diffdriver

Qirui Zhang

2025-03-15

```r
setwd("/Users/ld-zhangqr/Documents/diffdriver-main")
genef = system.file("extdata", "example_gene.txt", package = "diffdriver")
head(read.table(genef, header = F))
```

```
##      V1
## 1   CHD4
## 2 PIK3CA
```

```r
mutf = system.file("extdata/", "example_mutations.txt", package = "diffdriver")
head(read.table(mutf, header = T))
```

```
##   Chromosome  Position Ref Alt                SampleID
## 1         19  55653236   C   T TCGA-N6-A4VE-01A-11D-A28R-08
## 2         17  65134211   C   T TCGA-NA-A4R1-01A-11D-A28R-08
## 3         20  30354424   G   T TCGA-N8-A4PM-01A-11D-A28R-08
## 4          6  18215312   G   C TCGA-N8-A4PO-01A-11D-A28R-08
## 5          1 154186393   C   G TCGA-NA-A4R0-01A-11D-A28R-08
## 6         10  23003128   C   A TCGA-NF-A4X2-01A-11D-A28R-08
```

```r
phenof = system.file("extdata/", "example_phenotypes.txt", package = "diffdriver")
head(read.table(phenof, header = T))
```

```
##                    SampleID SmokingCessation Nsyn
## 1 TCGA-N5-A4R8-01A-11D-A28R-08        0.5319630    7
## 2 TCGA-N5-A4RD-01A-11D-A28R-08        0.0448991    8
## 3 TCGA-N5-A4RF-01A-11D-A28R-08       -0.3140750    7
## 4 TCGA-N5-A4RJ-01A-11D-A28R-08        0.4229920    8
## 5 TCGA-N5-A4RM-01A-11D-A28R-08       -0.2830070   12
## 6 TCGA-N5-A4RN-01A-12D-A28R-08        0.7874080   14
```

```r
bmrf = system.file("extdata/", "example_BMRlist.Rd", package = "diffdriver")
load(bmrf)
```

```r
hotf = system.file("extdata/", "example_hotspot.txt", package = "diffdriver")
head(read.table(hotf, header = T))
```

```
##   chrom   start
## 1 chr17 7577120
## 2 chr17 7577538
```

```
## 3 chr17 7577559
## 4 chr17 7578271
## 5 chr17 7578394
## 6 chr17 7578406
```

```
annodir = "/Users/ld-zhangqr/Documents/diffdriver-main/temp/annodir96"
list.files(annodir)
```

```
##  [1] "TCGA-UCS_nttype1_annodata.txt"  "TCGA-UCS_nttype10_annodata.txt"
##  [3] "TCGA-UCS_nttype11_annodata.txt" "TCGA-UCS_nttype12_annodata.txt"
##  [5] "TCGA-UCS_nttype13_annodata.txt" "TCGA-UCS_nttype14_annodata.txt"
##  [7] "TCGA-UCS_nttype15_annodata.txt" "TCGA-UCS_nttype16_annodata.txt"
##  [9] "TCGA-UCS_nttype17_annodata.txt" "TCGA-UCS_nttype18_annodata.txt"
## [11] "TCGA-UCS_nttype19_annodata.txt" "TCGA-UCS_nttype2_annodata.txt"
## [13] "TCGA-UCS_nttype20_annodata.txt" "TCGA-UCS_nttype21_annodata.txt"
## [15] "TCGA-UCS_nttype22_annodata.txt" "TCGA-UCS_nttype23_annodata.txt"
## [17] "TCGA-UCS_nttype24_annodata.txt" "TCGA-UCS_nttype25_annodata.txt"
## [19] "TCGA-UCS_nttype26_annodata.txt" "TCGA-UCS_nttype27_annodata.txt"
## [21] "TCGA-UCS_nttype28_annodata.txt" "TCGA-UCS_nttype29_annodata.txt"
## [23] "TCGA-UCS_nttype3_annodata.txt"  "TCGA-UCS_nttype30_annodata.txt"
## [25] "TCGA-UCS_nttype31_annodata.txt" "TCGA-UCS_nttype32_annodata.txt"
## [27] "TCGA-UCS_nttype33_annodata.txt" "TCGA-UCS_nttype34_annodata.txt"
## [29] "TCGA-UCS_nttype35_annodata.txt" "TCGA-UCS_nttype36_annodata.txt"
## [31] "TCGA-UCS_nttype37_annodata.txt" "TCGA-UCS_nttype38_annodata.txt"
## [33] "TCGA-UCS_nttype39_annodata.txt" "TCGA-UCS_nttype4_annodata.txt"
## [35] "TCGA-UCS_nttype40_annodata.txt" "TCGA-UCS_nttype41_annodata.txt"
## [37] "TCGA-UCS_nttype42_annodata.txt" "TCGA-UCS_nttype43_annodata.txt"
