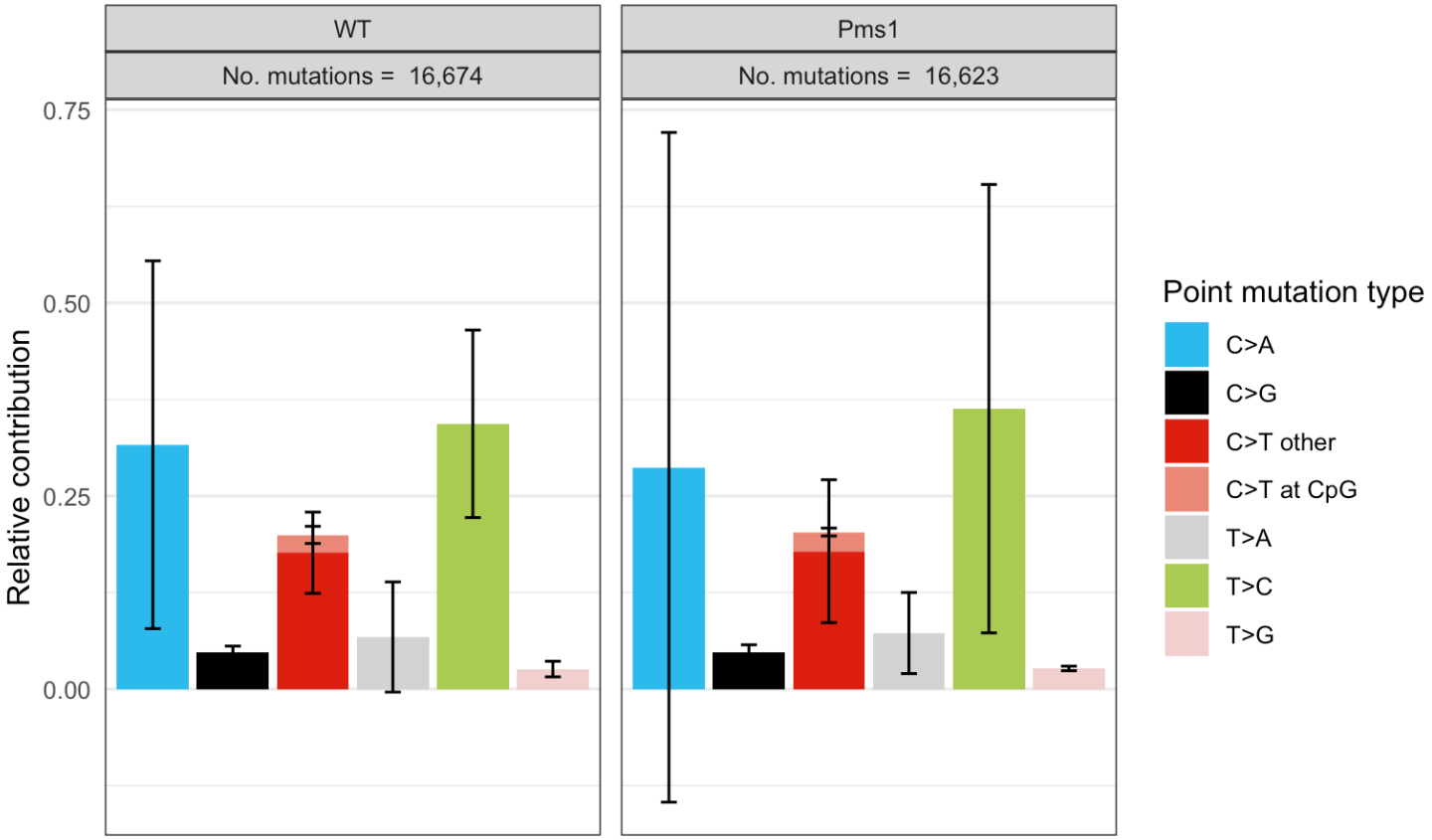
```
library("gridExtra")
grid.arrange(p1, p2, p3, ncol = 3, widths = c(3, 3, 1.75))
```




It's also possible to create a facet per sample group, e.g. plot the spectrum for each tissue separately:

Hide

```
p4 <- plot_spectrum(type_occurrences_norm, by = genotype, CT = TRUE, legend = TRUE)
p4
```




Or you could use the standard deviation instead of a 95% confidence interval:

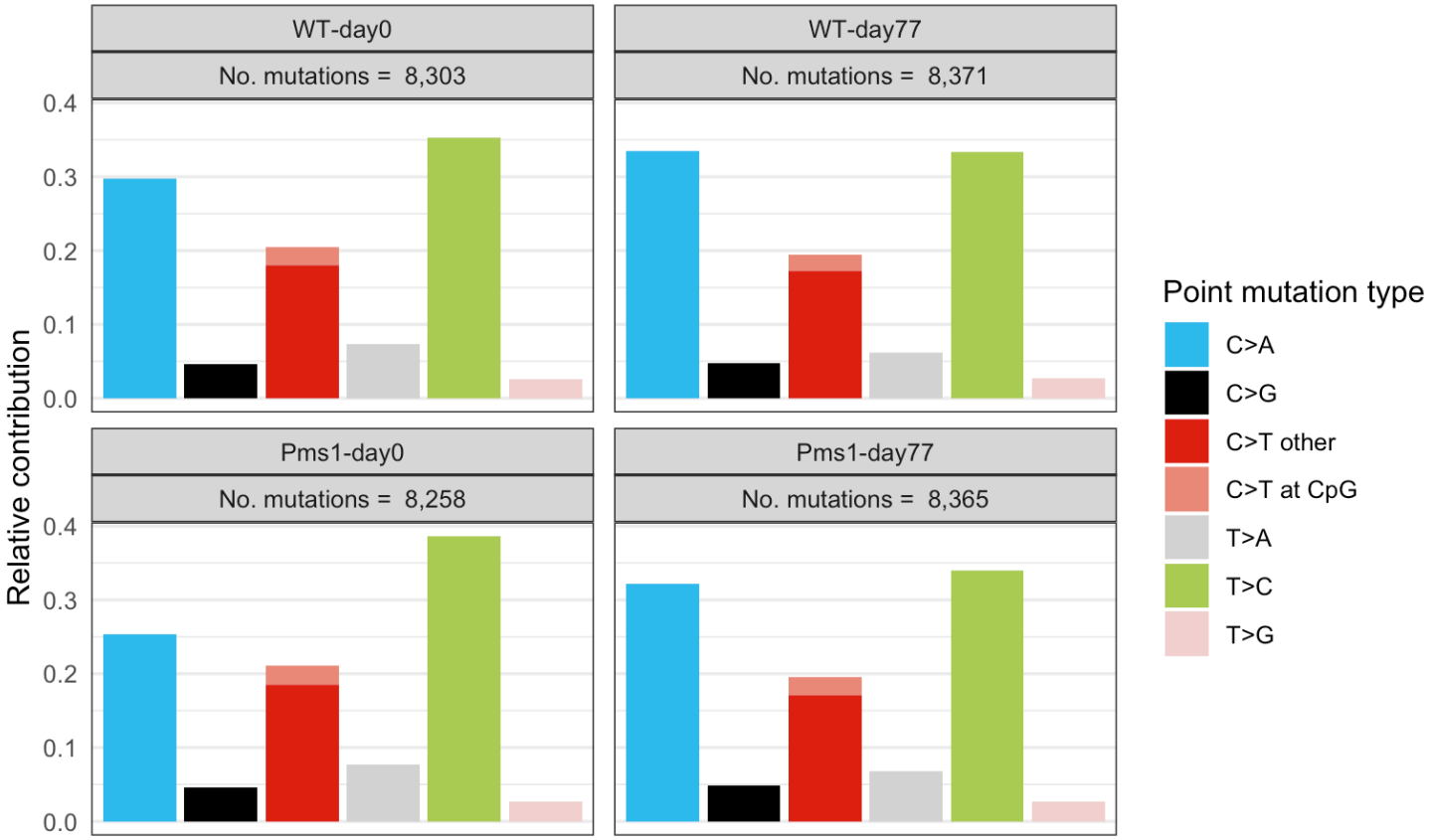
Hide

```
p5 <- plot_spectrum(type_occurrences_norm, by = groups, CT = TRUE,
                    legend = TRUE, error_bars = "stdev")
```

```
Warning: No error bars can be plotted, because there is only one sample per mutation
spectrum.
              Use the argument: `error_bars = 'none'`, if you want to avoid this warn
ing.
```

Hide

```
p5
```




96 mutational profile

First you should make a 96 trinucleodide mutation count matrix. (In contrast to previous versions this also works for single samples.)

