GWAS_vis_vignette

Arcadio

12/13/2018

Loading up packages

```
library(rJava)
library(GenomicRanges)
```

```
## Loading required package: stats4
```

```
## Loading required package: BiocGenerics
```

```
## Loading required package: parallel
```

```
##
## Attaching package: 'BiocGenerics'
```

```
## The following objects are masked from 'package:parallel':
##
##     clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
##     clusterExport, clusterMap, parApply, parCapply, parLapply,
##     parLapplyLB, parRapply, parSapply, parSapplyLB
```

```
## The following objects are masked from 'package:rJava':
##
##     anyDuplicated, duplicated, sort, unique
```

```
## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs
```

```
## The following objects are masked from 'package:base':
##
##     anyDuplicated, append, as.data.frame, basename, cbind,
##     colMeans, colnames, colSums, dirname, do.call, duplicated,
##     eval, evalq, Filter, Find, get, grep, grepl, intersect,
##     is.unsorted, lapply, lengths, Map, mapply, match, mget, order,
##     paste, pmax, pmax.int, pmin, pmin.int, Position, rank, rbind,
##     Reduce, rowMeans, rownames, rowSums, sapply, setdiff, sort,
##     table, tapply, union, unique, unsplit, which, which.max,
##     which.min
```

```
## Loading required package: S4Vectors
```

```
##
## Attaching package: 'S4Vectors'
```

```
## The following object is masked from 'package:base':
##
##     expand.grid
```

```
## Loading required package: IRanges
```

```
## Loading required package: GenomeInfoDb
```

```
library(SummarizedExperiment)
```

```
## Loading required package: Biobase
```

```
## Welcome to Bioconductor
##
##     Vignettes contain introductory material; view with
##     'browseVignettes()'. To cite Bioconductor, see
##     'citation("Biobase")', and for packages 'citation("pkgname")'.
```

```
## Loading required package: DelayedArray
```

```
## Loading required package: matrixStats
```

```
##
## Attaching package: 'matrixStats'
```

```
## The following objects are masked from 'package:Biobase':
##
##     anyMissing, rowMedians
```

```
## Loading required package: BiocParallel
```

```
##
## Attaching package: 'DelayedArray'
```

```
## The following objects are masked from 'package:matrixStats':
##
##     colMaxs, colMins, colRanges, rowMaxs, rowMins, rowRanges
```

```
## The following objects are masked from 'package:base':
##
##     aperm, apply
```

```
library(GWASpoly)
library(ggplot2)
library(ggrepel)
library(stringr)

multiple_trait_manhattan <-function(traitGWASresults){
  ggplot(traitGWASresults) + geom_point(aes(Position, markerLogPVal, col = trait), alpha=0.3) + geom_po
}
```

Set directory, loading up start functions and rTassel

```
is_experimental <- TRUE

#set workdir for rtassel
setwd("~/myBins/bucklerlabBitbucket/rtassel/")

path_exp_tassel <- paste0(getwd(),"/inst/java/sTASSEL.jar")
path_exp_tassel_libs <- paste0(getwd(),"/inst/java/lib")

## jinit
rJava::.jinit(parameters="-Xmx6g")
.jcall(.jnew("java/lang/Runtime"), "J", "totalMemory")
```

```
## [1] 257425408
```

```
.jcall(.jnew("java/lang/Runtime"), "J", "maxMemory")
```

```
## [1] 5726797824
```

```
## Add class paths
if(is_experimental == TRUE) {
  tasselPath <- path_exp_tassel
  tasselLibs <- path_exp_tassel_libs
}

rJava::.jaddClassPath(tasselPath)
rJava::.jaddClassPath(tasselLibs)
print(.jclassPath())
```

```
## [1] "/Users/jav246/myBins/R-packages/rJava/java"
## [2] "/Users/jav246/myBins/bucklerlabBitbucket/rtassel/inst/java/sTASSEL.jar"
## [3] "/Users/jav246/myBins/bucklerlabBitbucket/rtassel/inst/java/lib"
```

```
## Source files
source("R/AllGenerics.R")
source("R/AllClasses.R")
source("R/TasselPluginWrappers.R")
source("R/PullFunctions.R")
source("R/gwasPolyObjectCreator.R")
```

Load up genotypes implementing Tassel code through rJava

```
geno_fileName <- "/Users/jav246/myBins/bucklerlabBitbucket/rtassel/data/mdp_genotype.hmp.txt"

## Make genotype table from tasses sample data
tasGenoTable <- readGenotypeTable(geno_fileName)

## Make summarized experiment from genotypetable
tas_se <- summarizeExperimentFromGenotypeTable(tasGenoTable) # not working right now, but
```

```
## Extracting chromosome names for each postion...
## ...is there a quicker way to get this? (~ Brandon)
```

```
tas_se
```

```
## class: RangedSummarizedExperiment
## dim: 3093 281
## metadata(0):
## assays(1): ''
## rownames: NULL
## rowData names(0):
## colnames: NULL
## colData names(1): Sample
```

Load phenotype data

```
###straight load as dataframe, skpping first two rows on tassel specific phenotype table format
pheno_fileName <- "/Users/jav246/myBins/bucklerlabBitbucket/rtassel/data/mdp_phenotype.txt"
phenos <- read.table(file = pheno_fileName, skip = 2, header = T, sep = "\t", na.strings = "NaN")
summary(phenos)
```

```
##       Taxa     location     EarHT            dpoll           EarDia
##  33-16  :  2   A:283    Min.   :  6.40   Min.   :52.60   Min.   :23.72
##  38-11  :  2   B:280    1st Qu.: 48.50   1st Qu.:63.50   1st Qu.:34.35
##  4226   :  2            Median : 60.20   Median :67.50   Median :37.00
##  4722   :  2            Mean   : 61.58   Mean   :67.78   Mean   :37.06
```

```
##  A188    :  2           3rd Qu.: 72.50   3rd Qu.:71.50   3rd Qu.:40.09
##  A214N   :  2           Max.   :138.80   Max.   :85.80   Max.   :49.30
##  (Other):551           NA's   :4        NA's   :7       NA's   :37
##        Q1               Q2               Q3
##  Min.   :0.0010   Min.   :0.0010   Min.   :0.0000
##  1st Qu.:0.0020   1st Qu.:0.0050   1st Qu.:0.0020
##  Median :0.0100   Median :0.5700   Median :0.0190
##  Mean   :0.1744   Mean   :0.5011   Mean   :0.3245
##  3rd Qu.:0.1205   3rd Qu.:0.9680   3rd Qu.:0.7940
##  Max.   :0.9990   Max.   :0.9980   Max.   :0.9980
##
```

```
### select sinlge location, as GWASpoly requires single entries for taxa.
phenosOneLoc <- phenos[phenos$location == "A",]
rownames(phenosOneLoc) <- phenosOneLoc$Taxa
###remove location as it is now redundant.
###Also, GWASpoly expects all traits as initial columns, and fixed effect covariates last
phenosOneLoc <- phenosOneLoc[,-c(2)]

summary(phenosOneLoc)
```

```
##       Taxa         EarHT            dpoll           EarDia
##  33-16  :  1   Min.   :  8.00   Min.   :54.50   Min.   :23.72
##  38-11  :  1   1st Qu.: 48.12   1st Qu.:64.00   1st Qu.:34.86
##  4226   :  1   Median : 60.50   Median :67.50   Median :37.32
##  4722   :  1   Mean   : 61.75   Mean   :67.75   Mean   :37.20
##  A188   :  1   3rd Qu.: 73.00   3rd Qu.:71.50   3rd Qu.:40.02
##  A214N  :  1   Max.   :136.00   Max.   :85.00   Max.   :46.35
##  (Other):277   NA's   :1        NA's   :3       NA's   :33
##        Q1               Q2               Q3
##  Min.   :0.0010   Min.   :0.001   Min.   :0.0000
##  1st Qu.:0.0020   1st Qu.:0.005   1st Qu.:0.0020
##  Median :0.0090   Median :0.579   Median :0.0230
##  Mean   :0.1728   Mean   :0.502   Mean   :0.3253
##  3rd Qu.:0.1160   3rd Qu.:0.968   3rd Qu.:0.7940
##  Max.   :0.9990   Max.   :0.998   Max.   :0.9980
##
```