## [39] "TCGA-UCS_nttype44_annodata.txt" "TCGA-UCS_nttype45_annodata.txt"
## [41] "TCGA-UCS_nttype46_annodata.txt" "TCGA-UCS_nttype47_annodata.txt"
## [43] "TCGA-UCS_nttype48_annodata.txt" "TCGA-UCS_nttype49_annodata.txt"
## [45] "TCGA-UCS_nttype5_annodata.txt"  "TCGA-UCS_nttype50_annodata.txt"
## [47] "TCGA-UCS_nttype51_annodata.txt" "TCGA-UCS_nttype52_annodata.txt"
## [49] "TCGA-UCS_nttype53_annodata.txt" "TCGA-UCS_nttype54_annodata.txt"
## [51] "TCGA-UCS_nttype55_annodata.txt" "TCGA-UCS_nttype56_annodata.txt"
## [53] "TCGA-UCS_nttype57_annodata.txt" "TCGA-UCS_nttype58_annodata.txt"
## [55] "TCGA-UCS_nttype59_annodata.txt" "TCGA-UCS_nttype6_annodata.txt"
## [57] "TCGA-UCS_nttype60_annodata.txt" "TCGA-UCS_nttype61_annodata.txt"
## [59] "TCGA-UCS_nttype62_annodata.txt" "TCGA-UCS_nttype63_annodata.txt"
## [61] "TCGA-UCS_nttype64_annodata.txt" "TCGA-UCS_nttype65_annodata.txt"
## [63] "TCGA-UCS_nttype66_annodata.txt" "TCGA-UCS_nttype67_annodata.txt"
## [65] "TCGA-UCS_nttype68_annodata.txt" "TCGA-UCS_nttype69_annodata.txt"
## [67] "TCGA-UCS_nttype7_annodata.txt"  "TCGA-UCS_nttype70_annodata.txt"
## [69] "TCGA-UCS_nttype71_annodata.txt" "TCGA-UCS_nttype72_annodata.txt"
## [71] "TCGA-UCS_nttype73_annodata.txt" "TCGA-UCS_nttype74_annodata.txt"
## [73] "TCGA-UCS_nttype75_annodata.txt" "TCGA-UCS_nttype76_annodata.txt"
## [75] "TCGA-UCS_nttype77_annodata.txt" "TCGA-UCS_nttype78_annodata.txt"
## [77] "TCGA-UCS_nttype79_annodata.txt" "TCGA-UCS_nttype8_annodata.txt"
## [79] "TCGA-UCS_nttype80_annodata.txt" "TCGA-UCS_nttype81_annodata.txt"
## [81] "TCGA-UCS_nttype82_annodata.txt" "TCGA-UCS_nttype83_annodata.txt"
## [83] "TCGA-UCS_nttype84_annodata.txt" "TCGA-UCS_nttype85_annodata.txt"
## [85] "TCGA-UCS_nttype86_annodata.txt" "TCGA-UCS_nttype87_annodata.txt"
## [87] "TCGA-UCS_nttype88_annodata.txt" "TCGA-UCS_nttype89_annodata.txt"
## [89] "TCGA-UCS_nttype9_annodata.txt"  "TCGA-UCS_nttype90_annodata.txt"
## [91] "TCGA-UCS_nttype91_annodata.txt" "TCGA-UCS_nttype92_annodata.txt"
```

```
## [93] "TCGA-UCS_nttype93_annodata.txt" "TCGA-UCS_nttype94_annodata.txt"
## [95] "TCGA-UCS_nttype95_annodata.txt" "TCGA-UCS_nttype96_annodata.txt"
```

```r
annodir = "/Users/ld-zhangqr/Documents/diffdriver-main/temp/annodir9"
list.files(annodir, pattern = "_annodata.txt")
```

```
## [1] "nttype1_annodata.txt" "nttype2_annodata.txt" "nttype3_annodata.txt"
## [4] "nttype4_annodata.txt" "nttype5_annodata.txt" "nttype6_annodata.txt"
## [7] "nttype7_annodata.txt" "nttype8_annodata.txt" "nttype9_annodata.txt"
```

```r
library(diffdriver)
outputdir <- "/Users/ld-zhangqr/Documents/diffdriver-main/temp/output"
```

```r
res <- diffdriver(genef, mutf, phenof, bmrf, j = 2, hotf, annodir = "/Users/ld-zhangqr/Documents/diffdr
                  BMRmode = "signature", outputdir = outputdir, outputname = "testdiffdriver_sig")
```

```
## Warning in dir.create(outputdir):
## '/Users/ld-zhangqr/Documents/diffdriver-main/temp/output' already exists
```

```
## [1] "Infer parameters in background mutation rate model, adjusting for inter individual mutational s:
```

```
## Warning in fastTopics::fit_topic_model(ymatrix, k = k): One or more columns of
## X are all zero; after removing all-zero columns, 81 columns will be used for
## model fitting
```

