Statistical methods for mapping quantitative traits using high density SNPs in family samples

Josée Dupuis

Department of Biostatistics

Boston University School of Public Health

Outline

- Motivating dataset: Framingham Heart Study
- Quantitative trait mapping in family samples
- Common framework for linkage and association analysis
  - $\Rightarrow$ Application to the Framingham Heart Study
- Association analysis: value of family samples
- Linkage analysis: extensions to basic model
  - $\Rightarrow$ Results from the Framingham Heart Study

Outline

- **Motivating dataset: Framingham Heart Study**
- Quantitative trait mapping in family samples
- Common framework for linkage and association analysis

  $\Rightarrow$ Application to the Framingham Heart Study
- Association analysis: value of family samples
- Linkage analysis: extensions to basic model

  $\Rightarrow$ Results from the Framingham Heart Study

FHS Families

- **Original cohort enrolled 5,209 men and women from the town of Framingham**
  - ⇒ **Examination biennially since 1948**
- **The Framingham Offspring cohort (n=5,124) enrolled children of the original cohort and their spouses**
  - ⇒ **Examination roughly every four years since 1971**
- **Third cohort (Gen3; n=4,095) enrolled in 2002**
  - ⇒ **Grandchildren of original cohort members**
  - ⇒ **Data from a single exam available; 2nd exam on-going**

FHS: Phenotypes

- Multiple phenotypes collected at each exam
- Example of qualitative traits collected:
  - ⇒ Cardiovascular disease
  - ⇒ Hypertension
  - ⇒ Diabetes, etc.
- Example of quantitative traits collected:
  - ⇒ Blood pressure
  - ⇒ Lipid profile
  - ⇒ Fasting glucose, etc.

FHS: Genotypes

- > 400 polymorphic genetic markers (single tandem repeats or STRs) covering the genome
  - ⇒ Available on all three generations
- 100K Affy SNP array for a subset of 1345 individuals from original and offspring cohorts
  - 258 from original cohort + 1087 from offspring cohort
- 550K Affy SNP arrays for 9274 individuals from all three cohorts
  - SNP Health Association REsource (SHARe) project
  - Imputed SNP genotypes (~2.5 millions) using MACH software

FHS: Genetic analyses

- **FHS cohorts were not ascertained for any particular traits or diseases**

  $\Rightarrow$ **Very few families with 2 or more affected individuals for a particular qualitative trait**

- **Most genetic linkage analyses have been performed on quantitative phenotypes**
- **Correlation between relatives must be accounted for when performing population-based association analysis**

Outline

- Motivating dataset: Framingham Heart Study
- **Quantitative trait mapping in family samples**
- Common framework for linkage and association analysis

  $\Rightarrow$ Application to the Framingham Heart Study
- Association analysis: value of family samples
- Linkage analysis: extensions to basic model

  $\Rightarrow$ Results from the Framingham Heart Study

Quantitative Trait Mapping

- **Ultimate goal is to identify quantitative trait loci (QTL) defined as genetic variants influencing quantitative traits of interest**
  - **⇒ Blood pressure, body mass index, etc.**
- **Great success has been achieved in mapping disease with simple Mendelian inheritance**
  - **⇒ Cystic Fibrosis, Huntington's Disease**
- **Success in mapping complex and quantitative traits has been slower, despite the increase knowledge of the human genome**
  - **⇒ Most genetic variants identified to date explain only a very small portion of the trait variance**

Mapping Quantitative Traits: Then

- **Linkage analysis**

  ⇒ Relatives who have similar traits should have higher than expected levels of sharing of genetic material near the genes that influence those traits

  - **At a locus unrelated to the trait, amount of shared genetic material between relatives is determined (randomly) as a function of family relationship and is independent of trait**
  - **At a locus that affects the trait, expect greater sharing among people who have similar trait values and less sharing among people with dissimilar trait values**
  - **Because the sharing is driven by historical recombination events, the effects of increased sharing at the trait locus will also be seen at nearby loci**

Mapping Quantitative Traits: Now

- **Association analysis**

  ⇒ **Genetic variant is tested for association with phenotype**

  - **For example, for a SNP with two alleles, A and T, there are three possible genotype groups: AA, AT and TT**
  - **In unrelated individuals, analysis of variance may be used to determine if mean trait levels differ between genotype groups**
  - **Mixed effect model in related individuals**

  ⇒ **Requires genotypes at the true (but unknown) QTL**

  - **Association may also be detected at a "proxy", which is a variant correlated (i.e. in linkage disequilibrium (LD)) with the true quantitative trait locus**

General Framework

- Two approaches (linkage and association), same goal: Mapping quantitative traits
- Linkage and association analysis have been treated separately
  - $\Rightarrow$ Association analysis has been done mainly in unrelated samples
  - $\Rightarrow$ Linkage analysis requires collection of families
- In the next few slides, a unified framework for both linkage and association analysis is presented

Outline

- Motivating dataset: Framingham Heart Study
- Quantitative trait mapping in family samples
- **Common framework for linkage and association analysis**

  $\Rightarrow$ Application to the Framingham Heart Study
- Association analysis: value of family samples
- Linkage analysis: extensions to basic model

  $\Rightarrow$ Results from the Framingham Heart Study

Methods for Mapping Quantitative Traits: Basic Model

$$E[Y_i] = \eta_i(\beta) = \beta_z z_i + \beta_Q Q_i + \beta_G G_i + \beta_E E_i$$

- Y_i = phenotype of interest
  - $\Rightarrow$ Could be multivariate or longitudinal
  - $\Rightarrow$ i refers to an individual, which is part of a pedigree
  - $\Rightarrow$ When all pedigrees are of size 1, this corresponds to an unrelated sample
- z_i = (non-genetic) covariates
- Q_i = QTL covariates (trait loci)
- G_i = latent genetic factor (polygenic)
- E_i = Environmental & other non-genetic factors

Basic Model: Assumptions

Same model for linkage and association, but different assumptions!

$$E[Y_i] = \eta_i(\beta) = \beta_z z_i + \beta_Q Q_i + \beta_G G_i + \beta_E E_i$$

- **Common assumptions**
  - $\Rightarrow$ **Q_i (true quantitative loci) are never observable**
  - $\Rightarrow$ **Location of trait loci are unknown**
- **Association: Observe U_i a proxy for Q_i**
- **Linkage: Observe segregation of Q_i in pedigrees**
  - $\Rightarrow$ **Called "inheritance vectors"**

Linkage and Association Testing

Common assumption: phenotypes are normally distributed

- **Phenotypes are rarely normally distributed**

  ⇒ **E.g. Blood pressure has a non-symmetric distribution**

- **Departure from normality may cause spurious evidence for linkage**

- **One solution: robust score statistics**

  ⇒ **Test statistic is derived based on log likelihood conditional on observed phenotypes**

Log Likelihood Function

- Let $\theta_i = \eta_i(\beta)$
- Log likelihood of one observation:

$$\ell_i(\beta) = -\frac{(Y_i - \theta_i)^2}{2\phi} \propto \frac{\left(Y\theta_i - \theta_i^2/2\right)}{\phi}$$

- Log likelihood of all observations (conditional on covariates, both observed and unobserved)

$$\ell(\beta) = \sum_{i=1}^{N} \frac{\left(Y_i\theta_i - \theta_i^2/2\right)}{\phi}$$

assuming phenotypic measurements are conditionally independent, given the covariates.

Log Likelihood Function

- Phenotype (Y) and covariate z are observable
- Q is not observable
- In association, covariate U, a proxy for Q, is observable
- In linkage, υ_t, the inheritance vectors indicating the segregation of Q within a pedigree, are observable

  $\Rightarrow Q_i$ is a function of υ_t and g_t (the founders genotypes)
- G and E are unobserved latent variables
- Inference is based on marginal and conditional likelihoods of the observable covariates

Log Likelihood Function: Association

- **When *Y* and *U* are observed, marginal log-likelihood of β (omitting dependence of fixed covariate *z*):**

$$\ell(\beta; Y, U) = \log E[e^{\ell(\beta)} \mid Y, U]$$

- **Conditional log-likelihood of the genetic covariate, given the phenotypes (and other unobserved covariates):**

$$\ell(\beta; U \mid Y) = \log E[e^{\ell(\beta)} \mid Y, U] - \log E[e^{\ell(\beta)} \mid Y]$$

Note: Inference is based on likelihoods conditional on observed phenotype (Y) so that derived test statistics are robust to departure from normality assumption

Log Likelihood Function: Linkage

- **In linkage, U in unobserved, but one observed υ_t, the segregation of Q/U within the pedigree.**
- **Conditional likelihood**

$$\ell(\beta;\nu_t \mid Y) = \log E[e^{\ell(\beta)} \mid Y,\nu_t] - \log E[e^{\ell(\beta)} \mid Y]$$

- **Expectation over unobserved variables**
- **First expectation also over random distribution of founders genotypes (g_t)**

Efficient Score Statistic: Association

- **Obtained by taking first derivative of log likelihood with respect to β_U, evaluated at $\beta_U=0$**
- **For normal log likelihood function:**

$$\dot{\ell}(\beta_0; U \mid Y) = (Y - \beta z)' \sum^{-1} (U - E(U))$$

$$\beta_0 = (\beta_z, \beta_U = 0, \beta_G, \beta_E)$$

$$\sum = \sigma_G^2 \Phi + \sigma_D^2 \Delta + \sigma_e^2 I =$$ **(σ^2 I in unrelated samples)**

$\Phi =$ **Kinship coefficient matrix**

$\Delta =$ **Matrix of probabilities of pairs sharing 2 alleles IBD**

Φ and Δ are known and depend on pedigree structure

- **Association score statistic is a weighted sum of the genotype counts**

Efficient Score Statistic: Linkage

- First derivative of log likelihood for linkage with respect to β_U, evaluated at $\beta_U=0$ is 0

  $\Rightarrow$ Take second derivative (also evaluated at $\beta_U=0$)

- For normal log likelihood function:

$$\ddot{\ell}(\beta_0;\nu_t \mid Y) = tr\left[\Sigma^{-1}(Y-\beta z)(Y-\beta z)'\Sigma^{-1}-\Sigma^{-1}\right]\left[\left(\hat{\Phi}-\Phi\right)\sigma^2_{A:U}+\left(\hat{\Delta}-\Delta\right)\sigma^2_{D:U}\right]$$

$\hat{\Phi}=$ Observed pairwise identity-by-descent (IBD) proportion

$\hat{\Delta}=$ Indicator that pairs share 2 alleles IBD

$\Phi=$ Kinship coefficient (Expected IBD)

$\Delta=$ Probability that pair shares 2 alleles IBD

- Linkage score statistic is a weighted sum of the centered IBD probabilities

Efficient Score Statistic: Linkage

- Why need to take the second derivation?
- Likelihood of υ_t contains no information about effect of β_U on phenotype
- Likelihood of υ_t has information about effect of β_U^2 on the covariance of phenotypes
- Because the likelihood is parametrized in terms of β_U^2, first derivative vanishes and second derivative must be used
- Same statistic obtained if one reparametrized the likelihood in terms of β_U^2 and takes first derivative

Variance of Efficient Scores

- To construct a test statistic robust to departure from normality assumption, compute variance conditional on the observed phenotype *Y*
- U correlated within family: $E_0(UU') = \Phi\sigma^2_{A:U} + \Delta\sigma^2_{D:U}$
- Covariance matrix depends on "coding" used for U
- Typical coding for U:
  - $\Rightarrow$ Additive (U=0,1 or 2 depending on number of minor alleles)
    - For additive coding, $\sigma_{D:U} = 0$
  - $\Rightarrow$ Dominant (U = 0 if no minor alleles, 1 otherwise)
  - $\Rightarrow$ Recessive (U = 1 if two minor alleles, 0 otherwise)

Variance of Efficient Score: Association

- **Association statistic robust to departure from normality assumption: compute the variance conditional on the observed phenotype *Y***

$$V\left[\left(Y-\beta z\right)'\Sigma^{-1}\left(U-E(U)\right)\mid Y\right]=$$

$$\left(Y-\beta z\right)'\Sigma^{-1}\Phi\Sigma^{-1}\left(Y-\beta z\right)\sigma^2_{A:U}+\left(Y-\beta z\right)'\Sigma^{-1}\Delta\Sigma^{-1}\left(Y-\beta z\right)\sigma^2_{D:U}$$

- **Conditional on observed phenotype *Y***
- **Depends on relationship between individuals**
- **For unrelated individuals, Φ and Δ are identity matrices**

