

Allele-specific ChIPseq analysis. Check in on Genotype data

Anastasiia Hryhorzhevskia

1. Load data

2. Convert STARR and GR SNP data into GRange object

2.1. Convert STARR

```
starr.df$chr <- substr(starr.df$chr, 4, length(starr.df$chr))

starr <- GRanges(
  starr.df$chr,
  IRanges(start = starr.df$position, end = starr.df$position),
  strand = "*"
)

names(starr) <- starr.df$name
starr
```

```
## GRanges object with 234 ranges and 0 metadata columns:
##           seqnames      ranges strand
##           <Rle> <IRanges>  <Rle>
##    1-148409926      1 150143302      *
##    1-148443906      1 150177282      *
##    1-148541113      1 150274489      *
##    1-148575582      1 150308958      *
##   12-111668114     12 113183731      *
##           ...      ...      ...
##      rs246913       5 102560370      *
##   rs34509699       7  43773235      *
##      rs384097       3  41126731      *
##      rs9271171       6  32577907      *
##      rs948465      11 117966178      *
## -----
##   seqinfo: 20 sequences from an unspecified genome; no seqlengths
```

2.2. Convert GR SNP

```
grsnp <- GRanges(
  grsnp.df$CHROM,
  IRanges(start = grsnp.df$POS, end = grsnp.df$POS),
```

```

strand = "*"
)

names(grsnp) <- grsnp.df$ID
grsnp

## GRanges object with 3662 ranges and 0 metadata columns:
##           seqnames      ranges strand
##           <Rle> <IRanges>  <Rle>
## 1-148268256      1 150001632      *
## 1-148271849      1 150005225      *
## 1-148275208      1 150008584      *
## 1-148275840      1 150009216      *
## 1-148281790      1 150015166      *
##      ...      ...      ...
## rs9984686      21 46088370      *
## rs9994839       4 88393780      *
## rs9997288       4 56347227      *
## rs9997384       4 47976085      *
## rs9998146       4 56397977      *
## -----
## seqinfo: 22 sequences from an unspecified genome; no seqlengths

```

3. Overlap STARR data with peaksets

No overlaps for veh or veh-dex peaksets

```

r1 <- starr
r2 <- peaks.list$dex
overlap <- findOverlaps(r1, r2, select = "all")
overlap

```

```

## Hits object with 1 hit and 0 metadata columns:
##      queryHits subjectHits
##      <integer>  <integer>
## [1]      104      18123
## -----
## queryLength: 234 / subjectLength: 24212

```

Make vector of SNPs to peaks

```

hits <- names(r2)[subjectHits(overlap)]
names(hits) <- names(r1)[queryHits(overlap)]
hits

```

```

## rs419099
## "33559"

```

```
r1[names(hits),]
```

```
## GRanges object with 1 range and 0 metadata columns:
##           seqnames      ranges strand
##           <Rle> <IRanges> <Rle>
## rs419099      5 102089442      *
## -----
## seqinfo: 20 sequences from an unspecified genome; no seqlengths
```

```
r2[hits, ]
```

```
## GRanges object with 1 range and 3 metadata columns:
##           seqnames      ranges strand |      score      scoreA      scoreB
##           <Rle>           <IRanges> <Rle> | <numeric> <numeric> <numeric>
## 33559      5 102089398-102089591      * | 0.00738297      NA      NA
## -----
## seqinfo: 131 sequences from an unspecified genome; no seqlengths
```

