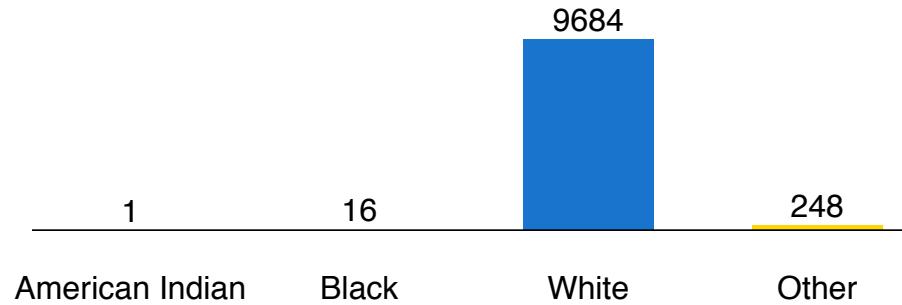


ADSP project

ADSP WES data: 9949 samples by 3 sequencing centers

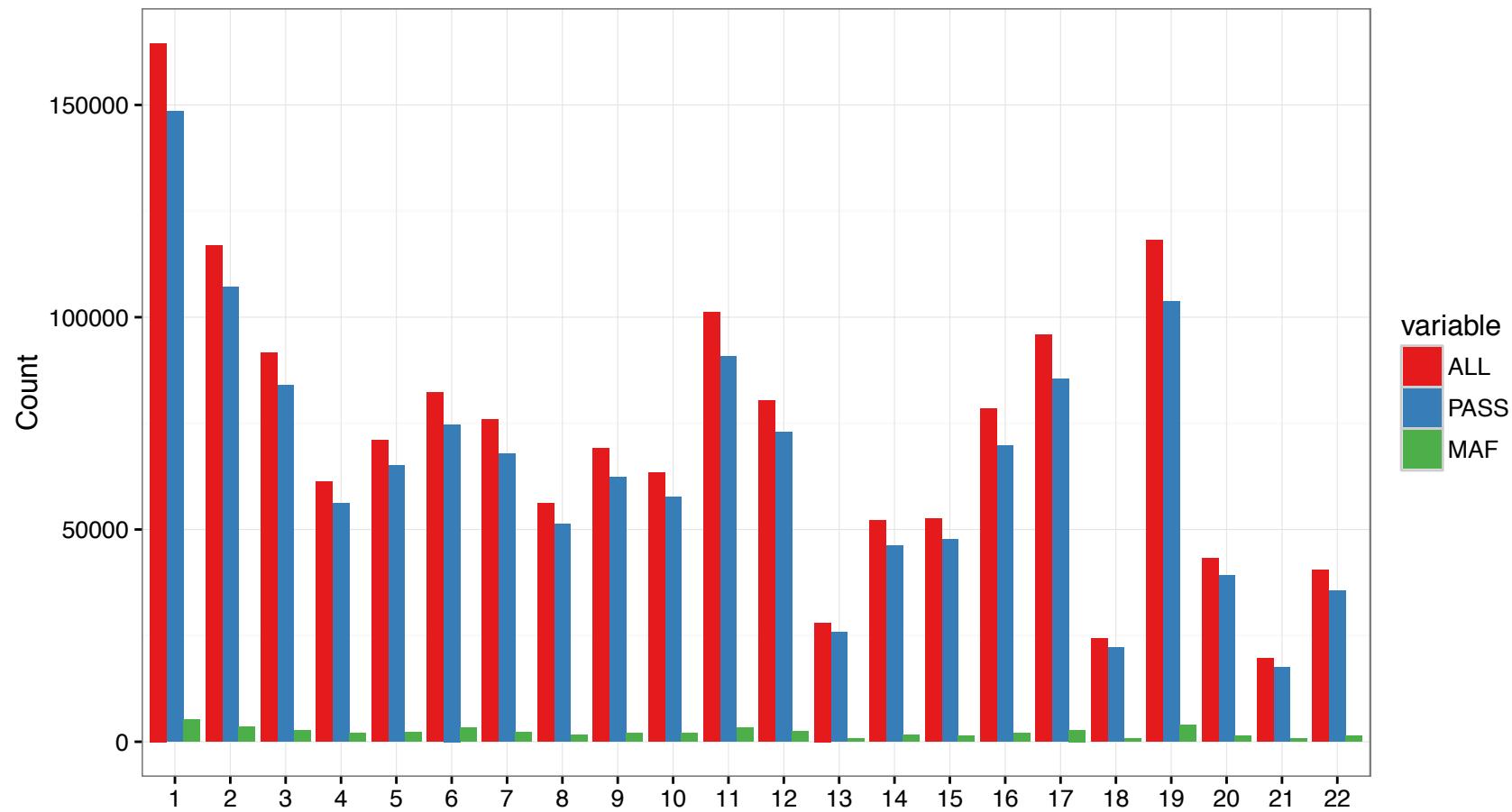


- 10 of the 9949 samples were sequenced independently by the 3 centers. This led to 20 redundant sequencing results, which were removed.
- The 9949 samples were each diagnosed as case or control in Alzheimer's disease status.
- The 9949 samples came from two groups: case control, and enriched. The 816 enriched samples had 645 case and 171 controls.

Variants by ADSP consortium

```
##FILTER=<ID=DISTINCT_REF_ALT,Description="FAIL: present in both Broad and Baylor, but alleles differ">
##FILTER=<ID=BOTH_failed,Description="FAIL: failed in both Broad and Baylor">
##FILTER=<ID=BOTH_passed,Description="PASS: passed in both Broad and Baylor">
##FILTER=<ID=BAYLOR_failed_BROAD_uncalled,Description="PASS: present in Broad, passed; uncalled in Baylor">