Variance of Efficient Score: Linkage

- **Variance of efficient score for linkage:**

$$V\left[\ddot{\ell}(\beta_0;\nu_t \mid Y)\right]=$$

$$\left(\hat{w}'\Sigma_{\hat{\Phi}\hat{\Phi}}\,\hat{w}\right)\sigma^4_{A:U}+2\left(\hat{w}'\Sigma_{\hat{\Phi}\hat{\Delta}}\,\hat{w}\right)\sigma^2_{A:U}\sigma^2_{D:U}+\left(\hat{w}'\Sigma_{\hat{\Delta}\hat{\Delta}}\,\hat{w}\right)\sigma^4_{D:U}$$

$\hat{w}$ = **vec of the matrix W**

$$W=\Sigma^{-1}(Y-\beta z)(Y-\beta z)'\Sigma^{-1}+\sigma_e^{-2}I-\Sigma^{-1}$$

$$\Sigma_{\hat{\Phi}\hat{\Phi}}=Cov_0(vec(\hat{\Phi})) \qquad \Sigma_{\hat{\Phi}\hat{\Delta}}=Cov_0(vec(\hat{\Phi}),vec(\hat{\Delta}))$$

$$\Sigma_{\hat{\Delta}\hat{\Delta}}=Cov_0(vec(\hat{\Delta}))$$

Variance of Efficient Score: Linkage

- When data from nuclear families, $\Sigma_{\hat{\Phi}\hat{\Phi}}$ and $\Sigma_{\hat{\Delta}\hat{\Delta}}$ reduce to identity matrix

  $\Rightarrow$ IBD counts between siblings are independent

- In extended pedigree, pairwise IBD counts may be correlated

  $\Rightarrow$ IBD count between individual and grand-maternal grandmother is negatively correlated with IBD count between individuals and grand-maternal grandfather

- In practice, one may substitute the empirical covariance $(\hat{\nu}_{ij} - \phi_{ij})(\hat{\nu}_{kl} - \phi_{kl})$ for the IBD count covariance between pairs *ij* and *kl*

Robust Score Statistics: Association

- **To test for association between proxy *U* and phenotype *Y*, use Z statistic:**

$$Z_a = \frac{\dot{\ell}(\beta_0; U \mid Y)}{\sqrt{V\left[\dot{\ell}(\beta_0; U \mid Y) \mid Y\right]}}$$

$$= \frac{(Y - \beta z)' \Sigma^{-1} (U - E(U))}{\sqrt{(Y - \beta z)' \Sigma^{-1} \Phi \Sigma^{-1} (Y - \beta z) \sigma^2_{A:U} + (Y - \beta z)' \Sigma^{-1} \Delta \Sigma^{-1} (Y - \beta z) \sigma^2_{D:U}}}$$

- **Similar statistic (Z_l) can be formed to test for linkage**

Robust Score Statistics

- **Conditional on observed phenotype Y, linkage and association score statistics are uncorrelated**

  ⇒ Square of association statistic has chi-square distribution with 1 (additive) or 2 (general model) degrees of freedom

  ⇒ Square of linkage statistic is a 50-50 mixture of chi-square distribution with 0 and 1 degrees of freedom

  - **One sided statistic because negative Z gives evidence against linkage**

Robust Score Statistics

- Combining these two statistics may improve power to detect genetic variants influencing trait of interest
- Optimal combination will take variance of both statistics into consideration
- Naïve way: sum of square of the score statistics

  $\Rightarrow$ Resulting statistic has a mixture of chi-square distribution

Outline

- Motivating dataset: Framingham Heart Study
- Quantitative trait mapping in family samples
- **Common framework for linkage and association analysis**

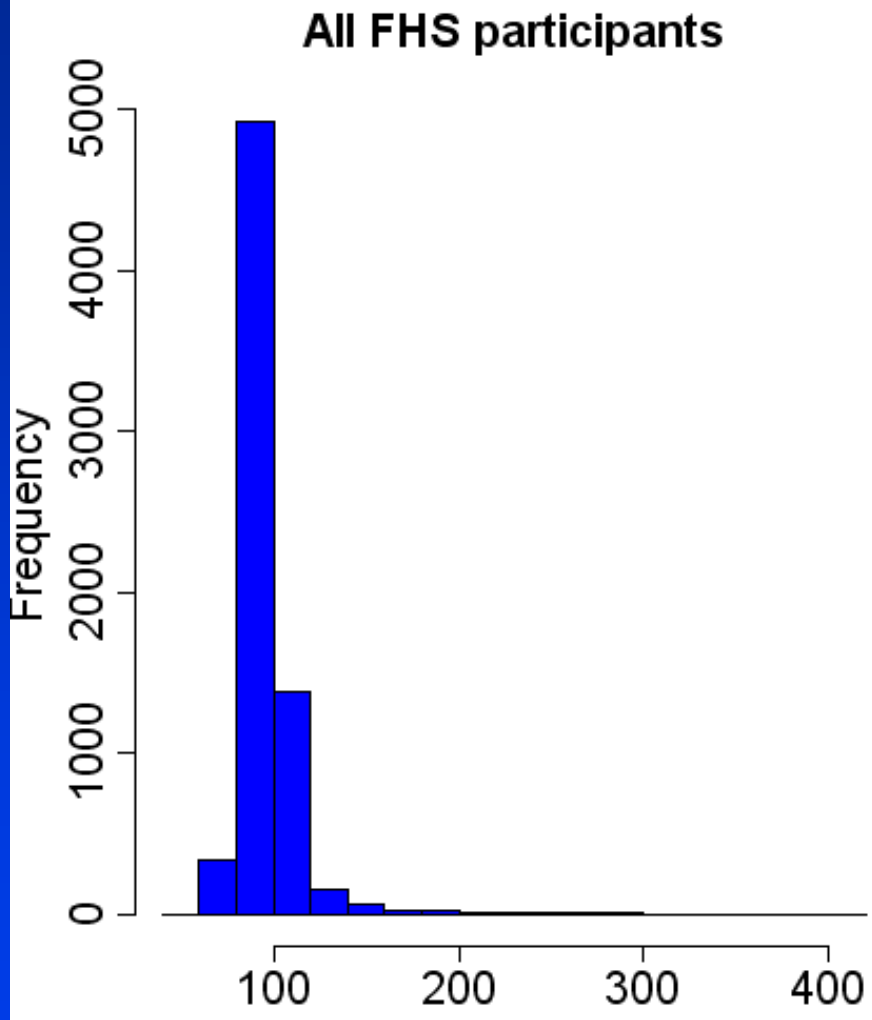
  $\Rightarrow$ **Application to the Framingham Heart Study**
- Association analysis: value of family samples
- Linkage analysis: extensions to basic model

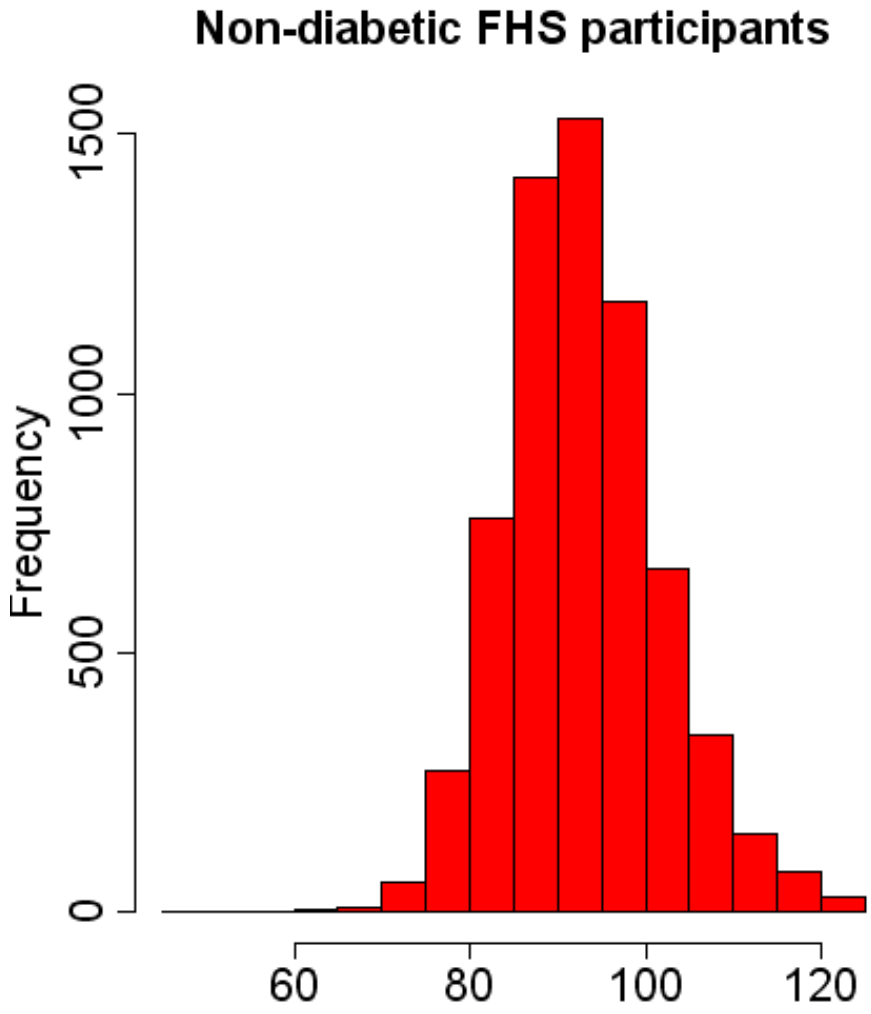
  $\Rightarrow$ Results from the Framingham Heart Study