Run GWASpoly

```
## create GWASpoly object with coopted read.GWASpoly function
data_gwasPoly <- se_createGWASpolyObject(ploidy = 2, phenoDF = phenosOneLoc,
                                         SummarizedExperimentObject =  tas_se,
                                         format = "numeric", n.traits = 3)
```

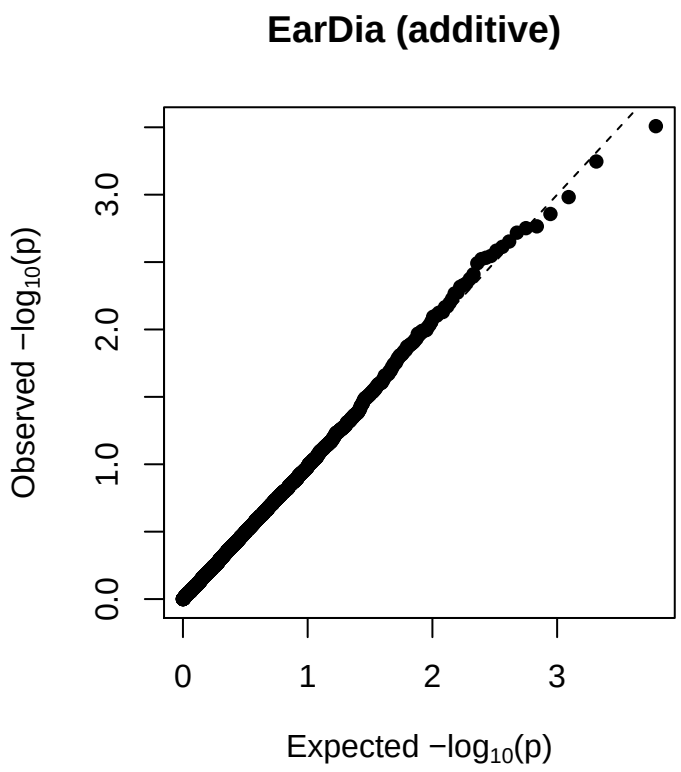
```
## Number of polymorphic markers: 3093
## Missing marker data imputed with population mode
## N = 264 individuals with phenotypic and genotypic information
## Detected following fixed effects:
## Q1
## Q2
## Q3
## Detected following traits:
## EarHT
## dpoll
## EarDia
```

```
## add kinship information to object
data_gwasPoly <- set.K(data_gwasPoly)

## set parameters for mixed model
params <- set.params(fixed=unlist(strsplit("Q1,Q2,Q3", ",")),
                     fixed.type=rep("numeric",3))

## run gwas
data_gwasPoly_res <- GWASpoly(data = data_gwasPoly, models = "additive",
                              params = params)
```

```
## Analyzing trait: EarHT
## P3D approach: Estimating variance components...Completed
## Testing markers for model: additive
## Analyzing trait: dpoll
## P3D approach: Estimating variance components...Completed
## Testing markers for model: additive
## Analyzing trait: EarDia
## P3D approach: Estimating variance components...Completed
## Testing markers for model: additive
```

Create GWASpoly plots and set thresholds to get QTLs

```
qq.plot(data_gwasPoly_res, trait = "EarDia", model = "additive")
```

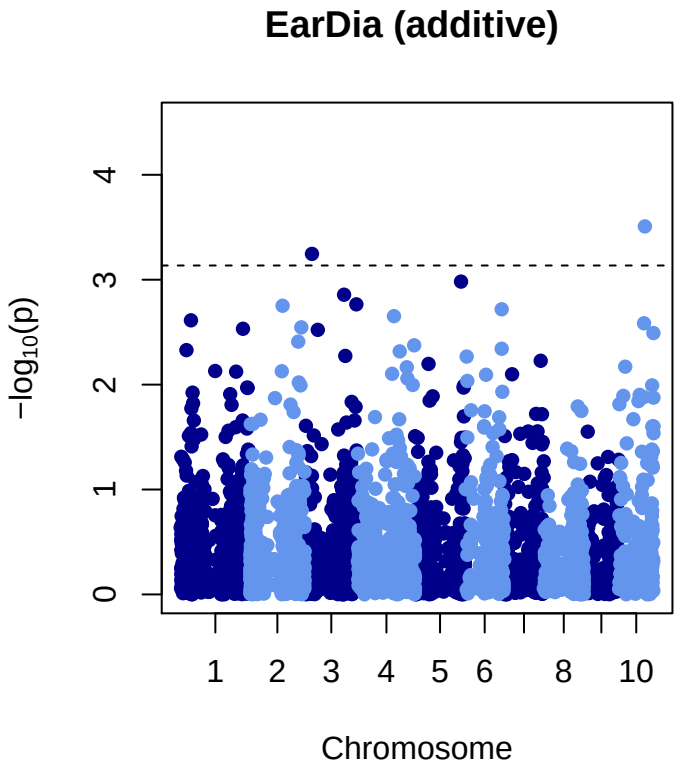



```
## NULL
```

```
#can set Bonferroni and own pvalue
data_gwasPoly_res <- set.threshold(data_gwasPoly_res, method = "FDR", level=0.05)

get.QTL(data = data_gwasPoly_res)
```

```
##       Trait    Model Threshold     Marker Chrom  Position Ref Alt Score
## 987  EarDia additive      3.14  dummy-987     3  31695534   0   1  3.25
## 2987 EarDia additive      3.14 dummy-2987    10 111608788   0   1  3.51
##      Effect
## 987   -1.08
## 2987  -1.15
```

```
#can set any of the 3 traits
manhattan.plot(data_gwasPoly_res, trait = "EarDia", model = "additive")
```




```
## NULL
```

Unwrap gwaspoly results class object

```
traitGWASresults <- data.frame()
for(trait in names(data_gwasPoly_res@scores)){
  message(paste("Getting results for:", trait))
  traitMarkerpScores <- as.data.frame(data_gwasPoly_res@scores[trait])
  Marker <- rownames(traitMarkerpScores)
  colnames(traitMarkerpScores) <- "markerLogPVal"
  traitMarkerpVals <- exp(traitMarkerpScores*-1)
  colnames(traitMarkerpVals) <- "markerpVal"
  traitMarkerEffects <-  as.data.frame(data_gwasPoly_res@effects[trait])
  colnames(traitMarkerEffects) <- "markerEffect"
  sigTreshold <- data_gwasPoly_res@threshold[rownames(data_gwasPoly_res@threshold)==trait]
  traitGWASresults <- rbind(traitGWASresults, data.frame(Marker, traitMarkerEffects, traitMarkerpVals,
}
```

```
## Getting results for: EarHT
```

```
## Getting results for: dpoll
```

```
## Getting results for: EarDia
```

```
summary(traitGWASresults)
```

