variantBedOverlap

Leland Taylor

2018-03-12

This document shows a demo of how to use variantBedOverlap. There is also a command line script.

```
# get the lib dir for variantBedOverlap
install_dir=$(R --slave -e 'cat(find.package("variantBedOverlap"))')

# see help options of command line script
Rscript "$install_dir/exec/variant_bed_overlap.R" --help
```

(1) Get variants in LD

Get proxies from the 1000 Genomes Project via proxysnps.

```
snps_q <- proxysnps::get_proxies(query = "rs2072014", pop = "FIN")
snps <- subset(snps_q, R.squared >= 0.8)
```

```
#knitr::kable( head(snps, 10) )
snps # enabled by setting --> df_print: paged
#> # A tibble: 8 x 9
#>   CHROM      POS ID          REF   ALT     MAF R.squared D.prime CHOSEN
#> * <int>    <int> <chr>       <chr> <chr> <dbl>     <dbl>   <dbl> <lgl>
#> 1    22 26231688 rs2072014   G     A     0.343      1.00      1. TRUE
#> 2    22 26228757 rs35045598  TA    T     0.348     0.978     -1. FALSE
#> 3    22 26227039 rs13056293  T     C     0.364     0.915     -1. FALSE
#> 4    22 26228540 rs6004788   T     C     0.364     0.915     -1. FALSE
#> 5    22 26226815 rs12166250  A     G     0.369     0.896     -1. FALSE
#> 6    22 26226935 rs12166358  A     G     0.374     0.877     -1. FALSE
#> 7    22 26227108 rs13056441  C     T     0.374     0.877     -1. FALSE
#> 8    22 26228281 rs9624913   A     G     0.374     0.877     -1. FALSE
```

(2) Get BED overlaps

Get the overlaps of each variant (row) and genomic regions from a list of BED files. Here we load example BED files included in this package taken from Varsheny et al. 2017 (https://doi.org/10.1073/pnas.1621192114).

```
# few pre-packaged bed files from
# https://theparkerlab.med.umich.edu/data/papers/doi/10.1073/pnas.1621192114/
dir <- system.file("extdata", package = "variantBedOverlap", mustWork = TRUE)

# get overlaps with all bed files in directory
snps_overlap <- variantBedOverlap::get_bed_overlaps(
  df = snps,
  dir = dir,
  col_itemRgb = 5
)
```

processing file: Adipose processing file: InferiorTemporalLobe processing file: Islets processing file: Liver processing file: SkeletalMuscle

1

```
snps_overlap
#> # A tibble: 8 x 14
#>   ID         CHROM       POS REF   ALT     MAF R.squared D.prime CHOSEN
#>   <chr>      <int>     <int> <chr> <chr> <dbl>     <dbl>   <dbl> <lgl>
#> 1 rs12166250    22 26226815 A     G     0.369     0.896     -1. FALSE
#> 2 rs12166358    22 26226935 A     G     0.374     0.877     -1. FALSE
#> 3 rs13056293    22 26227039 T     C     0.364     0.915     -1. FALSE
#> 4 rs13056441    22 26227108 C     T     0.374     0.877     -1. FALSE
#> 5 rs2072014     22 26231688 G     A     0.343     1.00       1. TRUE
#> 6 rs35045598    22 26228757 TA    T     0.348     0.978     -1. FALSE
#> 7 rs6004788     22 26228540 T     C     0.364     0.915     -1. FALSE
#> 8 rs9624913     22 26228281 A     G     0.374     0.877     -1. FALSE
#> # ... with 5 more variables: bed.Adipose <chr>,
#> #   bed.InferiorTemporalLobe <chr>, bed.Islets <chr>, bed.Liver <chr>,
#> #   bed.SkeletalMuscle <chr>
```