##FILTER=<ID=BAYLOR_passed_BROAD_uncalled,Description="PASS: present in Baylor, passed; uncalled in Broad">
##FILTER=<ID=BROAD_failed_BAYLOR_passed,Description="PASS: present in both, failed in Broad, passed in Baylor">
##FILTER=<ID=BROAD_passed_BAYLOR_failed,Description="PASS: present in both, passed in Broad, failed in Baylor">
##FILTER=<ID=BROAD_passed_BAYLOR_uncalled,Description="PASS: present in Broad, passed; uncalled in Baylor">
##FILTER=<ID=BROAD_failed_BAYLOR_uncalled,Description="FAIL: present in Broad, failed; uncalled in Baylor">
```

Variants by ADSP consortium



- ALL: 1,586,687; PASS: 1,431,108; MAF: 49,235
- PASS/ALL: 90.2%; MAF/PASS: 3.4%

Variants by ADSP consortium

Category	Count
Variants processed	49235
Variants remaining after filtering	49235
Novel / existing variants	316 (0.6%) / 48919 (99.4%)
Overlapped genes	20512
Overlapped transcripts	108678
Overlapped regulatory features	8297

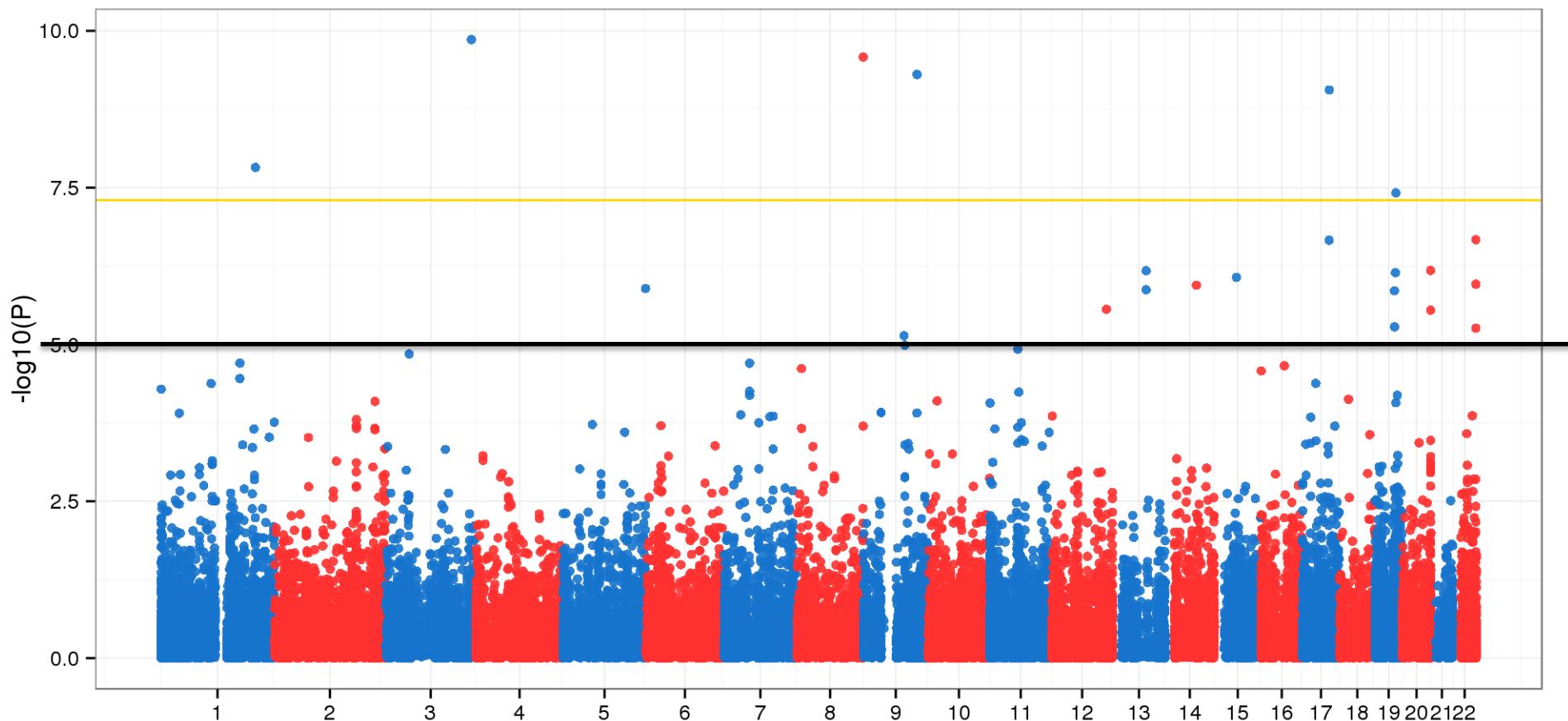
Consequences (all)