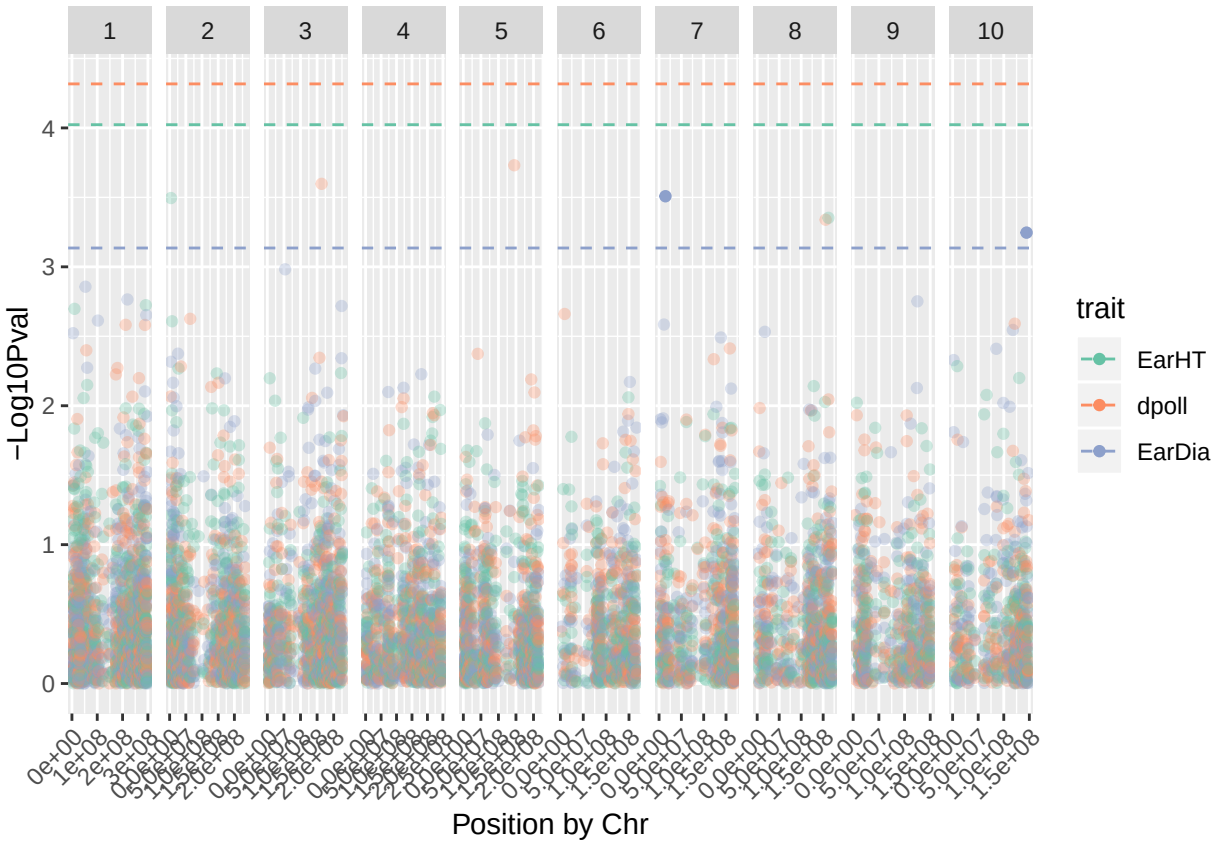
```
##         Marker      markerEffect        markerpVal      markerLogPVal
##  dummy-1    :   3   Min.   :-4.37836   Min.   :0.02396   Min.   :0.000007
##  dummy-10   :   3   1st Qu.:-0.29295   1st Qu.:0.54946   1st Qu.:0.127346
##  dummy-100  :   3   Median : 0.02879   Median :0.73557   Median :0.307111
##  dummy-1000:   3   Mean   : 0.07314   Mean   :0.69642   Mean   :0.434249
##  dummy-1001:   3   3rd Qu.: 0.36503   3rd Qu.:0.88043   3rd Qu.:0.598824
##  dummy-1002:   3   Max.   : 5.34599   Max.   :0.99999   Max.   :3.731254
##  (Other)   :9261
##     trait       sigTreshold
##  EarHT :3093   Min.   :3.135
##  dpoll :3093   1st Qu.:3.135
##  EarDia:3093   Median :4.023
##                Mean   :3.825
##                3rd Qu.:4.317
##                Max.   :4.317
##
```

```
traitGWASresults <- merge(traitGWASresults, data_gwasPoly_res@map, by = "Marker")

summary(traitGWASresults)
```

```
##         Marker      markerEffect        markerpVal      markerLogPVal
##  dummy-1    :   3   Min.   :-4.37836   Min.   :0.02396   Min.   :0.000007
##  dummy-10   :   3   1st Qu.:-0.29295   1st Qu.:0.54946   1st Qu.:0.127346
##  dummy-100  :   3   Median : 0.02879   Median :0.73557   Median :0.307111
##  dummy-1000:   3   Mean   : 0.07314   Mean   :0.69642   Mean   :0.434249
##  dummy-1001:   3   3rd Qu.: 0.36503   3rd Qu.:0.88043   3rd Qu.:0.598824
##  dummy-1002:   3   Max.   : 5.34599   Max.   :0.99999   Max.   :3.731254
##  (Other)   :9261
##     trait       sigTreshold        Chrom         Position
##  EarHT :3093   Min.   :3.135   1      :1620   Min.   :   139753
##  dpoll :3093   1st Qu.:3.135   2      :1179   1st Qu.: 43868122
##  EarDia:3093   Median :4.023   5      :1071   Median :128402775
##                Mean   :3.825   3      :1065   Mean   :119893324
##                3rd Qu.:4.317   4      : 957   3rd Qu.:175159119
##                Max.   :4.317   8      : 768   Max.   :299170077
##                                (Other):2619
##       Ref          Alt
##  Min.   :0   Min.   :1
##  1st Qu.:0   1st Qu.:1
##  Median :0   Median :1
##  Mean   :0   Mean   :1
##  3rd Qu.:0   3rd Qu.:1
##  Max.   :0   Max.   :1
##
```

Create simple manhattan like plot for all traits

```
multiple_trait_manhattan(traitGWASresults = traitGWASresults)
```



