

# 01.1\_\_Prepare-CredbileSet-T2D-file.Rmd

*Jason Torres*

*February 15, 2017*

Setup libraries and directories

```
"%&%&" <- function(a,b) paste0(a,b)
library("dplyr")
```

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library("data.table")
```

```
## -----
## data.table + dplyr code now lives in dtplyr.
## Please library(dtplyr)!
## -----
```

```
##
## Attaching package: 'data.table'
## The following objects are masked from 'package:dplyr':
##
##   between, first, last
```

```
library("ggplot2")
```

```
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```
serv.dir <- "/Users/jtorres/FUSE/"
rds.dir <- serv.dir %&% "projects/wtsa/enhancer_driven/probe_design/rds/"
fgwas.dir <- serv.dir %&% "projects/wtsa/enhancer_driven/probe_design/fgwas_output/"

fgwas.input.file <- serv.dir %&% "projects/t2d-integration/fgwas/" %&%
  "credsets_diagram_1KG/fgwas_input/fgwas_input27.txt.gz"
```

Will take as input credible sets from the DIAGRAM (imputed to 1K Genomes) fine-mapping study. Will compare probe designs from these input files: \* original file from Anubha \* fGWAS fine-mapping file (generated from my T2D-Integration pipeline)

```
cred.df <- fread("cat " %&% fgwas.input.file %&% " | zmore") %>% dplyr::select(one_of("SNPID","CHR","POS"))

fgwas.outfile <- fgwas.dir %&% "diagram_gwasmod.bfs.gz"

fgwas.df <- fread("cat " %&% fgwas.outfile %&% " | zmore") %>% dplyr::select(one_of("id","PPA","chunk"))
names(fgwas.df) <- c("SNPID","PPA.fgwas","SEGNUMBER")
```

```
fgwas.df$SEGNUMBER <- fgwas.df$SEGNUMBER + 1 # fgwas file is 0-based

cred.df <- inner_join(cred.df, fgwas.df, by="SNPID")
cred.df <- cred.df[(cred.df$SEGNUMBER.x == cred.df$SEGNUMBER.y),]
cred.df <- mutate(cred.df, change=PPA.fgwas-PPA) %>% arrange(desc(change))
cred.df <- dplyr::select(cred.df, -SEGNUMBER.y)
names(cred.df)[which(names(cred.df)=="SEGNUMBER.x")] = "SEGNUMBER"
```

Append Locus information

```
gwas.dir <- serv.dir%>% "reference/gwas/diagram_1Kgenomes/"
gwas.dir <- gwas.dir %>% "DIAGRAM_T2D_GWAS_1000G_99_Credsets_Unpublished_2017/"
locindex.file <- gwas.dir %>% "list.of.96.loci.used.for.finemapping.txt"
li.df <- fread(locindex.file, header=FALSE); names(li.df) <- c("SNPID", "LOCUS")
lookup.dir <- serv.dir%>% "projects/t2d-integration/fGWAS_files/diagram_1Kgenomes/"
lookup.df <- readRDS(lookup.dir %>% "diagram_1KGenomes_fGWAS-core.v2.df.RDS")

pb <- txtProgressBar(min = 0, max = dim(cred.df)[1], initial = 0, style = 3)
```

```
##
|
|                                     | 0%
```

```
LOCUS <- as.character(sapply(1:length(cred.df$SNPID), function(i){
  setTxtProgressBar(pb, i)
  s <- cred.df$SNPID[i]
  lead <- filter(lookup.df, SNPID==s)$LEADSNP
  s <- gsub("chr", "", lead)
  if (length(s) > 1){
    vec <- unique(as.character(sapply(1:length(s), function(e){
      loc <- filter(li.df, SNPID==s[e])$LOCUS
      loc <- gsub(" #", "", loc)
      return(loc)
    })))
    loc <- paste(vec, collapse=" & ")
  } else{
    loc <- filter(li.df, SNPID==s)$LOCUS
  }
  if (length(loc) == 0) loc <- NA
  loc <- gsub(" #", "", loc)
  if (length(loc) > 1) loc <- loc[1]
  return(loc)
}))
```

```
##
|
|                                     | 1%
|
|                                     | 1%
|=
|                                     | 2%
|
|=
|                                     | 2%
|
|=
|                                     | 3%
```

	==	4%
	===	4%
	===	5%
	====	5%
	====	6%
	====	7%
	=====	7%
	=====	8%
	=====	8%
	=====	9%
	=====	10%
	=====	10%
	=====	11%
	=====	12%
	=====	12%
	=====	13%
	=====	13%
	=====	14%
	=====	15%
	=====	15%
	=====	16%
	=====	16%
	=====	17%
	=====	18%
	=====	18%
	=====	19%
	=====	19%

=====		20%
=====		21%
=====		21%
=====		22%
=====		22%
=====		23%
=====		24%
=====		24%
=====		25%
=====		25%
=====		26%
=====		27%
=====		27%
=====		28%
=====		28%
=====		29%
=====		30%
=====		30%
=====		31%
=====		32%
=====		32%
=====		33%
=====		33%
=====		34%
=====		35%
=====		35%
=====		36%

=====		36%
=====		37%
=====		38%
=====		38%
=====		39%
=====		39%
=====		40%
=====		41%
=====		41%
=====		42%
=====		42%
=====		43%
=====		44%
=====		44%
=====		45%
=====		45%
=====		46%
=====		47%
=====		47%
=====		48%
=====		48%
=====		49%
=====		50%
=====		50%
=====		51%
=====		52%
=====		52%

	53%
=====	53%
=====	54%
=====	55%
=====	55%
=====	56%
=====	56%
=====	57%
=====	58%
=====	58%
=====	59%
=====	59%
=====	60%
=====	61%
=====	61%
=====	62%
=====	62%
=====	63%
=====	64%
=====	64%
=====	65%
=====	65%
=====	66%
=====	67%
=====	67%
=====	68%
=====	68%

=====			69%
=====			70%
=====			70%
=====			71%
=====			72%
=====			72%
=====			73%
=====			73%
=====			74%
=====			75%
=====			75%
=====			76%
=====			76%
=====			77%
=====			78%
=====			78%
=====			79%
=====			79%
=====			80%
=====			81%
=====			81%
=====			82%
=====			82%
=====			83%
=====			84%
=====			84%
=====			85%

=====		85%
=====		86%
=====		87%
=====		87%
=====		88%
=====		88%
=====		89%
=====		90%
=====		90%
=====		91%
=====		92%
=====		92%
=====		93%
=====		93%
=====		94%
=====		95%
=====		95%
=====		96%
=====		96%
=====		97%
=====		98%
=====		98%
=====		99%
=====		99%
=====		100%

```
cred.df <- cbind(cred.df,LOCUS)
cred.df$LOCUS <- as.character(cred.df$LOCUS)
saveRDS(cred.df,file=rds.dir%&%"fgwas.cred.diag.df.RDS")
```