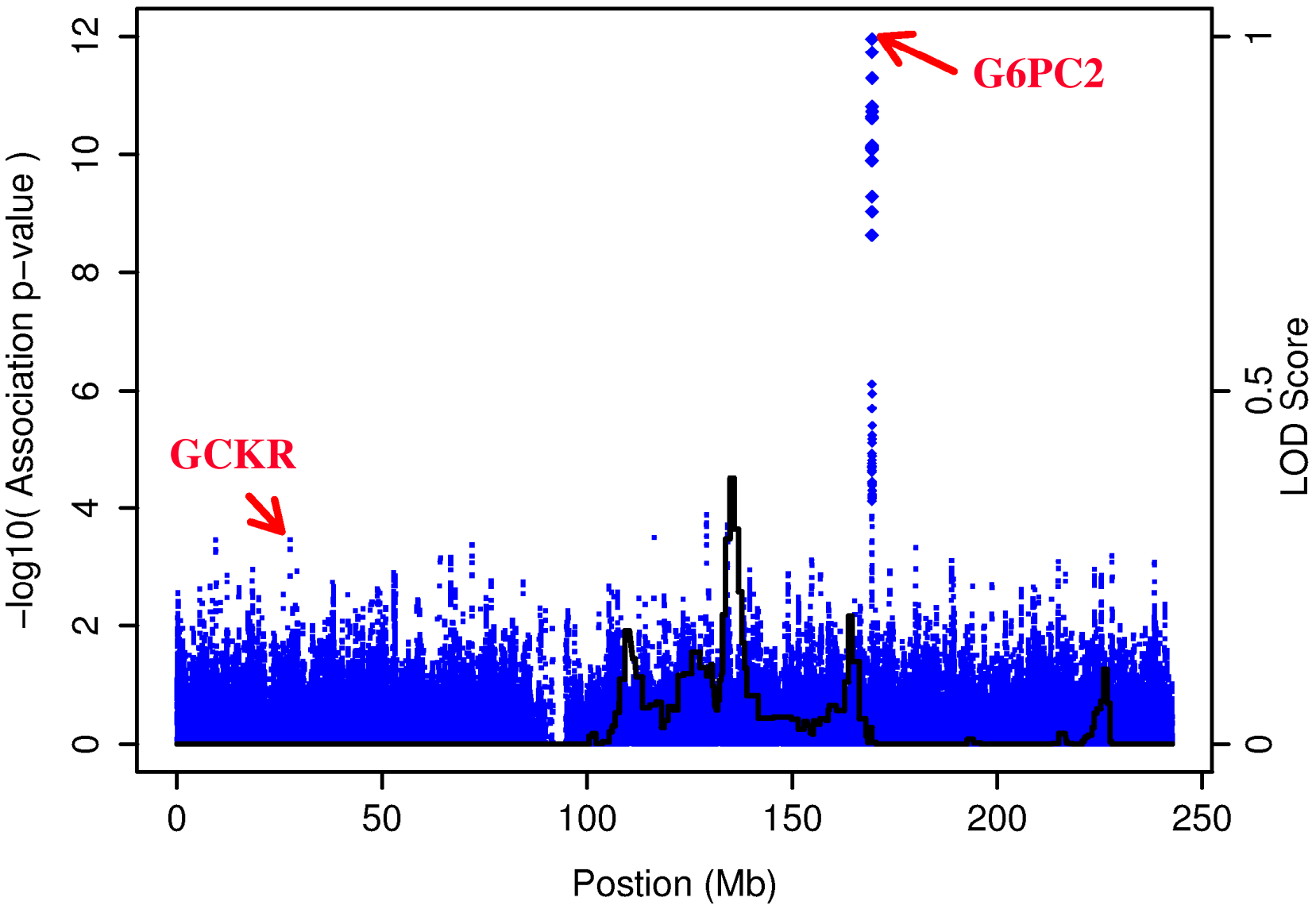
FHS: Phenotypes of Interest

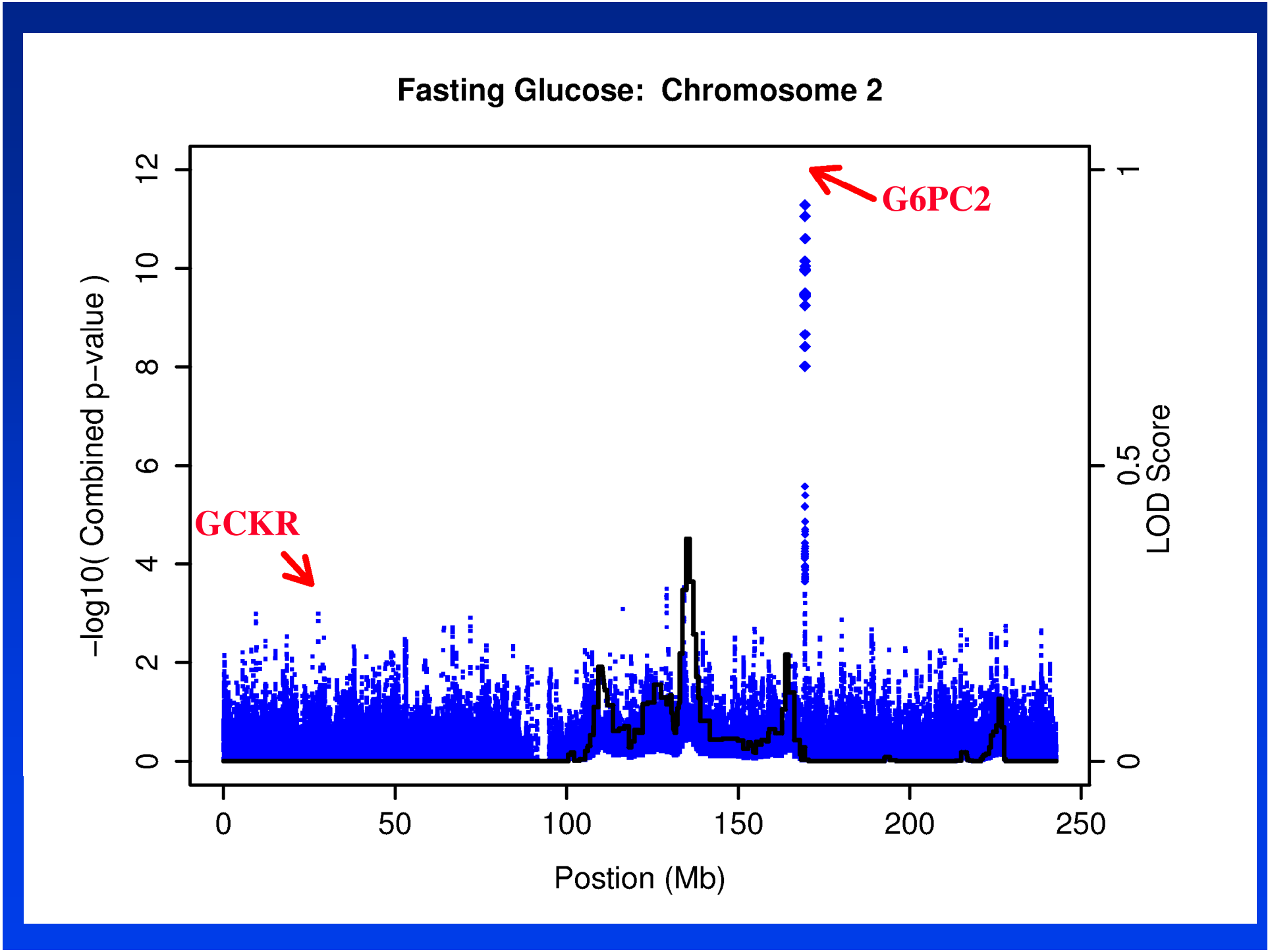
- **Type 2 diabetes and related quantitative traits**
  - ⇒ **Fasting plasma glucose level in non-diabetics**
    - ♦ **High level of fasting glucose is indicative of Type-2 diabetes**
- **Cardiovascular disease and related quantitative traits**
  - ⇒ **C-reactive protein (CRP) plasma level, a biomarker of inflammation**

Fasting Blood Glucose

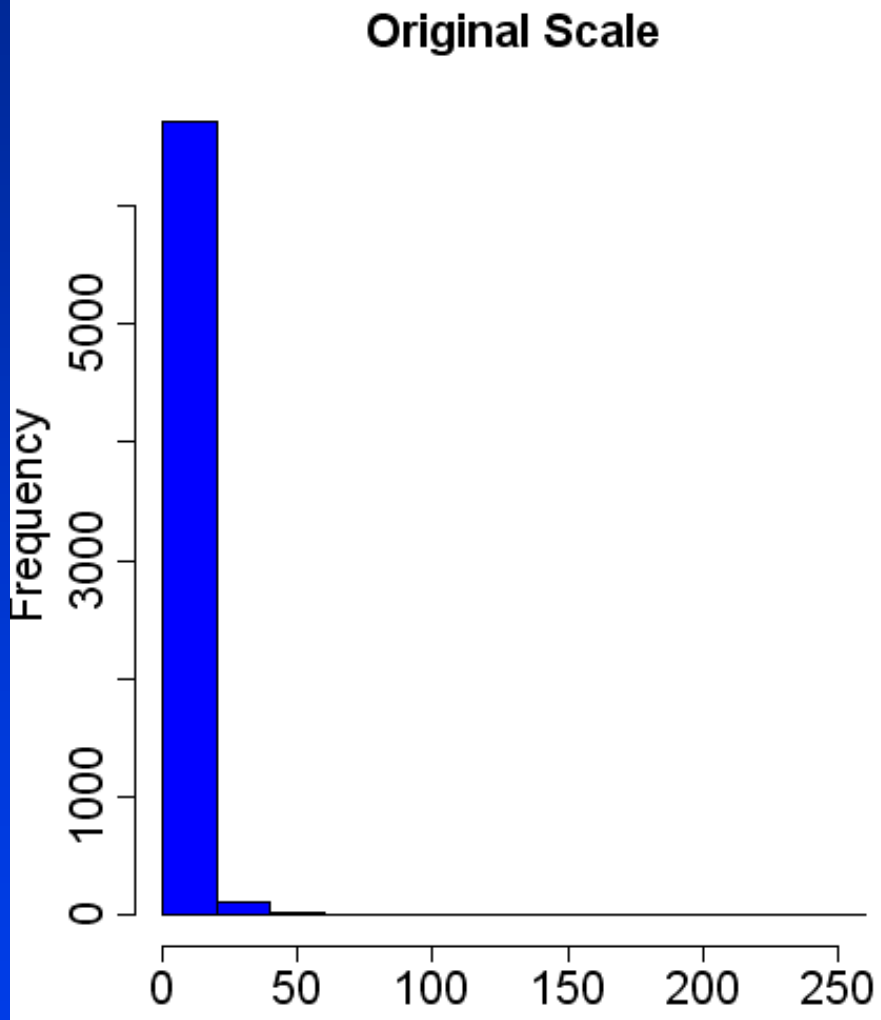


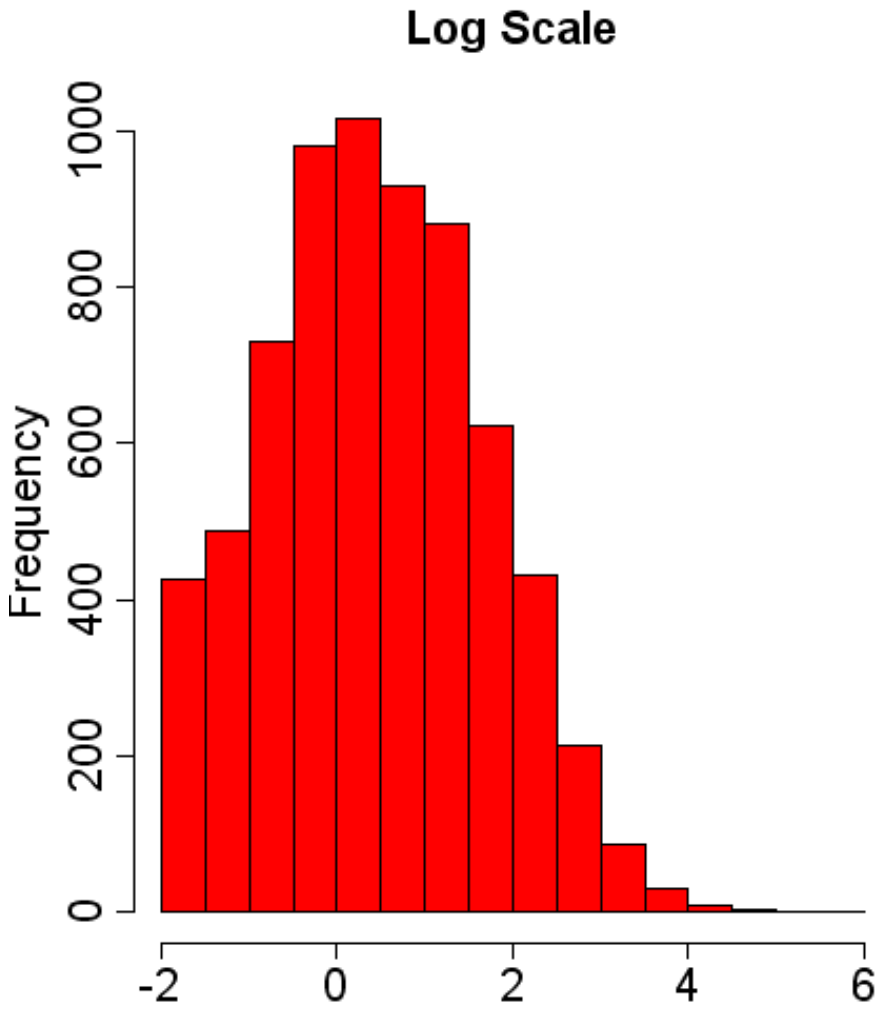


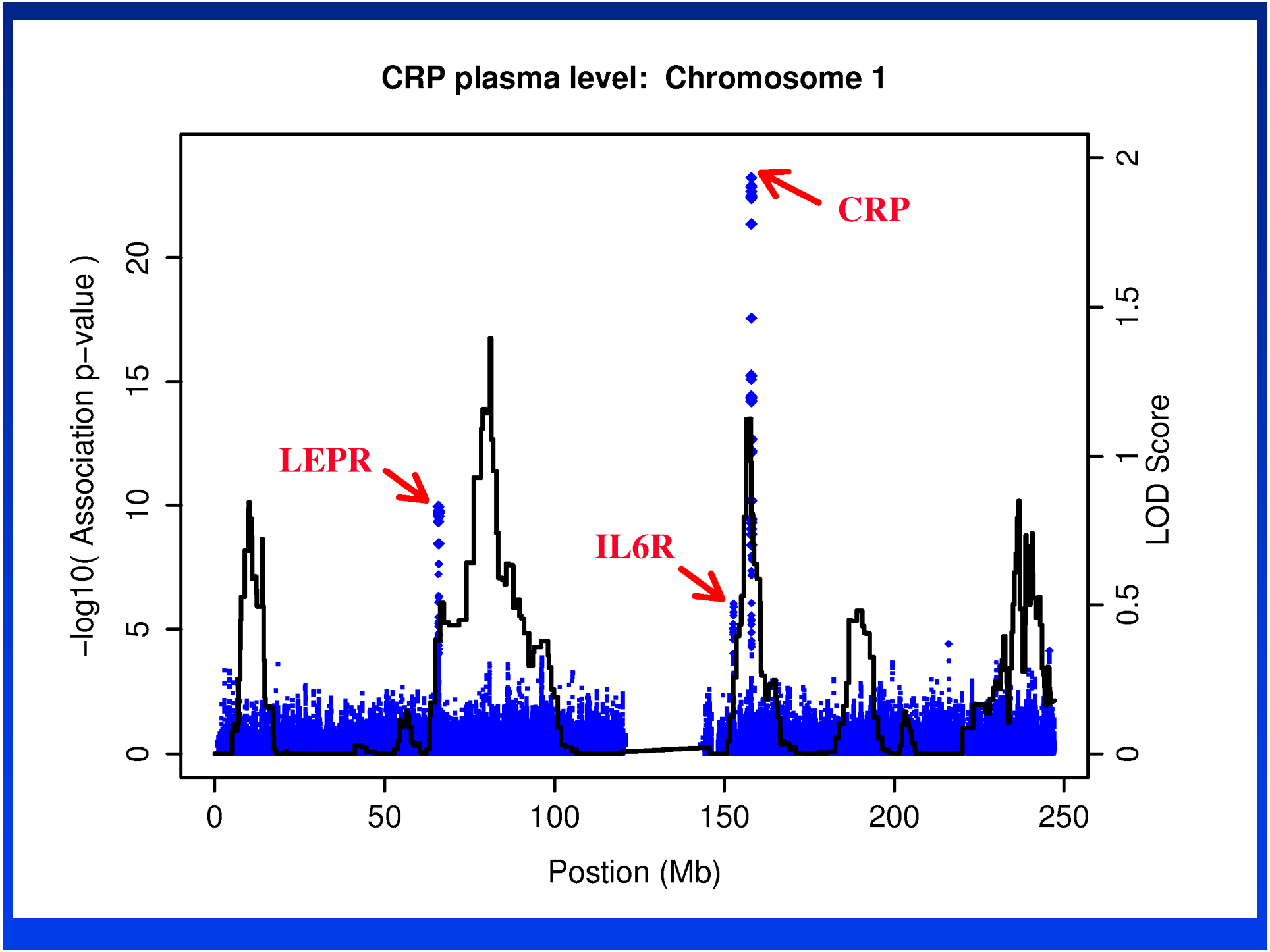

Fasting Glucose: Chromosome 2
G6PC2
GCKR
−log10(Association p−value)
0
2
4
6
8
10
12
0
0.5
1
LOD Score
0
50
100
150
200
250
Postion (Mb)