Hide

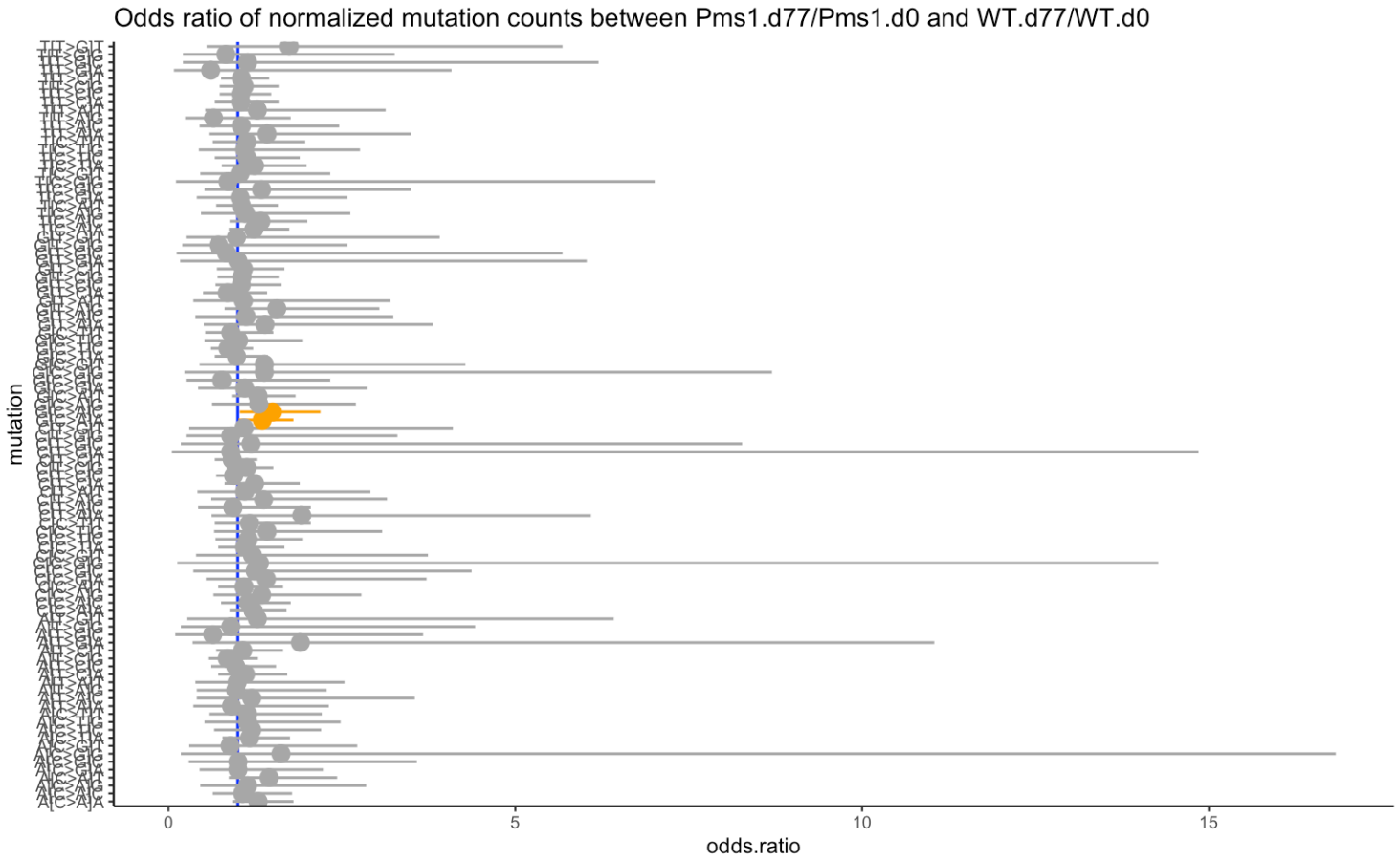
```
write.tsv(mut_mat_report, file = "mut_mat_report.txt")
```

```
Error in write.tsv(mut_mat_report, file = "mut_mat_report.txt") :
  could not find function "write.tsv"
```

Plot confidence intervals and odda.ratios

Hide

```
library(ggplot2)
library(cowplot)
library(extrafont)

mut_mat_report$odds.ratio <- as.numeric(mut_mat_report$odds.ratio)
mut_mat_report$conf.int.down <- as.numeric(mut_mat_report$conf.int.down)
mut_mat_report$conf.int.up <- as.numeric(mut_mat_report$conf.int.up)


p.coi <- ggplot(data = mut_mat_report, aes(x = mutation, y = odds.ratio, ymin = conf.
int.down, ymax = conf.int.up)) +
  geom_hline(yintercept = 1, color = "blue" ) +
  geom_pointrange(colour=ifelse(mut_mat_report$p.adj <= 0.05,"darkgreen",ifelse(mut_m
at_report$p.value <= 0.05,"orange","darkgrey"))) +
  ggtitle("Odds ratio of normalized mutation counts between Pms1.d77/Pms1.d0 and WT.d
77/WT.d0") +
  theme_classic(base_size = 8) +
  coord_flip()


ggsave2(filename = "mut_matrix_report.pdf", plot = p.coi, height = 11, width = 8)
p.coi
```



Green denotes mutations having a p.value <= 0.05.. In all cases adjusted p.values are > 0.05.

Next, you can use this matrix to plot the 96 profile of samples. In this example we do this for 2 samples:

Hide

```
p96_profile <- plot_96_profile(mut_mat_norm[, c(1:4)], condensed = TRUE, ymax = 0.05)
+ scale_y_continuous(breaks = c(0.01,0.02,0.03,0.04,0.05))
```

```
Scale for y is already present.
Adding another scale for y, which will replace the existing scale.
```

Hide

```
ggsave2(filename = "p96_profile_norm.pdf", plot = p96_profile)
```

```
Saving 7 x 7 in image
```

It's also possible to look at larger mutational contexts. However, this is only usefull if you have a large number of mutations.

Hide

```
mut_mat_ext_context_norm
```

```
          WT-day0 WT-day77 Pms1-day0 Pms1-day77
AA[C>A]AA      14        14          9          11
AA[C>A]AC       9         8          6          10
AA[C>A]AG       9        10         10          12
AA[C>A]AT       7         8          8          11
AA[C>A]CA      10         8          5           7
AA[C>A]CC       7         6          4           6
AA[C>A]CG       1         1          0           1
AA[C>A]CT       5         5          6           6
AA[C>A]GA       1         1          1           2
AA[C>A]GC       2         1          1           1
AA[C>A]GG       1         1          1           1
AA[C>A]GT       2         1          1           2
AA[C>A]TA       5         4          4           4
AA[C>A]TC       5         3          5           5
AA[C>A]TG       7         6          4           7
AA[C>A]TT       4         3          4           3
AC[C>A]AA      14        15          9          13
AC[C>A]AC      11        15          8          18
```

| | | | | |
|---|---|---|---|---|
| AC[C>A]AG | 13 | 14 | 9 | 17 |
| AC[C>A]AT | 8 | 13 | 9 | 13 |
| AC[C>A]CA | 10 | 12 | 10 | 15 |
| AC[C>A]CC | 6 | 8 | 5 | 8 |
| AC[C>A]CG | 1 | 2 | 2 | 1 |
| AC[C>A]CT | 8 | 9 | 6 | 8 |
| AC[C>A]GA | 2 | 4 | 2 | 3 |
| AC[C>A]GC | 4 | 3 | 1 | 2 |
| AC[C>A]GG | 2 | 3 | 2 | 3 |
| AC[C>A]GT | 3 | 3 | 3 | 5 |
| AC[C>A]TA | 4 | 5 | 3 | 5 |
| AC[C>A]TC | 8 | 9 | 3 | 8 |
| AC[C>A]TG | 8 | 11 | 7 | 9 |
| AC[C>A]TT | 5 | 7 | 5 | 5 |
| AG[C>A]AA | 23 | 20 | 19 | 24 |
| AG[C>A]AC | 16 | 19 | 17 | 21 |
| AG[C>A]AG | 26 | 24 | 23 | 29 |
| AG[C>A]AT | 20 | 19 | 16 | 19 |
| AG[C>A]CA | 20 | 20 | 16 | 21 |
| AG[C>A]CC | 15 | 13 | 9 | 15 |
| AG[C>A]CG | 2 | 2 | 2 | 2 |
| AG[C>A]CT | 14 | 14 | 10 | 17 |
| AG[C>A]GA | 3 | 3 | 4 | 4 |
| AG[C>A]GC | 5 | 4 | 1 | 3 |
| AG[C>A]GG | 4 | 2 | 2 | 4 |
| AG[C>A]GT | 3 | 3 | 2 | 2 |
| AG[C>A]TA | 10 | 10 | 7 | 11 |
| AG[C>A]TC | 12 | 14 | 9 | 12 |
| AG[C>A]TG | 19 | 19 | 12 | 19 |
| AG[C>A]TT | 10 | 12 | 6 | 11 |
| AT[C>A]AA | 6 | 8 | 6 | 11 |
| AT[C>A]AC | 9 | 13 | 7 | 13 |
| AT[C>A]AG | 9 | 10 | 11 | 14 |
| AT[C>A]AT | 9 | 10 | 6 | 11 |
| AT[C>A]CA | 11 | 10 | 6 | 10 |
| AT[C>A]CC | 9 | 9 | 6 | 9 |
| AT[C>A]CG | 1 | 1 | 1 | 1 |
| AT[C>A]CT | 9 | 8 | 7 | 10 |
| AT[C>A]GA | 1 | 2 | 1 | 2 |
| AT[C>A]GC | 2 | 1 | 2 | 1 |
| AT[C>A]GG | 1 | 2 | 1 | 2 |
| AT[C>A]GT | 2 | 2 | 2 | 2 |
| AT[C>A]TA | 3 | 5 | 4 | 5 |
| AT[C>A]TC | 6 | 8 | 4 | 7 |
| AT[C>A]TG | 7 | 9 | 5 | 10 |
| AT[C>A]TT | 5 | 7 | 4 | 7 |
| CA[C>A]AA | 12 | 11 | 7 | 8 |