- synonymous_variant: 16%
- downstream_gene_variant: 15%
- missense_variant: 14%
- intron_variant: 11%
- non_coding_transcript_variant: 11%
- upstream_gene_variant: 10%
- non_coding_transcript_exon_variant
- NMD_transcript_variant: 5%
- splice_region_variant: 4%
- Others

GWAS with all samples

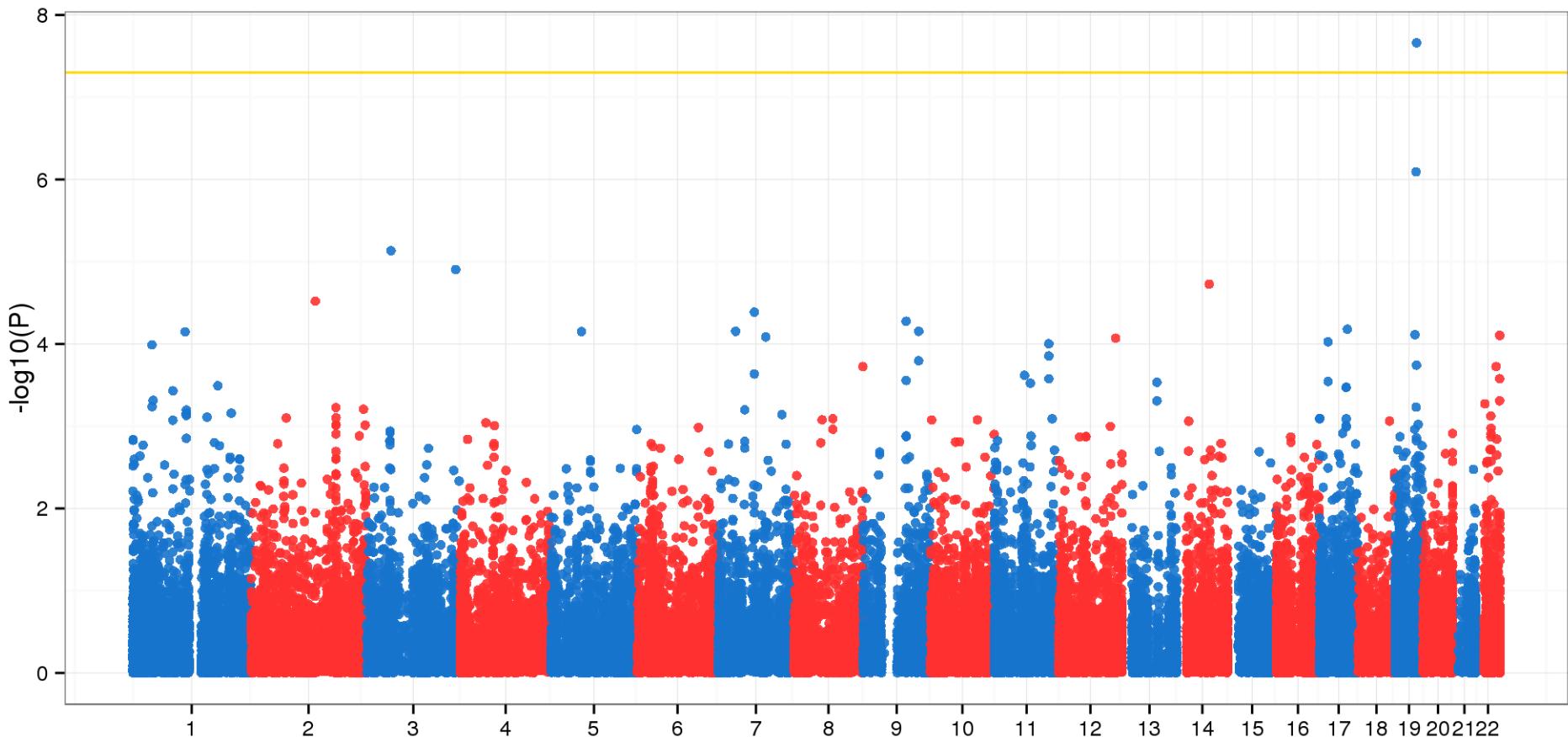
Method: GLMM in GMMAT with covariates: Age, Sex



