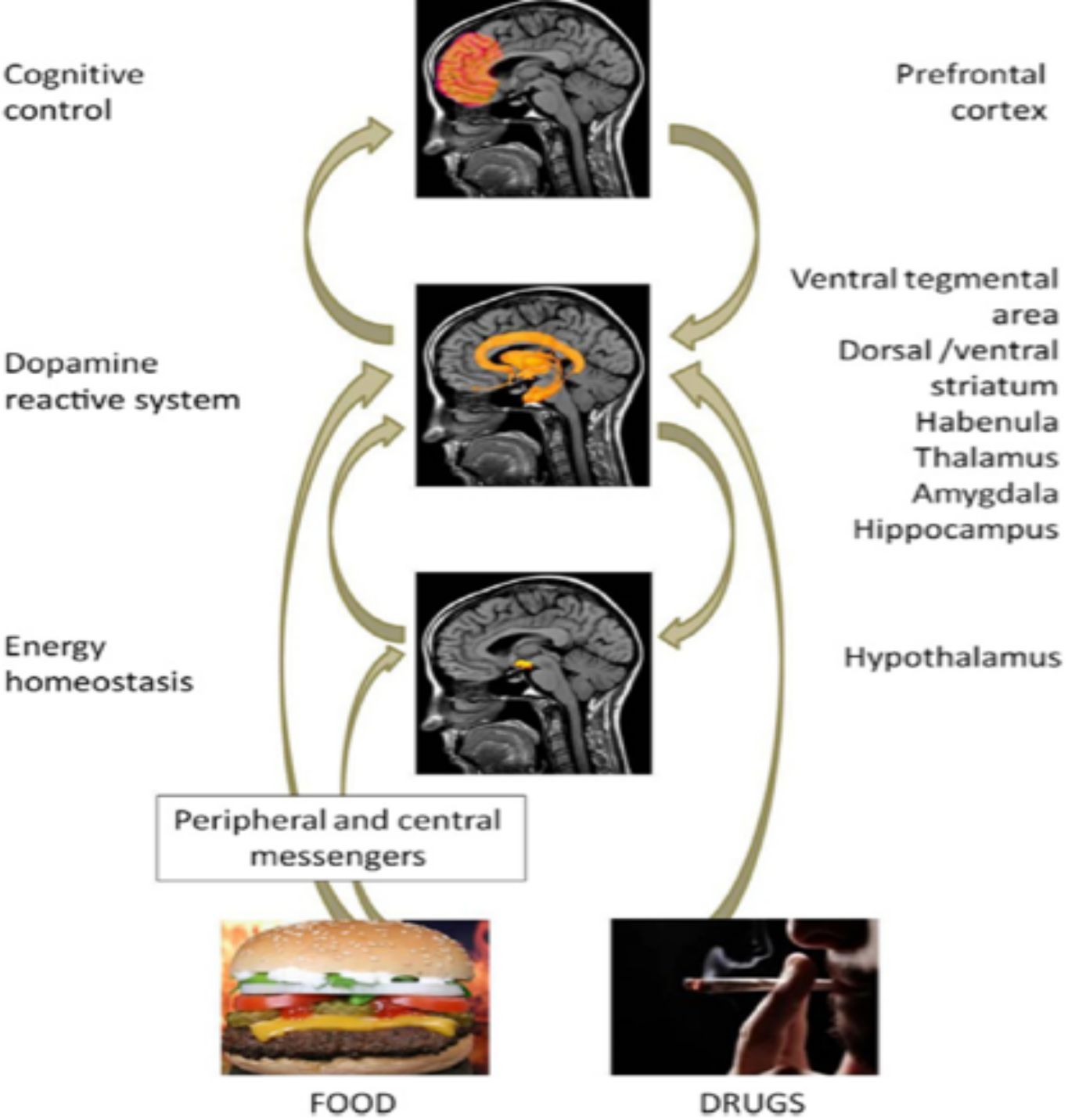
# Statistical Analysis of Pleiotropy between Obesity and Substance Dependence

Dan Zhao
Jiawei Zhang


Cognitive control
Prefrontal cortex
Dopamine reactive system
Ventral tegmental area
Dorsal /ventral striatum
Habenula
Thalamus
Amygdala
Hippocampus
Energy homeostasis
Hypothalamus
Peripheral and central messengers
FOOD
DRUGS

# Data

- **SSADDA**: 2379 European Americans
- **SAGE**: 2668 European Americans
- **Phenotype**: BMI, Substance dependence symptom score;
- **Genotype**: 988,306 SNPs (SSADDA)