```
## Initializing factors using Topic SCORE algorithm.
## Initializing loadings by running 10 SCD updates.
## Fitting rank-6 Poisson NMF to 52 x 81 dense matrix.
## Running at most 100 EM updates, without extrapolation (fastTopics 0.6-192).
## Refining model fit.
## Fitting rank-6 Poisson NMF to 52 x 81 dense matrix.
## Running at most 100 SCD updates, with extrapolation (fastTopics 0.6-192).
## [1] "coding..."
## [1] "processing ..."
## [1] "for qnvars, filling in missing values ..."
## [1] "for cvars (0/1 categories), filling in missing values ..."
## [1] "normalizing categorical variables in annotation matrix ..."
## [1] "Start to read in data for target genes ..."
## [1] "coding..."
## [1] "processing ..."
## [1] "for qnvars, filling in missing values ..."
## [1] "for cvars (0/1 categories), filling in missing values ..."
## [1] "normalizing categorical variables in annotation matrix ..."
## [1] "Start to process gene: CHD4"
## [1] "Start to process gene: PIK3CA"
## [1] "Finished."
```

```r
res
```

```
##              dd.p      mlr.p   mlr.v2.p  fisher.p  binom.p      lr.p     dd.fdr
## CHD4    1.00000000 0.85864512 0.83176808 1.0000000 1.000000 0.7509494 1.00000000
```

```
## PIK3CA 0.03993617 0.01370322 0.01268792 0.5576717 0.453489 0.1443223 0.07987234
##            mlr.fdr mlr.v2.fdr fisher.fdr binom.fdr    lr.fdr mut.E1 mut.E0 E1 E0
## CHD4    0.85864512 0.83176808          1 1.0000000 0.7509494      5      4 22 18
## PIK3CA 0.02740643 0.02537584          1 0.9069781 0.2886446      7      9 22 18
```

```r
res <- diffdriver(genef, mutf, phenof, bmrf, j = 2, hotf, annodir = "/Users/ld-zhangqr/Documents/diffdr
                  BMRmode = "regular", outputdir = outputdir, outputname = "testdiffdriver_reg")
```