- Chen et al., AJHG 2016 used penalized quasi-likelihood method as proposed by Breslow and Clayton 1993
- 22 above 5

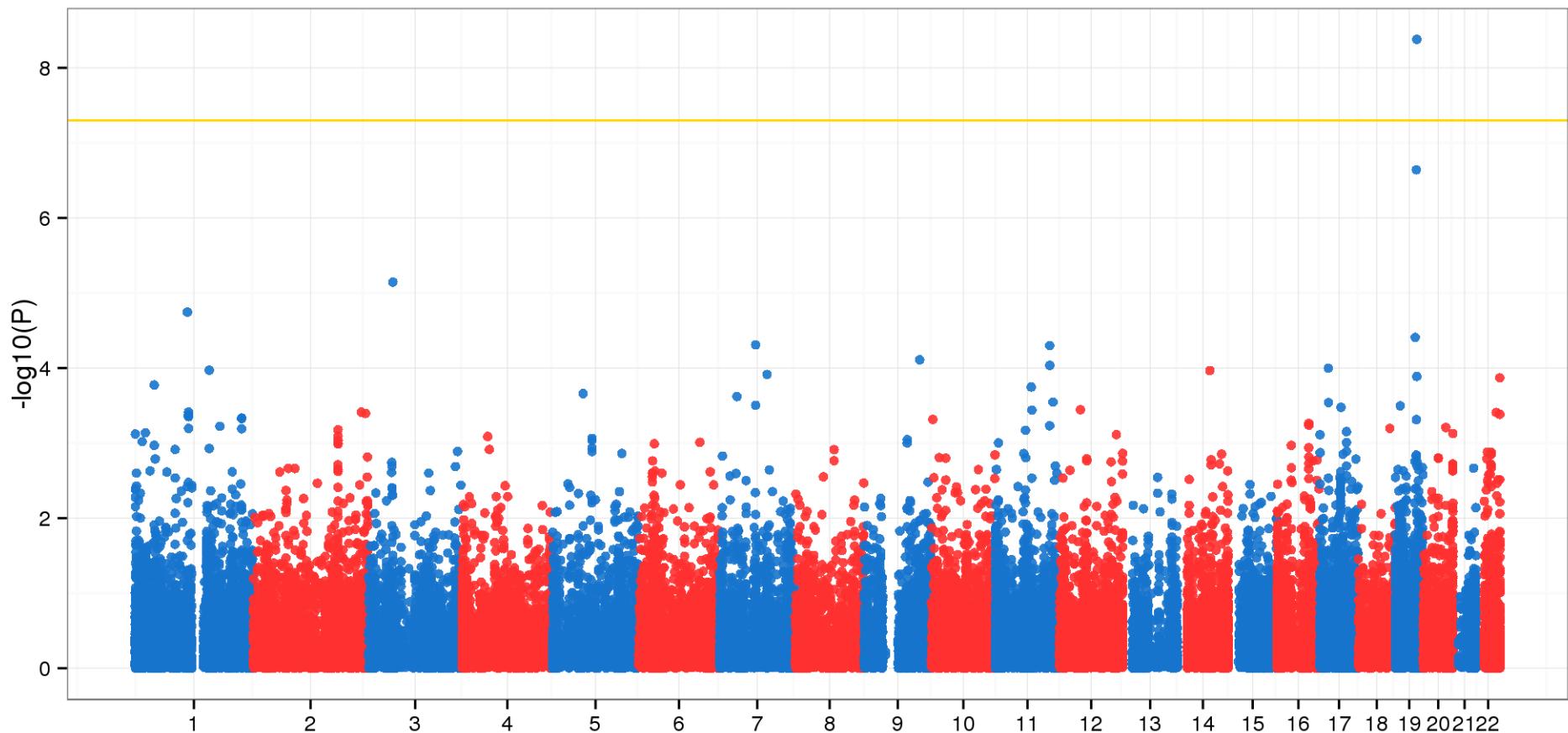
