r_assignment_markdown

Tom McBrien

October 14, 2015

Loading and Merging Data

```
d <- read.csv("Dataset_S1.txt", sep = ',', header = TRUE) #reading in Data
head(d)
```

```
##   start   end total.SNPs total.Bases depth unique.SNPs dhSNPs
## 1 55001 56000          0        1894  3.41           0      0
## 2 56001 57000          5        6683  6.68           2      2
## 3 57001 58000          1        9063  9.06           1      0
## 4 58001 59000          7       10256 10.26           3      2
## 5 59001 60000          4        8057  8.06           4      0
## 6 60001 61000          6        7051  7.05           2      1
##   reference.Bases  Theta     Pi Heterozygosity    X.GC Recombination
## 1             556  0.000  0.000          0.000 54.8096   0.009601574
## 2            1000  8.007 10.354          7.481 42.4424   0.009601574
## 3            1000  3.510  1.986          1.103 37.2372   0.009601574
## 4            1000  9.929  9.556          6.582 38.0380   0.009601574
## 5            1000 12.915  8.506          4.965 41.3413   0.009601574
## 6            1000  7.817  9.121          8.864 36.1361   0.009601574
##    Divergence Constraint SNPs
## 1 0.003006012          0    0
## 2 0.018018020          0    0
## 3 0.007007007          0    0
## 4 0.012012010          0    0
## 5 0.024024020          0    0
## 6 0.016016020          0    0
```

```
rcmb <- read.delim("motif_recombrates.txt", header = TRUE) #read in motif recomb rates data
rpts <- read.delim("motif_repeats.txt", header = TRUE) #read in motif repeat rates data
rcmb$pos <- paste(rcmb$chr, rcmb$motif_start, sep="-") #making column of specific positions per chromos
rpts$pos <- paste(rpts$chr, rpts$motif_start, sep="-") #same as above with repeats file
joined <- merge(rcmb, rpts, by.x="pos", by.y="pos") #mergin
head(joined)
```

```
##              pos chr.x motif_start.x motif_end    dist recomb_start
## 1 chr1-101890123  chr1     101890123 101890136 34154.0    101855215
## 2 chr1-101890123  chr1     101890123 101890136 35717.5    101853608
## 3 chr1-101890123  chr1     101890123 101890136  9704.0    101878637
## 4 chr1-101890123  chr1     101890123 101890136  7864.5    101882213
## 5 chr1-101890123  chr1     101890123 101890136 29463.0    101859577
## 6 chr1-101890123  chr1     101890123 101890136 37189.5    101852271
##   recomb_end  recom         motif chr.y     start       end  name
## 1  101856736 0.0700 CCTCCCTAGCCAC  chr1 101890032 101890381 THE1B
## 2  101855216 0.0722 CCTCCCTAGCCAC  chr1 101890032 101890381 THE1B
## 3  101882214 0.2445 CCTCCCTAGCCAC  chr1 101890032 101890381 THE1B
## 4  101882317 0.2445 CCTCCCTAGCCAC  chr1 101890032 101890381 THE1B
## 5  101861756 0.0691 CCTCCCTAGCCAC  chr1 101890032 101890381 THE1B
## 6  101853609 0.4441 CCTCCCTAGCCAC  chr1 101890032 101890381 THE1B
##   motif_start.y
## 1     101890123
## 2     101890123
## 3     101890123
## 4     101890123
## 5     101890123
## 6     101890123
```

