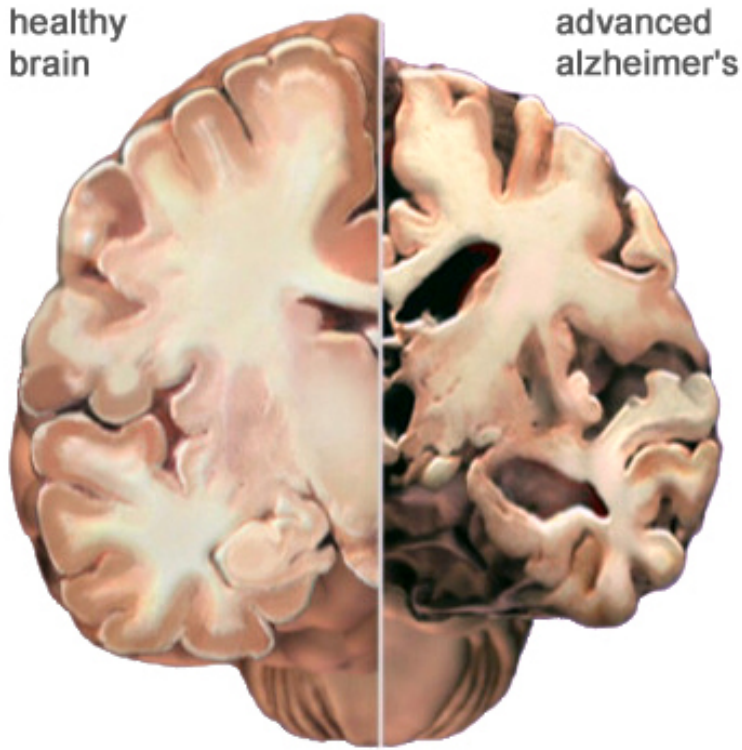
Novel Genomic Variants for Late-onset Alzheimer's Disease

Xulong Wang

August 25, 2014

Late-onset Alzheimer's Disease



Alzheimer's disease is the most common form of dementia affecting over 35 million people worldwide.

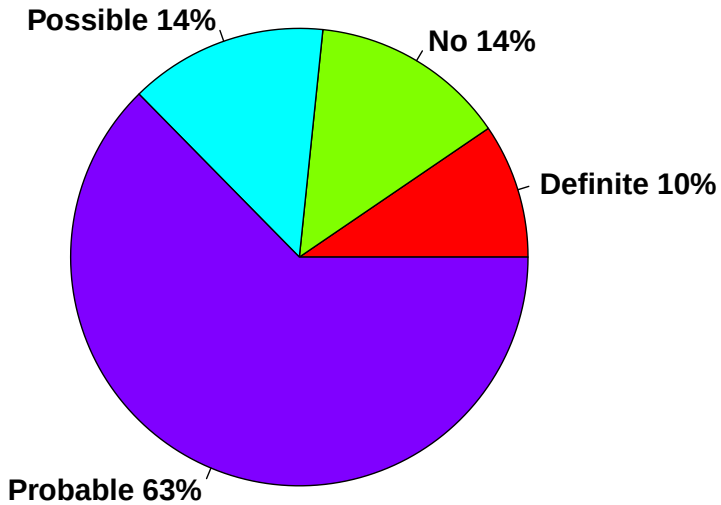
Late-onset Alzheimer's Disease

- Late-onset Alzheimer's disease usually occurs after age 65, and accounts for about 90% of total AD cases.
- Top LOAD associated genes by GWAS study: APOE $\epsilon 2/3/4$, BIN1, CLU, ABCA7, CR1, PICALM, MS4A6A, CD33, MS4A4E, CD2AP (Data from www.alzgene.org).

Alzheimer's Disease Sequencing Project

Objective: To identify new genomic variants that increase and decrease the risk of developing AD.