Find the distances to the nearest SNP for each peak for dex

```
nearest(r1, r2)[1:20]
```

```
## [1] 1168 1170 1172 1173 5556 18125 20754 20754 20842 20866 20756 20758
## [13] 18125 20755 20873 7557 7942 4869 6751 11987
```

```
dist.to.nearest.obj <- distanceToNearest(r1, r2)
dist.to.nearest.obj
```

```
## Hits object with 234 hits and 1 metadata column:
##           queryHits subjectHits | distance
##           <integer>  <integer> | <integer>
## [1]           1         1168 |      6218
## [2]           2         1170 |     14082
## [3]           3         1172 |      7028
## [4]           4         1173 |     14712
## [5]           5         5556 |     31027
## ...           ...         ... |      ...
## [230]        230        18126 |     80919
## [231]        231        20758 |     11589
## [232]        232        14905 |     63928
## [233]        233        19200 |      6069
## [234]        234         4326 |     13773
## -----
## queryLength: 234 / subjectLength: 24212
```

```
dists <- dist.to.nearest.obj@elementMetadata@listData$distance
summary(dists)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##         0   10015   30985   61067   75534   595803
```

```
# hist(dists,xlab = "dist to nearest peak", main = "Distances")
sort(dists)[1:20]
```

```
## [1]      0   130   154   234   369   409   771   818  1079  1172  1298  1481  1539  1768  1990
## [16] 1999  2335  2478  2563  2640
```

```
hits <- names(r2)[subjectHits(dist.to.nearest.obj)]
names(hits) <- names(r1)[queryHits(dist.to.nearest.obj)]
hits[1:3 ]
```

```
## 1-148409926 1-148443906 1-148541113
##      "2284"      "2287"      "2291"
```

```
r1[names(hits)[1], ]
```

```
## GRanges object with 1 range and 0 metadata columns:
##           seqnames      ranges strand
##           <Rle> <IRanges>  <Rle>
## 1-148409926      1 150143302      *
## -----
## seqinfo: 20 sequences from an unspecified genome; no seqlengths
```

```
r2[hits[1], ]
```

```
## GRanges object with 1 range and 3 metadata columns:
##           seqnames      ranges strand |      score      scoreA      scoreB
##           <Rle>          <IRanges> <Rle> | <numeric> <numeric> <numeric>
## 2284          1 150149521-150149887   * | 0.00188891      NA      NA
## -----
## seqinfo: 131 sequences from an unspecified genome; no seqlengths
```

Combine SNP, sites and distances into one dataframe

```
# dist.df <- cbind(as.data.frame(hits), dists)
# dist.df

starr.dist <- cbind(as.data.frame(r1[names(hits), ]), dists) %>%
  mutate(CHR = seqnames, POS = start, DIST = dists) %>%
  select(CHR, POS, DIST)

starr.bs.df <- cbind(starr.dist, as.data.frame(r2[hits, ]@seqnames), as.data.frame(r2[hits, ]@ranges))
colnames(starr.bs.df) <- c("CHR", "POS", "DIST", "PEAK_CHR", "PEAK_START", "PEAK_END", "PEAK_WIDTH", "PEAK_SCORE")
starr.bs.df["TREATMENT"] <- "DEX"
starr.bs.df <- starr.bs.df[order(starr.bs.df$DIST), ]
starr.bs.df[1:10, ]
```

```
##          CHR      POS DIST PEAK_CHR PEAK_START  PEAK_END  PEAK_WIDTH  PEAK_NAME
## rs419099    5 102089442    0         5  102089398 102089591        194    33559
## rs3134943   6  32147761   130        6   32147892  32148233        342    35658
## rs9273528   6  32628633   154        6   32628231  32628478        248    35673
## rs7298122  12   7870848   234       12    7870359   7870613        255     8600
## rs7439210   4   9971749   369        4    9970511   9971379        869   30329
## rs2263658  20   1520272   409       20    1520682   1521183        502   24659
## rs1968126   7   66057004  771        7    66055297  66056232        936   38649
## rs4242901  12   7871432   818       12    7870359   7870613        255     8600
## rs797263    1   9238553  1079        1    9239633   9240153        521     225
## rs9275314   6  32665909  1172        6   32667082  32668412       1331   35675
##          TREATMENT
## rs419099      DEX
## rs3134943      DEX
## rs9273528      DEX
## rs7298122      DEX
## rs7439210      DEX
## rs2263658      DEX
## rs1968126      DEX
## rs4242901      DEX
## rs797263       DEX
## rs9275314      DEX
```

```
# write.csv2(starr.bs.df, paste0(result.dir, "01_STARR_SNP_TO_ChIP_BINDING_SITES_DISTANCES_DEX.csv"), r
```

Find the distances to the nearest SNP for each peak for VEH

```
r1 <- starr
r2 <- peaks.list$veh
dist.to.nearest.obj <- distanceToNearest(r1, r2)
dists <- dist.to.nearest.obj@elementMetadata@listData$distance
summary(dist.to.nearest.obj)
```

```
## [1] "Hits object with 234 hits and 1 metadata column"
```

```
hits <- names(r2)[subjectHits(dist.to.nearest.obj)]
names(hits) <- names(r1)[queryHits(dist.to.nearest.obj)]
```