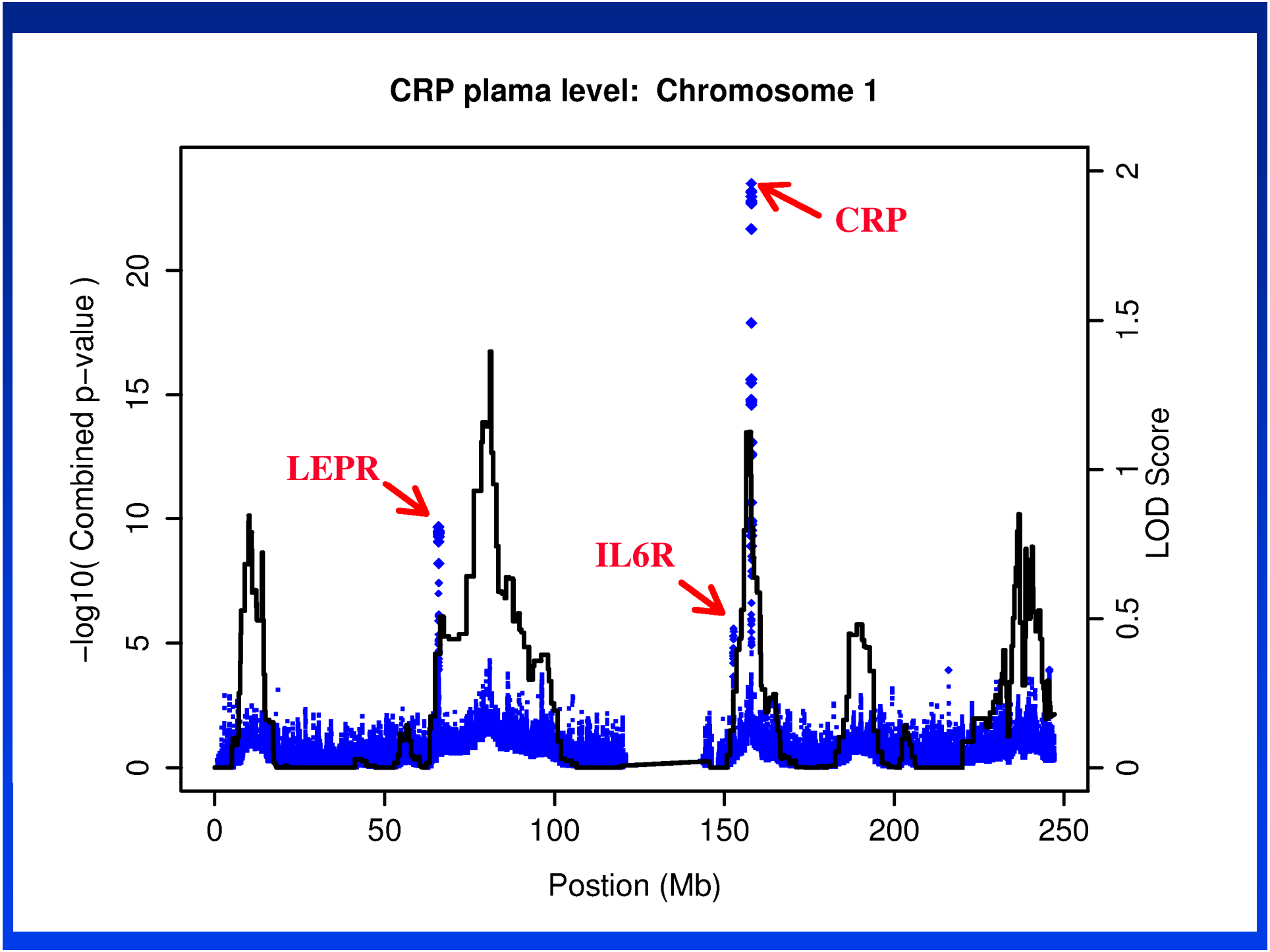

Fasting Glucose: Chromosome 2
G6PC2
GCKR
−log10(Combined p−value)
0
2
4
6
8
10
12
LOD Score
0
0.5
1
0
50
100
150
200
250
Postion (Mb)

Plasma CRP levels






CRP plasma level: Chromosome 1
CRP
LEPR
IL6R
−log10(Association p−value)
LOD Score
Postion (Mb)
0
5
10
15
20
0
0.5
1
1.5
2
0
50
100
150
200
250


CRP plama level: Chromosome 1
CRP
LEPR
IL6R
−log10(Combined p−value)
LOD Score
Postion (Mb)
0
5
10
15
20
0
0.5
1
1.5
2
0
50
100
150
200
250

Generalized Linear Model

- **Framework is general and not restricted to normal distribution**
- **May use generalized linear model likelihood, which includes the normal distribution**
- **Log likelihood becomes:** $\ell(\beta) = \sum_{i=1}^{N} \left[Y_i \theta_i - \psi(\theta_i) \right] / \phi$

  $\Rightarrow$ **ϕ is a scale parameter**

  $\Rightarrow$ **ψ is the cumulant generating function**

  $\Rightarrow \theta_i(\beta) = h\left(\eta_i(\beta)\right)$

  $\Rightarrow \eta_i(\beta) = \beta_z z_i + \beta_Q Q_i + \beta_G G_i + \beta_E E_i$

Outline

- Motivating dataset: Framingham Heart Study
- Quantitative trait mapping in family samples
- Common framework for linkage and association analysis

  $\Rightarrow$ Application to the Framingham Heart Study
- **Association analysis: value of family samples**
- Linkage analysis: extensions to basic model

  $\Rightarrow$ Results from the Framingham Heart Study

Power Approximation

- Can get explicit formulas for the non-centrality parameter (NCP) for both linkage and association score statistics
  - $\Rightarrow$ NCP = Expected value of test statistic
  - $\Rightarrow$ NCP may be used to approximate study power
- Common framework allows power comparisons
  - $\Rightarrow$ Linkage versus association (or joint statistic)
  - $\Rightarrow$ Different study designs
  - $\Rightarrow$ Effect of LD and model misspecification

Noncentrality Parameter

- **Approximation for non-centrality parameter for association statistic (ignoring dominance effect, for simplicity):**

$$E[Z_a \mid Y] \approx \frac{(Y-\beta z)' \Sigma^{-1} \Phi \Sigma^{-1} (Y-\beta z)\sigma^2_{A:U,\beta_Q}}{\sqrt{(Y-\beta z)' \Sigma^{-1} \Phi \Sigma^{-1} (Y-\beta z)\sigma^2_{A:U}}}$$

- $\sigma^2_{A:U,\beta_Q}$ **is the additive covariance between** $Q\beta_Q$ **and U**
  - $\Rightarrow$ Q **may be a matrix, but** $Q\beta_Q$ **is a vector**
  - $\Rightarrow$ **If U is a perfect proxy for Q,** $\sigma^2_{A:U,\beta_Q} = \beta_Q \sigma^2_{A:U}$
  - $\Rightarrow$ **Power of association test depends on correlation between U (proxy) and Q (trait loci)**
- **Equivalent approximation available for the linkage statistic**

Association Analysis in Family Samples

- **Advantages**
  - ⇒ **Able to perform additional quality control (inheritance checking)**
  - ⇒ **Estimate heritability of trait**
  - ⇒ **Able to perform family based association tests**
  - ⇒ **Able to perform joint linkage and association tests**
- **Disadvantages**
  - ⇒ **Need to account for familial correlation**
    - ♦ **Lose information/power by selecting a single individual per family**

Power of Additive Association test: Effect of study design

	Subset included in analysis		
Study Design	All family members	Only unrelateds	Equivalent # of unrelateds
3 sibs + 2 parents	95.2%	28.5%	98.9%
2 sibs + 2 parents	85.1%	28.5%	92.9%
3 sibs (no parents)	59.9%	2.3%	70.4%
2 sibs (no parents)	24.6%	2.3%	28.5%

Additive QTL explaining 1% of variance of trait with 30% heritability;

N family = 1200; Genome wide significance set at 5×10^{-8}

Power of Additive Association test: Effect of study design

	Subset included in analysis		
Study Design	All family members	Only unrelateds	Equivalent # of unrelateds
3 sibs + 2 parents	64.2%	7.3%	80.9%
2 sibs + 2 parents	44.1%	7.3%	58.1%
3 sibs (no parents)	21.4%	0.4%	29.0%
2 sibs (no parents)	5.9%	0.4%	7.3%

Recessive QTL explaining 2% of variance , MAF=20%.

N family = 1200; Genome wide significance set at 5×10^{-8}

Outline

- Motivating dataset: Framingham Heart Study
- Quantitative trait mapping in family samples
- Common framework for linkage and association analysis

  $\Rightarrow$ Application to the Framingham Heart Study
- Association analysis: value of family samples
- **Linkage analysis: extensions to basic model**

  $\Rightarrow$ **Results from the Framingham Heart Study**

Extension to Basic Model: Gene x Covariate Interaction

$$E[Y_i] = \eta_i(\beta) = \beta_z z_i + \beta_x x_i + \beta_U U_i + \beta_I x_i U_i + \beta_G G_i + \beta_E E_i$$

- Y_i = phenotype of interest
- z_i = (non-genetic) covariates
- U_i = proxy for QTL covariates
- x_i = covariate for which we want to test for interaction with genetic component (Q_i)
- G_i = latent genetic factor (polygenic)
- E_i = Environmental & other non-genetic factors

Efficient Linkage Score Statistic for Gene x Covariate Interaction

- First derivative of log likelihood for linkage with respect to β_U and β_I, evaluated at $\beta_U=\beta_I=0$ is 0

  $\Rightarrow$ Take second derivative (also evaluated at $\beta_U=\beta_I=0$)

