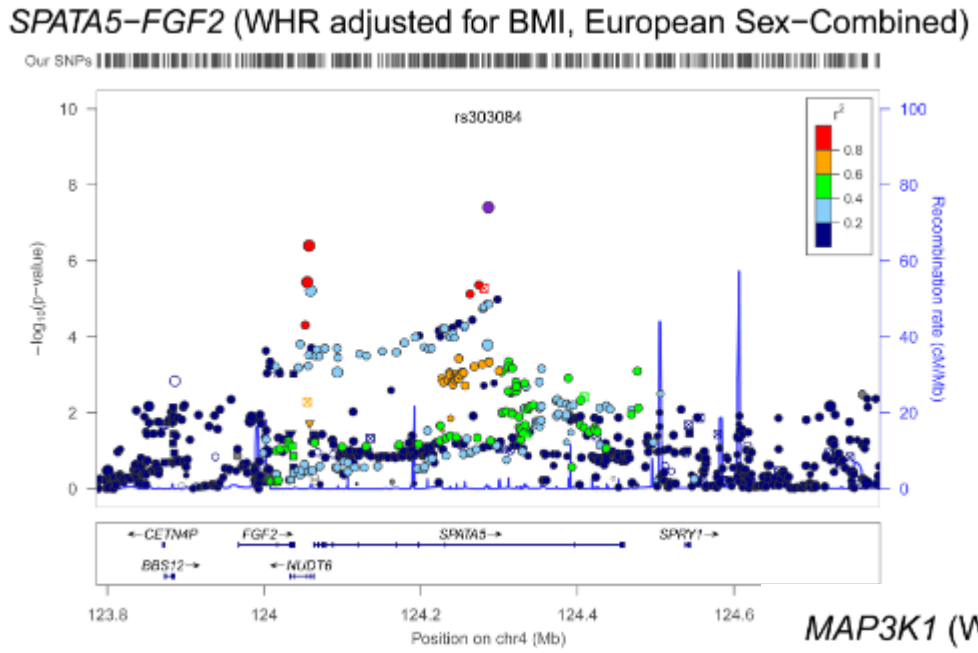
# Introduction to Genome-wide Association Studies (GWAS)

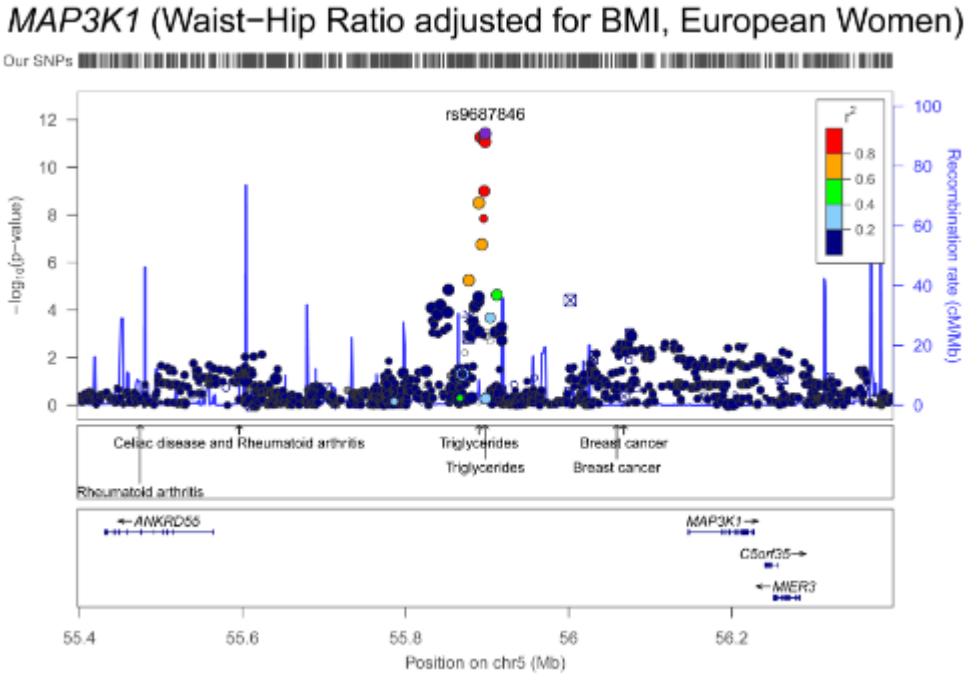
## Practical

## Part 2: Credible Sets

Teresa Ferreira and Nilufer Rahmioglu

Wellcome Centre for Human Genetics/Big Data Institute




Shungin et al (2015). *Nature* **518**: 187-196

# Refine location of causal variants

- Lead SNP commonly taken to represent an association signal observed in a region
  - often loci extend over hundreds of kilobases
  - at times span multiple genes with plausible functional impact on the trait of interest

- Goal: refine the localization of likely causal variants at significant loci

# Credible sets

**Goal**: define set of variants most likely to contain the causal variant (or tag an unobserved causal variant)

- Bayesian approach (Maller, J. B. et al. *Nature Genet*. 44, 1294–1301 (2012) )
- Bayes' factor in favour of association of a variant with the trait: estimated from the variant's effect estimate and SE (Wakefield, J. A. *Am. J. Hum. Genet*. 81, 208–227 (2007))
- Assumption: **single causal variant in the region**
- Uniform prior on any of the variants being causal