(3) Plot the data

Plot the overlap data.

```
# xid_solid_line = list of x-axis IDs to add a line through
# varshney_chrhmm = flag to say assume BED file names are from Varshney et al
#                   2017. Given that assumption clean up the names to make them
#                   publication ready.
lst <- variantBedOverlap::plot_overlaps(
    df = snps_overlap,
    xid_solid_line = c("rs2072014", "rs35045598"),
    varshney_chrhmm = TRUE
)
print( lst$plt )
```
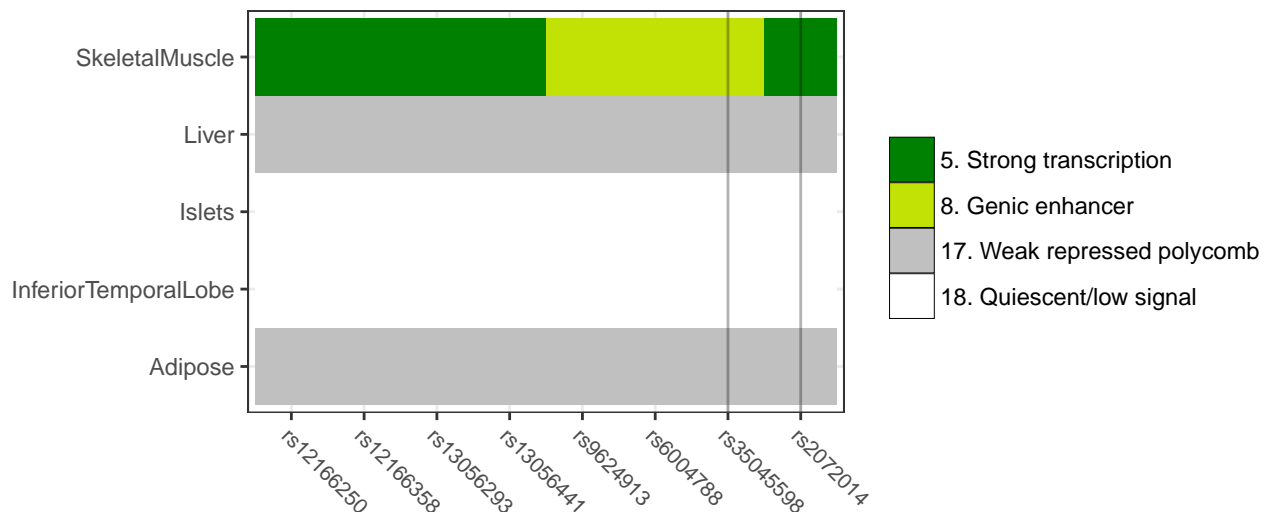

The output of plot_overlaps also contains the data underlying the plot. Note that *ID* is a factor now, sorted by *POS*. If varshney_chrhmm == TRUE, *bed_feature* will also be a factor sorted by *chrhmm_state*.

```
lst$df
#> # A tibble: 40 x 6
#>    ID            POS bed_file          bed_feature color chrhmm_state
```

```
#>       <fct>          <int> <chr>                    <fct>           <chr>        <dbl>
#>  1 rs12166250 26226815 Adipose             17. Weak r~ #C0C~          17.
#>  2 rs12166358 26226935 Adipose             17. Weak r~ #C0C~          17.
#>  3 rs13056293 26227039 Adipose             17. Weak r~ #C0C~          17.
#>  4 rs13056441 26227108 Adipose             17. Weak r~ #C0C~          17.
#>  5 rs2072014  26231688 Adipose             17. Weak r~ #C0C~          17.
#>  6 rs35045598 26228757 Adipose             17. Weak r~ #C0C~          17.
#>  7 rs6004788  26228540 Adipose             17. Weak r~ #C0C~          17.
#>  8 rs9624913  26228281 Adipose             17. Weak r~ #C0C~          17.
#>  9 rs12166250 26226815 InferiorTemporalLobe 18. Quiesc~ #FFF~          18.
#> 10 rs12166358 26226935 InferiorTemporalLobe 18. Quiesc~ #FFF~          18.
#> # ... with 30 more rows
```