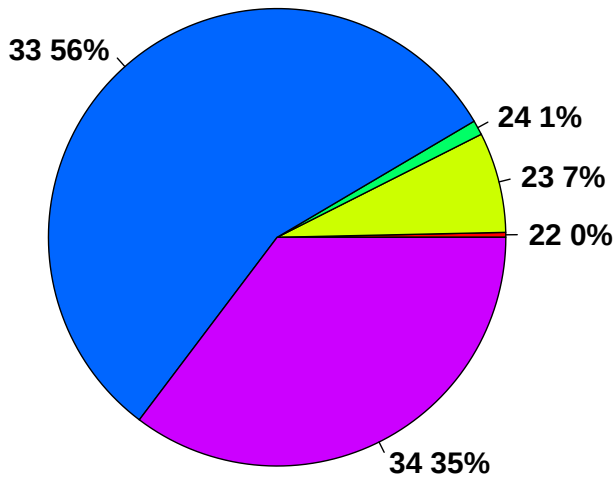
Alzheimer's Disease Sequencing Project



AD diagnosis for 584 people across 111 families. Results in this presentation are analyzed from 137 people across 32 families.

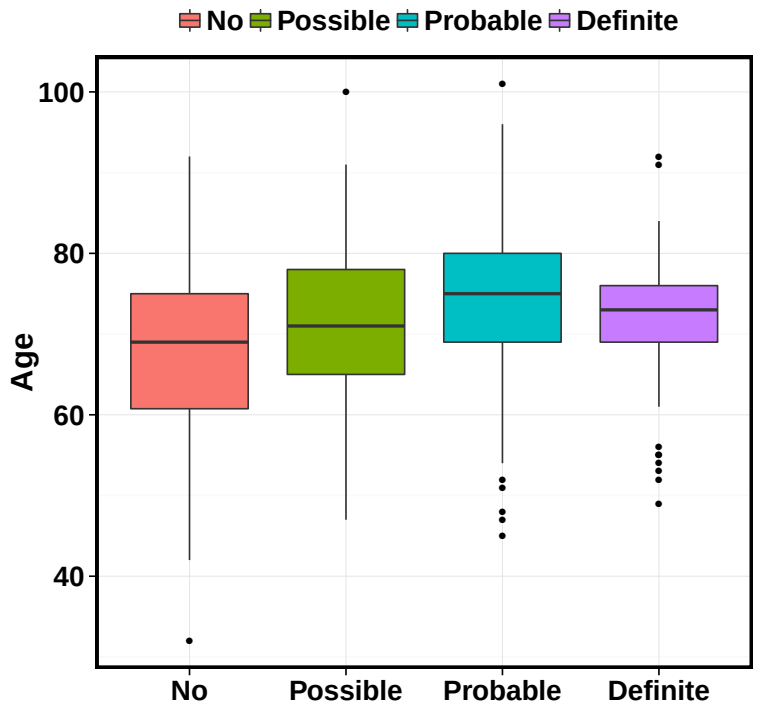
Alzheimer's Disease Sequencing Project

APOE (apolipoprotein E) on chromosome 19 makes protein that helps helps carry cholesterol in the bloodstream.

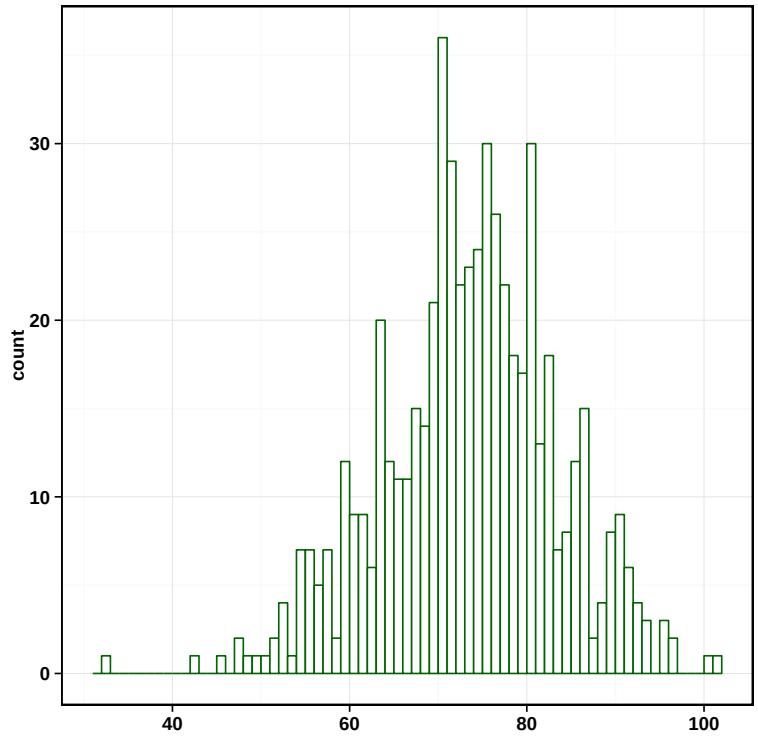


- APOE $\epsilon 2$ is rare and may protect against AD
- APOE $\epsilon 3$ is the most common allele and plays a neutral role
- APOE $\epsilon 4$ present in about 25 to 30 % of population and in about 40 % of people with LOAD.