Parse GFF file to get genes and create GenomicRanges object

```
maizeGFF <- read.table("~/Box/projectMaize/PHG/cimmyt_assemblies_analy/b73/Zea_mays.AGPv4.40.chr.gff3",
colnames(maizeGFF) <- c("Chr", "Source", "annotType", "Start", "End", "other", "Strand", "other2", "IDs

maizeGFF$Chr <- as.factor(maizeGFF$Chr)

maizeGFFgenes <- maizeGFF[maizeGFF$annotType=="gene",]

maizeGFFgenes$Gene <- str_split(str_split(maizeGFFgenes$IDs, ";", simplify = T)[,1],":", simplify = T)[

head(maizeGFFgenes)
```

```
##      Chr  Source annotType  Start    End other Strand other2
## 2      1 gramene      gene  44289  49837     .      +      .
## 24     1 gramene      gene  50877  55716     .      -      .
## 170    1 gramene      gene  92299  95134     .      -      .
## 184    1 gramene      gene 111655 118312     .      -      .
## 217    1 gramene      gene 118683 119739     .      -      .
## 231    1 gramene      gene 122120 122614     .      +      .
##
## 2   ID=gene:Zm00001d027230;biotype=protein_coding;description=Mitochondrial transcription terminatio
## 24                            ID=gene:Zm00001d027231;biotype=protein_coding;description=OSJNBa0093008
## 170                                                                          ID=gene:Zm00001d027232
## 184                                                                          ID=gene:Zm00001d027233
## 217                                                                          ID=gene:Zm00001d027234
## 231                  ID=gene:Zm00001d027235;biotype=protein_coding;description=Pentatricopeptide repe
##               Gene
```

```
## 2   Zm00001d027230
## 24  Zm00001d027231
## 170 Zm00001d027232
## 184 Zm00001d027233
## 217 Zm00001d027234
## 231 Zm00001d027235
```

```
maizeGFFgenesGR <- makeGRangesFromDataFrame(maizeGFFgenes, keep.extra.columns = T)

maizeGFFgenesGR
```

```
## GRanges object with 39179 ranges and 6 metadata columns:
##           seqnames        ranges strand |   Source annotType    other
##              <Rle>     <IRanges>  <Rle> | <factor>  <factor> <factor>
##         2        1   44289-49837      + |  gramene      gene        .
##        24        1   50877-55716      - |  gramene      gene        .
##       170        1   92299-95134      - |  gramene      gene        .
##       184        1 111655-118312      - |  gramene      gene        .
##       217        1 118683-119739      - |  gramene      gene        .
##       ...      ...           ...    ... .      ...       ...      ...
##   2804827       Pt 134341-134862      - |  gramene      gene        .
##   2804831       Pt 134923-135222      - |  gramene      gene        .
##   2804835       Pt 138323-139807      + |  gramene      gene        .
##   2804849       Pt 139824-140048      + |  gramene      gene        .
##   2804853       Pt 140068-140361      + |  gramene      gene        .
##            other2
##          <factor>
##         2        .
##        24        .
##       170        .
##       184        .
##       217        .
##       ...      ...
##   2804827        .
##   2804831        .
##   2804835        .
##   2804849        .
##   2804853        .
##
##
##         2                                  ID=gene:Zm00001d027230;biotype=protein_coding;description=
##        24                                                          ID=gene:Zm00001d027231;biotype=p
##       170
##       184
##       217
##       ...
##   2804827          ID=gene:GRMZM5G885905;Name=ycf73-A;biotype=protein_coding;description=Uncharacteri
##   2804831 ID=gene:GRMZM5G866761;Name=ycf15-A;biotype=protein_coding;description=Putative uncharacteri
##   2804835
##   2804849
##   2804853
##                    Gene
##             <character>
##         2 Zm00001d027230
##        24 Zm00001d027231
```

```
##       170 Zm00001d027232
##       184 Zm00001d027233
##       217 Zm00001d027234
##       ...            ...
##   2804827  GRMZM5G885905
##   2804831  GRMZM5G866761
##   2804835  GRMZM5G818111
##   2804849  GRMZM5G866064
##   2804853  GRMZM5G855343
##   -------
##   seqinfo: 12 sequences from an unspecified genome; no seqlengths
```

Make GenomicRanges object out of gwas results

```
traitGWASresultsGR <- makeGRangesFromDataFrame(traitGWASresults, seqnames.field = "Chrom", start.field =

traitGWASresultsGR
```

```
## GRanges object with 9279 ranges and 8 metadata columns:
##          seqnames    ranges strand |    Marker        markerEffect
##             <Rle> <IRanges>  <Rle> |  <factor>           <numeric>
##      [1]        1    157104      * |   dummy-1   0.343306778643111
##      [2]        1    157104      * |   dummy-1   0.454322818890773
##      [3]        1    157104      * |   dummy-1   0.357036991759809
##      [4]        1   3206090      * |  dummy-10   0.721473858204604
##      [5]        1   3206090      * |  dummy-10 -0.0265087099725783
##      ...      ...       ...    ... .       ...                ...
##   [9275]        3  43868043      * | dummy-998   0.571750799793797
##   [9276]        3  43868043      * | dummy-998    2.09836224957921
##   [9277]        3  43868067      * | dummy-999  -0.112171180072146
##   [9278]        3  43868067      * | dummy-999   0.686141690548524
##   [9279]        3  43868067      * | dummy-999    2.57171471115537
##                markerpVal      markerLogPVal    trait      sigTreshold
##                 <numeric>          <numeric> <factor>        <numeric>
##      [1] 0.912009425744584 0.0921049537149598    EarHT 4.02323519245304
##      [2] 0.495647822750053  0.701889639229406    dpoll 4.31685670024708
##      [3] 0.528853881117141  0.637043102448693   EarDia 3.13549015264627
##      [4] 0.240756564892385   1.42396896017148    dpoll 4.31685670024708
##      [5] 0.968201579430578  0.032314970171761   EarDia 3.13549015264627
##      ...               ...                ...      ...              ...
##   [9275] 0.391523703112522  0.937709221035299    dpoll 4.31685670024708
##   [9276] 0.438150807877916  0.825192117643301    EarHT 4.02323519245304
##   [9277] 0.865945729907151   0.1439330399271   EarDia 3.13549015264627
##   [9278] 0.292826155607051   1.22817617162555    dpoll 4.31685670024708
##   [9279] 0.329427758725234   1.11039819422075    EarHT 4.02323519245304
##               Ref       Alt
##         <numeric> <numeric>
##      [1]         0         1
##      [2]         0         1
##      [3]         0         1
##      [4]         0         1
##      [5]         0         1
##      ...       ...       ...
##   [9275]         0         1
##   [9276]         0         1
```

```
##   [9277]         0         1
##   [9278]         0         1
##   [9279]         0         1
##   -------
##   seqinfo: 10 sequences from an unspecified genome; no seqlengths
```

Subset geneRanges with gwasRanges. Finding genes with SNPs in them

```
geneGWAShits <- findOverlaps(maizeGFFgenesGR, traitGWASresultsGR)
geneGWAShits
```

```
## Hits object with 909 hits and 0 metadata columns:
##         queryHits subjectHits
##         <integer>   <integer>
##     [1]       145        2335
##     [2]       145        2336
##     [3]       145        2337
##     [4]       145        2668
##     [5]       145        2669
##     ...       ...         ...
##   [905]     38508        6254
##   [906]     38508        6255
##   [907]     38550        6271
##   [908]     38550        6272
##   [909]     38550        6273
##   -------
##   queryLength: 39179 / subjectLength: 9279
```

```
genesInGWASranges <- maizeGFFgenesGR[unique(queryHits(geneGWAShits)), ]

genesInGWASranges
```

```
## GRanges object with 157 ranges and 6 metadata columns:
##           seqnames              ranges strand |   Source annotType
##              <Rle>           <IRanges>  <Rle> | <factor>  <factor>
##     25628        1     4832874-4838381      + |  gramene      gene
##     25882        1     4909652-4914685      - |  gramene      gene
##     51549        1   18907931-18909955      + |  gramene      gene
##     56160        1   22599568-22606524      + |  gramene      gene
##     57715        1   23265205-23270017      + |  gramene      gene
##       ...      ...                 ...    ... .      ...       ...
##   2692010        9 109895426-109918891      + |  gramene      gene
##   2717909        9 130889521-130890480      - |  gramene      gene
##   2741832        9 142497782-142500049      - |  gramene      gene
##   2752056        9 147127843-147139955      - |  gramene      gene
##   2754518        9 148427300-148449087      + |  gramene      gene
##              other   other2
##           <factor> <factor>
##     25628        .        .
##     25882        .        .
##     51549        .        .
##     56160        .        .
##     57715        .        .
##       ...      ...      ...
##   2692010        .        .
##   2717909        .        .
```

```
##   2741832        .        .
##   2752056        .        .
##   2754518        .        .
##
##
##     25628         ID=gene:Zm00001d027415;Name=autophagy1a;biotype=protein_coding;description=Serine/
##     25882          ID=gene:Zm00001d027423;Name=flavone synthase typeI2;biotype=protein_coding;descri
##     51549                                                                          ID=gene:Zm00001
##     56160                           ID=gene:Zm00001d028088;biotype=protein_coding;description=Ubiquiti
##     57715                                  ID=gene:Zm00001d028114;biotype=protein_coding;description=n
##       ...
##   2692010                                                          ID=gene:Zm00001d046905;biotype=prote
##   2717909 ID=gene:Zm00001d047460;biotype=protein_coding;description=Haloacid dehalogenase-like hydro
##   2741832                                          ID=gene:Zm00001d047814;biotype=protein_coding;desc
##   2752056                      ID=gene:Zm00001d047983;biotype=protein_coding;description=LisH and Ran
##   2754518                    ID=gene:Zm00001d048031;biotype=protein_coding;description=Calmodulin-bin
##                     Gene
##               <character>
##     25628 Zm00001d027415
##     25882 Zm00001d027423
##     51549 Zm00001d027973
##     56160 Zm00001d028088
##     57715 Zm00001d028114
##       ...              ...
##   2692010 Zm00001d046905
##   2717909 Zm00001d047460
##   2741832 Zm00001d047814
##   2752056 Zm00001d047983
##   2754518 Zm00001d048031
##   -------
##   seqinfo: 12 sequences from an unspecified genome; no seqlengths
```