Combine SNP, sites and distances into one dataframe

```
# dist.df <- cbind(as.data.frame(hits), dists)
# dist.df

starr.dist <- cbind(as.data.frame(r1[names(hits), ]), dists) %>%
  dplyr::mutate(CHR = seqnames, POS = start, DIST = dists) %>%
  dplyr::select(CHR, POS, DIST)

starr.bs.df <- cbind(starr.dist, as.data.frame(r2[hits, ]@seqnames), as.data.frame(r2[hits, ]@ranges))
colnames(starr.bs.df) <- c("CHR", "POS", "DIST", "PEAK_CHR", "PEAK_START", "PEAK_END", "PEAK_WIDTH", "PEAK_NAME", "TREATMENT")
starr.bs.df["TREATMENT"] <- "VEH"
starr.bs.df <- starr.bs.df[order(starr.bs.df$DIST), ]
starr.bs.df[1:10, ]
```

```
##          CHR          POS DIST PEAK_CHR PEAK_START  PEAK_END PEAK_WIDTH
## rs6963745      7  29566161  123      7   29566285  29566648      364
## rs11769079     7  65842128 2069      7   65839852  65840058      207
## rs11788797     9 130941741 2191      9  130939346 130939549      204
## rs10793957     9 136087931 2302      9  136090234 136090443      210
## rs45585631     9 130942972 3422      9  130939346 130939549      204
## rs9890920     17  7400041  4712     17   7404754   7405095      342
## rs3826440     17  7411455  6359     17   7404754   7405095      342
## rs406589       3  41125313  7482      3  41117524  41117830      307
## rs10873133    14  61632115  8414     14  61640530  61640722      193
## rs384097       3  41126731  8900      3  41117524  41117830      307
##          PEAK_NAME TREATMENT
## rs6963745      38238      VEH
## rs11769079      38644      VEH
## rs11788797      43519      VEH
## rs10793957      43629      VEH
## rs45585631      43519      VEH
## rs9890920      16579      VEH
## rs3826440      16579      VEH
## rs406589       27921      VEH
## rs10873133     12484      VEH
## rs384097       27921      VEH
```

```
# write.csv2(starr.bs.df, paste0(result.dir, "02_STARR_SNP_TO_ChIP_BINDING_SITES_DISTANCES_VEH.csv"),
```

Find the distances to the nearest SNP for each peak for VEH

```
r1 <- starr
r2 <- peaks.list$`veh-dex`
dist.to.nearest.obj <- distanceToNearest(r1, r2)
dists <- dist.to.nearest.obj@elementMetadata@listData$distance
summary(dist.to.nearest.obj)
```

```
## [1] "Hits object with 234 hits and 1 metadata column"
```

```
hits <- names(r2)[subjectHits(dist.to.nearest.obj)]
names(hits) <- names(r1)[queryHits(dist.to.nearest.obj)]
```

Combine SNP, sites and distances into one dataframe

```
# dist.df <- cbind(as.data.frame(hits), dists)
# dist.df

starr.dist <- cbind(as.data.frame(r1[names(hits), ]), dists) %>%
  dplyr::mutate(CHR = seqnames, POS = start, DIST = dists) %>%
  dplyr::select(CHR, POS, DIST)

starr.bs.df <- cbind(starr.dist, as.data.frame(r2[hits, ]@seqnames), as.data.frame(r2[hits, ]@ranges))
colnames(starr.bs.df) <- c("CHR", "POS", "DIST", "PEAK_CHR", "PEAK_START", "PEAK_END", "PEAK_WIDTH", "PEAK_NAME")
starr.bs.df["TREATMENT"] <- "VEH-DEX"
starr.bs.df <- starr.bs.df[order(starr.bs.df$DIST), ]
starr.bs.df[1:10, ]
```

```
##          CHR      POS DIST PEAK_CHR PEAK_START  PEAK_END PEAK_WIDTH
## rs9861194      3 40309225   52        3  40309278  40310318      1041
## rs7798630      7  65771479  510        7   65770427  65770968       542
## rs9275314      6  32665909  557        6   32664592  32665351       760
## rs10274883     7  66116091  638        7   66114633  66115452       820
## 12-111668114  12 113183731 1123       12  113184855 113186089     1235
## rs4846085      1  12050634 1294        1   12051929  12052491       563
## rs17826037    12 113183359 1495       12  113184855 113186089     1235
## rs11788797     9 130941741 1773        9  130939589 130939967       379
## rs56125600     1 150012917 1877        1  150010128 150011039       912
## rs10235858     7  43652411 2098        7   43648189  43650312     2124
##          PEAK_NAME TREATMENT
## rs9861194      27912   VEH-DEX
## rs7798630      38641   VEH-DEX
## rs9275314      35674   VEH-DEX
## rs10274883     38651   VEH-DEX
## 12-111668114   10302   VEH-DEX
## rs4846085       295    VEH-DEX
## rs17826037     10302   VEH-DEX
## rs11788797     43520   VEH-DEX
## rs56125600      2281   VEH-DEX
## rs10235858     38419   VEH-DEX
```

```
# write.csv2(starr.bs.df, paste0(result.dir, "03_STARR_SNP_TO_ChIP_BINDING_SITES_DISTANCES_VEH-DEX.csv"))
```