Alzheimer's Disease Sequencing Project

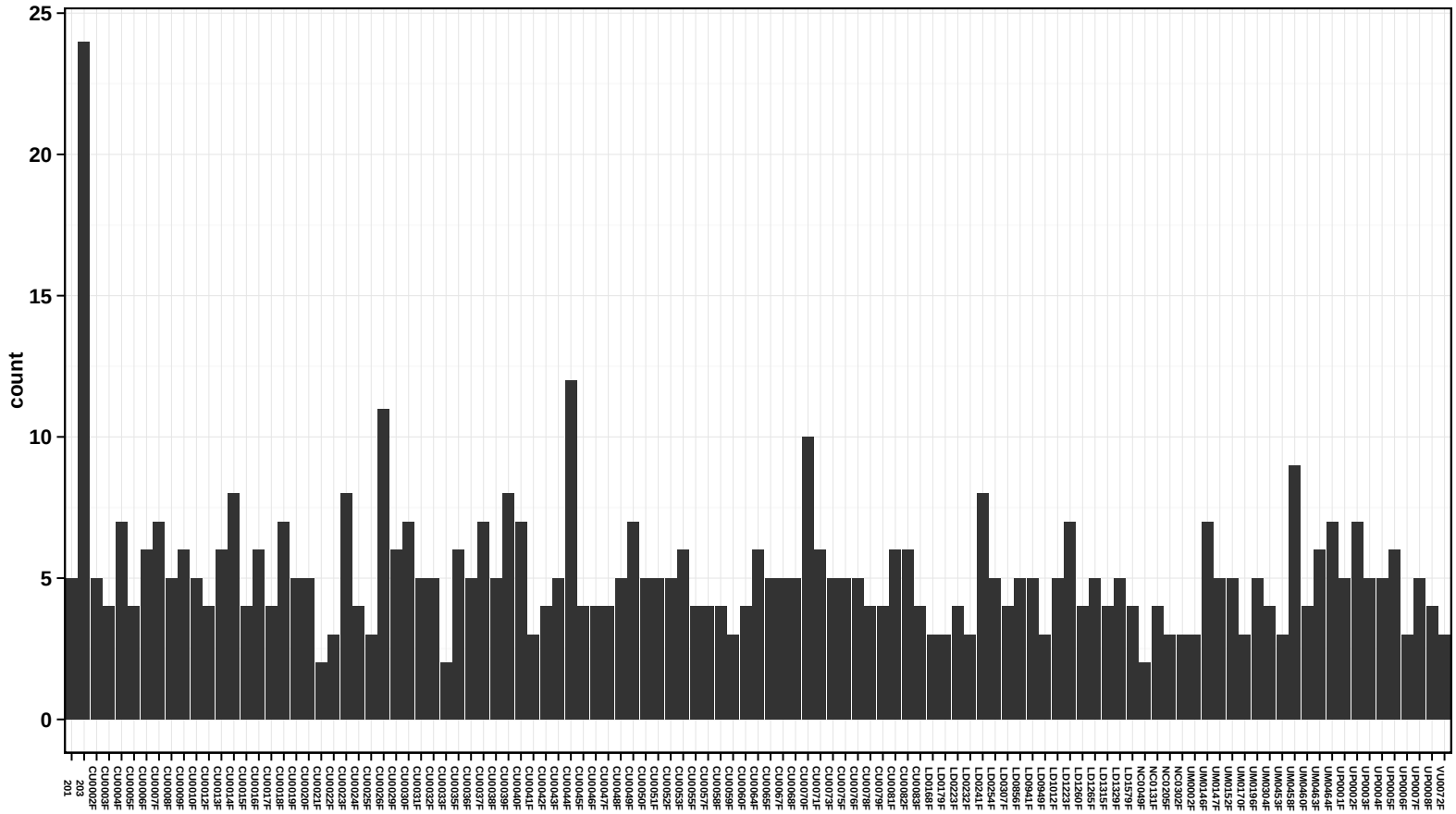


Age distribution by AD status.



Age distribution.