| | | | | |
|---|---|---|---|---|
| CA[C>A]AC | 21 | 17 | 18 | 20 |
| CA[C>A]AG | 13 | 15 | 10 | 15 |
| CA[C>A]AT | 11 | 9 | 9 | 14 |
| CA[C>A]CA | 6 | 10 | 7 | 8 |
| CA[C>A]CC | 5 | 5 | 6 | 4 |
| CA[C>A]CG | 1 | 1 | 1 | 0 |
| CA[C>A]CT | 5 | 5 | 7 | 6 |
| CA[C>A]GA | 2 | 1 | 1 | 1 |
| CA[C>A]GC | 2 | 2 | 1 | 2 |
| CA[C>A]GG | 3 | 3 | 3 | 3 |
| CA[C>A]GT | 2 | 3 | 3 | 3 |
| CA[C>A]TA | 3 | 3 | 3 | 4 |
| CA[C>A]TC | 4 | 3 | 3 | 4 |
| CA[C>A]TG | 10 | 7 | 7 | 10 |
| CA[C>A]TT | 5 | 7 | 2 | 7 |
| CC[C>A]AA | 6 | 9 | 8 | 9 |
| CC[C>A]AC | 12 | 12 | 7 | 10 |
| CC[C>A]AG | 12 | 14 | 11 | 14 |
| CC[C>A]AT | 5 | 10 | 6 | 7 |
| CC[C>A]CA | 10 | 12 | 7 | 10 |
| CC[C>A]CC | 5 | 7 | 4 | 7 |
| CC[C>A]CG | 1 | 3 | 1 | 2 |
| CC[C>A]CT | 5 | 6 | 4 | 6 |
| CC[C>A]GA | 2 | 3 | 2 | 3 |
| CC[C>A]GC | 3 | 2 | 2 | 2 |
| CC[C>A]GG | 3 | 3 | 2 | 3 |
| CC[C>A]GT | 2 | 2 | 2 | 2 |
| CC[C>A]TA | 3 | 5 | 2 | 4 |
| CC[C>A]TC | 7 | 8 | 6 | 5 |
| CC[C>A]TG | 8 | 10 | 8 | 8 |
| CC[C>A]TT | 5 | 7 | 5 | 5 |
| CG[C>A]AA | 1 | 1 | 0 | 1 |
| CG[C>A]AC | 3 | 1 | 2 | 1 |
| CG[C>A]AG | 2 | 1 | 2 | 2 |
| CG[C>A]AT | 1 | 1 | 1 | 1 |
| CG[C>A]CA | 1 | 1 | 1 | 2 |
| CG[C>A]CC | 2 | 2 | 0 | 3 |
| CG[C>A]CG | 0 | 0 | 0 | 0 |
| CG[C>A]CT | 1 | 1 | 0 | 1 |
| CG[C>A]GA | 0 | 0 | 0 | 0 |
| CG[C>A]GC | 1 | 1 | 0 | 1 |
| CG[C>A]GG | 1 | 1 | 0 | 0 |
| CG[C>A]GT | 0 | 1 | 0 | 1 |
| CG[C>A]TA | 1 | 1 | 0 | 1 |
| CG[C>A]TC | 1 | 1 | 1 | 1 |
| CG[C>A]TG | 1 | 1 | 2 | 2 |
| CG[C>A]TT | 1 | 1 | 1 | 0 |

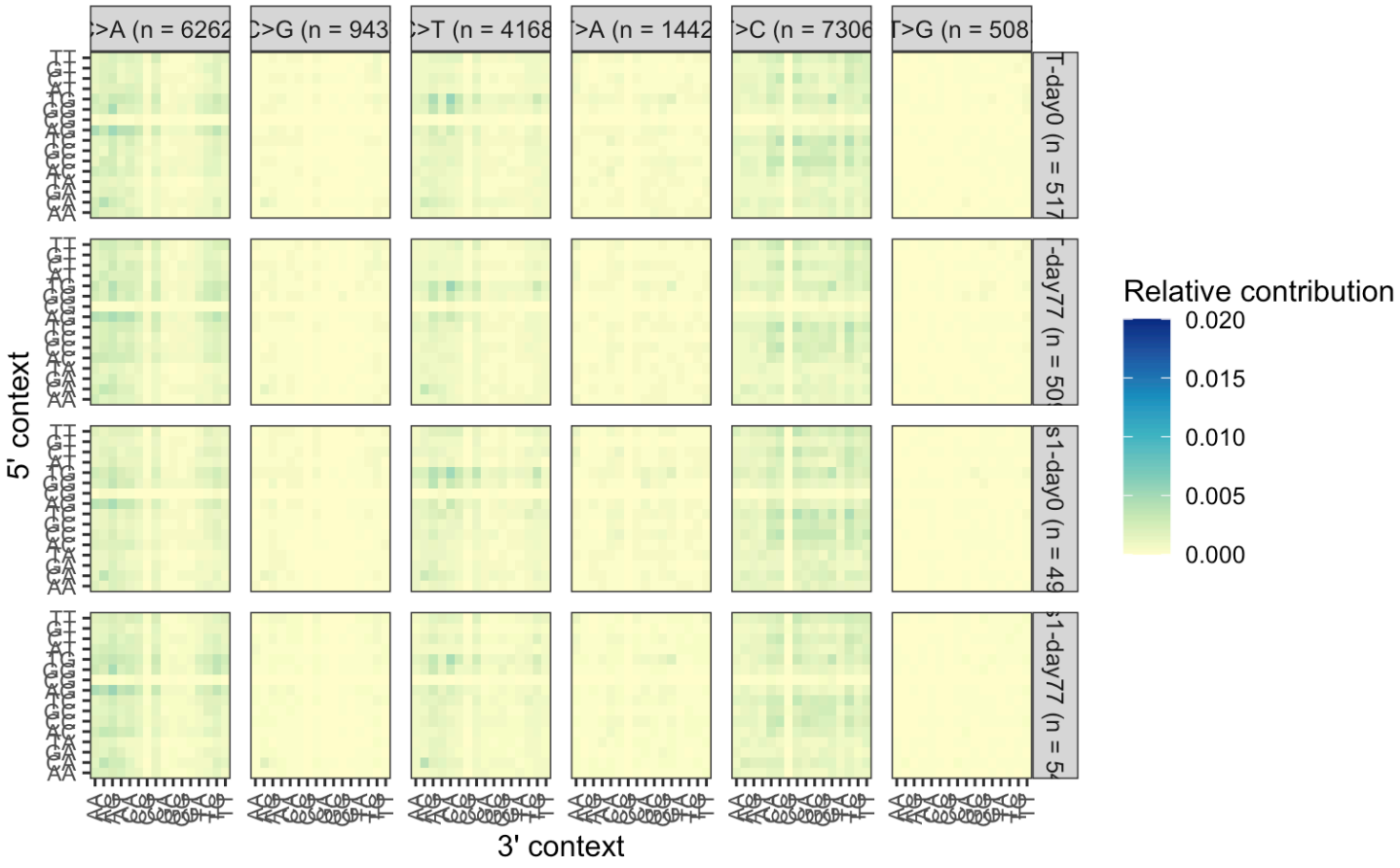
| | | | | |
|---|---|---|---|---|
| CT[C>A]AA | 7 | 7 | 5 | 7 |
| CT[C>A]AC | 10 | 8 | 8 | 8 |
| CT[C>A]AG | 12 | 15 | 9 | 14 |
| CT[C>A]AT | 5 | 8 | 4 | 8 |
| CT[C>A]CA | 6 | 12 | 7 | 11 |
| CT[C>A]CC | 8 | 9 | 7 | 9 |
| CT[C>A]CG | 1 | 2 | 1 | 2 |
| CT[C>A]CT | 6 | 9 | 6 | 7 |
| CT[C>A]GA | 2 | 2 | 1 | 2 |
| CT[C>A]GC | 2 | 1 | 2 | 3 |
| CT[C>A]GG | 2 | 4 | 1 | 3 |
| CT[C>A]GT | 1 | 3 | 1 | 2 |
| CT[C>A]TA | 2 | 4 | 5 | 4 |
| CT[C>A]TC | 4 | 9 | 8 | 8 |
| CT[C>A]TG | 10 | 10 | 8 | 11 |
| CT[C>A]TT | 4 | 6 | 5 | 6 |
| GA[C>A]AA | 7 | 8 | 6 | 9 |
| GA[C>A]AC | 9 | 8 | 5 | 8 |
| GA[C>A]AG | 10 | 13 | 10 | 13 |
| GA[C>A]AT | 7 | 7 | 5 | 9 |
| GA[C>A]CA | 9 | 6 | 4 | 5 |
| GA[C>A]CC | 5 | 5 | 4 | 4 |
| GA[C>A]CG | 1 | 1 | 0 | 1 |
| GA[C>A]CT | 5 | 4 | 6 | 6 |
| GA[C>A]GA | 2 | 2 | 2 | 1 |
| GA[C>A]GC | 1 | 1 | 1 | 2 |
| GA[C>A]GG | 1 | 1 | 1 | 1 |
| GA[C>A]GT | 2 | 1 | 1 | 1 |
| GA[C>A]TA | 2 | 2 | 4 | 3 |
| GA[C>A]TC | 4 | 4 | 3 | 6 |
| GA[C>A]TG | 5 | 6 | 4 | 6 |
| GA[C>A]TT | 4 | 3 | 4 | 5 |
| GC[C>A]AA | 7 | 9 | 5 | 8 |
| GC[C>A]AC | 5 | 11 | 6 | 8 |
| GC[C>A]AG | 14 | 14 | 8 | 11 |
| GC[C>A]AT | 6 | 7 | 4 | 8 |
| GC[C>A]CA | 8 | 9 | 6 | 10 |
| GC[C>A]CC | 6 | 6 | 4 | 6 |
| GC[C>A]CG | 2 | 1 | 1 | 2 |
| GC[C>A]CT | 4 | 7 | 4 | 7 |
| GC[C>A]GA | 1 | 1 | 1 | 1 |
| GC[C>A]GC | 1 | 2 | 1 | 2 |
| GC[C>A]GG | 2 | 2 | 2 | 2 |
| GC[C>A]GT | 2 | 3 | 0 | 2 |
| GC[C>A]TA | 3 | 4 | 3 | 3 |
| GC[C>A]TC | 7 | 7 | 6 | 9 |
| GC[C>A]TG | 10 | 10 | 8 | 10 |