- For normal log likelihood function:

$$\ddot{\ell}(\beta_0;\nu_t \mid Y) \propto [tr(WA_\pi), tr(WB_{x,\pi}), tr(WC_{x,\pi})]$$

$$W = \Sigma_x^{-1}(Y-\beta z)(Y-\beta z)' \Sigma_x^{-1} - \Sigma_x^{-1}$$

$A_\pi = \hat{\Phi} - \Phi$ = Centered observed pairwise IBD proportions

$B_{x,\pi}$ is formed by multiplying each entry of A_π by $x_i + x_j$

$C_{x,\pi}$ is formed by multiplying each entry of A_π by $x_i \times x_j$

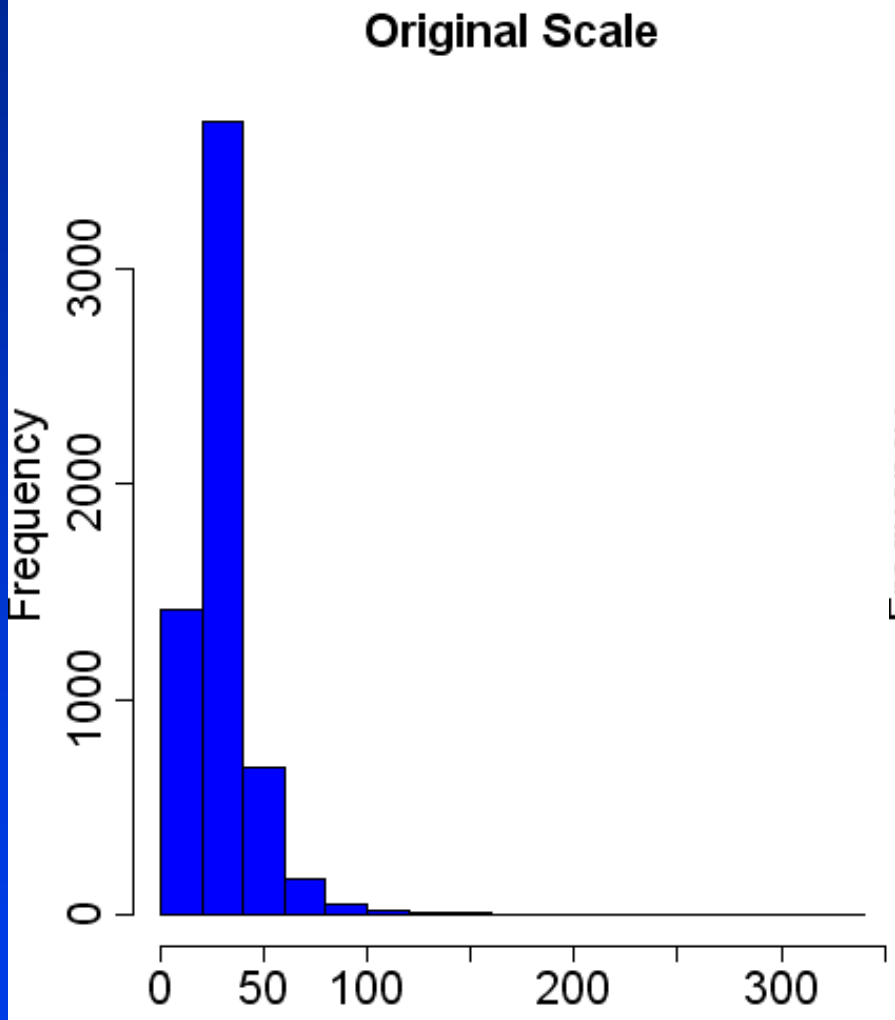
Extension to Basic Model: Gene x Covariate Interaction

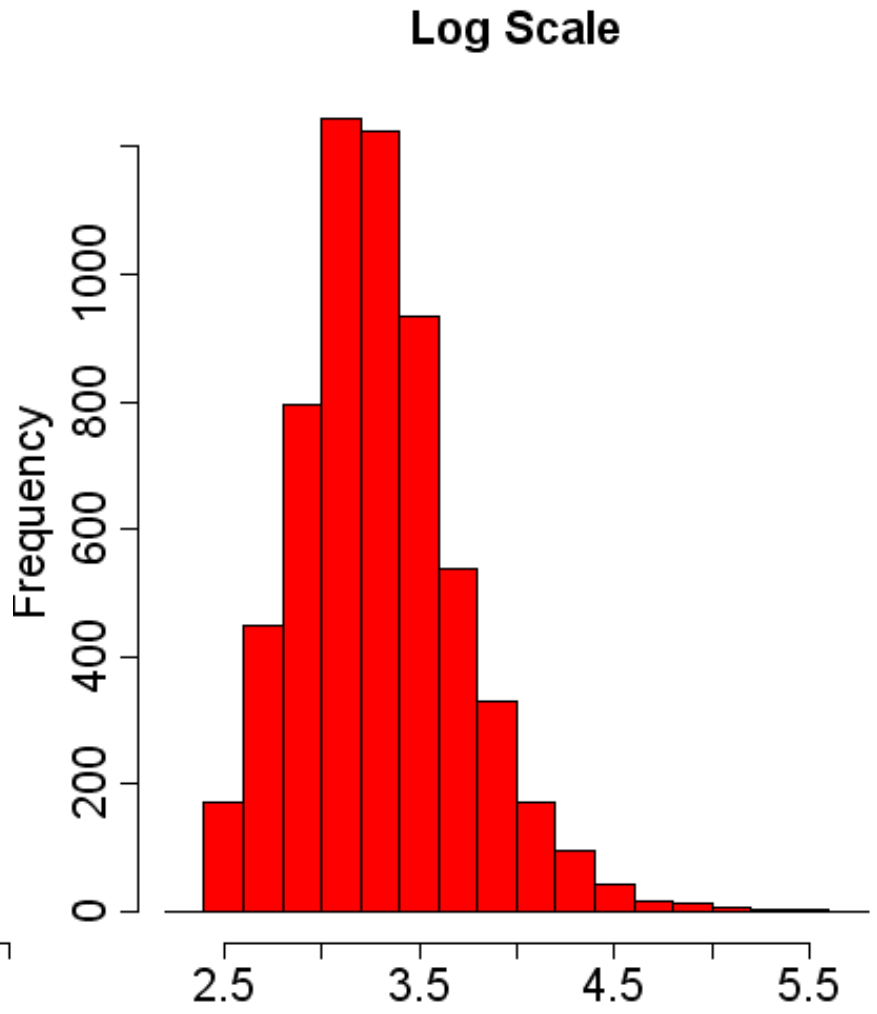
- Score statistic has three components
- Some constraints apply to the three components of the efficient score statistic
  - $\Rightarrow$ First and third components must be non-negative
  - $\Rightarrow$ Second component constrained by the first and third

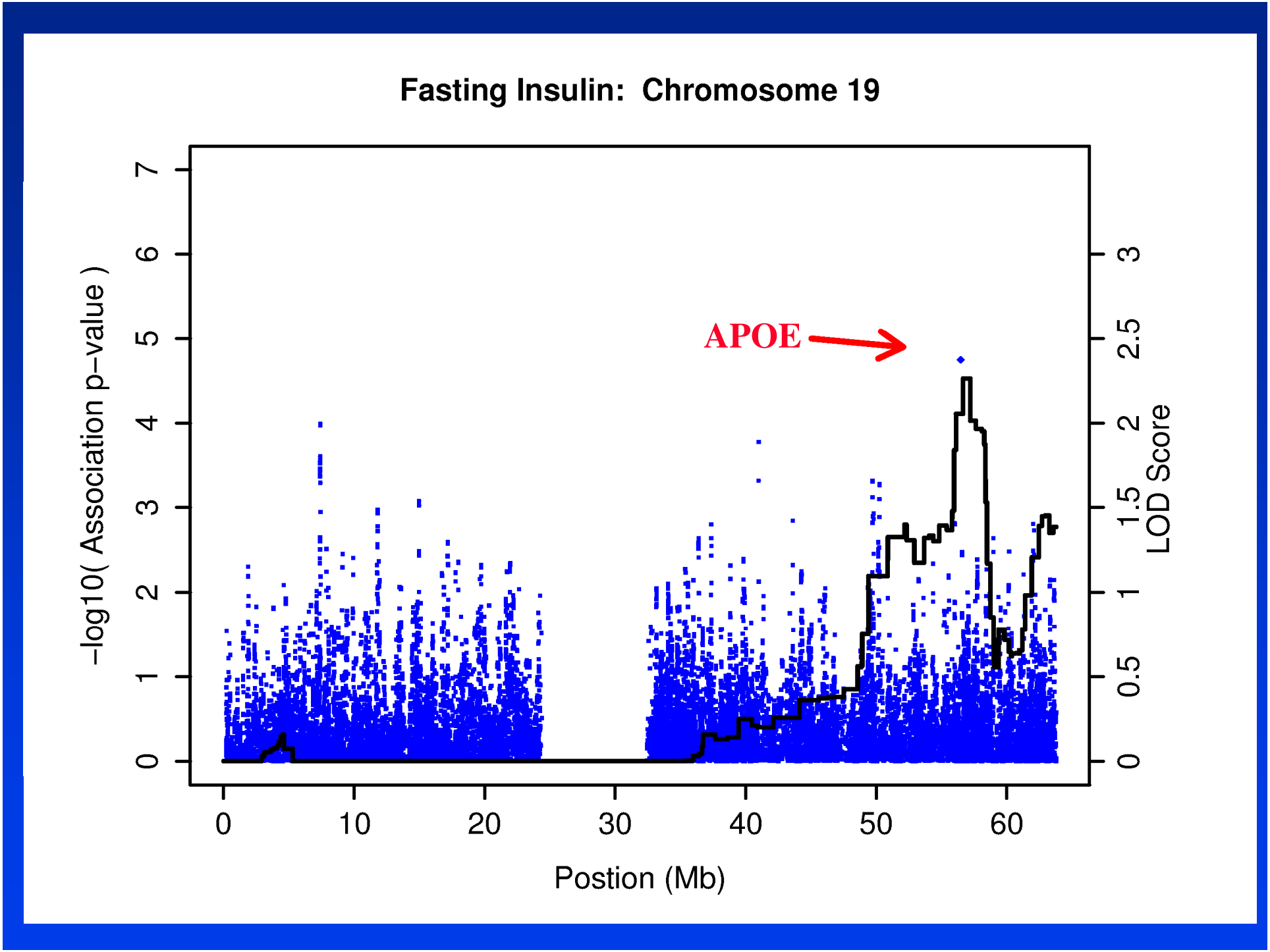
FHS: Phenotypes of Interest

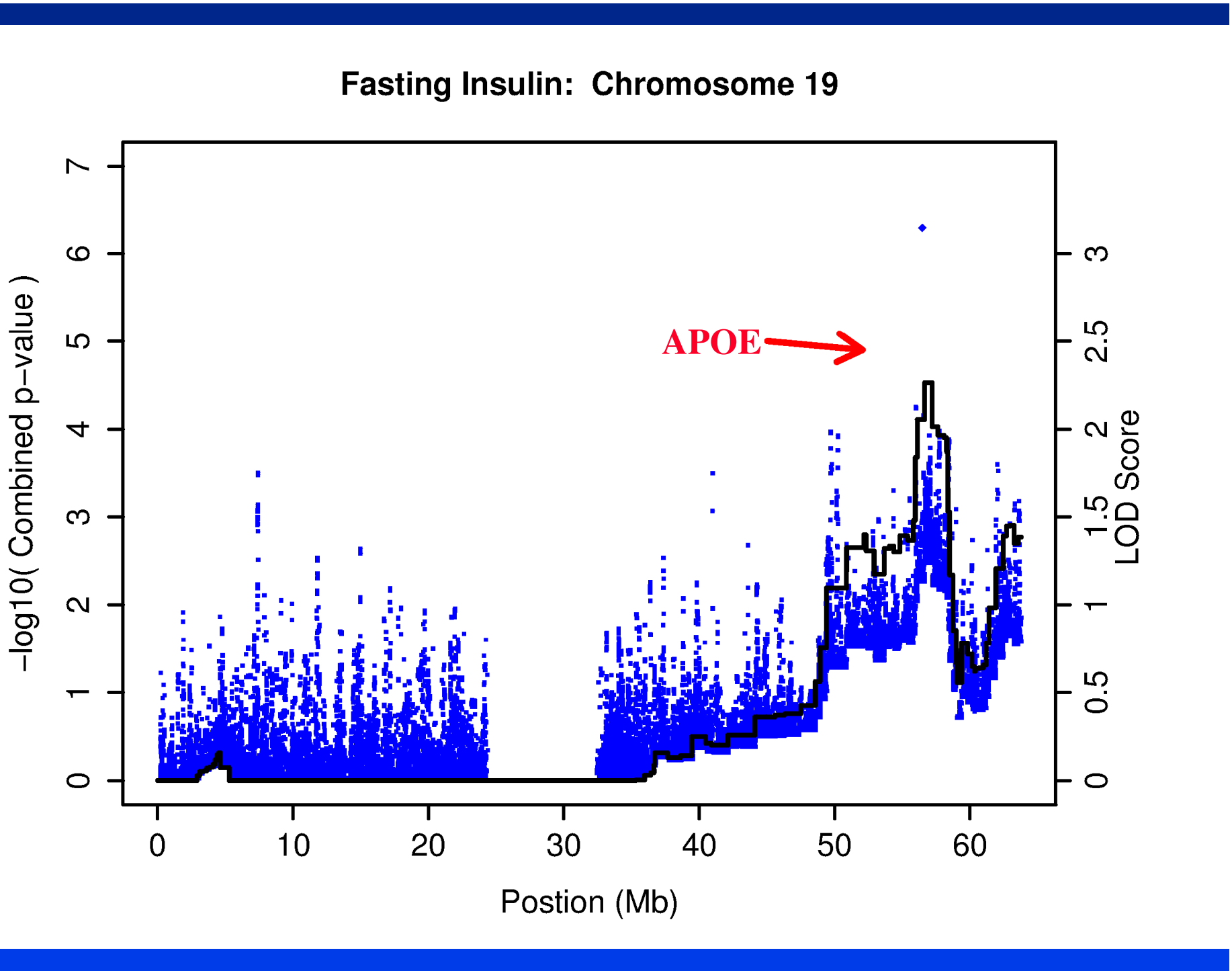
- **Type 2 diabetes and related quantitative traits**
  - ⇒ **Fasting insulin in non-diabetics**
  - ⇒ **Evidence for association with APOE gene located on chromosome 19**
  - ⇒ **One published report suggesting association is modified by body mass index (BMI)**