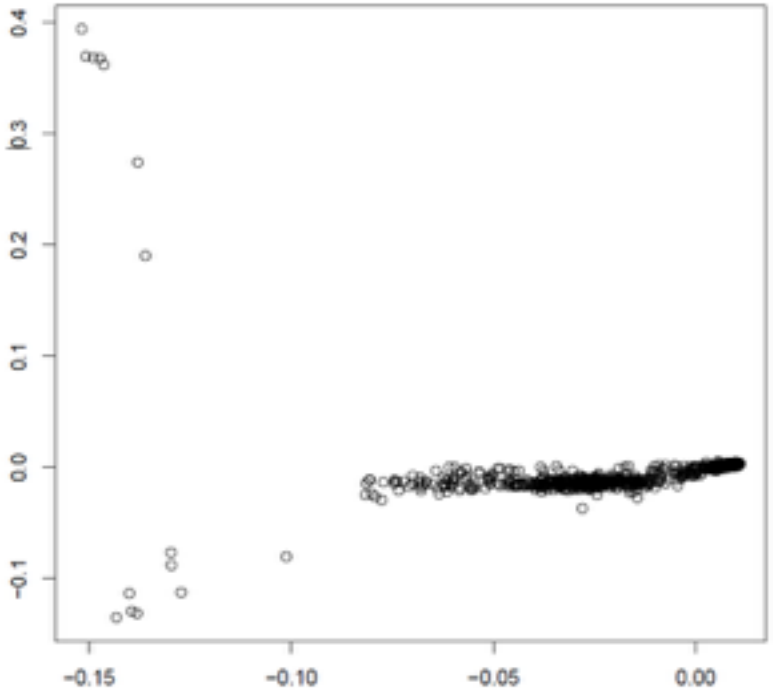
# Quality Control

a. Misidentified individuals
b. Genotype failure rate  0.02
c. Extreme heterozygosity (+/-3 sd)
d. Duplicated or related individuals

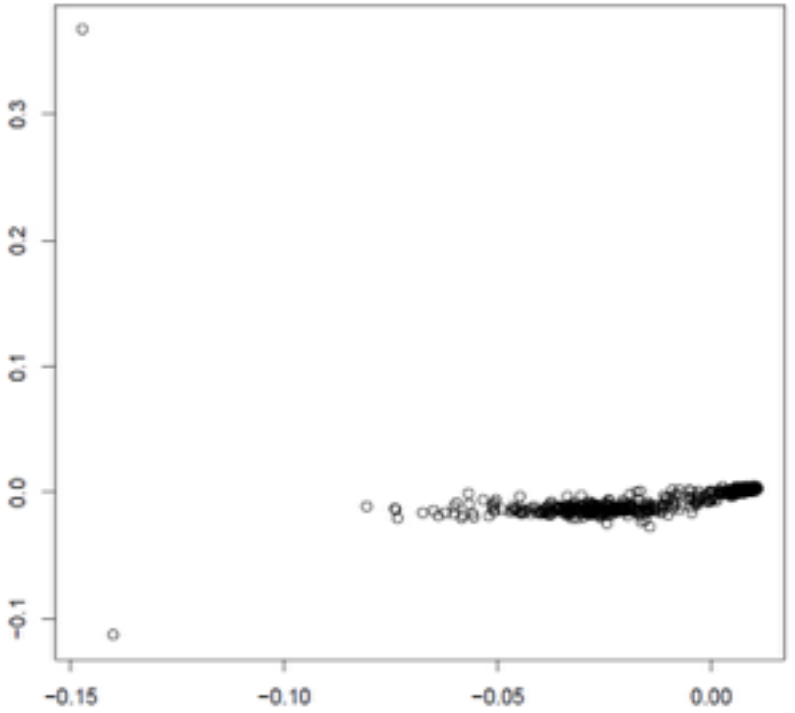
- Sample QC
- (2379=>1828)

a. MAF 0.01
b. HWE 1e-06
c. Genotype missing rate 0.02
d. Unbalanced genotype rates between case/control p=1e-05

- SNP QC
- (988,306=>805,782)




Before QC




After QC

From PCA plots, most suspected outliers have been removed in the quality control (QC) process.