| | | | | |
|---|---|---|---|---|
| GC[C>A]TT | 6 | 7 | 4 | 5 |
| GG[C>A]AA | 11 | 12 | 9 | 16 |
| GG[C>A]AC | 12 | 10 | 10 | 12 |
| GG[C>A]AG | 24 | 18 | 17 | 25 |
| GG[C>A]AT | 9 | 9 | 11 | 11 |
| GG[C>A]CA | 10 | 13 | 12 | 14 |
| GG[C>A]CC | 8 | 8 | 5 | 9 |
| GG[C>A]CG | 2 | 2 | 2 | 1 |
| GG[C>A]CT | 10 | 13 | 7 | 13 |
| GG[C>A]GA | 2 | 3 | 3 | 2 |
| GG[C>A]GC | 2 | 1 | 3 | 3 |
| GG[C>A]GG | 4 | 3 | 3 | 4 |
| GG[C>A]GT | 2 | 2 | 3 | 2 |
| GG[C>A]TA | 8 | 10 | 6 | 7 |
| GG[C>A]TC | 11 | 13 | 7 | 11 |
| GG[C>A]TG | 12 | 16 | 11 | 19 |
| GG[C>A]TT | 10 | 10 | 8 | 11 |
| GT[C>A]AA | 7 | 7 | 5 | 7 |
| GT[C>A]AC | 10 | 10 | 7 | 7 |
| GT[C>A]AG | 10 | 12 | 6 | 10 |
| GT[C>A]AT | 5 | 7 | 5 | 9 |
| GT[C>A]CA | 7 | 6 | 5 | 5 |
| GT[C>A]CC | 7 | 6 | 3 | 6 |
| GT[C>A]CG | 1 | 1 | 1 | 1 |
| GT[C>A]CT | 6 | 8 | 5 | 8 |
| GT[C>A]GA | 1 | 1 | 1 | 2 |
| GT[C>A]GC | 1 | 2 | 2 | 1 |
| GT[C>A]GG | 1 | 1 | 2 | 1 |
| GT[C>A]GT | 2 | 2 | 1 | 2 |
| GT[C>A]TA | 4 | 4 | 2 | 3 |
| GT[C>A]TC | 8 | 8 | 5 | 6 |
| GT[C>A]TG | 10 | 7 | 6 | 6 |
| GT[C>A]TT | 5 | 5 | 4 | 6 |
| TA[C>A]AA | 7 | 9 | 5 | 9 |
| TA[C>A]AC | 7 | 6 | 5 | 7 |
| TA[C>A]AG | 6 | 8 | 8 | 8 |
| TA[C>A]AT | 7 | 7 | 7 | 5 |
| TA[C>A]CA | 4 | 4 | 3 | 4 |
| TA[C>A]CC | 3 | 6 | 3 | 4 |
| TA[C>A]CG | 0 | 1 | 1 | 1 |
| TA[C>A]CT | 3 | 5 | 3 | 5 |
| TA[C>A]GA | 1 | 1 | 1 | 0 |
| TA[C>A]GC | 1 | 0 | 0 | 1 |
| TA[C>A]GG | 0 | 1 | 1 | 1 |
| TA[C>A]GT | 0 | 1 | 0 | 2 |
| TA[C>A]TA | 1 | 2 | 2 | 2 |
| TA[C>A]TC | 4 | 2 | 2 | 2 |

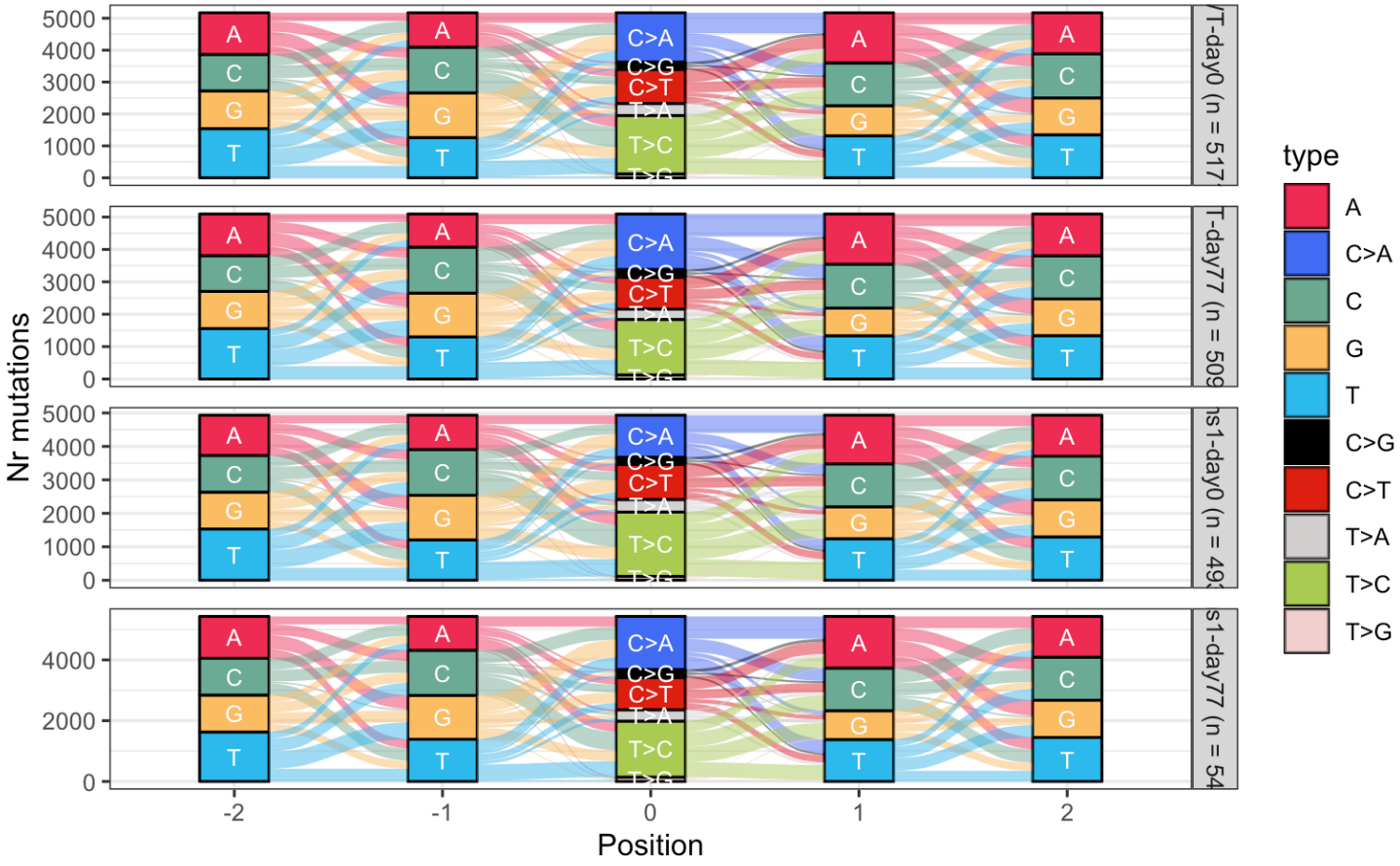
```
TA[C>A]TG       3        4         3          5
TA[C>A]TT       3        4         4          4
TC[C>A]AA       7        9         5          7
TC[C>A]AC       8       11         6         10
TC[C>A]AG      14       14        10         15
TC[C>A]AT       7       12         6          9
TC[C>A]CA      12       12         8          9
TC[C>A]CC       7        8         6          7
TC[C>A]CG       1        2         1          1
TC[C>A]CT       6       12         6          7
TC[C>A]GA       1        2         2          2
TC[C>A]GC       2        2         2          2
TC[C>A]GG       2        3         2          2
TC[C>A]GT       2        2         1          3
TC[C>A]TA       2        4         2          6
TC[C>A]TC       6        9         6          9
TC[C>A]TG       9       11         5          9
TC[C>A]TT       6       10         6          7
TG[C>A]AA      13       15        13         15
TG[C>A]AC      10        9        11         16
TG[C>A]AG      16       20        15         19
TG[C>A]AT      15       12        11         16
TG[C>A]CA      13       12         8         14
TG[C>A]CC      11        9         7         11
TG[C>A]CG       2        1         1          2
TG[C>A]CT      13       14        10         16
TG[C>A]GA       1        2         2          2
TG[C>A]GC       2        2         2          4
TG[C>A]GG       3        3         1          2
TG[C>A]GT       3        2         3          3
TG[C>A]TA       8        7         8         11
TG[C>A]TC      10       14         9         13
TG[C>A]TG      15       15        13         20
TG[C>A]TT      10       12         8         14
TT[C>A]AA       7       10         8          9
TT[C>A]AC       9       13         6         10
TT[C>A]AG      11       14         8         12
TT[C>A]AT       8        9         8          9
TT[C>A]CA      10       10         8         12
TT[C>A]CC       8       12         7         11
TT[C>A]CG       2        2         0          2
TT[C>A]CT      12       11         8         12
TT[C>A]GA       2        3         0          1
TT[C>A]GC       1        2         1          1
 [ reached getOption("max.print") -- omitted 1286 rows ]
```

The extension argument also works for the mut_context and type_context functions.