Fasting Insulin

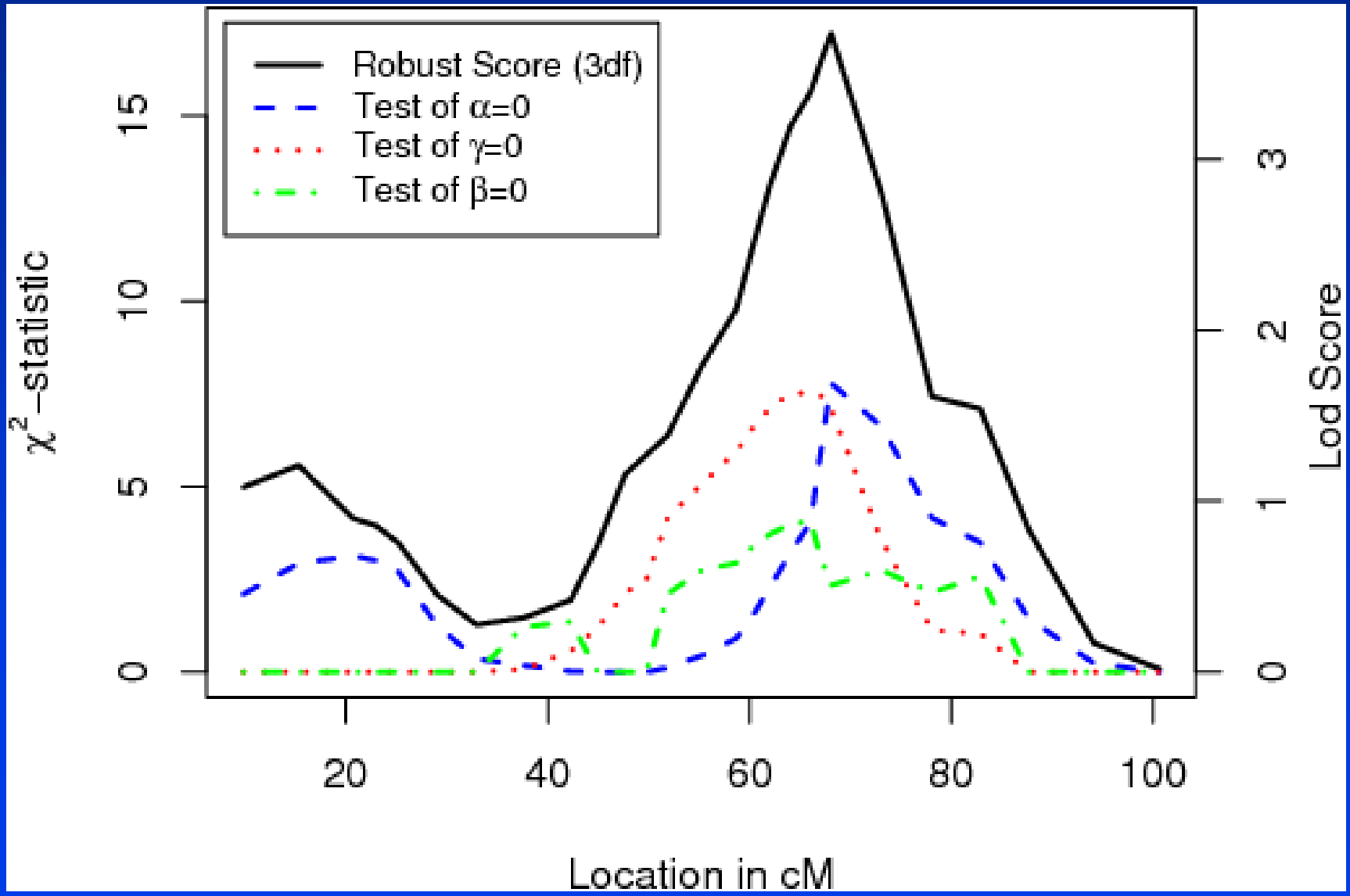




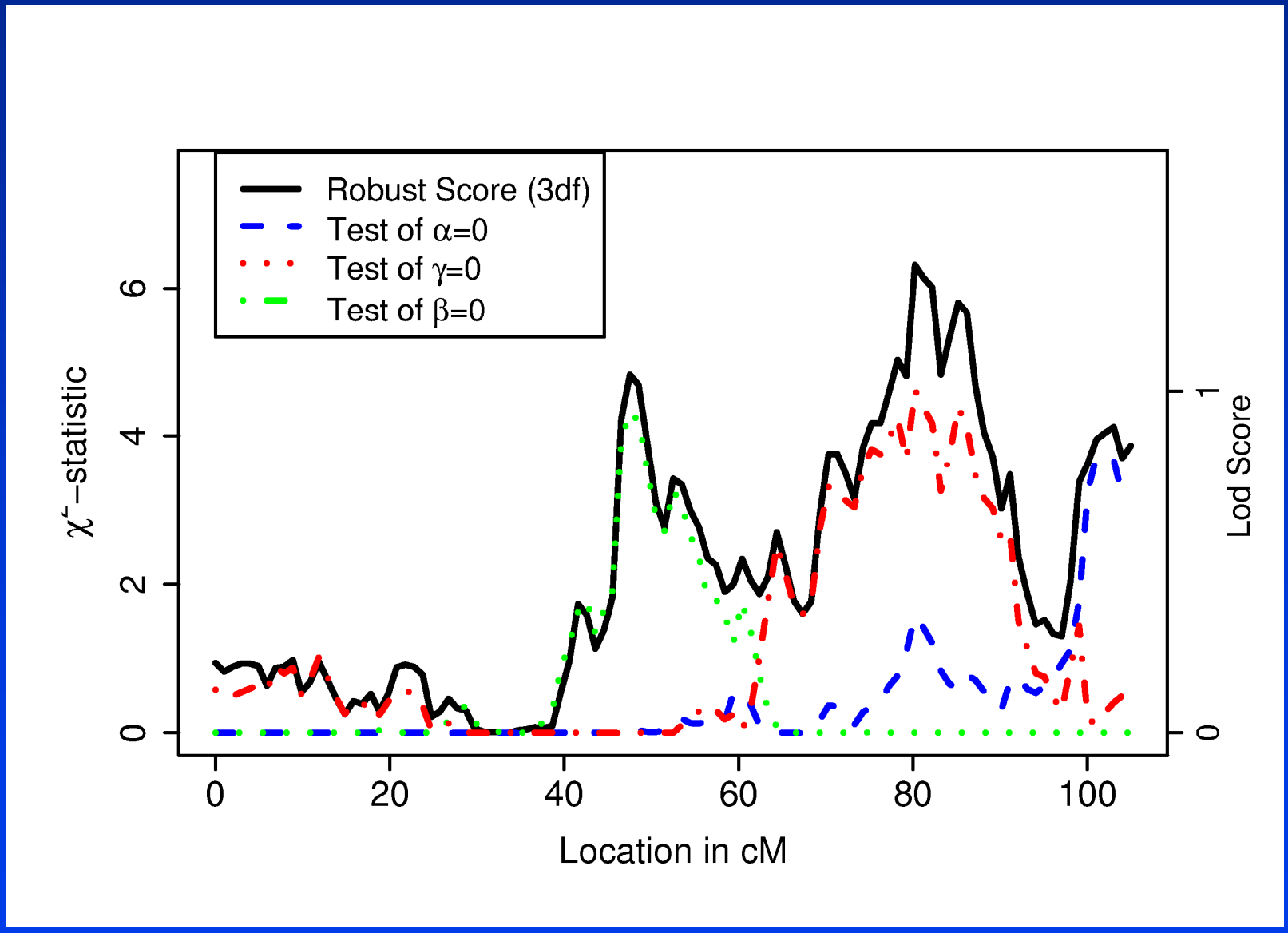

Fasting Insulin: Chromosome 19
APOE
−log10(Association p−value)
LOD Score
Postion (Mb)
0
1
2
3
4
5
6
7
0
0.5
1
1.5
2
2.5
3
0
10
20
30
40
50
60


Fasting Insulin: Chromosome 19
APOE
−log10(Combined p−value)
LOD Score
Postion (Mb)
0
1
2
3
4
5
6
7
0
0.5
1
1.5
2
2.5
3
0
10
20
30
40
50
60

Fasting Insulin with gene x BMI interaction: Offspring generation only



Fasting Insulin with gene x BMI interaction: Offspring and Generation 3



Extension to Basic Model Multivariate Phenotypes

- May be of interest to jointly analyze phenotypes that may be under similar genetic control

  $\Rightarrow$ Lipoprotein-associated phospholipase A_2 (LpPla2) plasma level **mass** and **activity**, biomarkers of inflammation

- Bivariate score statistic for linkage has 3 components

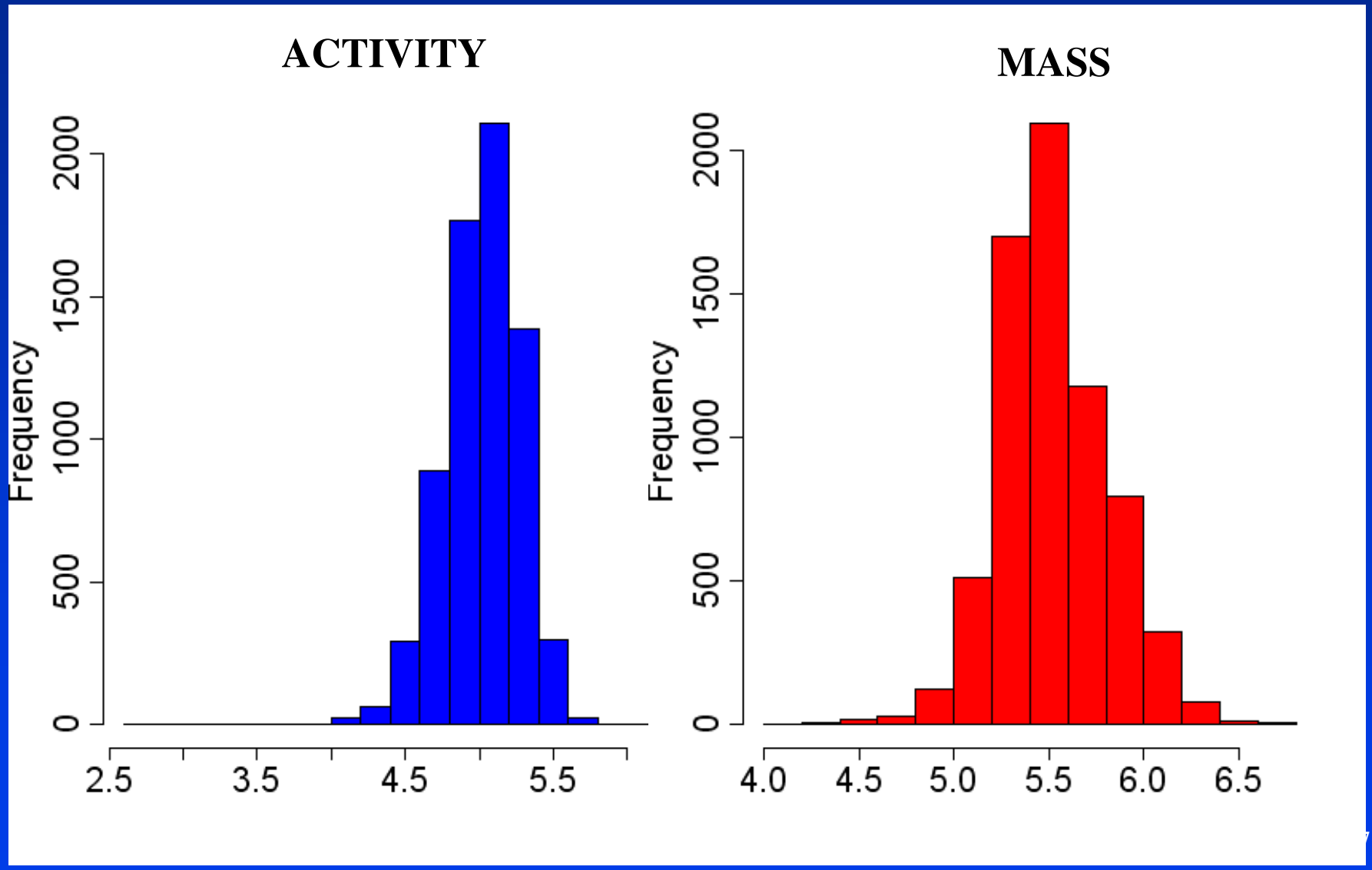
  $\Rightarrow$ Similar to linkage statistic for gene x covariate interaction