Annotating SNPs with their closest gene

```
genesNearestToGWASranges <- nearest(traitGWASresultsGR, maizeGFFgenesGR, ignore.strand = T)

distToNearest <- distanceToNearest(traitGWASresultsGR, maizeGFFgenesGR, ignore.strand = T)

genesNearestToGWASranges <- maizeGFFgenesGR[genesNearestToGWASranges]

genesNearestToGWASranges <- as.data.frame(genesNearestToGWASranges, row.names = NULL)

genesNearestToGWASranges$distanceToNearestAnnot <- distToNearest@elementMetadata$distance

traitGWASresults_temp <- traitGWASresults

colnames(traitGWASresults_temp) <- paste(colnames(traitGWASresults), "gwas", sep = "_")

genesNearestToGWASranges <- cbind(traitGWASresults_temp, genesNearestToGWASranges)

rm(traitGWASresults_temp)

genesNearestToGWASranges <- makeGRangesFromDataFrame(genesNearestToGWASranges, keep.extra.columns = T)

genesNearestToGWASranges
```

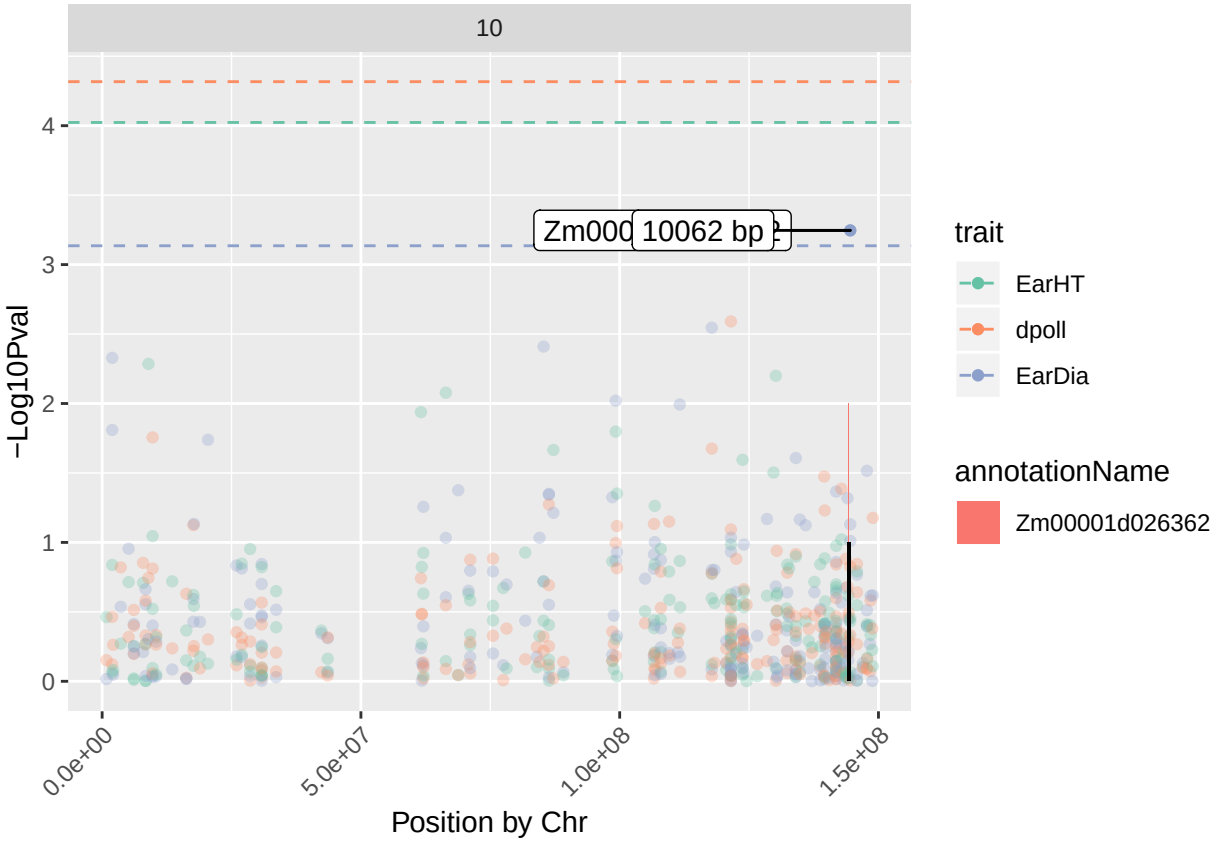
```
## GRanges object with 9279 ranges and 17 metadata columns:
##          seqnames            ranges strand | Marker_gwas
##             <Rle>         <IRanges>  <Rle> |    <factor>
##      [1]        1     138378-139043      - |     dummy-1
##      [2]        1     138378-139043      - |     dummy-1
##      [3]        1     138378-139043      - |     dummy-1
##      [4]        1   3224423-3230647      + |    dummy-10
##      [5]        1   3224423-3230647      + |    dummy-10
##      ...      ...               ...    ... .         ...
##   [9275]        3 43825329-43828852      - |   dummy-998
##   [9276]        3 43825329-43828852      - |   dummy-998
##   [9277]        3 43825329-43828852      - |   dummy-999
##   [9278]        3 43825329-43828852      - |   dummy-999
##   [9279]        3 43825329-43828852      - |   dummy-999
##            markerEffect_gwas   markerpVal_gwas markerLogPVal_gwas
##                    <numeric>         <numeric>          <numeric>
##      [1]   0.343306778643111 0.912009425744584 0.0921049537149598
##      [2]   0.454322818890773 0.495647822750053  0.701889639229406
##      [3]   0.357036991759809 0.528853881117141  0.637043102448693
##      [4]   0.721473858204604 0.240756564892385   1.42396896017148
##      [5] -0.0265087099725783 0.968201579430578  0.032314970171761
##      ...                 ...               ...                ...
##   [9275]   0.571750799793797 0.391523703112522  0.937709221035299
##   [9276]    2.09836224957921 0.438150807877916  0.825192117643301
##   [9277]  -0.112171180072146 0.865945729907151    0.1439330399271
##   [9278]   0.686141690548524 0.292826155607051   1.22817617162555
##   [9279]    2.57171471115537 0.329427758725234   1.11039819422075
##          trait_gwas sigTreshold_gwas Chrom_gwas Position_gwas  Ref_gwas
##            <factor>        <numeric>  <ordered>     <integer> <numeric>
##      [1]      EarHT 4.02323519245304          1        157104         0
##      [2]      dpoll 4.31685670024708          1        157104         0
##      [3]     EarDia 3.13549015264627          1        157104         0
##      [4]      dpoll 4.31685670024708          1       3206090         0
##      [5]     EarDia 3.13549015264627          1       3206090         0
##      ...        ...               ...        ...           ...       ...
##   [9275]      dpoll 4.31685670024708          3      43868043         0
##   [9276]      EarHT 4.02323519245304          3      43868043         0
##   [9277]     EarDia 3.13549015264627          3      43868067         0
##   [9278]      dpoll 4.31685670024708          3      43868067         0
##   [9279]      EarHT 4.02323519245304          3      43868067         0
##           Alt_gwas   Source annotType    other   other2
##          <numeric> <factor>  <factor> <factor> <factor>
##      [1]         1  gramene      gene        .        .
##      [2]         1  gramene      gene        .        .
##      [3]         1  gramene      gene        .        .
##      [4]         1  gramene      gene        .        .
##      [5]         1  gramene      gene        .        .
##      ...       ...      ...       ...      ...      ...
##   [9275]         1  gramene      gene        .        .
##   [9276]         1  gramene      gene        .        .
##   [9277]         1  gramene      gene        .        .
##   [9278]         1  gramene      gene        .        .
##   [9279]         1  gramene      gene        .        .
##
```

```
##
##      [1]                                                                 ID=gene:Zm00001d027236;bio
##      [2]                                                                 ID=gene:Zm00001d027236;bio
##      [3]                                                                 ID=gene:Zm00001d027236;bio
##      [4] ID=gene:Zm00001d027337;biotype=protein_coding;description=Di-glucose binding protein with K
##      [5] ID=gene:Zm00001d027337;biotype=protein_coding;description=Di-glucose binding protein with K
##      ...
##   [9275]       ID=gene:Zm00001d040449;biotype=protein_coding;description=protein kinase C substrate
##   [9276]       ID=gene:Zm00001d040449;biotype=protein_coding;description=protein kinase C substrate
##   [9277]       ID=gene:Zm00001d040449;biotype=protein_coding;description=protein kinase C substrate
##   [9278]       ID=gene:Zm00001d040449;biotype=protein_coding;description=protein kinase C substrate
##   [9279]       ID=gene:Zm00001d040449;biotype=protein_coding;description=protein kinase C substrate
##                   Gene distanceToNearestAnnot
##             <character>             <integer>
##      [1] Zm00001d027236                 18060
##      [2] Zm00001d027236                 18060
##      [3] Zm00001d027236                 18060
##      [4] Zm00001d027337                 18332
##      [5] Zm00001d027337                 18332
##      ...            ...                   ...
##   [9275] Zm00001d040449                 39190
##   [9276] Zm00001d040449                 39190
##   [9277] Zm00001d040449                 39214
##   [9278] Zm00001d040449                 39214
##   [9279] Zm00001d040449                 39214
##   -------
##   seqinfo: 12 sequences from an unspecified genome; no seqlengths
```