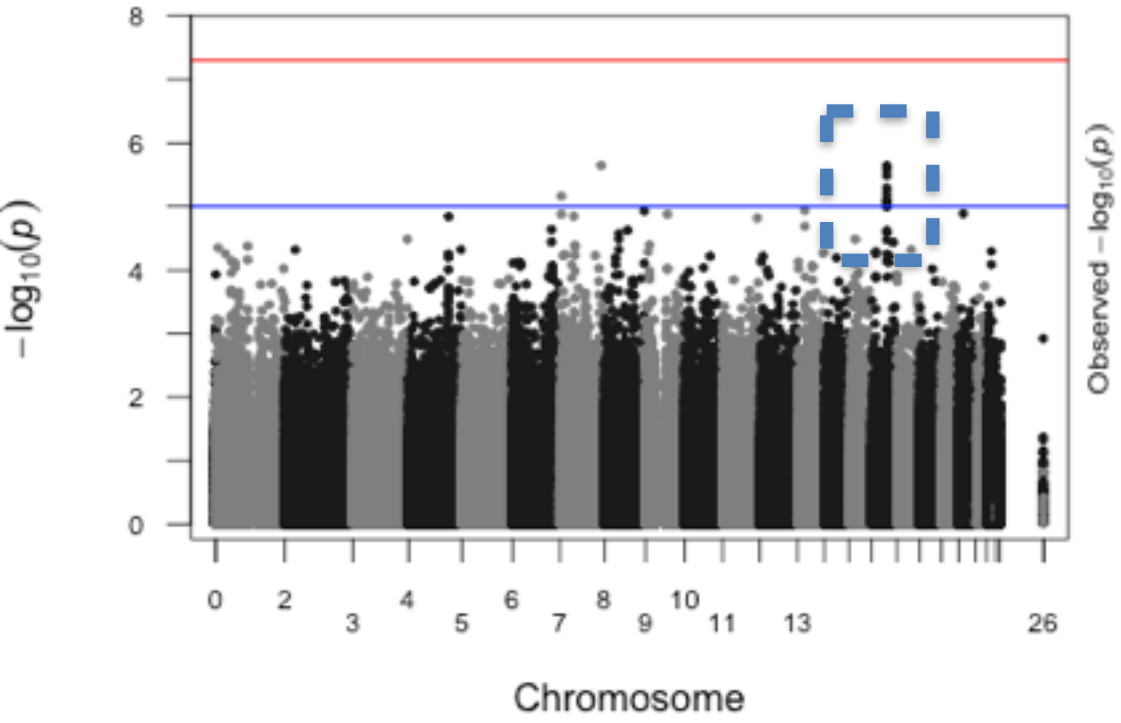
# Single Marker Association Analysis

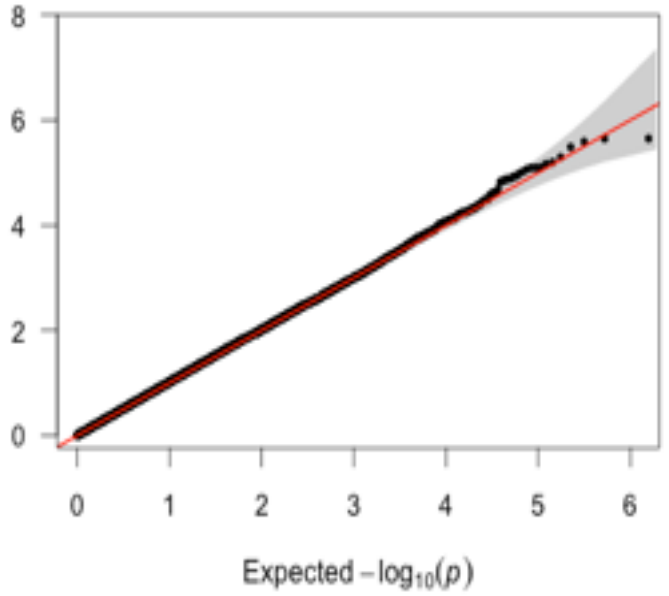
- **Genotype Model**: additive model
  - Assume there is a linear increase of risk with each additional risk allele.
- **Test Approach**: linear regression
- **Covariates**: adjusted in linear regression
  - Age and sex
  - first 4 scaling factors from MDS analysis (for population stratification)