You can visualize this matrix with a heatmap.




You can also visualize this with a riverplot.




Indels

First you should get the COSMIC indel contexts. This is done with get_indel_context, which adds the columns muttype and muttype_sub to the GRangesList. The muttype column contains the main type of indel. The muttype_sub column shows the number of repeat units. For microhomology (mh) deletions the mh length is shown.

Hide

```
head(indel_grl, n = 4)
```

```
$`WT-day0`
GRanges object with 5971 ranges and 7 metadata columns:
                         seqnames            ranges strand | paramRangeID
REF
                            <Rle>         <IRanges>  <Rle> |     <factor> <DNAStringS
et>
      chr1:3555815_GTC/G     chr1   3555815-3555817      * |           NA
GTC
       chr1:4009853_G/GA     chr1           4009853      * |           NA
G
       chr1:4624693_CA/C     chr1   4624693-4624694      * |           NA
```

```
CA
       chr1:4851624_TA/T     chr1   4851624-4851625      * |           NA
TA
      chr1:4894076_TTG/T     chr1   4894076-4894078      * |           NA
TTG
                     ...      ...               ...    ... .          ...
...
     chrY:39960747_AAC/A     chrY 39960747-39960749      * |           NA
AAC
     chrY:76250767_CTG/C     chrY 76250767-76250769      * |           NA
CTG
     chrY:86118528_GGA/G     chrY 86118528-86118530      * |           NA
GGA
      chrY:90744554_TC/T     chrY 90744554-90744555      * |           NA
TC
  chrY:90744588_CCCTAG/C     chrY 90744588-90744593      * |           NA         CCC
TAG
                                     ALT      QUAL      FILTER      muttype muttyp
e_sub
                         <DNAStringSetList> <numeric> <character>  <character>   <num
eric>
      chr1:3555815_GTC/G                  G     55.12        PASS 2bp_deletion
16
       chr1:4009853_G/GA                 GA     66.12        PASS  T_insertion
1
       chr1:4624693_CA/C                  C     47.12        PASS   T_deletion
11
       chr1:4851624_TA/T                  T     47.12        PASS   T_deletion
12
      chr1:4894076_TTG/T                  T     61.12        PASS 2bp_deletion
12
                     ...                ...       ...         ...          ...
...
     chrY:39960747_AAC/A                  A     60.98        PASS 2bp_deletion
11
     chrY:76250767_CTG/C                  C     44.12        PASS 2bp_deletion
26
     chrY:86118528_GGA/G                  G     58.12        PASS 2bp_deletion
14
      chrY:90744554_TC/T                  T    443.85        PASS   C_deletion
1
  chrY:90744588_CCCTAG/C                  C    204.85        PASS 5bp_deletion
2
  -------
  seqinfo: 21 sequences from mm10 genome

$`WT-day77`
```

```
GRanges object with 6568 ranges and 7 metadata columns:
                                  seqnames            ranges strand | paramRangeID
REF
                                     <Rle>         <IRanges>  <Rle> |     <factor> <D
NAStringSet>
                chr1:4351880_T/TA     chr1           4351880      * |           NA
T
               chr1:4611421_TTC/T     chr1   4611421-4611423      * |           NA
TTC
               chr1:4777182_TTC/T     chr1   4777182-4777184      * |           NA
TTC
          chr1:5455397_AGTGTGT/A     chr1   5455397-5455403      * |           NA
AGTGTGT
                chr1:5779323_T/TA     chr1           5779323      * |           NA
T
                              ...      ...               ...    ... .          ...
...
           chrY:82534962_TTCTC/T     chrY 82534962-82534966      * |           NA
TTCTC
  chrY:90739415_G/GAGAGTTTAAAAAGA     chrY          90739415      * |           NA
G
               chrY:90739472_AG/A     chrY 90739472-90739473      * |           NA
AG
               chrY:90739474_CG/C     chrY 90739474-90739475      * |           NA
CG
               chrY:90800381_TC/T     chrY 90800381-90800382      * |           NA
TC
                                                    ALT      QUAL      FILTER        mut
type
                                  <DNAStringSetList> <numeric> <character>    <charac
ter>
                chr1:4351880_T/TA                 TA     44.99        PASS    T_inser
tion
               chr1:4611421_TTC/T                  T     69.71        PASS   2bp_dele
tion
               chr1:4777182_TTC/T                  T     52.98        PASS   2bp_dele
tion
          chr1:5455397_AGTGTGT/A                   A     81.12        PASS   6bp_dele
tion
                chr1:5779323_T/TA                 TA     38.12        PASS    T_inser
tion
                              ...                ...       ...         ...
...
           chrY:82534962_TTCTC/T                  T     81.12        PASS   4bp_dele
tion
  chrY:90739415_G/GAGAGTTTAAAAAGA    GAGAGTTTAAAAAGA    119.40        PASS 14bp_inser
tion
```

```
               chrY:90739472_AG/A                  A     76.30        PASS     C_dele
tion
               chrY:90739474_CG/C                  C     76.30        PASS     C_dele
tion
               chrY:90800381_TC/T                  T     79.98        PASS     C_dele
tion
                                  muttype_sub
                                    <numeric>
                 chr1:4351880_T/TA          11
                chr1:4611421_TTC/T          27
                chr1:4777182_TTC/T          17
            chr1:5455397_AGTGTGT/A           7
                 chr1:5779323_T/TA          16
                               ...         ...
             chrY:82534962_TTCTC/T           7
  chrY:90739415_G/GAGAGTTTAAAAAGA            1
               chrY:90739472_AG/A            1
               chrY:90739474_CG/C            1
               chrY:90800381_TC/T            5
  -------
  seqinfo: 21 sequences from mm10 genome

$`Pms1-day0`
GRanges object with 5294 ranges and 7 metadata columns:
                         seqnames            ranges strand | paramRangeID            R
EF
                            <Rle>         <IRanges>  <Rle> |     <factor> <DNAStringSe
t>
     chr1:3045468_TCA/T     chr1   3045468-3045470      * |           NA            T
CA
     chr1:4702032_AAC/A     chr1   4702032-4702034      * |           NA            A
AC
      chr1:4759375_TG/T     chr1   4759375-4759376      * |           NA
TG
     chr1:6085755_TTG/T     chr1   6085755-6085757      * |           NA            T
TG
      chr1:6996884_GT/G     chr1   6996884-6996885      * |           NA
GT
                    ...      ...               ...    ... .          ...            .
..
      chrY:3824549_TA/T     chrY   3824549-3824550      * |           NA
TA
  chrY:3826886_C/CTGTTT     chrY            3826886      * |           NA
C
     chrY:29299377_GC/G     chrY 29299377-29299378      * |           NA
GC
    chrY:90813687_AAG/A     chrY 90813687-90813689      * |           NA            A
```

```
AG
    chrY:90813695_TAC/T     chrY 90813695-90813697      * |           NA            T
AC
                                      ALT      QUAL      FILTER                mutty
pe muttype_sub
                         <DNAStringSetList> <numeric> <character>            <characte
r>   <numeric>
     chr1:3045468_TCA/T                  T     49.03        PASS           2bp_deleti
on          14
     chr1:4702032_AAC/A                  A     44.12        PASS           2bp_deleti
on          28
      chr1:4759375_TG/T                  T     60.12        PASS             C_deleti
on           4
     chr1:6085755_TTG/T                  T     44.13        PASS           2bp_deleti
on          23
      chr1:6996884_GT/G                  G     55.33        PASS             T_deleti
on           2
                    ...                ...       ...         ...                   .
..         ...
      chrY:3824549_TA/T                  T     55.12        PASS             T_deleti
on           6
  chrY:3826886_C/CTGTTT             CTGTTT     81.12        PASS          5bp_inserti
on           9
     chrY:29299377_GC/G                  G     62.12        PASS             C_deleti
on           3
    chrY:90813687_AAG/A                  A     38.03        PASS           2bp_deleti
on           3
    chrY:90813695_TAC/T                  T     38.03        PASS 2bp_deletion_with_mi
..           1
  -------
  seqinfo: 21 sequences from mm10 genome

$`Pms1-day77`
GRanges object with 5789 ranges and 7 metadata columns:
                          seqnames            ranges strand | paramRangeID
REF
                             <Rle>         <IRanges>  <Rle> |     <factor> <DNAString
Set>
       chr1:4095024_ATG/A     chr1   4095024-4095026      * |           NA
ATG
        chr1:4850769_CA/C     chr1   4850769-4850770      * |           NA
CA
     chr1:4941825_CGTGT/C     chr1   4941825-4941829      * |           NA          C
GTGT
        chr1:6108783_GA/G     chr1   6108783-6108784      * |           NA
GA
       chr1:6706414_C/CTG     chr1           6706414      * |           NA
```

```
C
                      ...      ...              ...    ... .          ...
...
      chrY:89038622_TTG/T     chrY 89038622-89038624      * |           NA
TTG
  chrY:90800540_GGGAGAT/G     chrY 90800540-90800546      * |           NA        GGG
AGAT
    chrY:90805299_CAGAG/C     chrY 90805299-90805303      * |           NA          C
AGAG
  chrY:90813518_GACAGAC/G     chrY 90813518-90813524      * |           NA        GAC
AGAC
  chrY:90813555_A/AGGTTAG     chrY          90813555      * |           NA
A
                                      ALT      QUAL      FILTER               mut
type
                         <DNAStringSetList> <numeric> <character>           <charac
ter>
       chr1:4095024_ATG/A                A     50.12        PASS          2bp_dele
tion
        chr1:4850769_CA/C                C     66.12        PASS            T_dele
tion
     chr1:4941825_CGTGT/C                C     80.00        PASS          4bp_dele
tion
        chr1:6108783_GA/G                G     47.12        PASS            T_dele
tion
       chr1:6706414_C/CTG              CTG     50.12        PASS         2bp_inser
tion
                      ...              ...       ...         ...
...
      chrY:89038622_TTG/T                T     54.12        PASS          2bp_dele
tion
  chrY:90800540_GGGAGAT/G                G     76.02        PASS 6bp_deletion_with_
mi..
    chrY:90805299_CAGAG/C                C     37.02        PASS          4bp_dele
tion
  chrY:90813518_GACAGAC/G                G     73.02        PASS 6bp_deletion_with_
mi..
  chrY:90813555_A/AGGTTAG           AGGTTAG     73.02        PASS          6bp_inser
tion
                         muttype_sub
                           <numeric>
       chr1:4095024_ATG/A          19
        chr1:4850769_CA/C           2
     chr1:4941825_CGTGT/C          10
        chr1:6108783_GA/G          12
       chr1:6706414_C/CTG          17
                      ...         ...
```

```
     chrY:89038622_TTG/T          17
  chrY:90800540_GGGAGAT/G          1
    chrY:90805299_CAGAG/C          2
  chrY:90813518_GACAGAC/G          5
  chrY:90813555_A/AGGTTAG          1
  -------
  seqinfo: 21 sequences from mm10 genome
```