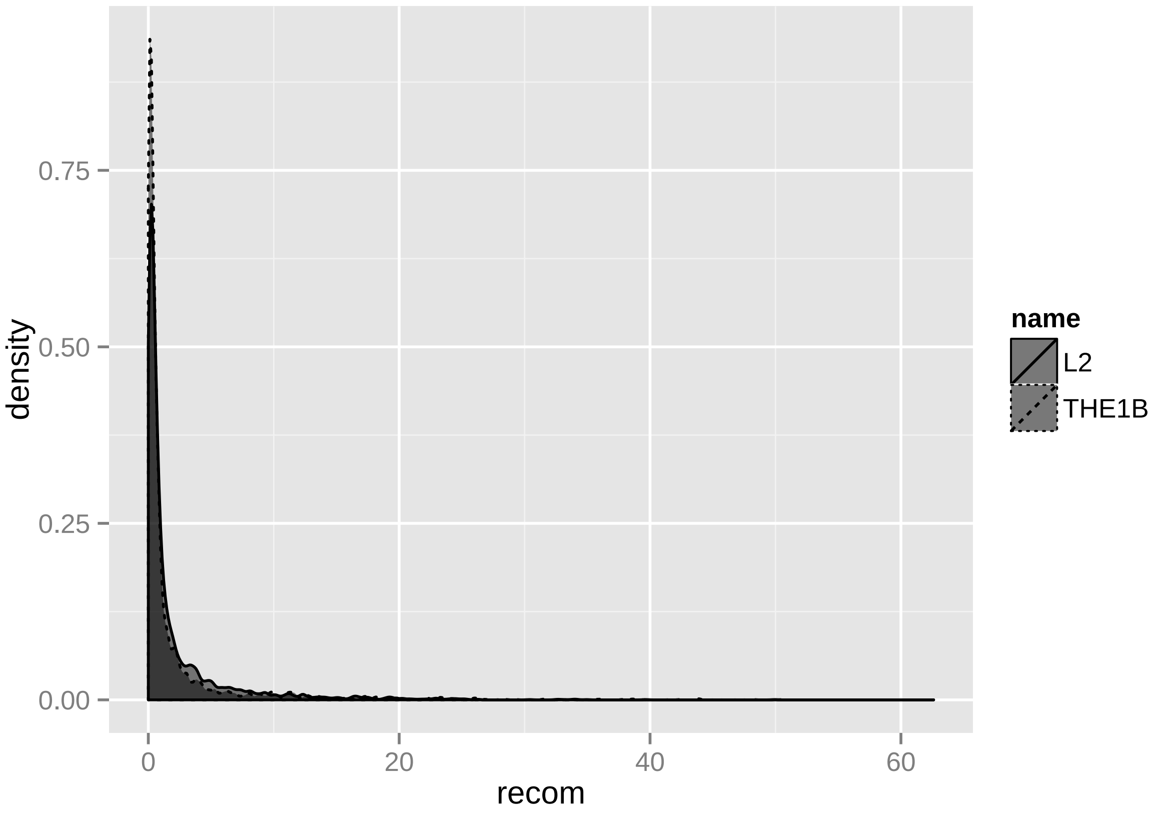
Analysing Data

```
aggregate(joined$recom, list(motif=joined$motif), mean) #this uses the aggregate function to give two s
```

```
##           motif        x
## 1 CCTCCCTAGCCAC 1.963472
## 2 CCTCCCTGACCAC 2.138344
```

Analyzing if Distributions of Recombination Rate Differs by Motif Type

```
library(ggplot2)
ggplot(joined) + geom_density(aes(x=recom, linetype=name), fill='black', alpha=0.5)
```