GWAS with all samples

Method: GLMM in GMMAT with covariates: Age, Sex, PC1



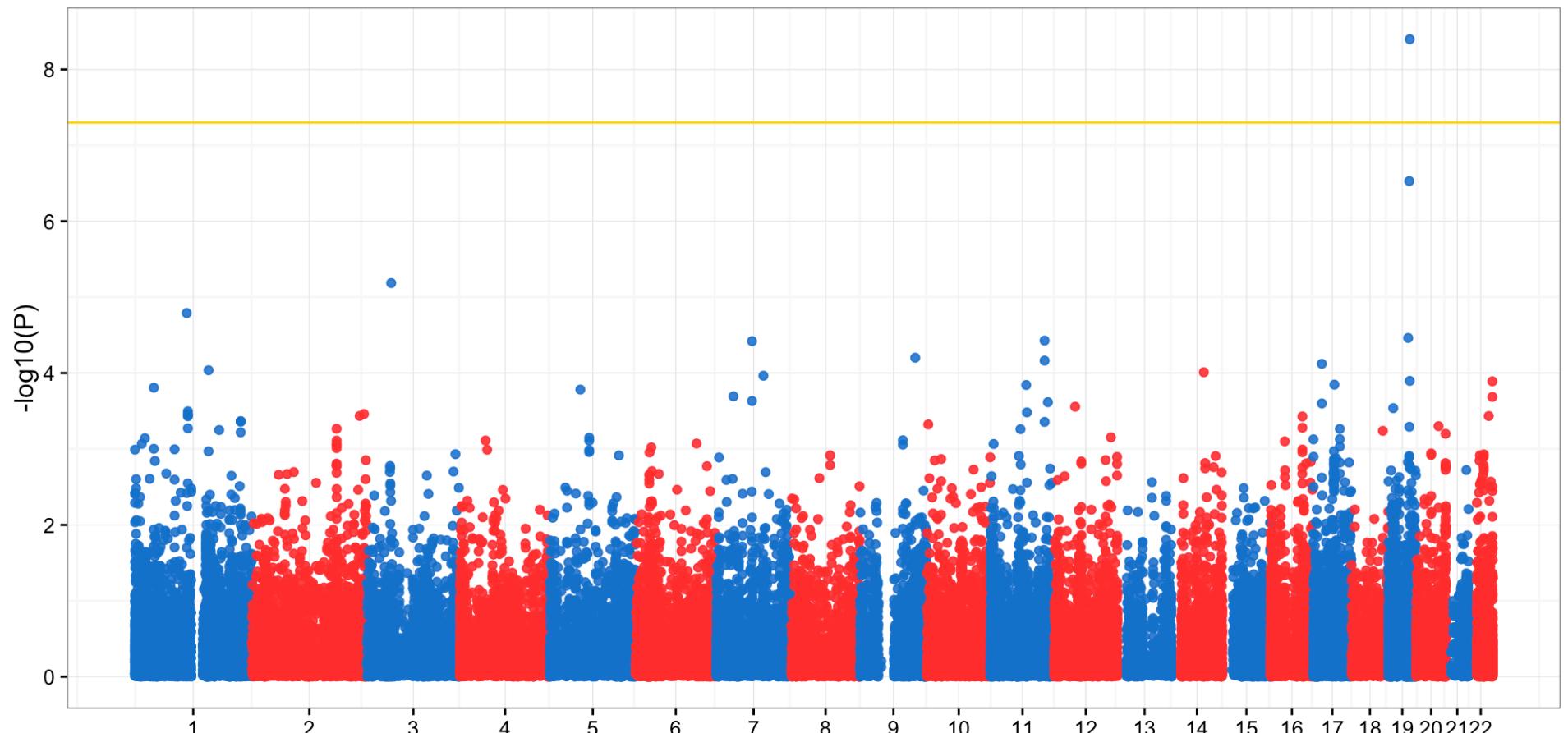
GWAS with all samples

Method: GLMM in GMMAT with