Subset GWAS hits by significance for plotting with genes.

```
traitGWASresultsGR_sig <- traitGWASresultsGR[traitGWASresultsGR$markerLogPVal > traitGWASresultsGR$sigT

traitGWASresultsGR_sig_df <- as.data.frame(traitGWASresultsGR_sig)

maxDistToAnnotation <- 3e5

genesInGWASranges_sigHits <- findOverlaps(genesInGWASranges, traitGWASresultsGR_sig, maxgap = maxDistTo

genesInGWASranges_sigGenesGR <-  genesInGWASranges[unique(queryHits(genesInGWASranges_sigHits)), ]

annotationPadding <- 1e3

annotationsDF <- data.frame()
for(seqName in as.character(seqnames(genesInGWASranges_sigGenesGR))){
  #working whole sequence present
  oneSeqRange <- genesInGWASranges_sigGenesGR[as.character(seqnames(genesInGWASranges_sigGenesGR)) == s
  seqRangeStart <- start(oneSeqRange) - annotationPadding  - maxDistToAnnotation
  seqRangeEnd <- end(oneSeqRange) + annotationPadding + maxDistToAnnotation

  annotationCount <- length(oneSeqRange)

  #working individual annotations
  annotationNum <- 0
  for(annotationName in oneSeqRange$Gene){
    annotationNum <- annotationNum + 1
```

```
    oneAnnotationRange <- oneSeqRange[oneSeqRange$Gene == annotationName,]

    annotationStart <- start(oneAnnotationRange)
    annotationEnd <- end(oneAnnotationRange)
    annotationStrand  <- strand(oneAnnotationRange)

    annotationsDF <- rbind(annotationsDF, data.frame(seqname = seqName, start=seqRangeStart, end=seqRang

  }
  ggplot(annotationsDF) + geom_rect(aes(xmin = seqRangeStart, xmax = seqRangeEnd, ymin = 0, ymax=1), fi
}

annotationsDF_GR <- makeGRangesFromDataFrame(annotationsDF, keep.extra.columns = T)

traitGWASresultsGR_annotations <- mergeByOverlaps(traitGWASresultsGR, annotationsDF_GR)

traitGWASresultsGR_annotations_df <- as.data.frame(traitGWASresultsGR_annotations)

traitGWASresultsGR_annotations_df$MinDistToAnnotation <- min(abs(traitGWASresultsGR_annotations_df$trai

traitGWASresultsGR_annotations_df$MinDistToAnnotation[traitGWASresultsGR_annotations_df$annotationsDF_GR


ggplot(subset(traitGWASresults, Chrom %in% traitGWASresultsGR_annotations_df$traitGWASresultsGR.seqname
```



Subset GWAS hits by significance for plotting with genes. Working to make multiple genes show up right

```r
genesNearestToGWASranges$Gene_label <- as.character(genesNearestToGWASranges$Gene)

bestHitByAnnot <- data.frame()
for(annotationName in unique(genesNearestToGWASranges$Gene)){
  annotationData <- genesNearestToGWASranges[genesNearestToGWASranges$Gene == annotationName,]
  bestHit <- min(annotationData$markerpVal_gwas)
  bestHitByAnnot <- rbind(bestHitByAnnot, data.frame(Gene = annotationName, bestHitPval = bestHit))
}

genesNearestToGWASranges <- merge(genesNearestToGWASranges, bestHitByAnnot, by = "Gene")


#to only plot gene names for best gwas SNP
genesNearestToGWASranges$Gene_label <- as.character(genesNearestToGWASranges$Gene)
genesNearestToGWASranges$Gene_label[genesNearestToGWASranges$bestHitPval != genesNearestToGWASranges$ma
genesNearestToGWASranges$Gene_label <- as.factor(genesNearestToGWASranges$Gene_label)
summary(genesNearestToGWASranges$Gene_label)
```