Alzheimer's Disease Sequencing Project



Sample number per family

What are the genomic variants that are significant associated with AD?

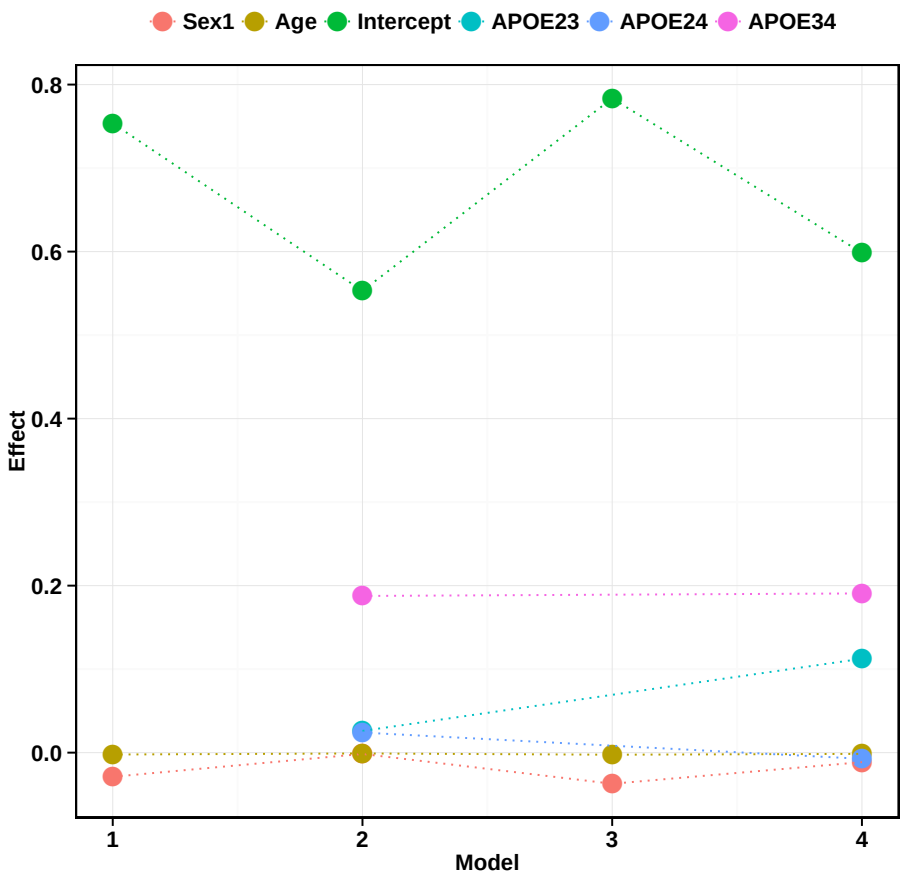
$$AD \sim Sex + Age + SNP \quad (1)$$