Next count the number of indels per type. This results in a matrix that is similar to the mut_mat matrix.

Hide

```
indel_counts <- count_indel_contexts(indel_grl)
head(indel_counts)
```

```
             WT-day0 WT-day77 Pms1-day0 Pms1-day77
C_deletion_1      99       78        93        101
C_deletion_2     132       99       129        143
C_deletion_3     104       92       103         93
C_deletion_4      48       39        55         55
C_deletion_5      39       24        21         20
C_deletion_6+    244      309       161        223
```

Hide

```
# Normalize Indel counts
total_indel <- colSums(indel_counts)
indel_counts_norm <- t(apply(indel_counts, 1, function(y) round(y*10000/total_indel))
)
indel_counts_norm
```

```
                               WT-day0 WT-day77 Pms1-day0 Pms1-day77
C_deletion_1                       166      119       176        174
C_deletion_2                       221      151       244        247
C_deletion_3                       174      140       195        161
C_deletion_4                        80       59       104         95
C_deletion_5                        65       37        40         35
C_deletion_6+                      409      470       304        385
T_deletion_1                       189      111       149        161
T_deletion_2                       298      269       278        297
T_deletion_3                       154      108       159        138
T_deletion_4                        60       52        64         78
T_deletion_5                        32       35        42         36
T_deletion_6+                     1015     1558       667       1273
C_insertion_0                       40       40        55         73
C_insertion_1                       39       46        34         31
```

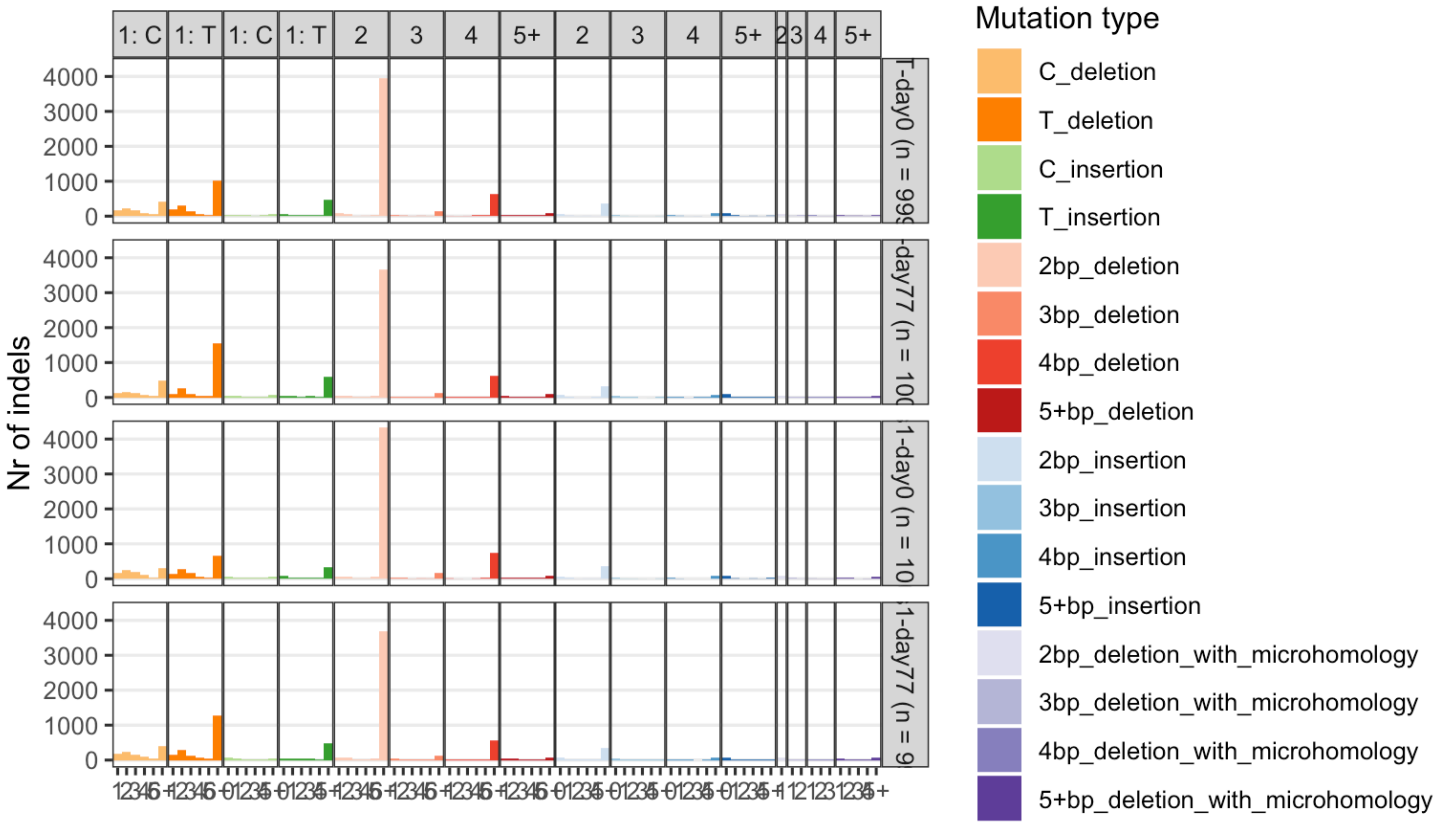
| | | | | |
|---|---|---|---|---|
| C_insertion_2 | 20 | 20 | 28 | 24 |
| C_insertion_3 | 10 | 17 | 25 | 10 |
| C_insertion_4 | 25 | 15 | 17 | 22 |
| C_insertion_5+ | 54 | 64 | 59 | 47 |
| T_insertion_0 | 55 | 43 | 83 | 55 |
| T_insertion_1 | 35 | 41 | 36 | 43 |
| T_insertion_2 | 28 | 18 | 36 | 31 |
| T_insertion_3 | 28 | 32 | 32 | 36 |
| T_insertion_4 | 18 | 18 | 23 | 28 |
| T_insertion_5+ | 471 | 601 | 323 | 485 |
| 2bp_deletion_1 | 99 | 53 | 64 | 76 |
| 2bp_deletion_2 | 57 | 43 | 57 | 69 |
| 2bp_deletion_3 | 7 | 14 | 15 | 14 |
| 2bp_deletion_4 | 13 | 12 | 15 | 10 |
| 2bp_deletion_5 | 44 | 40 | 51 | 38 |
| 2bp_deletion_6+ | 3957 | 3665 | 4318 | 3691 |
| 3bp_deletion_1 | 18 | 23 | 26 | 33 |
| 3bp_deletion_2 | 17 | 17 | 19 | 17 |
| 3bp_deletion_3 | 7 | 6 | 6 | 5 |
| 3bp_deletion_4 | 12 | 12 | 13 | 19 |
| 3bp_deletion_5 | 8 | 15 | 11 | 10 |
| 3bp_deletion_6+ | 154 | 131 | 161 | 124 |
| 4bp_deletion_1 | 7 | 15 | 13 | 19 |
| 4bp_deletion_2 | 5 | 18 | 2 | 7 |
| 4bp_deletion_3 | 7 | 15 | 0 | 16 |
| 4bp_deletion_4 | 23 | 20 | 11 | 16 |
| 4bp_deletion_5 | 20 | 24 | 26 | 26 |
| 4bp_deletion_6+ | 640 | 624 | 748 | 556 |
| 5+bp_deletion_1 | 27 | 40 | 34 | 54 |
| 5+bp_deletion_2 | 37 | 12 | 42 | 33 |
| 5+bp_deletion_3 | 22 | 29 | 34 | 19 |
| 5+bp_deletion_4 | 27 | 29 | 28 | 26 |
| 5+bp_deletion_5 | 25 | 26 | 43 | 24 |
| 5+bp_deletion_6+ | 85 | 93 | 96 | 81 |
| 2bp_insertion_0 | 52 | 61 | 55 | 74 |
| 2bp_insertion_1 | 17 | 12 | 17 | 16 |
| 2bp_insertion_2 | 5 | 2 | 4 | 7 |
| 2bp_insertion_3 | 8 | 3 | 6 | 5 |
| 2bp_insertion_4 | 3 | 12 | 6 | 14 |
| 2bp_insertion_5+ | 362 | 314 | 363 | 333 |
| 3bp_insertion_0 | 37 | 38 | 21 | 43 |
| 3bp_insertion_1 | 8 | 12 | 8 | 9 |
| 3bp_insertion_2 | 5 | 5 | 6 | 5 |
| 3bp_insertion_3 | 2 | 0 | 2 | 2 |
| 3bp_insertion_4 | 2 | 2 | 0 | 2 |
| 3bp_insertion_5+ | 7 | 14 | 9 | 21 |
| 4bp_insertion_0 | 22 | 23 | 21 | 26 |

