

## Stat 261, lab 5

The data for this lab are in the file `MLLT3_small.vcf`. These are data on genomic variants in a gene called `MLLT3`. The file includes many meta-data lines that start with `##`, and then a header for the data columns that starts with `#`. The separator for the data columns is the tab character.

1. Use the appropriate `read_*` function to read these data into R, skipping the meta-data lines.

```
mm <- read_tsv("MLLT3_small.vcf",comment="##")

## Parsed with column specification:
## cols(
##   .default = col_character(),
##   `#CHROM` = col_double(),
##   POS = col_double(),
##   QUAL = col_double()
## )

## See spec(...) for full column specifications.
```

```
mm <- read_tsv("MLLT3_small.vcf",skip=255)
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   `#CHROM` = col_double(),
##   POS = col_double(),
##   QUAL = col_double()
## )

## See spec(...) for full column specifications.
```

2. Use the `spec()` function to print a list of the column specifications. Copy this and make changes so that the `REF` and `ALT` variables are read in as factors.

```
# spec(mm)
mm <- read_tsv("MLLT3_small.vcf",comment="##",
  col_types= cols(
    `#CHROM` = col_double(),
    POS = col_double(),
    ID = col_character(),
    REF = col_factor(),
    ALT = col_factor(),
    QUAL = col_double(),
    FILTER = col_character(),
    INFO = col_character(),
    FORMAT = col_character(),
    HG00096 = col_character(),
    HG00097 = col_character(),
    HG00099 = col_character(),
```

```

HG00100 = col_character(),
HG00101 = col_character(),
HG00102 = col_character(),
HG00103 = col_character(),
HG00105 = col_character(),
HG00106 = col_character(),
HG00107 = col_character(),
HG00108 = col_character(),
HG00109 = col_character(),
HG00110 = col_character()
)
)

```

3. Rename the first column **CHROM**. Hint: Beware the #-sign in the column name. This is what the text calls a “nonsyntactic” name. How do you refer to such a name?

```

mm <- rename(mm, CHROM = `#CHROM`)
# or mm <- rename(mm, CHROM = "#CHROM")

```