$$AD \sim Sex + Age + APOE + SNP \quad (2)$$

$$AD \sim Sex + Age + +SNP + (1|Family) \quad (3)$$

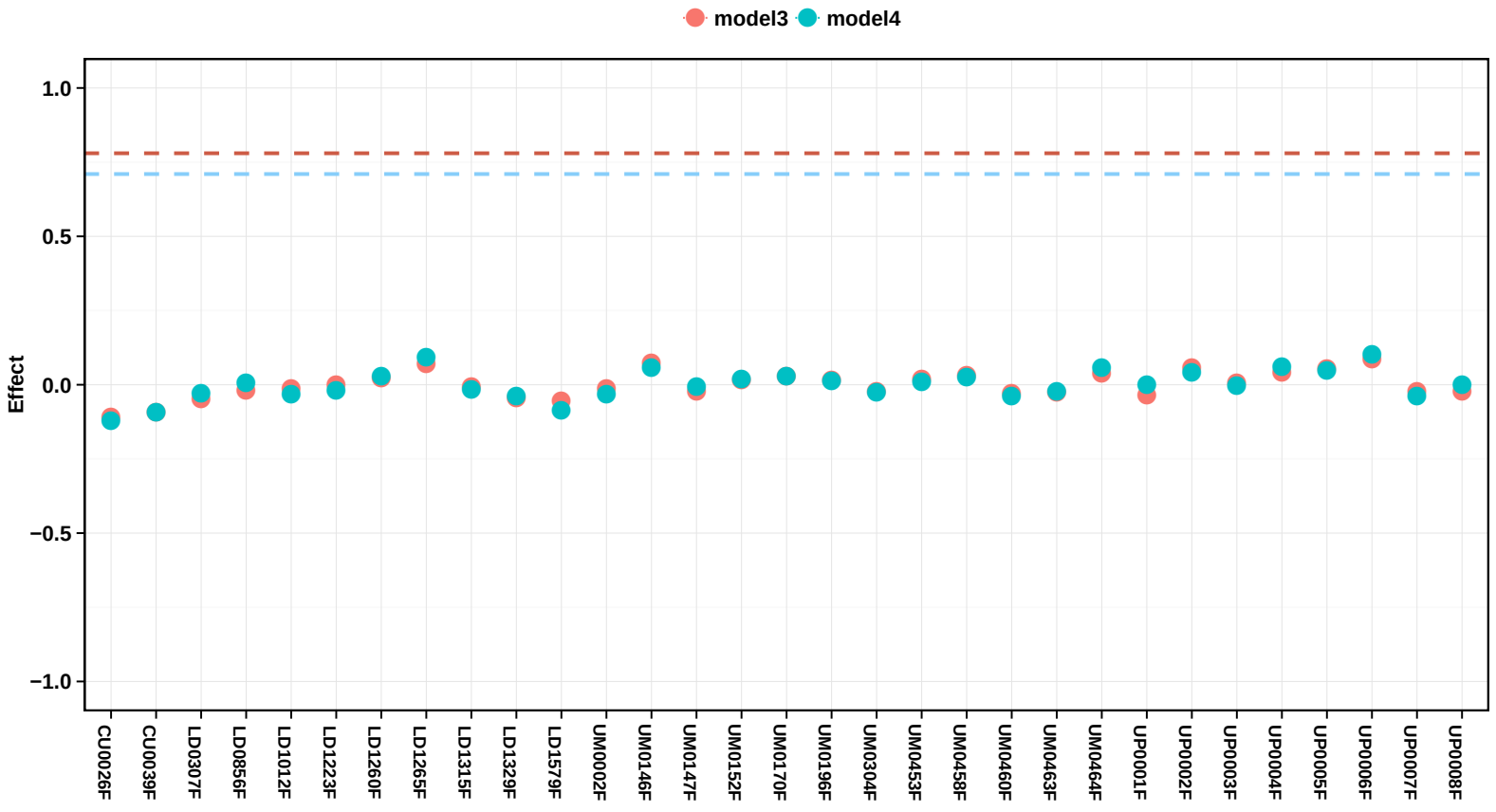
$$AD \sim Sex + Age + APOE + SNP + (1|Family) \quad (4)$$