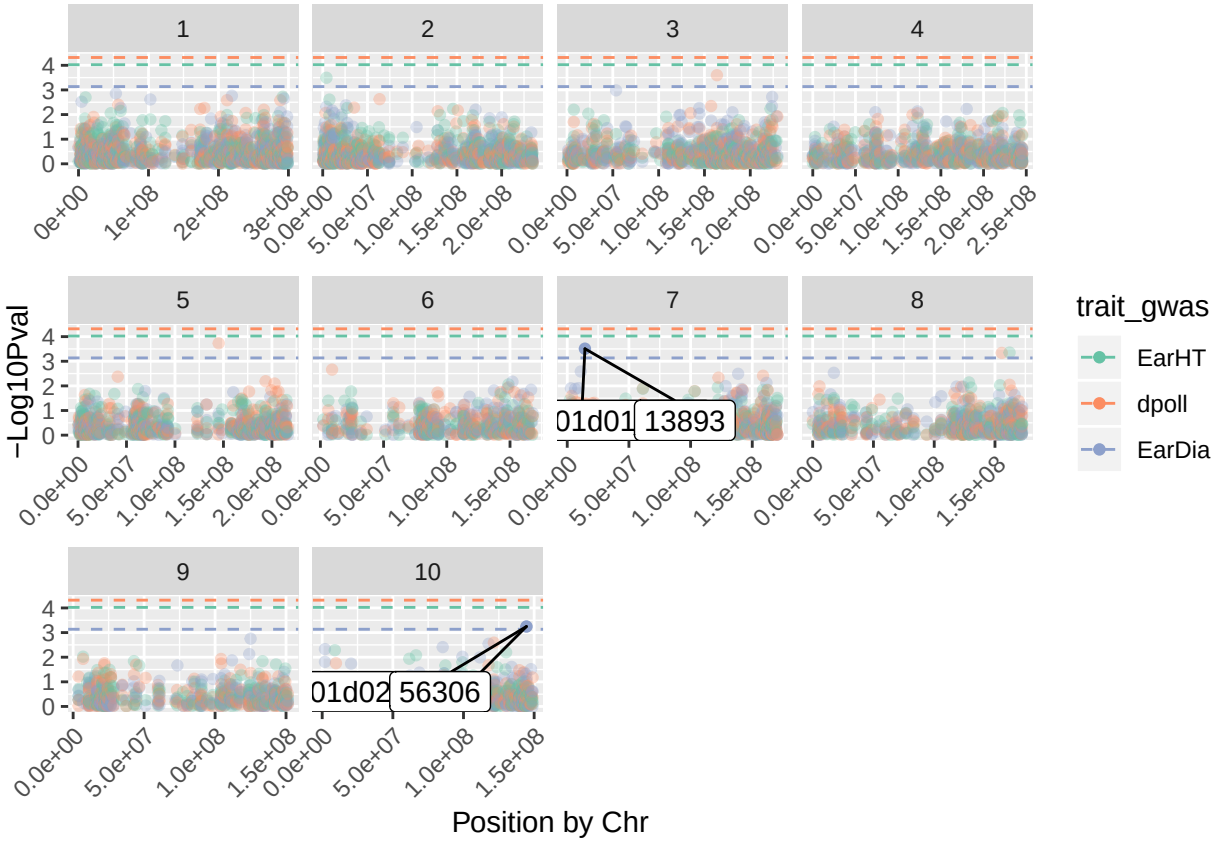
```
##                Zm00001d001781 Zm00001d001798 Zm00001d001800 Zm00001d001802
##           7791              1              1              1              1
## Zm00001d001803 Zm00001d001887 Zm00001d001896 Zm00001d001912 Zm00001d001939
##              1              1              1              1              1
## Zm00001d001945 Zm00001d002004 Zm00001d002031 Zm00001d002058 Zm00001d002090
##              1              1              1              1              1
## Zm00001d002096 Zm00001d002109 Zm00001d002130 Zm00001d002135 Zm00001d002184
##              1              1              1              1              1
## Zm00001d002278 Zm00001d002307 Zm00001d002328 Zm00001d002336 Zm00001d002370
##              1              1              1              1              1
## Zm00001d002390 Zm00001d002396 Zm00001d002431 Zm00001d002456 Zm00001d002457
##              1              1              1              1              1
## Zm00001d002488 Zm00001d002524 Zm00001d002573 Zm00001d002621 Zm00001d002628
##              1              1              1              1              1
## Zm00001d002649 Zm00001d002706 Zm00001d002714 Zm00001d002735 Zm00001d002737
##              1              1              1              1              1
## Zm00001d002742 Zm00001d002762 Zm00001d002773 Zm00001d002778 Zm00001d002793
##              1              1              1              1              1
## Zm00001d002862 Zm00001d002873 Zm00001d002889 Zm00001d002937 Zm00001d002944
##              1              1              1              1              1
## Zm00001d002969 Zm00001d003033 Zm00001d003081 Zm00001d003093 Zm00001d003123
##              1              1              1              1              1
## Zm00001d003150 Zm00001d003188 Zm00001d003205 Zm00001d003231 Zm00001d003237
##              1              1              1              1              1
## Zm00001d003266 Zm00001d003292 Zm00001d003343 Zm00001d003357 Zm00001d003394
##              1              1              1              1              1
## Zm00001d003403 Zm00001d003423 Zm00001d003445 Zm00001d003452 Zm00001d003543
##              1              1              1              1              1
## Zm00001d003592 Zm00001d003646 Zm00001d003730 Zm00001d003735 Zm00001d003772
##              1              1              1              1              1
## Zm00001d003774 Zm00001d003794 Zm00001d003852 Zm00001d003855 Zm00001d003863
##              1              1              1              1              1
## Zm00001d003925 Zm00001d003929 Zm00001d003993 Zm00001d004105 Zm00001d004178
##              1              1              1              1              1
## Zm00001d004327 Zm00001d004342 Zm00001d004376 Zm00001d004560 Zm00001d004599
##              1              1              1              1              1
```

```
## Zm00001d004649 Zm00001d004667 Zm00001d004669 Zm00001d004689 Zm00001d004753
##              1              1              1              1              1
## Zm00001d004794 Zm00001d004800 Zm00001d004804 Zm00001d004829        (Other)
##              1              1              1              1           1390
```