Recombination Rates of Motif Types Vs. Background

```
joined_with_background <- merge(rcmb, rpts, by.x = "pos", by.y = "pos",
    all.x = TRUE)  #merging with left outer join so i get all data
head(joined_with_background)
```

```
##              pos chr.x motif_start.x motif_end    dist recomb_start
## 1 chr1-101890123  chr1     101890123 101890136 34154.0    101855215
## 2 chr1-101890123  chr1     101890123 101890136 35717.5    101853608
## 3 chr1-101890123  chr1     101890123 101890136  9704.0    101878637
## 4 chr1-101890123  chr1     101890123 101890136  7864.5    101882213
## 5 chr1-101890123  chr1     101890123 101890136 29463.0    101859577
## 6 chr1-101890123  chr1     101890123 101890136 37189.5    101852271
##   recomb_end  recom         motif chr.y     start       end  name
## 1  101856736 0.0700 CCTCCCTAGCCAC  chr1 101890032 101890381 THE1B
## 2  101855216 0.0722 CCTCCCTAGCCAC  chr1 101890032 101890381 THE1B
## 3  101882214 0.2445 CCTCCCTAGCCAC  chr1 101890032 101890381 THE1B
## 4  101882317 0.2445 CCTCCCTAGCCAC  chr1 101890032 101890381 THE1B
## 5  101861756 0.0691 CCTCCCTAGCCAC  chr1 101890032 101890381 THE1B
## 6  101853609 0.4441 CCTCCCTAGCCAC  chr1 101890032 101890381 THE1B
##   motif_start.y
## 1     101890123
## 2     101890123
## 3     101890123
## 4     101890123
## 5     101890123
## 6     101890123
```

```
joined_with_background$category <- ifelse(joined_with_background$name ==
    "THE1B", 1, 2)  #I am making a new column that will call all THE1B '1', L2 '2', and <NA> 'NA' becau
head(joined_with_background[, c("chr.x", "motif", "chr.y", "name",
    "category")], 50)
```

```
##    chr.x         motif chr.y  name category
## 1   chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 2   chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 3   chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 4   chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 5   chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 6   chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 7   chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 8   chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 9   chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 10  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 11  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 12  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 13  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 14  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 15  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 16  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 17  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 18  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 19  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 20  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 21  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 22  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 23  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 24  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 25  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 26  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 27  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 28  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 29  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 30  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 31  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 32  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 33  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 34  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 35  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 36  chr1 CCTCCCTAGCCAC  chr1 THE1B        1
## 37  chr1 CCTCCCTAGCCAC  <NA>  <NA>       NA
## 38  chr1 CCTCCCTAGCCAC  <NA>  <NA>       NA
## 39  chr1 CCTCCCTAGCCAC  <NA>  <NA>       NA
## 40  chr1 CCTCCCTAGCCAC  <NA>  <NA>       NA
## 41  chr1 CCTCCCTAGCCAC  <NA>  <NA>       NA
## 42  chr1 CCTCCCTAGCCAC  <NA>  <NA>       NA
## 43  chr1 CCTCCCTAGCCAC  <NA>  <NA>       NA
## 44  chr1 CCTCCCTAGCCAC  <NA>  <NA>       NA
## 45  chr1 CCTCCCTAGCCAC  <NA>  <NA>       NA
## 46  chr1 CCTCCCTAGCCAC  <NA>  <NA>       NA
## 47  chr1 CCTCCCTAGCCAC  <NA>  <NA>       NA
## 48  chr1 CCTCCCTAGCCAC  <NA>  <NA>       NA
## 49  chr1 CCTCCCTAGCCAC  <NA>  <NA>       NA
## 50  chr1 CCTCCCTAGCCAC  <NA>  <NA>       NA
```

```r
joined_with_background[c("category")][is.na(joined_with_background[c("category")])] <- 0  #this will ma
joined_with_background$newname <- ifelse(joined_with_background$category ==
    0, joined_with_background$newname <- "NA", ifelse(joined_with_background$category ==
    1, joined_with_background$newname <- "THE1B", joined_with_background$newname <- "L2"))
# GGPLOT should now be able to separate out linetypes because
# not using a continual number variable
head(joined_with_background[, c("chr.x", "motif", "chr.y", "name",
    "category", "newname")], 100)
```

```
##     chr.x         motif chr.y  name category newname
## 1    chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 2    chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 3    chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 4    chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 5    chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 6    chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 7    chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 8    chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 9    chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 10   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 11   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 12   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 13   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 14   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 15   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 16   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 17   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 18   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 19   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 20   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 21   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 22   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 23   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 24   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 25   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 26   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 27   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 28   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 29   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 30   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 31   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 32   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 33   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 34   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 35   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 36   chr1 CCTCCCTAGCCAC  chr1 THE1B        1   THE1B
## 37   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 38   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 39   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 40   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 41   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 42   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 43   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 44   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 45   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 46   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 47   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 48   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 49   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 50   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 51   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 52   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 53   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 54   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 55   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 56   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 57   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 58   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 59   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 60   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 61   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 62   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 63   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 64   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 65   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 66   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 67   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 68   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 69   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 70   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 71   chr1 CCTCCCTAGCCAC  <NA>  <NA>        0      NA
## 72   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 73   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 74   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 75   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 76   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 77   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 78   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 79   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 80   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 81   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 82   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 83   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 84   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 85   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 86   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 87   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 88   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 89   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 90   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 91   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 92   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 93   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 94   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 95   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 96   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 97   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 98   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 99   chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
## 100  chr1 CCTCCCTGACCAC  <NA>  <NA>        0      NA
```

```r
## SUMMARY OF NON-BACKGROUND RECOMB RATES##
summary(joined_with_background$recom[joined_with_background$category >=
    1])
```