- Similar constraints apply to the three components of the efficient score statistic

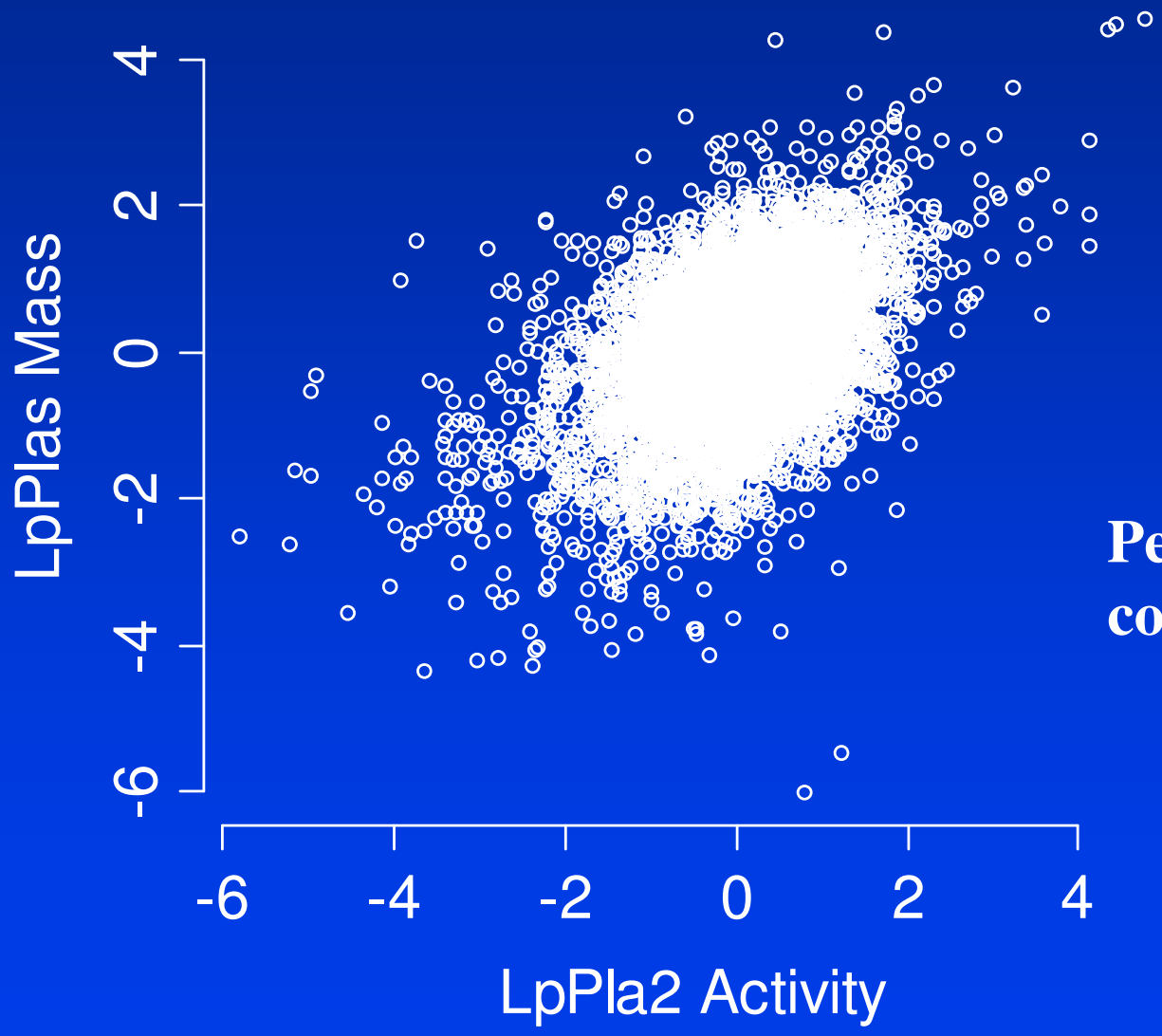
  $\Rightarrow$ First and third component must be non-negative