4. Overlap all GR-SNPs data with peaksets

```
r1 <- grsnp
r2 <- peaks.list$dex
overlap <- findOverlaps(r1, r2, select = "all")
overlap
```

```
## Hits object with 25 hits and 0 metadata columns:
```

```
##          queryHits subjectHits
##          <integer>  <integer>
##    [1]          280          18787
##    [2]          301          20756
##    [3]          393          20757
##    [4]          487          23595
##    [5]          722           1267
##    ...          ...           ...
##   [21]         3497          19203
##   [22]         3498          19203
##   [23]         3499          19203
##   [24]         3500          19203
##   [25]         3651           7358
## -----
## queryLength: 3662 / subjectLength: 24212
```

Make vector of SNPs to peaks

```
hits <- names(r2)[subjectHits(overlap)]
names(hits) <- names(r1)[queryHits(overlap)]
hits
```

```
## rs1022549 rs10236317 rs10480185 rs10987905 rs12128066 rs13438566 rs1762
## "34804" "38421" "38422" "43518" "2489" "38424" "35672"
## rs1786186 rs2256520 rs2395891 rs2942904 rs377325 rs419099 rs454744
## "8137" "24659" "19416" "22640" "33559" "33559" "33559"
## rs584356 rs8033385 rs9271353 rs9271354 rs9275334 rs9275338 rs9275356
## "8136" "14155" "35669" "35669" "35675" "35675" "35675"
## rs9275360 rs9275365 rs9275373 rs994402
## "35675" "35675" "35675" "13571"
```

```
r1[names(hits),]
```

```
## GRanges object with 25 ranges and 0 metadata columns:
##           seqnames      ranges strand
##           <Rle> <IRanges> <Rle>
## rs1022549      6 1714677      *
## rs10236317      7 43735644     *
## rs10480185      7 43756916     *
## rs10987905      9 130931586    *
## rs12128066      1 156127368    *
## ...           ...           ...
## rs9275356      6 32667850      *
## rs9275360      6 32667937      *
## rs9275365      6 32668125      *
## rs9275373      6 32668411      *
## rs994402      15 40161267      *
## -----
## seqinfo: 22 sequences from an unspecified genome; no seqlengths
```

```
r2[hits, ]
```

```
## GRanges object with 25 ranges and 3 metadata columns:
##           seqnames      ranges strand |      score      scoreA      scoreB
##           <Rle> <IRanges> <Rle> | <numeric> <numeric> <numeric>
## 34804      6 1714393-1714901      * | 0.00343692      NA      NA
## 38421      7 43734990-43736205      * | 0.00651783      NA      NA
## 38422      7 43756647-43756959      * | 0.00392346      NA      NA
## 43518      9 130931475-130931906      * | 0.00288559      NA      NA
## 2489      1 156126432-156127605      * | 0.00397314      NA      NA
## ...           ...           ... | ...
## 35675      6 32667082-32668412      * | 0.00610844      NA      NA
## 35675      6 32667082-32668412      * | 0.00610844      NA      NA
## 35675      6 32667082-32668412      * | 0.00610844      NA      NA
## 35675      6 32667082-32668412      * | 0.00610844      NA      NA
## 13571     15 40160980-40161614      * | 0.00198467      NA      NA
## -----
## seqinfo: 131 sequences from an unspecified genome; no seqlengths
```