```
## Warning in dir.create(outputdir):
## '/Users/ld-zhangqr/Documents/diffdriver-main/temp/output' already exists
```
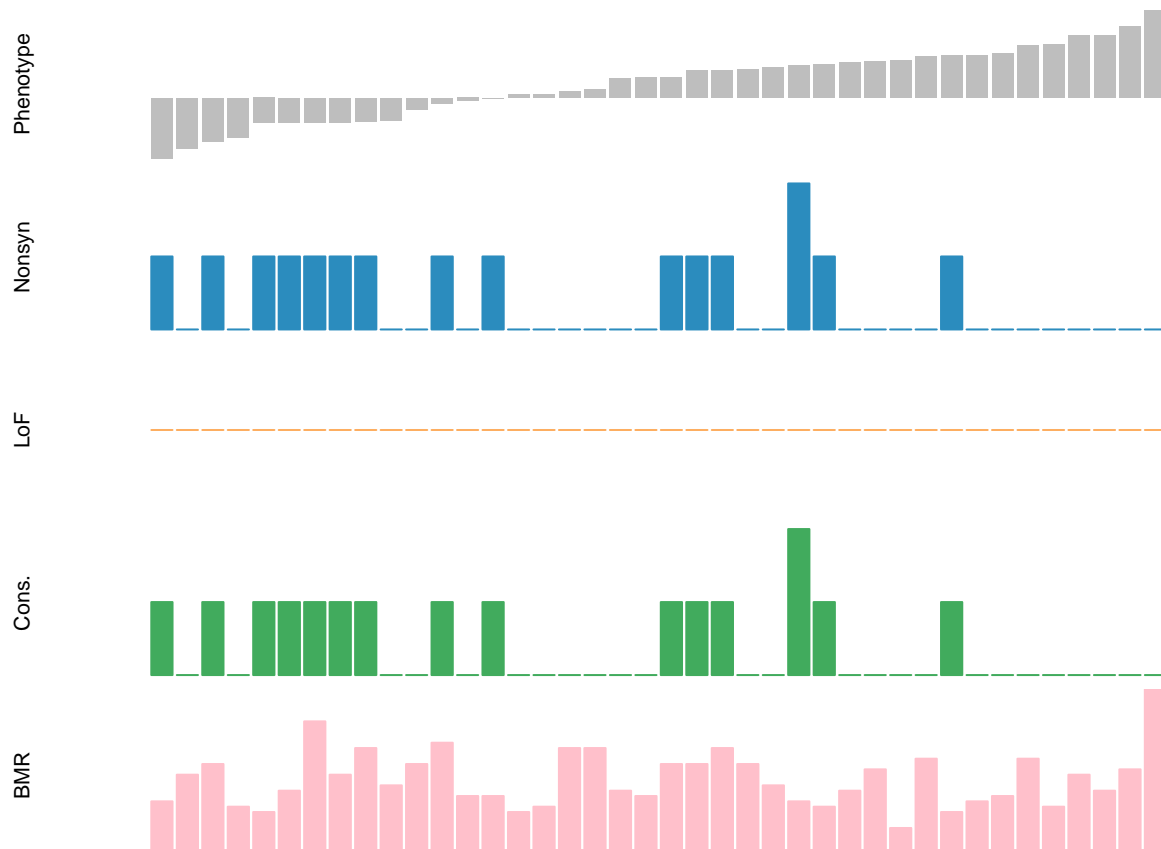
```
## [1] "Start to read in data for target genes ..."
## [1] "coding..."
## [1] "processing ..."
## [1] "for qnvars, filling in missing values ..."
## [1] "for cvars (0/1 categories), filling in missing values ..."
## [1] "normalizing categorical variables in annotation matrix ..."
## [1] "Start to process gene: CHD4"
## [1] "Start to process gene: PIK3CA"
## [1] "Finished."
```

```r
res
```

```
##               dd.p      mlr.p   mlr.v2.p  fisher.p  binom.p      lr.p     dd.fdr
## CHD4    1.00000000 0.85864512 0.83176808 1.0000000 1.000000 0.7509494 1.00000000
## PIK3CA 0.03061515 0.01370322 0.01268792 0.5576717 0.453489 0.1443223 0.06123031
##            mlr.fdr mlr.v2.fdr fisher.fdr binom.fdr    lr.fdr mut.E1 mut.E0 E1 E0
## CHD4    0.85864512 0.83176808          1 1.0000000 0.7509494      5      4 22 18
## PIK3CA 0.02740643 0.02537584          1 0.9069781 0.2886446      7      9 22 18
```

```r
load(file.path(outputdir,"testdiffdriver_reg_SmokingCessation_resdd.Rd"))
g <- "PIK3CA"
rig <- riallg[[g]]
rig$ridx <- 1:dim(rig)[1]
muts <- data.table::fread(mutf, header = T)
if (!grepl('chr', muts$Chromosome[1], fixed = T)) {muts$Chromosome <- paste0("chr",muts$Chromosome)}
muti <- na.omit(ci[rig[muts, on = c("chrom"= "Chromosome", "start" = "Position",  "ref" = "Ref",  "alt"=
                on = "SampleID"])
mutmtx <- Matrix::sparseMatrix(i = muti$ridx, j = muti$cidx, dims = c(max(rig$ridx), max(ci$cidx)))
bmrmtx= bmrallg[[g]]
ganno <- fannoallg[[g]]

diffdriver::plot_mut(mutmtx=mutmtx, canno=e, bmrmtx=bmrmtx, ganno=ganno)
```

4

```r
sessionInfo()
```

```