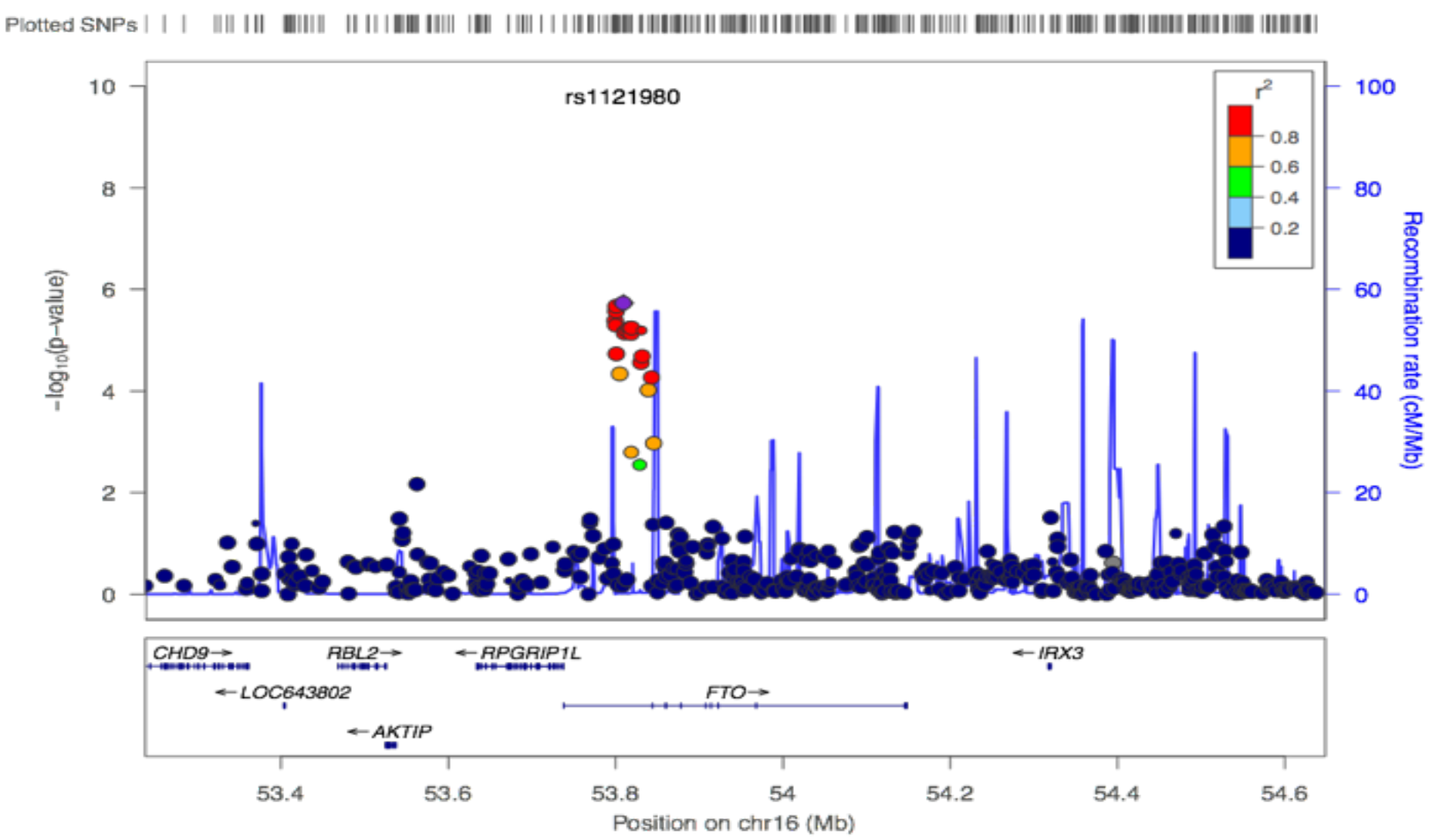
# Outcome: BMI

rs1121980

Inflation factor λ=1.02



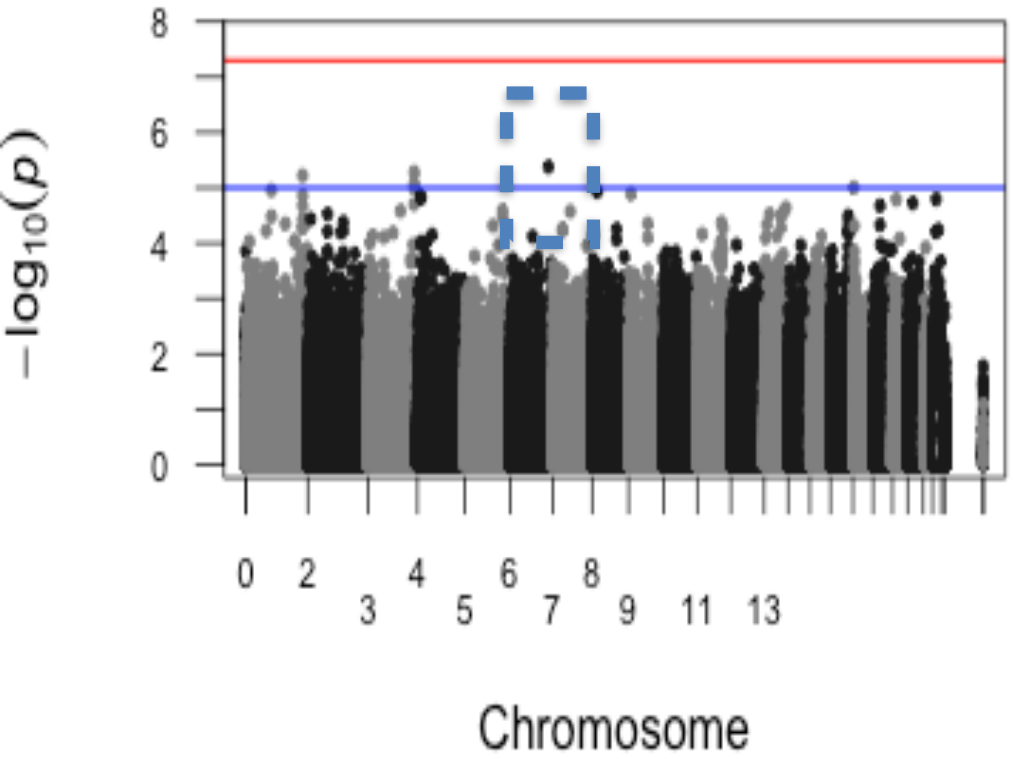