Find the distances to the nearest SNP for each peak for DEX

```
r1 <- grsnp
r2 <- peaks.list$dex
dist.to.nearest.obj <- distanceToNearest(r1, r2)
dists <- dist.to.nearest.obj@elementMetadata@listData$distance
summary(dist.to.nearest.obj)
```

```
## [1] "Hits object with 3662 hits and 1 metadata column"
```

```
hits <- names(r2)[subjectHits(dist.to.nearest.obj)]
names(hits) <- names(r1)[queryHits(dist.to.nearest.obj)]
hits[1:10]
```

```
## 1-148268256 1-148271849 1-148275208 1-148275840 1-148281790 1-148283858
##      "2280"      "2280"      "2280"      "2280"      "2280"      "2280"
## 1-148285660 1-148290924 1-148291801 1-148313743
##      "2280"      "2280"      "2280"      "2283"
```

Combine SNP, sites and distances into one dataframe

```
# dist.df <- cbind(as.data.frame(hits), dists)
# dist.df

grsnp.dist <- cbind(as.data.frame(r1[names(hits), ]), dists) %>%
  dplyr::mutate(CHR = seqnames, POS = start, DIST = dists) %>%
  dplyr::select(CHR, POS, DIST)

grsnp.bs.df <- cbind(grsnp.dist, as.data.frame(r2[hits, ]@seqnames), as.data.frame(r2[hits, ]@ranges))
colnames(grsnp.bs.df) <- c("CHR", "POS", "DIST", "PEAK_CHR", "PEAK_START", "PEAK_END", "PEAK_WIDTH", "PEAK_ID")
grsnp.bs.df["TREATMENT"] <- "DEX"
grsnp.bs.df <- grsnp.bs.df[order(grsnp.bs.df$DIST), ]
grsnp.bs.df[1:30, ]
```

##		CHR	POS	DIST	PEAK_CHR	PEAK_START	PEAK_END	PEAK_WIDTH
##	rs1022549	6	1714677	0	6	1714393	1714901	509
##	rs10236317	7	43735644	0	7	43734990	43736205	1216
##	rs10480185	7	43756916	0	7	43756647	43756959	313
##	rs10987905	9	130931586	0	9	130931475	130931906	432
##	rs12128066	1	156127368	0	1	156126432	156127605	1174
##	rs13438566	7	43784862	0	7	43784825	43785184	360
##	rs1762	6	32627777	0	6	32627567	32627866	300
##	rs1786186	11	117945632	0	11	117944896	117945802	907
##	rs2256520	20	1520883	0	20	1520682	1521183	502
##	rs2395891	19	2032148	0	19	2032117	2032899	783
##	rs2942904	2	101306999	0	2	101306722	101307344	623
##	rs377325	5	102089491	0	5	102089398	102089591	194
##	rs419099	5	102089442	0	5	102089398	102089591	194
##	rs454744	5	102089523	0	5	102089398	102089591	194
##	rs584356	11	117930428	0	11	117930261	117930895	635
##	rs8033385	15	68820489	0	15	68820455	68820728	274

```
## rs9271353      6 32584014    0      6 32583977 32584552    576
## rs9271354      6 32584066    0      6 32583977 32584552    576
## rs9275334      6 32667107    0      6 32667082 32668412   1331
## rs9275338      6 32667343    0      6 32667082 32668412   1331
## rs9275356      6 32667850    0      6 32667082 32668412   1331
## rs9275360      6 32667937    0      6 32667082 32668412   1331
## rs9275365      6 32668125    0      6 32667082 32668412   1331
## rs9275373      6 32668411    0      6 32667082 32668412   1331
## rs994402      15 40161267    0     15 40160980 40161614    635
## 1-148347283    1 150080659   20      1 150080680 150081225    546
## rs2295281      1 12059412   20      1 12059433 12060203    771
## rs3129757      6 32584609   56      6 32583977 32584552    576
## rs7681628      4 38856384   58      4 38856443 38856927    485
## rs11773628     7 65982631   59      7 65981813 65982571    759
##
##          PEAK_NAME TREATMENT
## rs1022549      34804      DEX
## rs10236317     38421      DEX
## rs10480185     38422      DEX
## rs10987905     43518      DEX
## rs12128066      2489      DEX
## rs13438566     38424      DEX
## rs1762         35672      DEX
## rs1786186       8137      DEX
## rs2256520     24659      DEX
## rs2395891     19416      DEX
## rs2942904     22640      DEX
## rs377325      33559      DEX
## rs419099      33559      DEX
## rs454744      33559      DEX
## rs584356       8136      DEX
## rs8033385     14155      DEX
## rs9271353     35669      DEX
## rs9271354     35669      DEX
## rs9275334     35675      DEX
## rs9275338     35675      DEX
## rs9275356     35675      DEX
## rs9275360     35675      DEX
## rs9275365     35675      DEX
## rs9275373     35675      DEX
## rs994402     13571      DEX
## 1-148347283    2283      DEX
## rs2295281       297      DEX
## rs3129757     35669      DEX
## rs7681628     30600      DEX
## rs11773628    38648      DEX
```

```
# write.csv2(grsnp.bs.df, paste0(result.dir, "04_GRSNP_SNP_TO_ChIP_BINDING_SITES_DISTANCES_DEX.csv"), r
```