| | | | | |
|---|---|---|---|---|
| 4bp_insertion_1 | 12 | 12 | 6 | 5 |
| 4bp_insertion_2 | 2 | 3 | 0 | 5 |
| 4bp_insertion_3 | 3 | 6 | 0 | 0 |
| 4bp_insertion_4 | 0 | 5 | 0 | 7 |
| 4bp_insertion_5+ | 74 | 79 | 87 | 81 |
| 5+bp_insertion_0 | 85 | 94 | 83 | 78 |
| 5+bp_insertion_1 | 23 | 15 | 15 | 17 |
| 5+bp_insertion_2 | 5 | 9 | 2 | 5 |
| 5+bp_insertion_3 | 10 | 6 | 11 | 9 |
| 5+bp_insertion_4 | 3 | 5 | 4 | 3 |
| 5+bp_insertion_5+ | 15 | 27 | 25 | 21 |
| 2bp_deletion_with_microhomology_1 | 59 | 35 | 77 | 69 |
| 3bp_deletion_with_microhomology_1 | 13 | 9 | 28 | 16 |
| 3bp_deletion_with_microhomology_2 | 18 | 9 | 11 | 12 |
| 4bp_deletion_with_microhomology_1 | 22 | 14 | 11 | 21 |
| 4bp_deletion_with_microhomology_2 | 7 | 11 | 6 | 9 |
| 4bp_deletion_with_microhomology_3 | 8 | 8 | 6 | 7 |
| 5+bp_deletion_with_microhomology_1 | 23 | 27 | 17 | 36 |
| 5+bp_deletion_with_microhomology_2 | 15 | 23 | 23 | 14 |
| 5+bp_deletion_with_microhomology_3 | 15 | 8 | 2 | 14 |
| 5+bp_deletion_with_microhomology_4 | 10 | 15 | 6 | 5 |
| 5+bp_deletion_with_microhomology_5+ | 42 | 44 | 60 | 60 |

Now you can plot the Indel spectra. The facets at the top show the indel types. First the C and T deletions. Then the C and T insertions. Next are the multi base deletions and insertions. Finally the deletions with microhomology are shown. The x-axis at the bottom shows the number of repeat units. For mh deletions the microhomology length is shown.

Hide

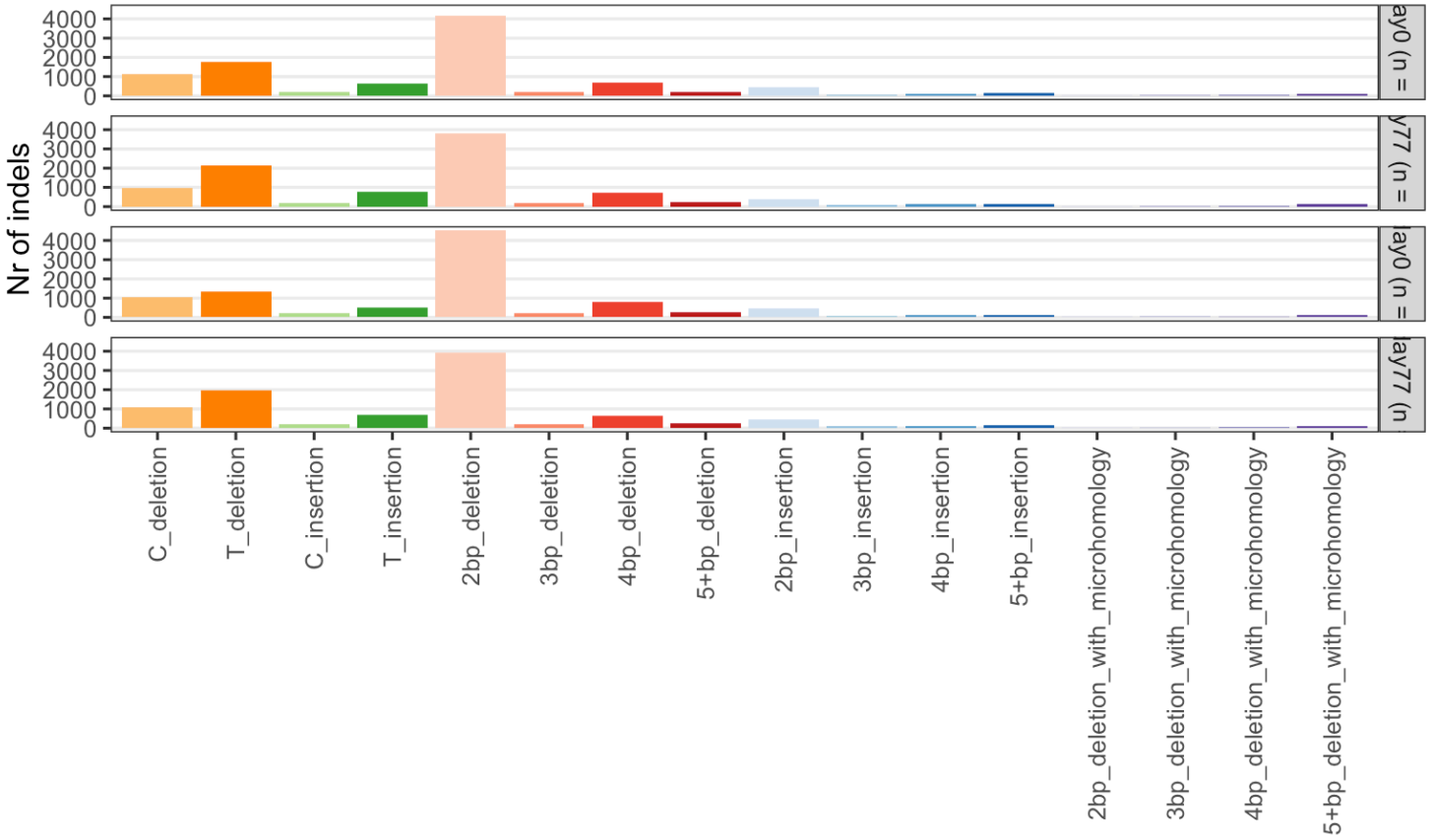
```
p_indel_context <- plot_indel_contexts(indel_counts_norm[,c(1:4)], condensed = TRUE,
same_y = TRUE)
ggsave2(filename = "p_indel_context_norm.pdf", plot = p_indel_context, width = 11, he
ight = 8 )
p_indel_context
```




You can also choose to only plot the main contexts, without taking the number of repeat units or microhomology length into account.

Hide

```
plot_main_indel_contexts(indel_counts_norm[,c(1:4)], same_y = TRUE)
```




Fisher's exact test for normalized Indels

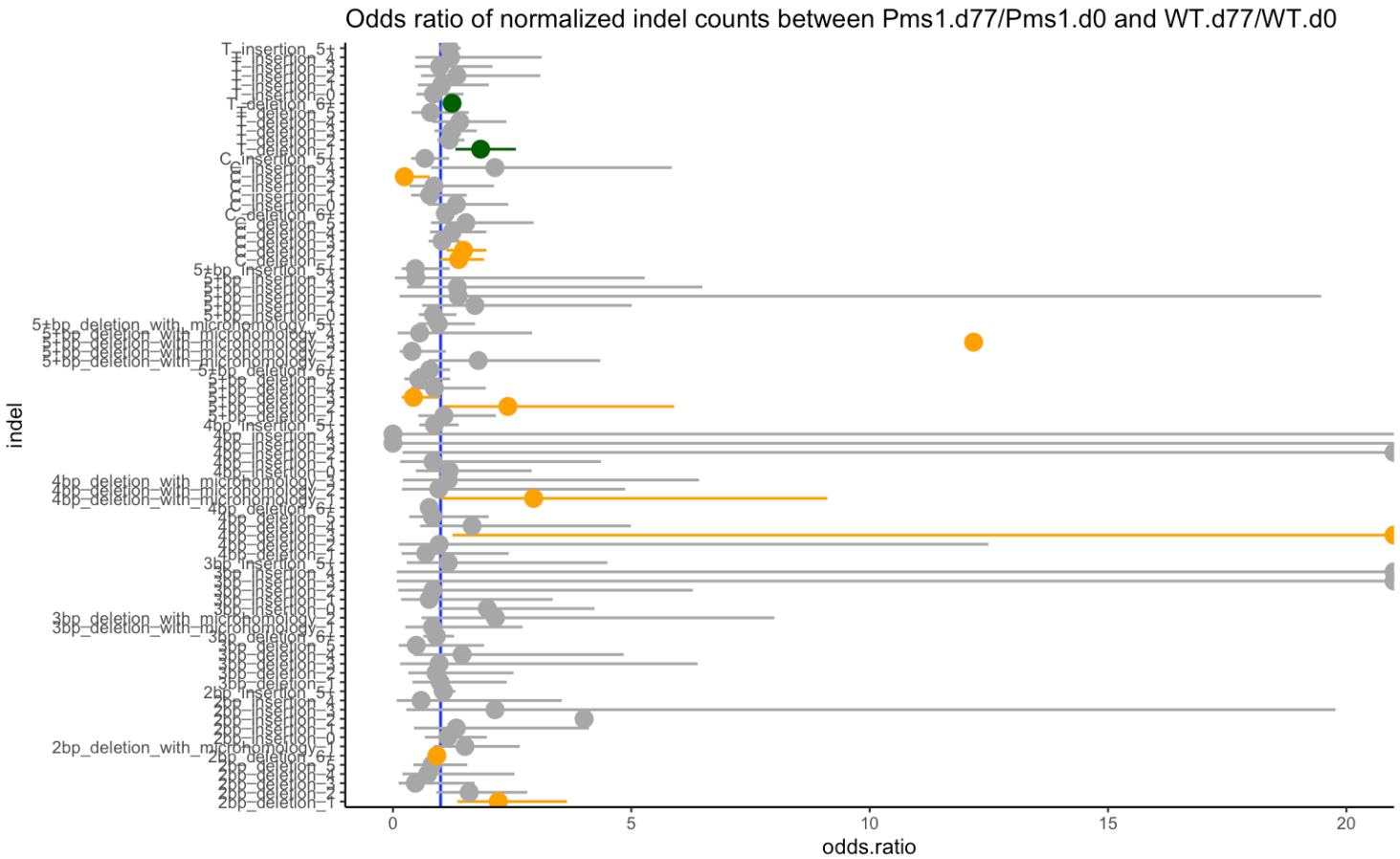
| indel <chr> | WT-day0 <chr> | WT-day77 <chr> | Pms1-day0 <chr> | Pms1-day77 <chr> | p.value <chr> | p.adj <chr> | conf.int.down <chr> |
|---|---|---|---|---|---|---|---|
| T_deletion_1 | 189 | 111 | 149 | 161 | 2e-04 | 0.0201 | 1.31 |
| T_deletion_6+ | 1015 | 1558 | 667 | 1273 | 5e-04 | 0.0413 | 1.1 |
| 2bp_deletion_1 | 99 | 53 | 64 | 76 | 0.001 | 0.08 | 1.35 |
| 4bp_deletion_6+ | 640 | 624 | 748 | 556 | 7e-04 | 0.0547 | 0.65 |