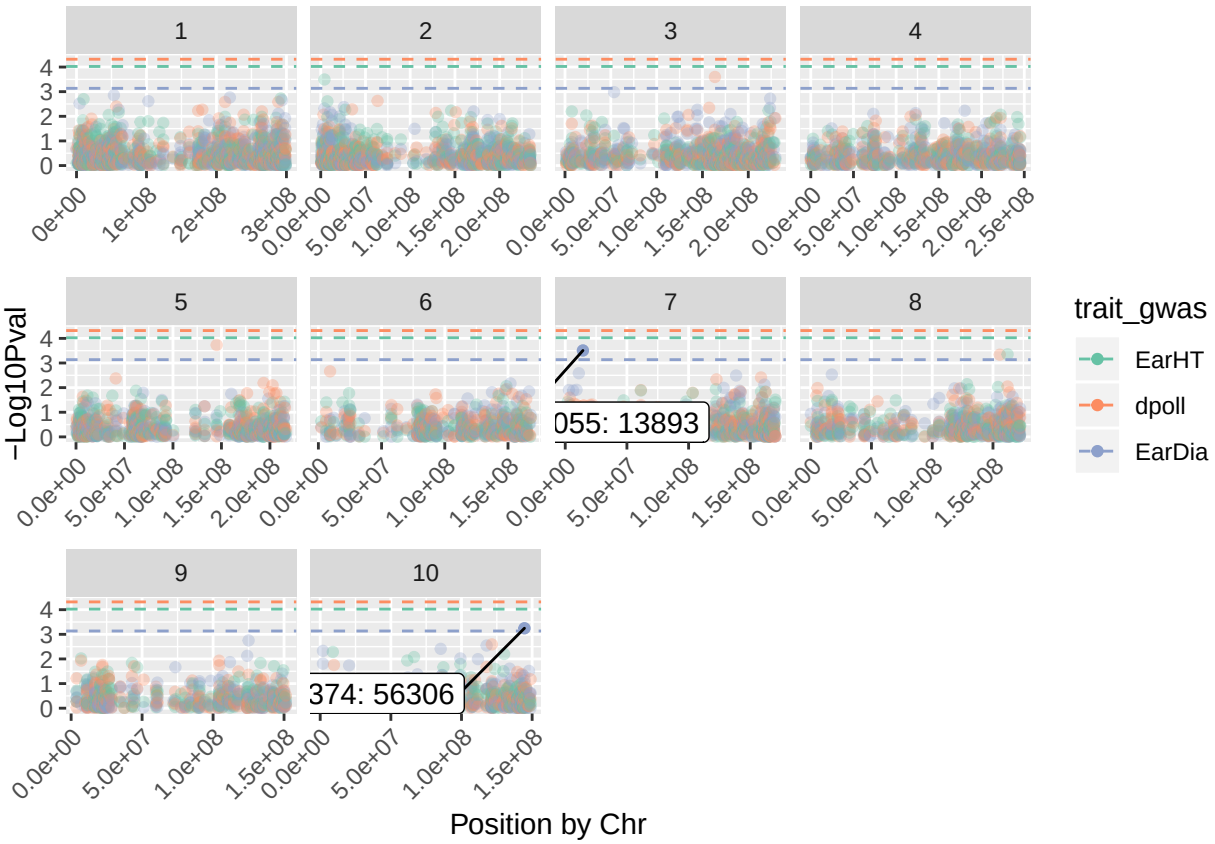
```
genesNearestToGWASranges$distanceToNearestAnnot_label <- genesNearestToGWASranges$distanceToNearestAnno
genesNearestToGWASranges$distanceToNearestAnnot_label[genesNearestToGWASranges$bestHitPval != genesNeare

#to only plot gene names for significant SNPs
genesNearestToGWASranges$Gene_label <- as.character(genesNearestToGWASranges$Gene)
genesNearestToGWASranges$Gene_label[genesNearestToGWASranges$markerLogPVal_gwas < genesNearestToGWASrang
genesNearestToGWASranges$Gene_label <- as.factor(genesNearestToGWASranges$Gene_label)
summary(genesNearestToGWASranges$Gene_label)
```

```
##                Zm00001d019055 Zm00001d026374
##           9277              1              1
```

```
genesNearestToGWASranges$distanceToNearestAnnot_label <- genesNearestToGWASranges$distanceToNearestAnno
genesNearestToGWASranges$distanceToNearestAnnot_label[genesNearestToGWASranges$markerLogPVal_gwas < gene

ggplot(genesNearestToGWASranges) + geom_point(aes(Position_gwas, markerLogPVal_gwas, col = trait_gwas),
```




```
genesNearestToGWASranges$single_label <- paste(genesNearestToGWASranges$Gene_label, genesNearestToGWASra

genesNearestToGWASranges$single_label <- gsub("^: $", "" ,genesNearestToGWASranges$single_label)

ggplot(genesNearestToGWASranges) + geom_point(aes(Position_gwas, markerLogPVal_gwas, col = trait_gwas),
```




```
#+ geom_rect(aes(xmin = seqRangeStart, xmax = seqRangeEnd, ymin = 0, ymax=1), fill = "black", data=anno
```

Subset GWAS hits by significance for plotting with genes. Working to make multiple genes show up right.
Temp work not working

```
traitGWASresultsGR_sig <- traitGWASresultsGR[traitGWASresultsGR$markerLogPVal > traitGWASresultsGR$sigT

traitGWASresultsGR_sig_df <- as.data.frame(traitGWASresultsGR_sig)

maxDistToAnnotation <- 3e5

genesInGWASranges_sigHits <- findOverlaps(genesInGWASranges, traitGWASresultsGR_sig, maxgap = maxDistTo

genesInGWASranges_sigGenesGR <-  genesInGWASranges[unique(queryHits(genesInGWASranges_sigHits)), ]

annotationPadding <- 1e3

annotationsDF <- data.frame()
for(seqName in as.character(seqnames(genesInGWASranges_sigGenesGR))){
  #working whole sequence present
  oneSeqRange <- genesInGWASranges_sigGenesGR[as.character(seqnames(genesInGWASranges_sigGenesGR)) == s
  seqRangeStart <- start(oneSeqRange) - annotationPadding  - maxDistToAnnotation
  seqRangeEnd <- end(oneSeqRange) + annotationPadding + maxDistToAnnotation

  annotationCount <- length(oneSeqRange)

  #working individual annotations
  annotationNum <- 0
```

```
  for(annotationName in oneSeqRange$Gene){
    annotationNum <- annotationNum + 1
    oneAnnotationRange <- oneSeqRange[oneSeqRange$Gene == annotationName,]

    annotationStart <- start(oneAnnotationRange)
    annotationEnd <- end(oneAnnotationRange)
    annotationStrand  <- strand(oneAnnotationRange)

    annotationsDF <- rbind(annotationsDF, data.frame(seqname = seqName, start=seqRangeStart, end=seqRang

  }
  #ggplot(annotationsDF) + geom_rect(aes(xmin = seqRangeStart, xmax = seqRangeEnd, ymin = 0, ymax=1), f
}

annotationsDF_GR <- makeGRangesFromDataFrame(annotationsDF, keep.extra.columns = T)

traitGWASresultsGR_annotations <- mergeByOverlaps(traitGWASresultsGR, annotationsDF_GR)

traitGWASresultsGR_annotations_df <- as.data.frame(traitGWASresultsGR_annotations)

traitGWASresultsGR_annotations_df$MinDistToAnnotation <- min(abs(traitGWASresultsGR_annotations_df$trai

traitGWASresultsGR_annotations_df$MinDistToAnnotation[traitGWASresultsGR_annotations_df$annotationsDF_GR

bestHitByAnnot <- data.frame()
for(annotationName in unique(traitGWASresultsGR_annotations_df$annotationsDF_GR.annotationName)){
  message(annotationName)
  annotationData <- traitGWASresultsGR_annotations_df[traitGWASresultsGR_annotations_df$annotationsDF_GR
  bestHit <- min(annotationData$traitGWASresultsGR.markerpVal)
  message(bestHit)
  bestHitByAnnot <- rbind(bestHitByAnnot, data.frame(annotationsDF_GR.annotationName = annotationName, b
}

traitGWASresultsGR_annotations_df <- merge(traitGWASresultsGR_annotations_df, bestHitByAnnot, by = "anno

traitGWASresultsGR_annotations_df$annotationsDF_GR.annotationName_label <- as.character(traitGWASresults

traitGWASresultsGR_annotations_df$annotationsDF_GR.annotationName_label[traitGWASresultsGR_annotations_d

traitGWASresultsGR_annotations_df$annotationsDF_GR.annotationName_label <- as.factor(traitGWASresultsGR_

summary(traitGWASresultsGR_annotations_df$annotationsDF_GR.annotationName_label)

traitGWASresultsGR_annotations_df$MinDistToAnnotation_label <- paste(traitGWASresultsGR_annotations_df$M


traitGWASresultsGR_annotations_df$MinDistToAnnotation_label[traitGWASresultsGR_annotations_df$bestHitPva

#ggplot(subset(traitGWASresults, Chrom %in% traitGWASresultsGR_annotations_df$traitGWASresultsGR.seqnam

ggplot(traitGWASresultsGR_annotations_df) + geom_point(aes(traitGWASresultsGR.start, traitGWASresultsGR_

#+ geom_rect(aes(xmin = seqRangeStart, xmax = seqRangeEnd, ymin = 0, ymax=1), fill = "black", data=anno
```