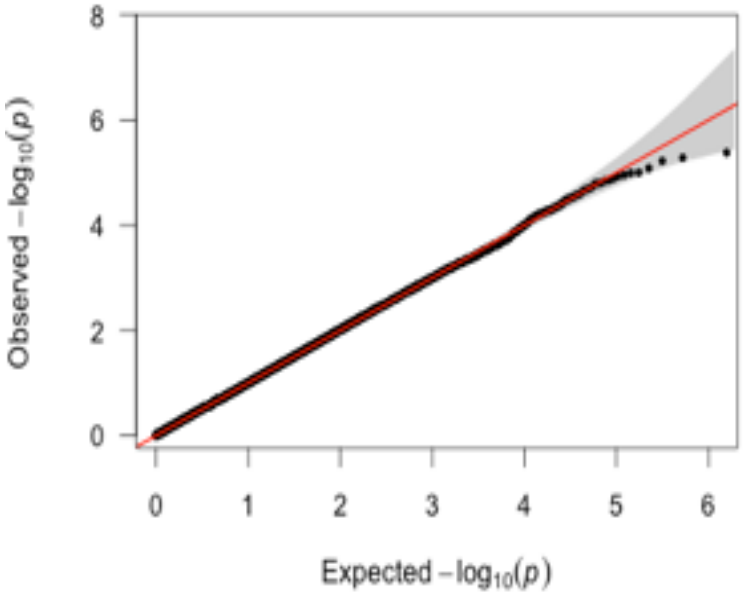

| SNP | CHR | Nearest Gene | Beta | P-value |
|---|---|---|---|---|
| rs1121980 | 16 | FTO | 0.9207 | 2.26E-06 |

