Obesity susceptibility genes in a Spanish population using sequencing data

Isaac De la Hoz

9 de abril de 2019

Introduction

Obesity

Obesity is defined as an increase in fat mass that is sufficient to adversely affect health.According World Health Organization, people with a body mass index (BMI; weight in kg/height in m^2) higher than 30 $\frac{kg}{m^2}$ are considered obese. Nowadays, obesity is considered as a worldwide epidemic associated with increased morbidity and mortality that imposes an enormous burden on individual and public health (Xu and Tong 2011). In the Europe population, for instance, 10%-20% of people are classified as obese (Klaauw and Farooqi 2015).

Obesity and genetics

Around 40-70% of inter-individual variability in BMI, commonly used to assess obesity, has been attributed to genetic factors(Xu and Tong 2011). The evidence for genetic contributions to body weight comes from family, twin, and adoption studies. Other studies cumulatively demonstrate that the heritability (fraction of the total phenotypic variance of a quantitative trait attributable to genes in a specified environment) of BMI is between 0.71 and 0.86 (Klaauw and Farooqi 2015).

Objectives

In order to expand the catalogue od BMI suceptibility SNPs we perform an study on whole exome sequence data from 16 different obese individuals.

Methodology

From already aligned data, a pipeline which include variant calling, variant annotation and statistical analysis was performed.

Variant Calling

In order to find the best way to obtain variant from the alignment files, two diiferent variant callers were proved and compared. One of them was selected for being used in this analysis.

R package: VariantTools

VariantTools is R package which allows to perfom the variant calling using R. The following code is the used to perform the alignment:

First of all, the libraries needed are loaded

```
library(GenomicAlignments)
library(VariantAnnotation)
library(Rsamtools)
library(VariantTools)
library(GenomicRanges)
```

```
library(BiocParallel)
library(BSgenome.Hsapiens.UCSC.hg38)
library(gmapR)
library(org.Hs.eg.db)
library(TxDb.Hsapiens.UCSC.hg38.knownGene)
library(AnnotationHub)
```

Once libraries are loaded, the Gmap genome of human hg38 version (the human genome version used to perform the alignment) is created:

```
#Gmap genome object creation. The human genome is indexed
## IMPORTANT: the following lines have to be executed 1 time,
## the needed to create the dependency.
setwd("/scratch")
chrs <- standardChromosomes(Hsapiens)
hs <- getSeq(Hsapiens, chrs)
##genome seqlevels style correction so that it is equal to bam files' seqlevels
seqlevels(hs)[25]<-"chrMT"
names(hs)[25]<-"chrMT"
##Gmap genome creation
gmapGenomePath <- file.path(getwd(), "HSG")
gmapGenomeDirectory <- GmapGenomeDirectory(gmapGenomePath, create = TRUE)
gmapGenome <- GmapGenome(genome=hs, directory=gmapGenomeDirectory,
                         name="hg38", create=TRUE, k = 14L)
```

This last code have to be run only one time because once run, a gmap object is created and stored in the directory selected. The following code performs the variant calling and genotyping and it has to be exucuted once per Bam file you have, changing each time the object `FILE` by the correpondent bam file name.

```
#data loading
##filename
FILE <- "<file name>"

bamFile <- sprintf("~/data/WES_obesity/Data/%s.bam", FILE)

#Tallies creation (VRanges object with variant information)

chrs <- standardChromosomes(Hsapiens)
hs <- getSeq(Hsapiens, chrs)
seqlevels(hs)[25]<-"chrMT"
names(hs)[25]<-"chrMT"
gmapGenomePath <- file.path("/scratch/HSG")
gmapGenomeDirectory <- GmapGenomeDirectory(gmapGenomePath)
HGmapGenome <- GmapGenome(genome=hs, directory=gmapGenomeDirectory, name="hg38")
tiles <- tileGenome(seqinfo(HGmapGenome), ntile = 100)
param <- TallyVariantsParam(HGmapGenome, which = unlist(tiles), indels = TRUE)
bpparam <- MulticoreParam(workers = 5)
tallies <- tallyVariants(bamFile, param, BPPARAM = bpparam)
mcols(tallies) <- NULL
sampleNames(tallies) <- FILE

#Calling and filtering
##Call genotypes
cov <- coverage(bamFile)
params <- CallGenotypesParam(HGmapGenome, p.error = 1/1000, which = tiles)
```

```
genotypes <- callGenotypes(tallies, cov, params, BPPARAM = bpparam)

##The default variant calling filters:
##VariantCallingFilters(read.count = 2L, p.lower = 0.2, p.error = 1/1000)
calling.filters <- VariantCallingFilters()
post.filters <- VariantPostFilters()
variants <- callVariants(genotypes, calling.filters, post.filters)

### Saving as a R data
vcf <- asVCF(sort(variants))
save(vcf, file=sprintf("~/data/WES_obesity/genotypedVariants/%s.rda", FILE))
```

The resulting files are saved as R data to be easily readed by R once we perform the later analisis.

GATK haplotype caller

Variant annotation

Statistic analysis

Klaauw, Agatha A Van Der, and I Sadaf Farooqi. 2015. "Review The Hunger Genes : Pathways to Obesity." *Cell* 161 (1). Elsevier Inc.: 119–32. doi:10.1016/j.cell.2015.03.008.

Xu, Yuanzhong, and Qingchun Tong. 2011. "Expanding neurotransmitters in the hypothalamic neurocircuitry for energy balance regulation." *Protein & Cell* 2 (10): 800–813. doi:10.1007/s13238-011-1112-4.