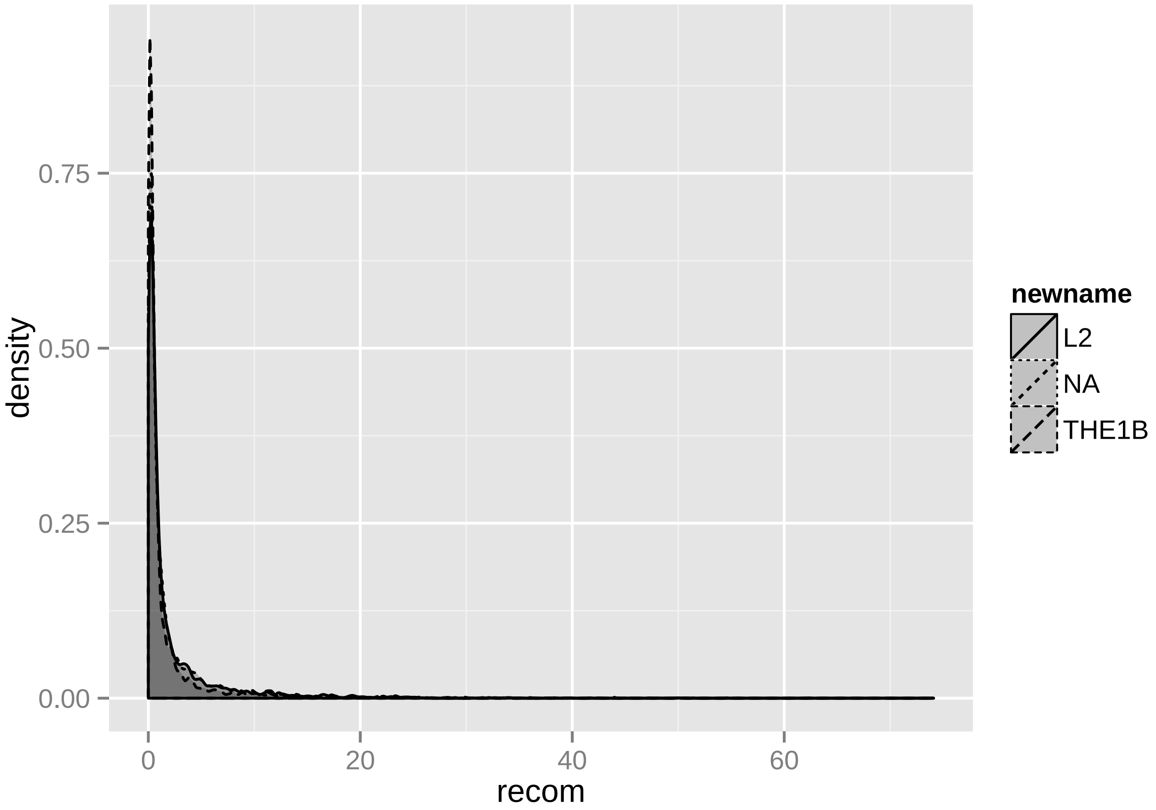
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0001  0.1770  0.5129  2.0410  1.6860 62.6100
```

```r
## SUMMARY OF BACKGROUND RECOMB RATES##
summary(joined_with_background$recom[joined_with_background$category ==
    0])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0002  0.2100  0.5869  2.0930  1.9890 74.1000
```

```r
## PLOT OF DIFFERENCES IN RECOM RATES##

ggplot(joined_with_background) + geom_density(aes(x = recom,
    linetype = newname), fill = "black", alpha = 0.2)
```


0.75
0.50
0.25
0.00
density
0
20
40
60
recom
newname
L2
NA
THE1B