  $\Rightarrow$ Second component constrained by the first and third

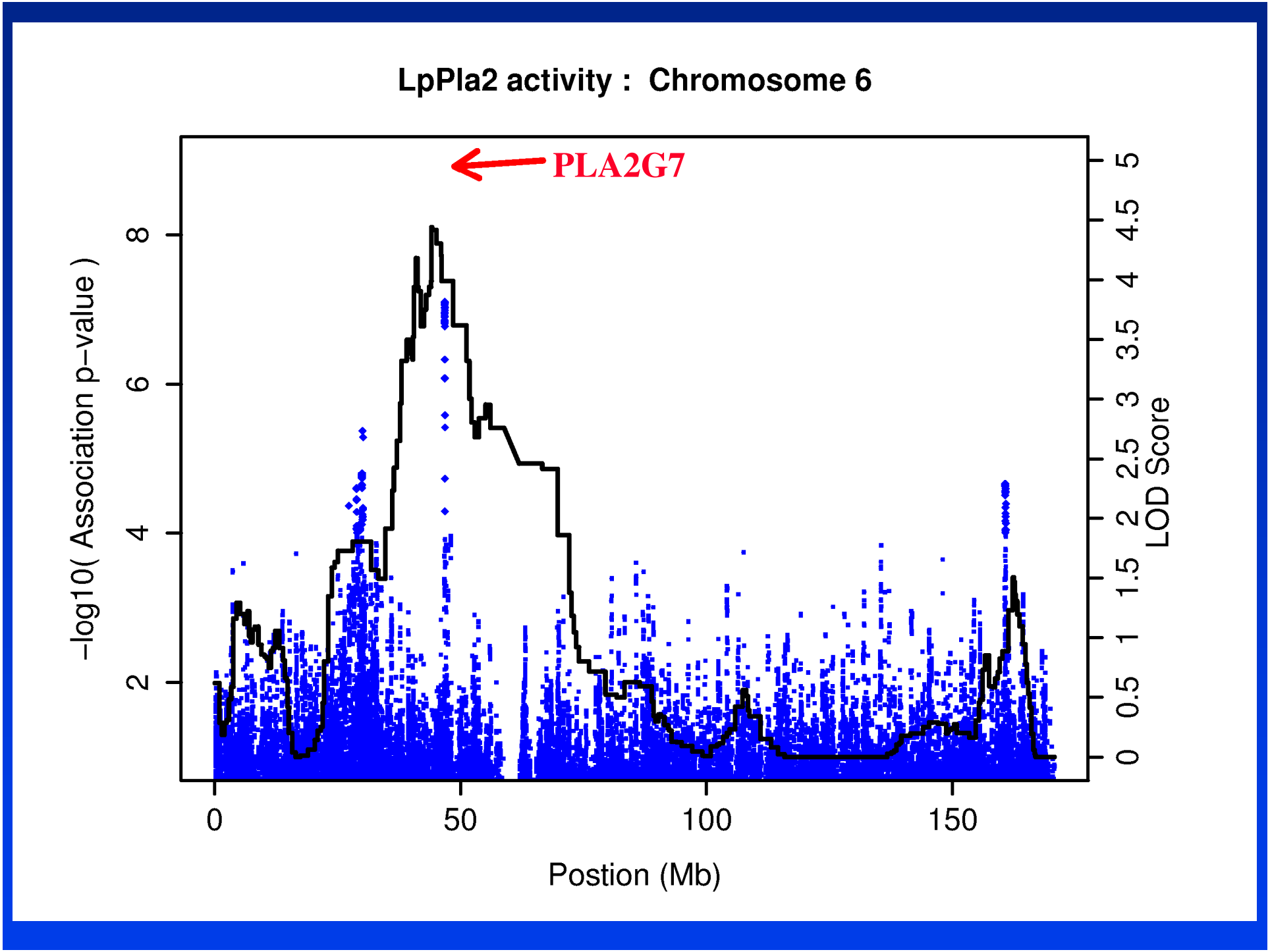
LpPla2: Log Scale

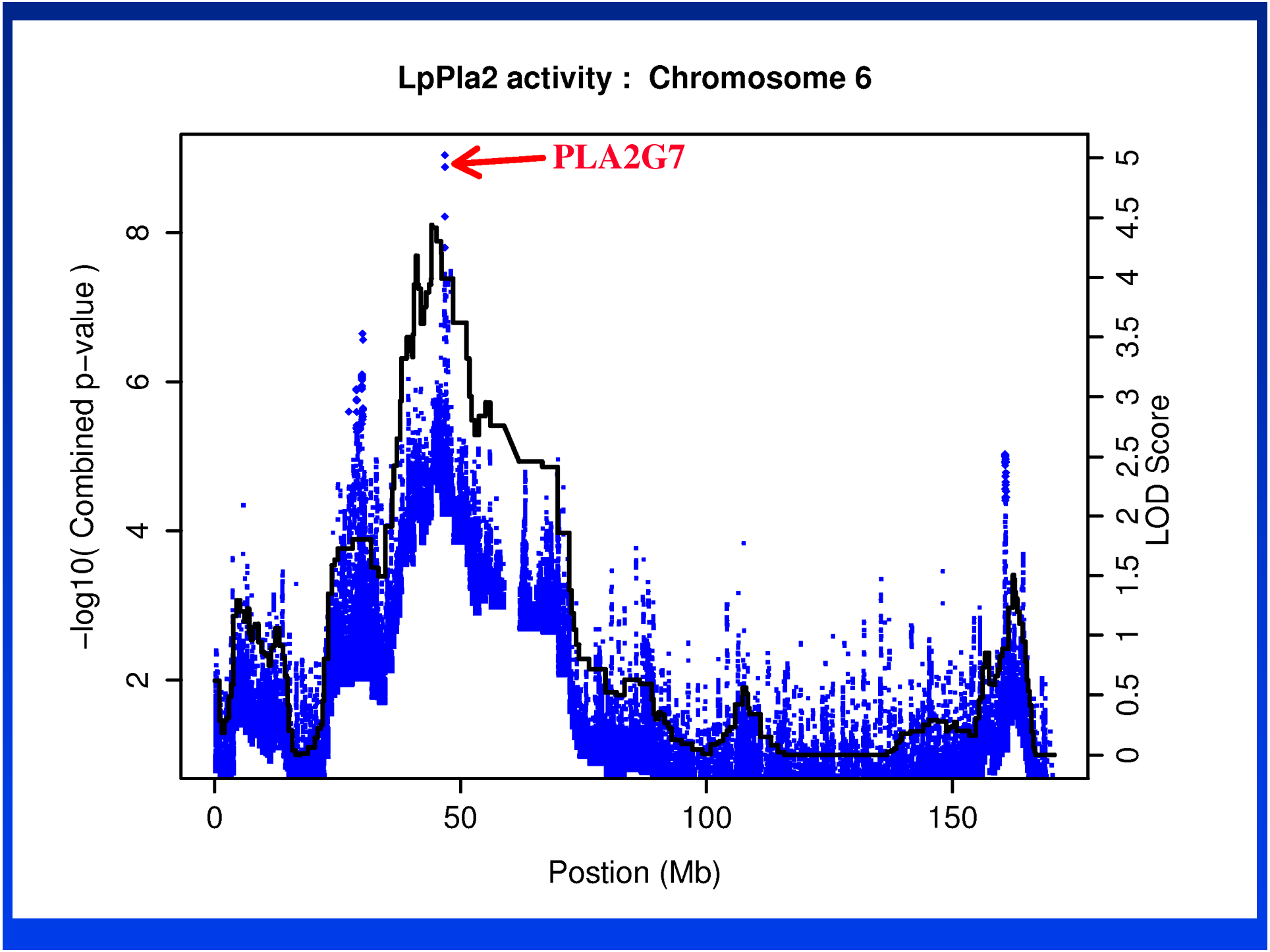


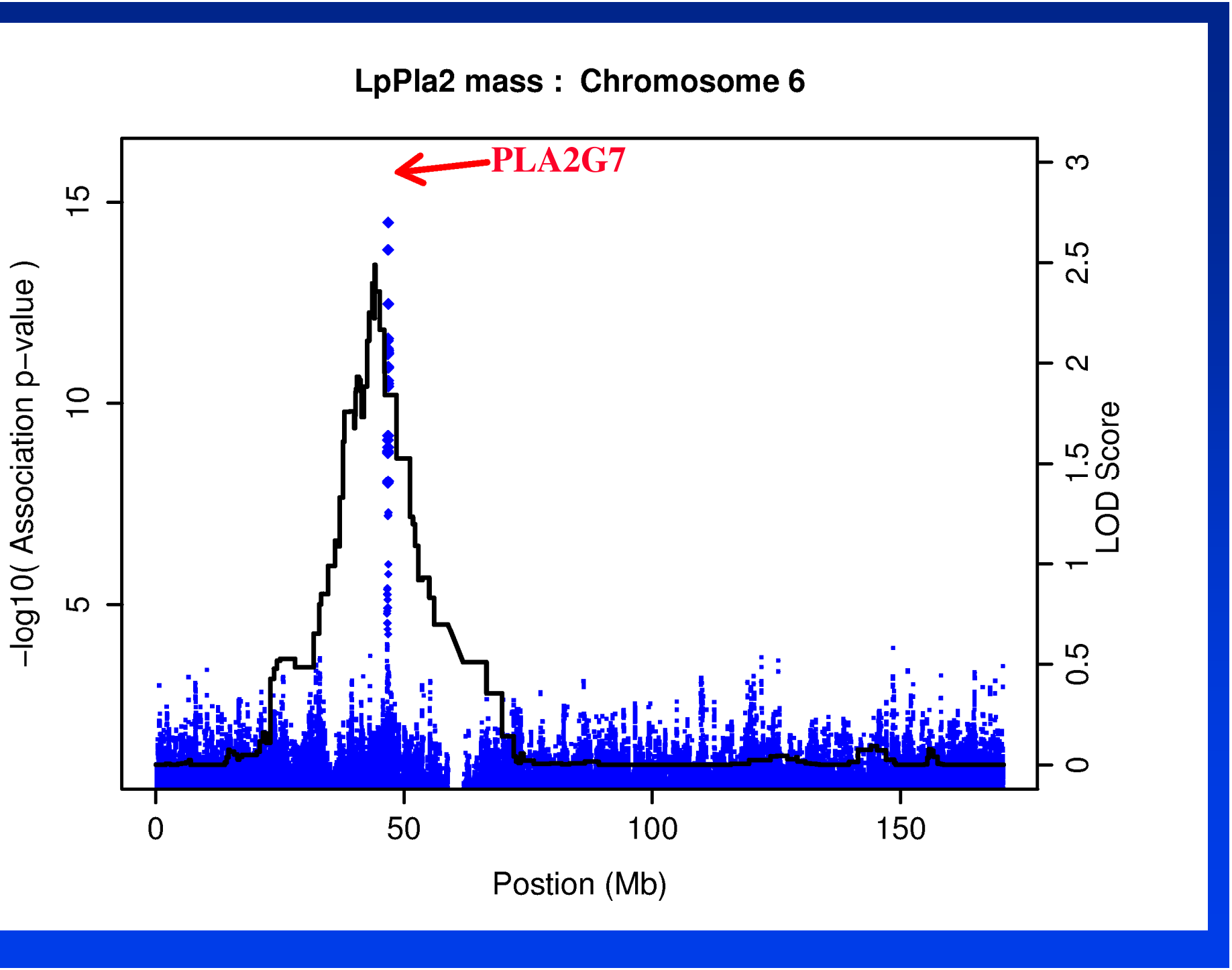
LpPla2: Log Scale

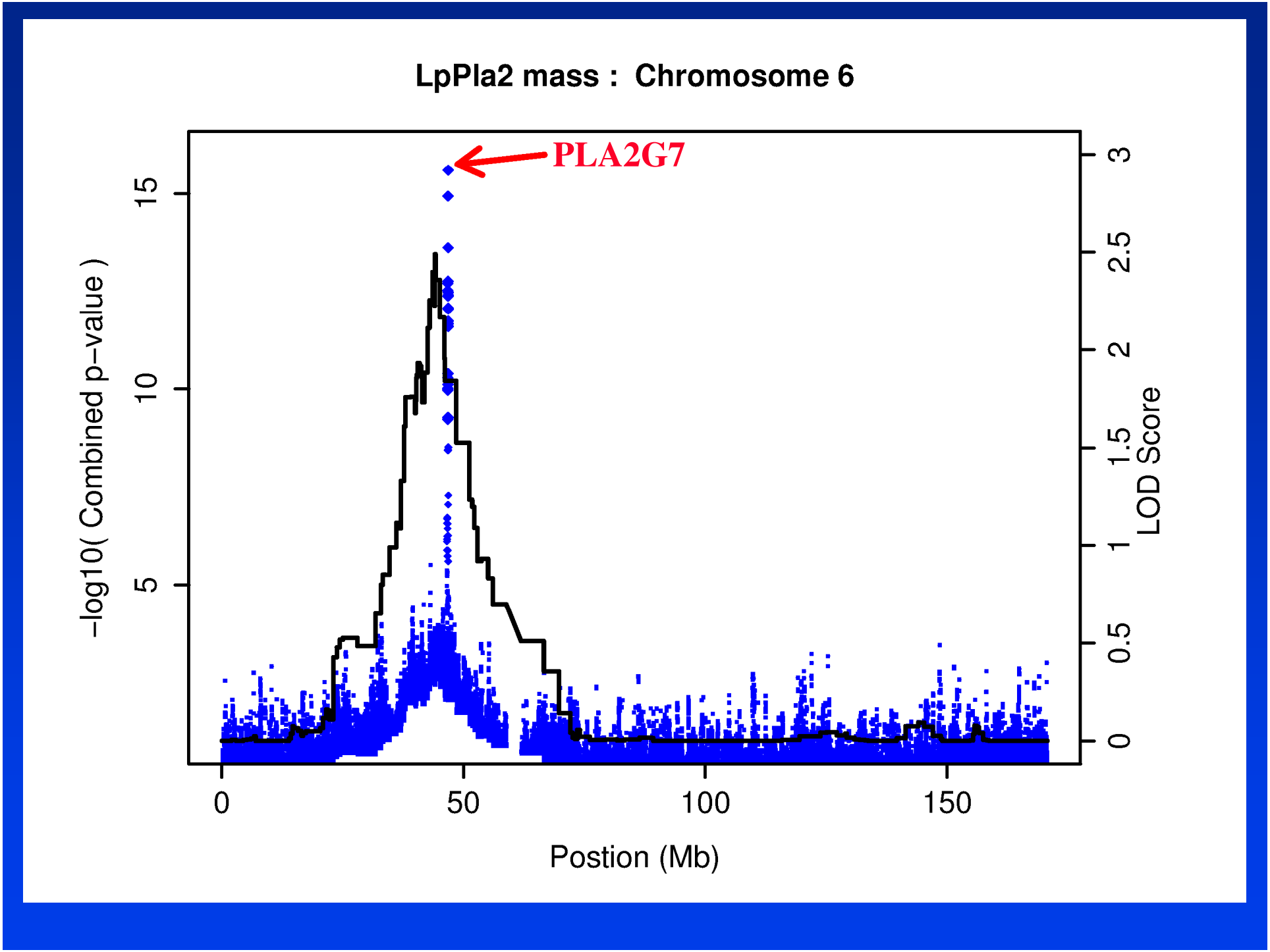


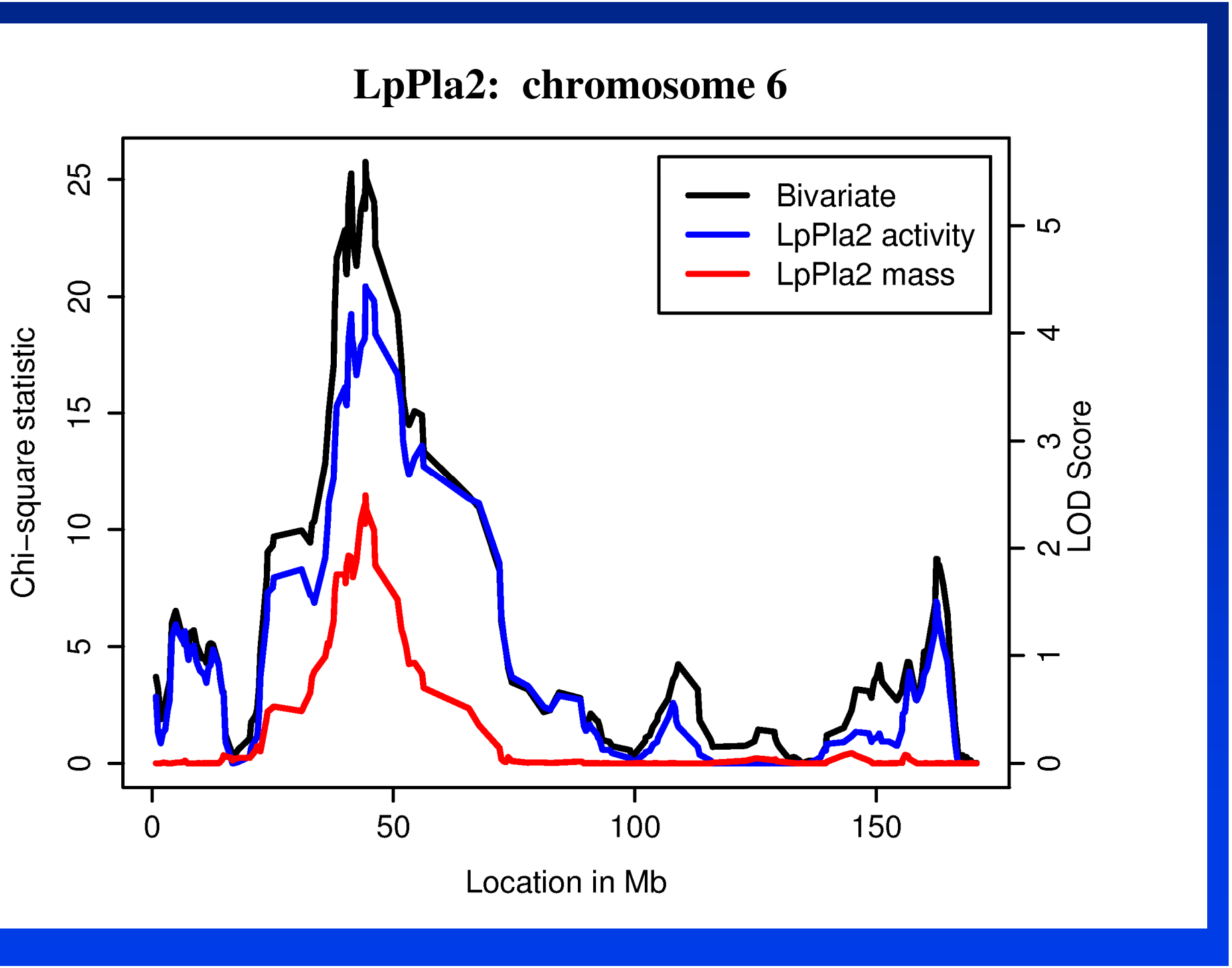
Pearson correlation coefficient = 0.46


LpPla2 activity : Chromosome 6
PLA2G7
−log10(Association p−value)
2
4
6
8
LOD Score
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
0
50
100
150
Postion (Mb)


LpPla2 activity : Chromosome 6
PLA2G7
−log10(Combined p−value)
2
4
6
8
0
50
100
150
Postion (Mb)
LOD Score
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5


LpPla2 mass : Chromosome 6
PLA2G7
−log10(Association p−value)
5
10
15
0
50
100
150
Postion (Mb)
0
0.5
1
1.5
2
2.5
3
LOD Score


LpPla2 mass : Chromosome 6
PLA2G7
−log10(Combined p−value)
5
10
15
LOD Score
0
0.5
1
1.5
2
2.5
3
0
50
100
150
Postion (Mb)


LpPla2: chromosome 6
Bivariate
LpPla2 activity
LpPla2 mass
Chi−square statistic
LOD Score
Location in Mb
0
5
10
15
20
25
0
1
2
3
4
5
0
50
100
150

Concluding Remarks

- **Several advantages to conducting association analysis in family samples**
  - ⇒ **Power reduction from relatedness is modest and depends on heritability of trait**
  - ⇒ **Better quality control**
  - ⇒ **Can obtain estimate of heritability of trait**
- **General framework allows comparison of different study designs**
  - ⇒ **Family versus unrelated individuals (association)**
  - ⇒ **Nuclear families versus extended pedigrees (association & linkage)**

Concluding Remarks

- General framework can be easily extended
  - ⇒ Survival phenotypes
  - ⇒ Gene x covariate interaction
  - ⇒ Multivariate phenotypes
- Methods combining linkage and association may offer increased power over either approaches alone
  - ⇒ Loss of power is small in region of no linkage
  - ⇒ Future research needed to determine how/when/if it is advisable to combine linkage and association statistics

Acknowledgement/Collaborators

Stanford University

David Siegmund

The Hebrew University of Jerusalem

Benjamin Yakir

Boston University/Framingham Heart Study

James B. Meigs

José Florez

Alisa Manning

Emelia Benjamin

Renate Schnabel

Chen Lu