- 99% credible set: set of variants that accounts for 99% of posterior probability of association in the region

# Credible sets

- Input
  - Regions defined by chromosome and position interval
  - Association analysis summary statistics (betas and SEs)

**REMINDER**: region should have a single causal variant

# WHRadjBMI data

- https://portals.broadinstitute.org/collaboration/giant/images/6/6e/Whradjbmi.giant-ukbb.meta-analysis.combined.23May2018.HapMap2_only.txt.gz
  - Pulit et al (2018). *Hum Mol Genet*. 2019 Jan 1;28(1):166-174.
  - Sex-combined association analysis
    - `regions.for.credible.sets.txt`
    - 206 regions for which is reasonable to assume there is a single causal variant given the data
    - Index SNPs reported but which is the set of variants likely to contain (or tag) the causal variant?

- **Objective**: determine the credible set of SNPs for each of those regions

  R script `crediblesets.R`

```
Rscript --vanilla crediblesets.R alpha=0.99
data.file="results.file.name"
regions.file="region.file.name" path="/well/abc/"
```

# crediblesets.R

- Outputs
  - Whradjbmi.giant-ukbb.meta-analysis.combined.23May2018.HapMap2_only.txt.credibleset.summary

```
-bash-4.2$ head Whradjbmi.giant-ukbb.meta-analysis.combined.23May2018.HapMap2_only.txt.credibleset.summary
SNP     Locus   chr     startPos        endPos  n_snps  distance        start   end     total_posterior
rs717795:T:C    1       1       2470464 3470464 14      35311   2938265 2973576 0.99264155864637
rs9435732:C:T   2       1       16808158        17808158        2       29233   17308158        17337391        0.995027296923281
rs7521902:C:A   3       1       21990724        22990724        6       36964   22436446        22473410        0.996982515624467
rs2997447:G:A   4       1       25887423        26887423        22      397524  26120367        26517891        0.990059772376607
rs11205773:T:C  5       1       50784905        51784905        36      600520  50906366        51506886        0.991006636753516
rs598104:A:T    6       1       85766109        86766109        7       14273   86251836        86266109        0.991281948648392
rs2061708:G:C   7       1       102917203       103917203       20      226007  103345744       103571751       0.99045138685687
rs11204762:A:G  8       1       150499737       151499737       3       8516    150991221       150999737       0.996765519955423
rs7554947:C:T   9       1       159913460       160913460       14      21163   160407025       160428188       0.993299352535928
-bash-4.2$
```

  - Whradjbmi.giant-ukbb.meta-analysis.combined.23May2018.HapMap2_only.txt.credibleset.results

```
-bash-4.2$ head Whradjbmi.giant-ukbb.meta-analysis.combined.23May2018.HapMap2_only.txt.credibleset.results
CHR     POS     SNP     Tested_Allele   Other_Allele    Freq_Tested_Allele      BETA    SE      P       N       INFO    z       ABF     posterior       Locus
1       2943183 rs7412983:A:C   A       C       0.8024  -0.0196 0.0023  2.088e-17       627163  0.995425        -8.52173913043478       67270517063784.3        0.131391046586611       1_rs717795:T:C
1       2970464 rs717795:T:C    T       C       0.8014  -0.0196 0.0023  6.858e-18       627060  0.991382        -8.52173913043478       67270517063784.3        0.131391046586611       1_rs717795:T:C
1       2949028 rs1890338:C:T   T       C       0.1905  0.0195  0.0023  4.169e-17       627216  0.996493        8.47826086956522        46488534372212.5        0.0908002116239275      1_rs717795:T:C
1       2957600 rs12409277:T:C  T       C       0.8097  -0.0195 0.0023  3.547e-17       627220  1       -8.47826086956522       46488534372212.5        0.0908002116239275      1_rs717795:T:C
1       2961402 rs10909867:G:T  T       G       0.1902  0.0195  0.0023  3.197e-17       627244  0.997545        8.47826086956522        46488534372212.5        0.0908002116239275      1_rs717795:T:C
1       2973433 rs2993481:A:T   A       T       0.6703  -0.0195 0.0023  1.62e-17        627043  0.994582        -8.47826086956522       46488534372212.5        0.0908002116239275      1_rs717795:T:C
1       2973576 rs1109251:G:A   A       G       0.1779  0.0203  0.0024  1.695e-17       627088  0.994179        8.45833333333333        40960544152039.6        0.0800030830711563      1_rs717795:T:C
1       2950334 rs6661908:G:A   A       G       0.1904  0.0194  0.0023  5.507e-17       627126  0.996483        8.43478260869565        32187541569277.6        0.0628678796957581      1_rs717795:T:C
1       2951244 rs10489589:T:C  T       C       0.8096  -0.0194 0.0023  5.997e-17       627216  0.997049        -8.43478260869565       32187541569277.6        0.0628678796957581      1_rs717795:T:C
-bash-4.2$
```

- 18 regions where 1 marker was selected
- 40 regions that cover <10kb