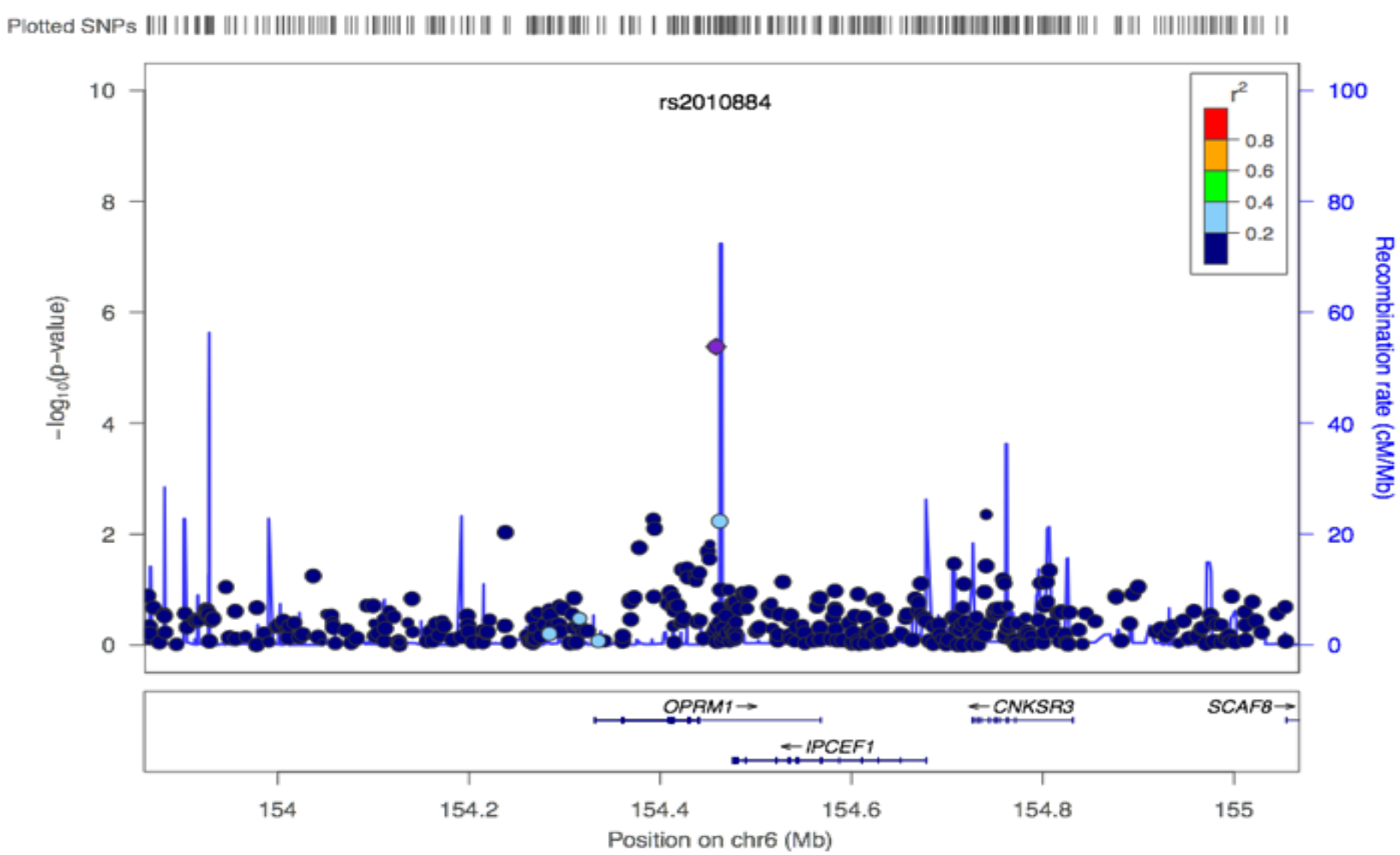
# Outcome: Sub_Dep

Inflation factor λ=1.007

rs2010884






| SNP | CHR | Nearest Gene | Beta | P-value |
|---|---|---|---|---|
| rs2010884 | 6 | OPRM1 | -1.39 | 4.18E-06 |