Find the distances to the nearest SNP for each peak for VEH

```
r1 <- grsnp
r2 <- peaks.list$veh
```

```
dist.to.nearest.obj <- distanceToNearest(r1, r2)
dists <- dist.to.nearest.obj@elementMetadata@listData$distance
summary(dist.to.nearest.obj)
```

```
## [1] "Hits object with 3662 hits and 1 metadata column"
```

```
hits <- names(r2)[subjectHits(dist.to.nearest.obj)]
names(hits) <- names(r1)[queryHits(dist.to.nearest.obj)]
hits[1:10]
```

```
## 1-148268256 1-148271849 1-148275208 1-148275840 1-148281790 1-148283858
## "2276" "2276" "2276" "2276" "2276" "2276"
## 1-148285660 1-148290924 1-148291801 1-148313743
## "2276" "2276" "2276" "2276"
```

Combine SNP, sites and distances into one dataframe

```
# dist.df <- cbind(as.data.frame(hits), dists)
# dist.df

grsnp.dist <- cbind(as.data.frame(r1[names(hits), ]), dists) %>%
  dplyr::mutate(CHR = seqnames, POS = start, DIST = dists) %>%
  dplyr::select(CHR, POS, DIST)

grsnp.bs.df <- cbind(grsnp.dist, as.data.frame(r2[hits, ]@seqnames), as.data.frame(r2[hits, ]@ranges))
colnames(starr.bs.df) <- c("CHR", "POS", "DIST", "PEAK_CHR", "PEAK_START", "PEAK_END", "PEAK_WIDTH", "PEAK_VALUE")
grsnp.bs.df["TREATMENT"] <- "VEH"
grsnp.bs.df <- grsnp.bs.df[order(grsnp.bs.df$DIST), ]
grsnp.bs.df[1:20, ]
```

##	CHR	POS	DIST	value	start	end	width	names	TREATMENT
##	rs8075218	17	7405074	0	17	7404754	7405095	342 16579	VEH
##	rs4815437	20	25584788	67	20	25584856	25585730	875 25007	VEH
##	rs60492890	19	1275368	99	19	1274970	1275268	299 19373	VEH
##	rs6963745	7	29566161	123	7	29566285	29566648	364 38238	VEH
##	rs6963739	7	29566147	137	7	29566285	29566648	364 38238	VEH
##	rs34771359	7	29565995	289	7	29566285	29566648	364 38238	VEH
##	rs10121371	9	2790559	310	9	2790870	2791199	330 41981	VEH
##	rs8033385	15	68820489	351	15	68820841	68821042	202 14156	VEH
##	rs778680	7	65840414	355	7	65839852	65840058	207 38644	VEH
##	rs10782443	14	61641164	441	14	61640530	61640722	193 12484	VEH
##	rs4075428	9	130940121	571	9	130939346	130939549	204 43519	VEH
##	rs442115	3	41118450	619	3	41117524	41117830	307 27921	VEH
##	rs7315258	12	48572670	620	12	48571757	48572049	293 9218	VEH
##	rs6597610	9	136091096	652	9	136090234	136090443	210 43629	VEH
##	rs4454354	9	136089529	704	9	136090234	136090443	210 43629	VEH
##	rs12979308	19	1275987	718	19	1274970	1275268	299 19373	VEH
##	rs59464952	17	5585519	742	17	5584486	5584776	291 16528	VEH
##	rs62364874	5	102392179	822	5	102391055	102391356	302 33563	VEH
##	rs62364875	5	102392190	833	5	102391055	102391356	302 33563	VEH
##	rs56361326	5	102389888	1166	5	102391055	102391356	302 33563	VEH

```
# write.csv2(grsnp.bs.df, paste0(result.dir, "05_GRSNP_SNP_TO_ChIP_BINDING_SITES_DISTANCES_VEH.csv"), r
```