4 rows | 1-9 of 10 columns

Plot COI for Indels

Hide

```
indel_mat_report$odds.ratio <- as.numeric(indel_mat_report$odds.ratio)
indel_mat_report$conf.int.down <- as.numeric(indel_mat_report$conf.int.down)
indel_mat_report$conf.int.up <- as.numeric(indel_mat_report$conf.int.up)


p.indel.coi <- ggplot(data = indel_mat_report, aes(x = indel, y = odds.ratio, ymin =
conf.int.down, ymax = conf.int.up)) +
  geom_hline(yintercept = 1, color = "blue" ) +
  geom_pointrange(colour=ifelse(indel_mat_report$p.adj <= 0.05,"darkgreen",ifelse(ind
el_mat_report$p.value <= 0.05,"orange","darkgrey"))) +
  ggtitle("Odds ratio of normalized indel counts between Pms1.d77/Pms1.d0 and WT.d77/
WT.d0") +
  theme_classic(base_size = 8) +
  coord_flip() +
  ylim(0,20)


ggsave2(filename = "indel_mat_report.pdf", plot = p.indel.coi, height = 11, width = 8
)
p.indel.coi
```




Hide

```
library("TxDb.Mmusculus.UCSC.mm10.knownGene")
genes_mm10 <- genes(TxDb.Mmusculus.UCSC.mm10.knownGene)
```

```
  66 genes were dropped because they have exons located on both strands of the same r
eference
  sequence or on more than one reference sequence, so cannot be represented by a sing
le genomic
  range.
  Use 'single.strand.genes.only=FALSE' to get all the genes in a GRangesList object,
or use
  suppressMessages() to suppress this message.
```

Hide

```
mut_strand(snp_grl[[1]], genes_mm10, mode = "transcription")
```

```
  [1] untranscribed -            untranscribed untranscribed untranscribed untranscr
ibed untranscribed
  [8] transcribed   transcribed  transcribed   transcribed   transcribed   untranscr
ibed transcribed
 [15] transcribed   untranscribed transcribed   untranscribed untranscribed -
-
 [22] -            -            untranscribed transcribed   untranscribed transcrib
ed   transcribed
 [29] untranscribed untranscribed transcribed   transcribed   transcribed   transcrib
ed   transcribed
 [36] transcribed   transcribed  untranscribed untranscribed untranscribed transcrib
ed   untranscribed
 [43] transcribed   transcribed  untranscribed transcribed   transcribed   transcrib
ed   untranscribed
 [50] transcribed   untranscribed transcribed   untranscribed untranscribed untranscr
ibed -
 [57] untranscribed -            -            -            -            -
-
 [64] untranscribed transcribed  transcribed   untranscribed transcribed   -
-
 [71] -            untranscribed untranscribed untranscribed untranscribed transcrib
ed   transcribed
 [78] untranscribed transcribed  untranscribed untranscribed -            -
-
 [85] untranscribed -            -            -            -            -
-
 [92] -            -            -            -            -            -
-
 [99] -            -            -            untranscribed transcribed   untranscr
```

```
ibed transcribed
[106] transcribed   untranscribed untranscribed -             transcribed   transcrib
ed   -
[113] untranscribed -             -             -             -             -
-
[120] -             -             -             -             transcribed   transcrib
ed   transcribed
[127] -             untranscribed transcribed   -             -             untranscr
ibed -
[134] -             untranscribed -             -             -             -
-
[141] -             -             -             -             -             -
-
[148] -             -             -             -             -             -
-
[155] -             -             -             -             -             -
-
[162] -             -             -             -             -             -
-
[169] -             -             -             -             -             -
-
[176] -             transcribed   untranscribed transcribed   transcribed   untranscr
ibed transcribed
[183] transcribed   untranscribed untranscribed transcribed   transcribed   untranscr
ibed -
[190] -             -             -             -             -             -
-
[197] -             -             untranscribed untranscribed transcribed   transcrib
ed   transcribed
[204] transcribed   untranscribed transcribed   untranscribed -             -
-
[211] -             -             -             untranscribed untranscribed untranscr
ibed transcribed
[218] -             -             -             untranscribed transcribed   transcrib
ed   -
[225] -             -             -             -             -             transcrib
ed   -
[232] -             -             -             -             -             transcrib
ed   transcribed
[239] transcribed   untranscribed transcribed   transcribed   untranscribed transcrib
ed   untranscribed
[246] transcribed   transcribed   untranscribed untranscribed transcribed   transcrib
ed   transcribed
[253] untranscribed untranscribed -             -             -             -
-
[260] transcribed   untranscribed untranscribed untranscribed transcribed   -
-
```

```
[267] transcribed   transcribed   transcribed   untranscribed untranscribed untranscr
ibed untranscribed
[274] untranscribed transcribed   transcribed   transcribed   transcribed   transcrib
ed   -
[281] untranscribed transcribed   -             -             -             -
-
[288] -             -             -             -             -             -
-
[295] -             -             -             -             -             -
-
[302] -             -             -             -             -             -
-
[309] -             -             -             -             -             -
-
[316] -             -             -             -             -             -
-
[323] -             -             -             -             transcribed   transcrib
ed   transcribed
[330] untranscribed transcribed   transcribed   untranscribed -             -
-
[337] -             transcribed   transcribed   untranscribed transcribed   -
-
[344] untranscribed transcribed   transcribed   transcribed   untranscribed untranscr
ibed transcribed
[351] untranscribed transcribed   untranscribed transcribed   transcribed   transcrib
ed   transcribed
[358] transcribed   untranscribed transcribed   transcribed   untranscribed transcrib
ed   transcribed
[365] transcribed   transcribed   transcribed   transcribed   untranscribed transcrib
ed   transcribed
[372] -             -             -             -             -             -
untranscribed
[379] untranscribed transcribed   transcribed   untranscribed -             -
untranscribed
[386] untranscribed transcribed   -             untranscribed -             -
-
[393] untranscribed untranscribed -             -             -             -
-
[400] transcribed   transcribed   untranscribed transcribed   untranscribed untranscr
ibed untranscribed
[407] transcribed   -             transcribed   -             -             -
-
[414] -             -             -             -             -             -
-
[421] -             -             -             -             -             -
-
[428] -             -             -             -             -             -
```

```
-
[435] -             -             -             -             -             -
-
[442] -             transcribed   -             transcribed   -             untranscr
ibed untranscribed
[449] untranscribed transcribed   untranscribed untranscribed transcribed   transcrib
ed   transcribed
[456] transcribed   transcribed   untranscribed transcribed   transcribed   transcrib
ed   untranscribed
[463] -             -             untranscribed -             -             -
-
[470] -             -             -             -             untranscribed transcrib
ed   transcribed
[477] untranscribed untranscribed transcribed   untranscribed untranscribed untranscr
ibed transcribed
[484] transcribed   -             transcribed   untranscribed transcribed   transcrib
ed   untranscribed
[491] untranscribed untranscribed untranscribed untranscribed untranscribed -
-
[498] -             -             -             -             transcribed   transcrib
ed   -
[505] -             transcribed   -             -             -             -
-
[512] -             -             untranscribed -             untranscribed untranscr
ibed -
[519] transcribed   untranscribed untranscribed transcribed   -             transcrib
ed   transcribed
[526] -             -             -             -             -             untranscr
ibed transcribed
[533] transcribed   -             -             -             -
Levels: untranscribed transcribed -
```