# Mixed Effects Model Based Analysis

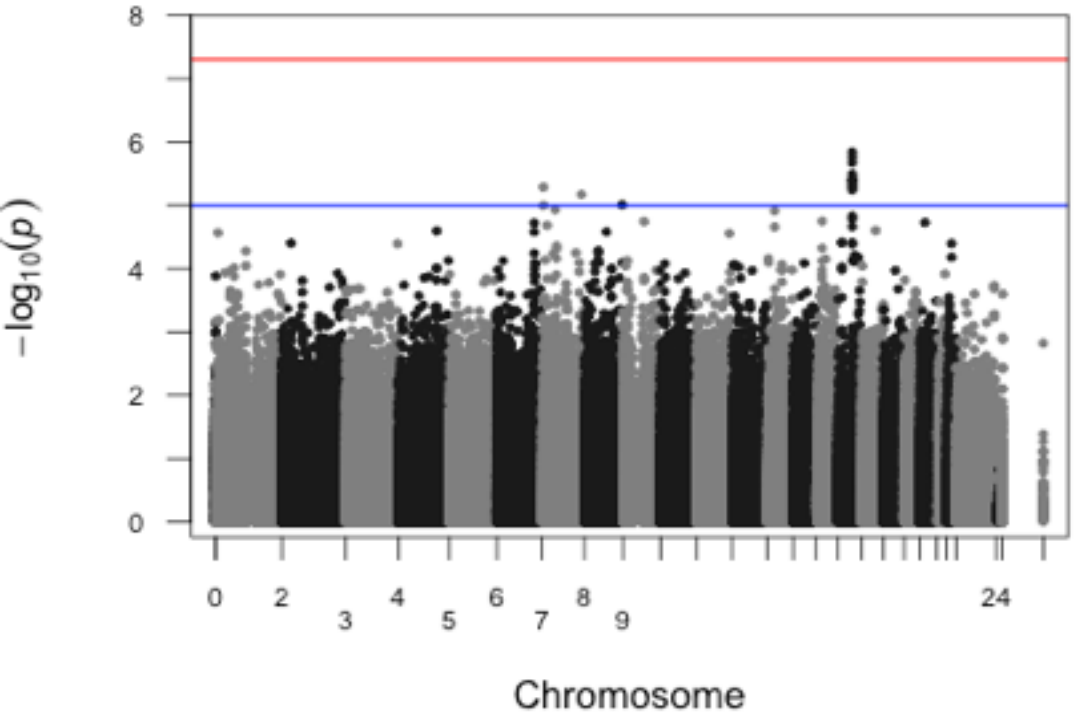
$$Y = X\beta + u + \varepsilon$$

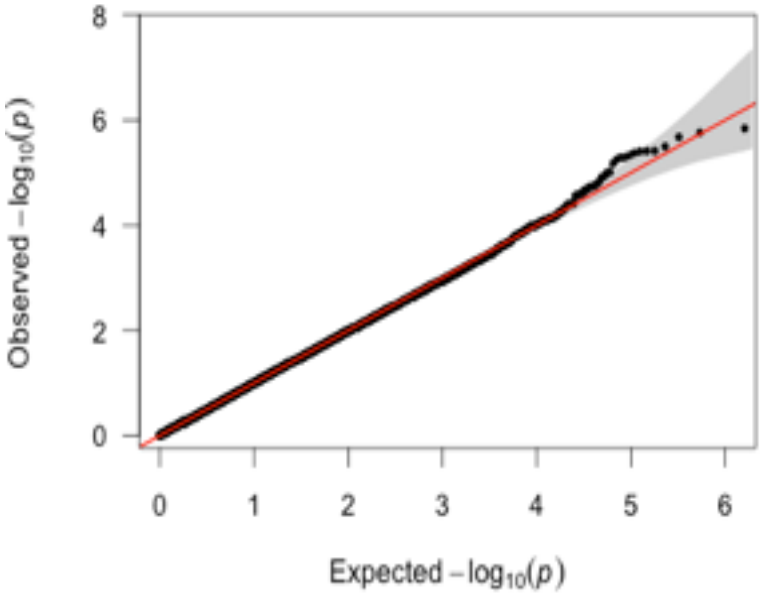
- Y=phenotypes
- X=SNP genotypes (+covariates)
- $Var(u) = \sigma_g^2 A$, where A is the genetic relationship matrix (GRM)

$$A_{jk} = \frac{1}{M} \sum_i \frac{(g_{ij} - 2p_i)(g_{ik} - 2p_i)}{2p_i(1 - 2p_i)}$$