Summary of Null Models



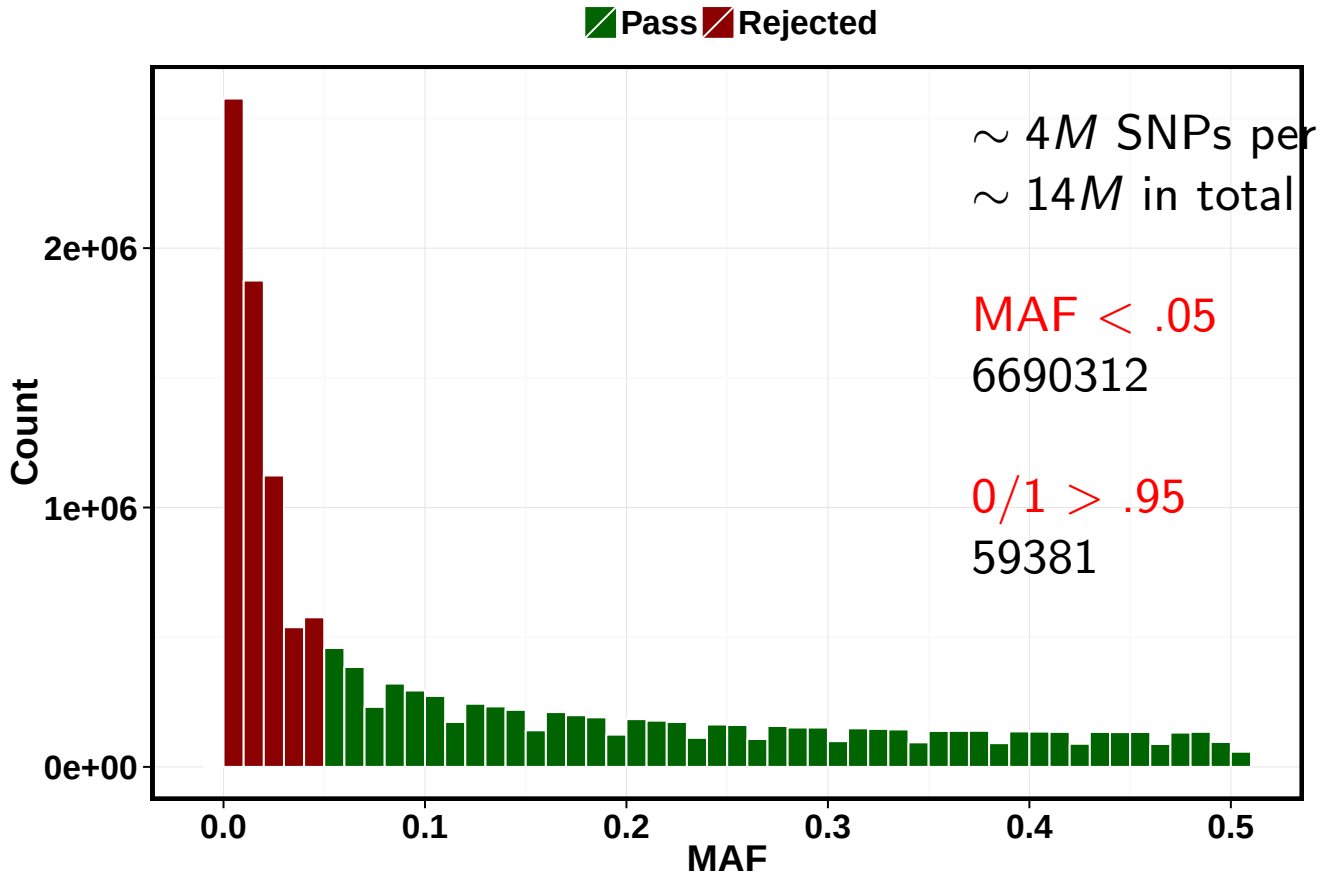
- Intercepts high. Most samples are probable (0.5).
- APOE $\epsilon 34$ drops intercepts, increases AD risk.
- APOE $\epsilon 24$ and $\epsilon 33$ decrease AD risk in model4.
- Age and Sex have little effects

Family level effects on intercepts



Family level effects on intercepts generally small.