## R version 4.3.1 (2023-06-16)
## Platform: x86_64-apple-darwin20 (64-bit)
## Running under: macOS 15.2
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.3-x86_64/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.3-x86_64/Resources/lib/libRlapack.dylib;  LAPACK
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: America/New_York
## tzcode source: internal
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] diffdriver_0.1.3
##
## loaded via a namespace (and not attached):
##  [1] tidyr_1.3.1        plotly_4.10.4      generics_0.1.3
##  [4] ashr_2.2-63        gtools_3.9.5       lattice_0.22-6
##  [7] hms_1.1.3          digest_0.6.37      magrittr_2.0.3
```

```
## [10] evaluate_1.0.3      grid_4.3.1         fastmap_1.2.0
## [13] jsonlite_1.9.1      Matrix_1.6-5       progress_1.2.3
## [16] ggrepel_0.9.6       profileModel_0.6.1 mixsqp_0.3-54
## [19] httr_1.4.7          purrr_1.0.4        viridisLite_0.4.2
## [22] scales_1.3.0        truncnorm_1.0-9    RhpcBLASctl_0.23-42
## [25] pbapply_1.7-2       invgamma_1.1       lazyeval_0.2.2
## [28] cli_3.6.4           brglm_0.7.2        crayon_1.5.3
## [31] rlang_1.1.5         uwot_0.2.3         cowplot_1.1.3
## [34] munsell_0.5.1       yaml_2.3.10        fastTopics_0.6-192
## [37] Rtsne_0.17          tools_4.3.1        SQUAREM_2021.1
## [40] parallel_4.3.1      dplyr_1.1.4        colorspace_2.1-1
## [43] ggplot2_3.5.1       vctrs_0.6.5        R6_2.6.1
## [46] lifecycle_1.0.4     htmlwidgets_1.6.4  irlba_2.3.5.1
## [49] pkgconfig_2.0.3     RcppParallel_5.1.10 pillar_1.10.1
## [52] gtable_0.3.6        data.table_1.17.0  glue_1.8.0
## [55] Rcpp_1.0.14         xfun_0.51          tibble_3.2.1
## [58] tidyselect_1.2.1    rstudioapi_0.17.1  knitr_1.49
## [61] htmltools_0.5.8.1   rmarkdown_2.29     compiler_4.3.1
## [64] quadprog_1.5-8      prettyunits_1.2.0
```