# Outcome: BMI

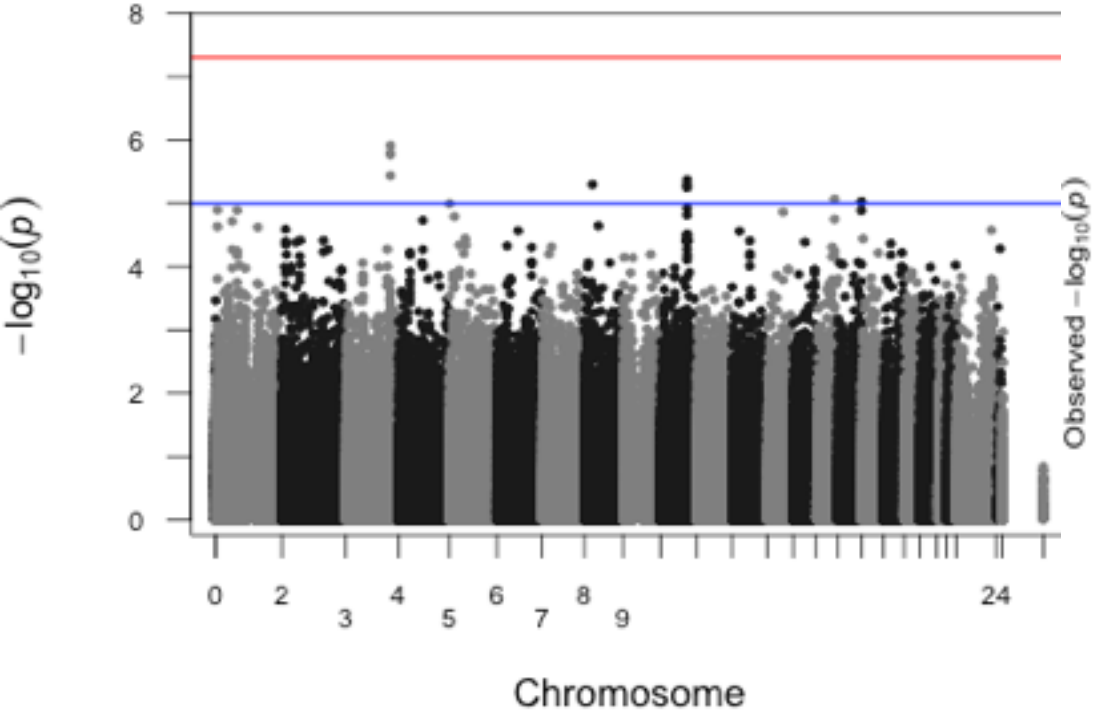
rs1121980

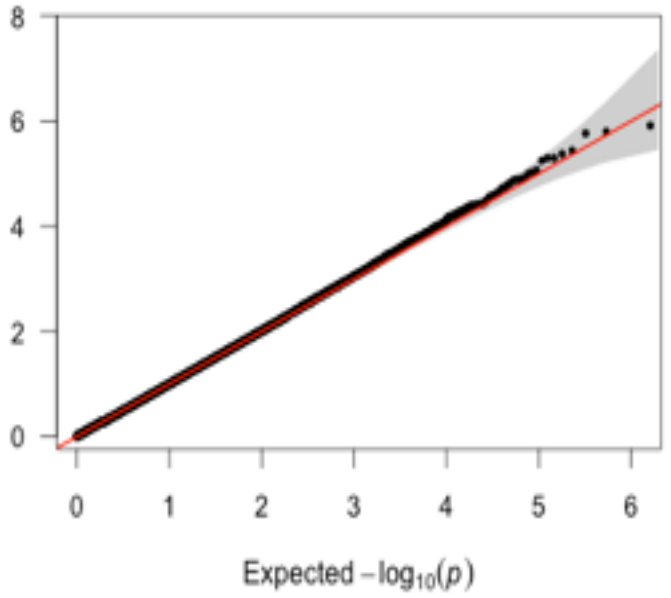


# Outcome: Sub_Dep

rs2010884



# Heritability Estimates

$$Y = X\beta + u + \varepsilon$$

- $Var(u) = \sigma_g^2 A, \sigma_g^2$ is the variance explained by all the SNPs
- Estimated by the restricted maximum likelihood (REML) approach

| Phenotype | N | Hg | SE | LRT | P-value |
|---|---|---|---|---|---|
| BMI | 1828 | 0.2595 | 0.16 | 2.917 | 0.0438 |
| Sub_Dep | 1828 | 0.2156 | 0.16 | 1.890 | 0.0846 |

# SNP Coheritabilities

- The variance-covariance matrix across the two traits is:

$$V = \begin{pmatrix} Z_1 A Z_1' + I\sigma_{g1}^2 & Z_2 A Z_1' \sigma_{g12} \\ Z_1 A Z_2' \sigma_{g12} & Z_2 A Z_2' + I\sigma_{g2}^2 \end{pmatrix}$$

- The genetic correlation coefficient is:

$$r_g SNP = \frac{\sigma_{g12}}{\sigma_{g1} + \sigma_{g2}}$$

- Estimated by the Bivariate REML approach

| | N | $r_G$ | S.E. | P-value |
|---|---|---|---|---|
| BMI:Sub_Dep | 1828 | 0.2408 | 0.41 | 0.71 |

Integrative Analysis of Two GWAS Datasets with Functional Annotations

- We have P-values from two independent GWAS datasets
- Indicator variable $Z_j=[Z_{j00}, Z_{j10}, Z_{j01}, Z_{j11}]$ for the j-th SNP: e.g, $Z_{j11}$ means the j-th SNP is associated with both BMI and Sub-Dep
- Functional annotation data: $A \in \overrightarrow{M}$ , where $A_j \in \{0,1\}$ indicates whether the j-th SNP is functionally annotation.

- Model the relationship between $Z_j$ and $A_j$ as:

$$q_{00} = \Pr(A_j = 1 \mid Z_{j00} = 1)$$

$$q_{10} = \Pr(A_j = 1 \mid Z_{j10} = 1)$$

$$q_{01} = \Pr(A_j = 1 \mid Z_{j01} = 1)$$

$$q_{11} = \Pr(A_j = 1 \mid Z_{j11} = 1)$$

- The joint distribution of Pr (P, A) can be estimated by EM algorithm

$$\Pr(P, A) = \prod_{j=1}^{M} \left[ \sum_{l \in \{00,10,01,11\}} \Pr(Z_{jl} = 1) \Pr(P_j, A_j \mid Z_{ij} = 1) \right]$$

- The summary statistics of two phenotypes:
  BMI: 805,782 p-values;
  Substance Dependence: 845,871 p-values;
- Overlapping SNPs of two phenotypes is 466,115
- Using the central neural system gene as annotation data, 63,274 (13.6%) of the SNPs were annotated

| | 00 | 10 | 01 | 11 |
|---|---|---|---|---|
| $\hat{\pi}$ | 0.911(0.086) | 0.046(0.053) | 0.04(0.084) | 0.02(0.049) |
| $\hat{q}$ | 0.126(0.013) | 0.213(0.094) | 0.268(0.086) | 0.288(1.843) |

# Conclusion

- The strongest association signal for obesity: FTO gene;
- The strongest association signal for substance dependence: OPRM1 gene
- $h_g^2$ estimated for obesity was 0.26, 0.22 for substance dependence
- No evidence suggests pleiotropy between obesity and substance dependence in this data set.