Summary statistics for SNP



$\sim 4M$ SNPs per sample

$\sim 14M$ in total

MAF $< .05$

6690312 7820739

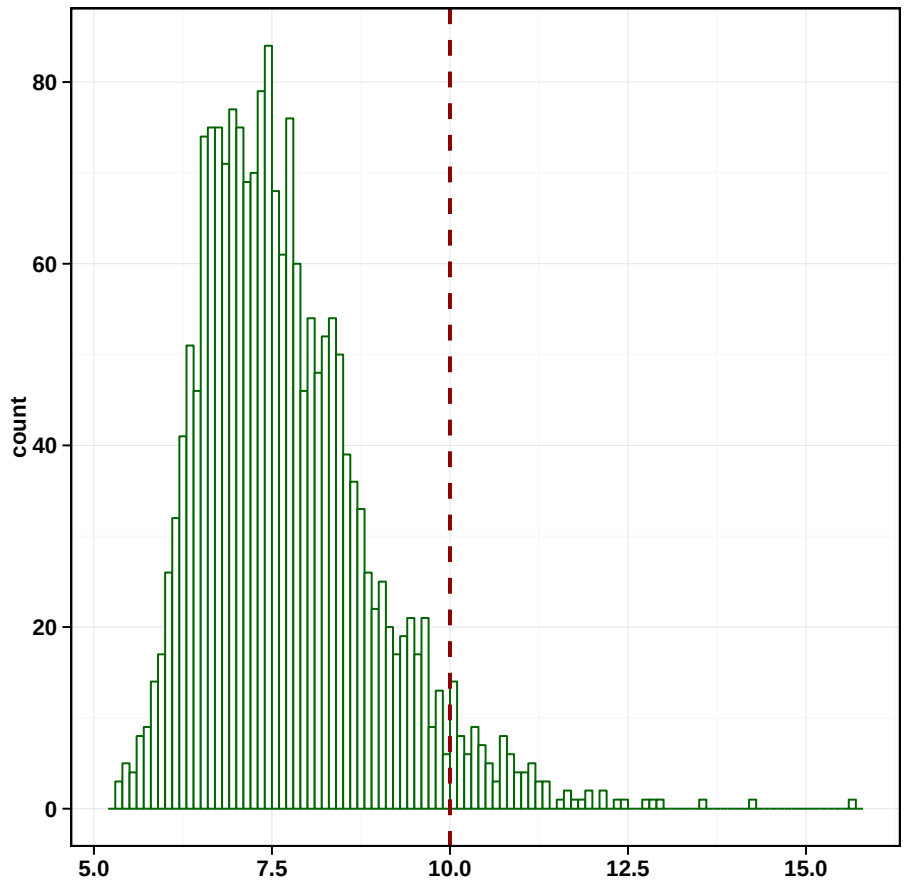
$0/1 > .95$

59381 7743948

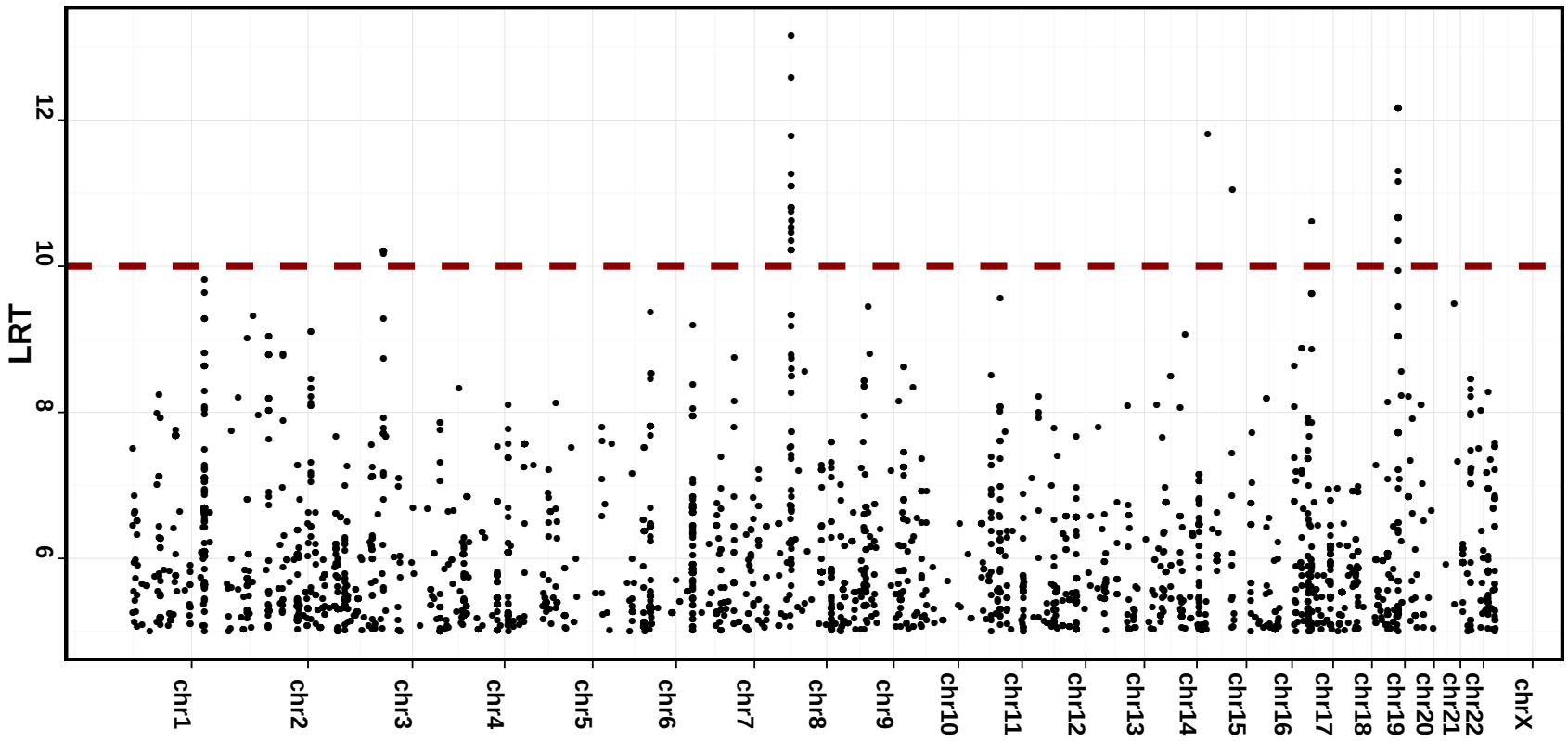
Permutation

$$AD \sim Sex + Age + APOE + p(SNP) + (1|Family)$$

1. Randomly pick 1M SNP.
2. Run full model on permuted SNP.
3. Save maximal LRT value.
4. Repeat 1-3 2000 times
5. Set threshold as .95 quantile of maximal LRTs