Find the distances to the nearest SNP for each peak for DEX-VEH

```
r1 <- grsnp
r2 <- peaks.list$`veh-dex`
dist.to.nearest.obj <- distanceToNearest(r1, r2)
dists <- dist.to.nearest.obj@elementMetadata@listData$distance
summary(dist.to.nearest.obj)
```

```
## [1] "Hits object with 3662 hits and 1 metadata column"
```

```
hits <- names(r2)[subjectHits(dist.to.nearest.obj)]
names(hits) <- names(r1)[queryHits(dist.to.nearest.obj)]
hits[1:10]
```

```
## 1-148268256 1-148271849 1-148275208 1-148275840 1-148281790 1-148283858
##      "2279"      "2281"      "2281"      "2281"      "2281"      "2281"
## 1-148285660 1-148290924 1-148291801 1-148313743
##      "2281"      "2281"      "2281"      "2282"
```

Combine SNP, sites and distances into one dataframe

```
# dist.df <- cbind(as.data.frame(hits), dists)
# dist.df

grsnp.dist <- cbind(as.data.frame(r1[names(hits), ]), dists) %>%
  dplyr::mutate(CHR = seqnames, POS = start, DIST = dists) %>%
  dplyr::select(CHR, POS, DIST)

grsnp.bs.df <- cbind(grsnp.dist, as.data.frame(r2[hits, ]@seqnames), as.data.frame(r2[hits, ]@ranges))
colnames(starr.bs.df) <- c("CHR", "POS", "DIST", "PEAK_CHR", "PEAK_START", "PEAK_END", "PEAK_WIDTH", "PI")
grsnp.bs.df["TREATMENT"] <- "VEH-DEX"
grsnp.bs.df <- grsnp.bs.df[order(grsnp.bs.df$DIST), ]
grsnp.bs.df[1:20, ]
```

##	CHR	POS	DIST	value	start	end	width	names	TREATMENT
##	2-98508072	2	99141640	0	2	99140879	99141865	987 22607	VEH-DEX
##	6-32592568	6	32484590	0	6	32484312	32484830	519 35667	VEH-DEX
##	rs10009593	4	56436681	0	4	56434907	56436740	1834 30838	VEH-DEX
##	rs1006717	19	2841360	0	19	2840900	2841508	609 19483	VEH-DEX
##	rs10226623	7	43729316	0	7	43729181	43729972	792 38420	VEH-DEX
##	rs10243963	7	43729899	0	7	43729181	43729972	792 38420	VEH-DEX
##	rs10252175	7	43649141	0	7	43648189	43650312	2124 38419	VEH-DEX
##	rs10282065	7	43649248	0	7	43648189	43650312	2124 38419	VEH-DEX
##	rs12126312	1	150364447	0	1	150363666	150365702	2037 2294	VEH-DEX
##	rs12979308	19	1275987	0	19	1275430	1276352	923 19374	VEH-DEX
##	rs1584135	15	68818569	0	15	68818423	68819580	1158 14153	VEH-DEX
##	rs17137491	6	4018568	0	6	4018095	4018881	787 34874	VEH-DEX

##	rs28445925	4	56435849	0	4	56434907	56436740	1834	30838	VEH-DEX
##	rs3132946	6	32190028	0	6	32189846	32190559	714	35660	VEH-DEX
##	rs34235717	7	43648749	0	7	43648189	43650312	2124	38419	VEH-DEX
##	rs34485459	6	32623277	0	6	32622432	32623960	1529	35671	VEH-DEX
##	rs34794727	7	43648574	0	7	43648189	43650312	2124	38419	VEH-DEX
##	rs36113818	7	43648723	0	7	43648189	43650312	2124	38419	VEH-DEX
##	rs60157471	2	113792252	0	2	113791830	113792531	702	22815	VEH-DEX
##	rs61776516	1	12047028	0	1	12046985	12048020	1036	294	VEH-DEX

```
# write.csv2(grsnp.bs.df, paste0(result.dir, "06_GRSNP_SNP_TO_ChIP_BINDING_SITES_DISTANCES_VEH_DEX.csv"))
```