VCF Coverage Stats

Nate Olson

7/24/2018

Calculating median depth for callableLoci maxcov filter using FreeBayes vcf files. Downloaded vcf files from DNAnexus. Generated *.ldepth files using `vcftools` and bash script `get_depth.sh`.

```
library(tidyverse); packageVersion("tidyverse")
```

```
## [1] '1.2.1'
```

```
depth_files <- list.files(path = ".", pattern = "ldepth")
names(depth_files) <- depth_files %>% str_remove(".vcf.gz.ldepth")
```

```
depth_df <- depth_files %>%
    map_dfr(read_tsv, col_types = "ciic", .id = "dataset")
```

```
summary_df <- depth_df  %>% group_by(dataset) %>%
    summarise(med_depth = median(SUM_DEPTH),
              min_depth = min(SUM_DEPTH),
              max_depth = max(SUM_DEPTH),
              mean_depth = mean(SUM_DEPTH)) %>%
    mutate(maxcov = 2*med_depth)
```

```
summary_df %>%
    mutate(dataset = str_remove(dataset, "_novoalign"))
```

```
## # A tibble: 8 x 6
##   dataset                  med_depth min_depth max_depth mean_depth maxcov
##   <chr>                        <dbl>     <dbl>     <dbl>      <dbl>  <dbl>
## 1 HG006_1_GRCh37_6Kb_Mate~        19         2    186310       36.4     38
## 2 HG006_1_GRCh37_Hiseq100~        91         2     15171       93.5    182
## 3 HG006_1_GRCh38_6Kb_Mate~        18         2     22685       24.2     36
## 4 HG006_1_GRCh38_Hiseq100~        89         1    455338      112.     178
## 5 HG007_1_GRCh37_6Kb_Mate~        18         1    202736       38.1     36
## 6 HG007_1_GRCh37_Hiseq100~        91         1      7321       94.2    182
## 7 HG007_1_GRCh38_6Kb_Mate~        17         1     26551       24.6     34
## 8 HG007_1_GRCh38_Hiseq100~        90         1    497810      119.     180
```

```
write_tsv(summary_df, "chrom1_variant_depth.tsv")
```