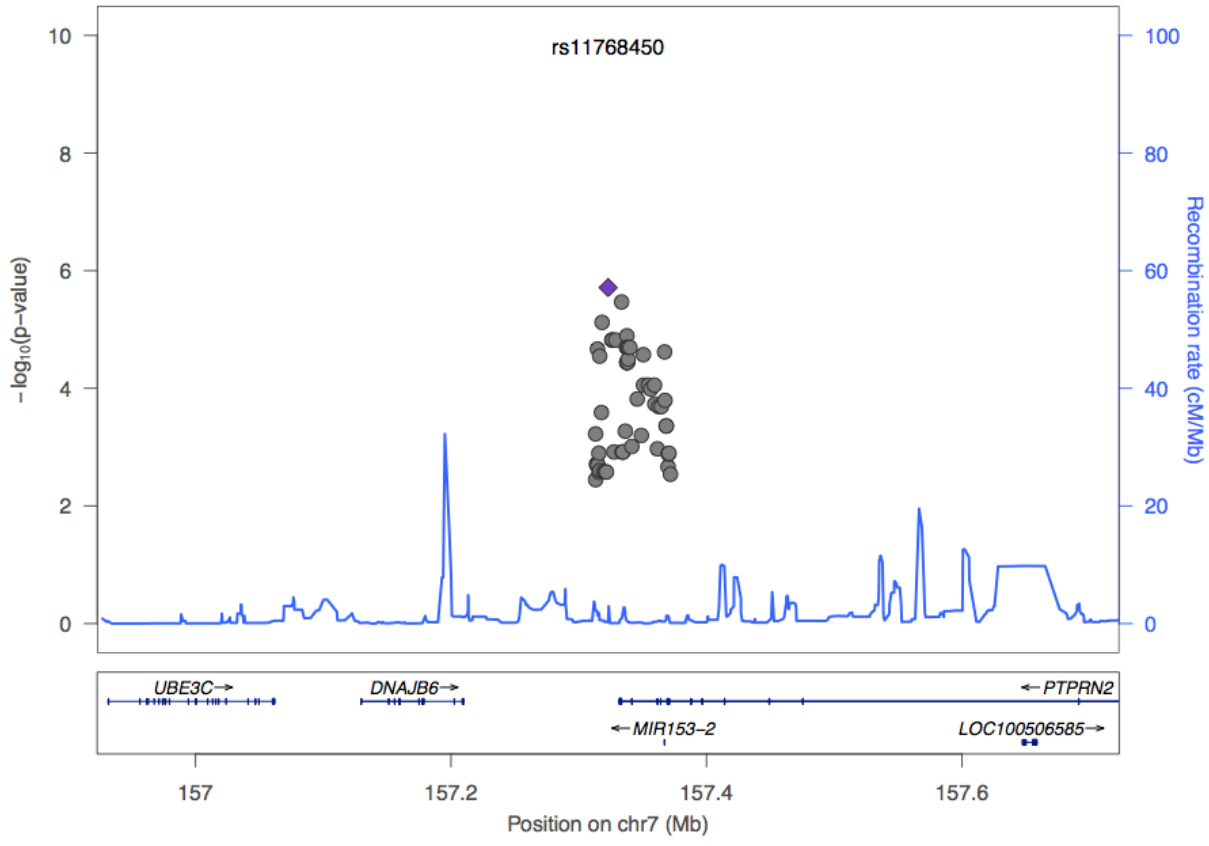


LRT line



$$AD \sim Sex + Age + APOE + SNP + (1|Family)$$

LocusZoom



$$AD \sim Sex + Age + APOE + SNP + (1|Family)$$

To do list

- Run models 1-4 on full data
- Bayesian inference on multilevel ordered categorical model
- Pair-wise regression