covariates: Age, Sex, PC1, PC2



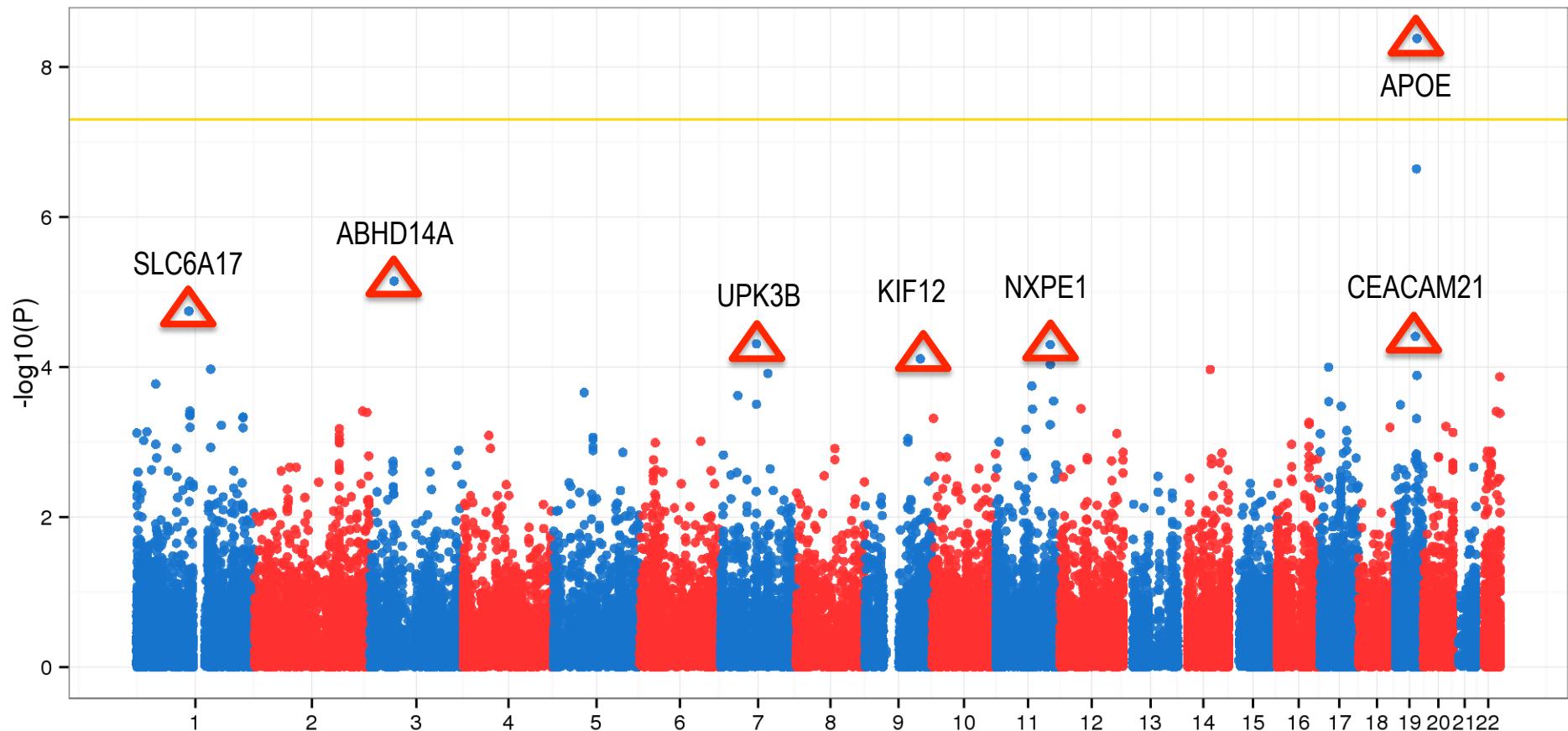
GWAS with all samples

Method: Logistic regression in Plink with covariates: Age, Sex, PC1, PC2

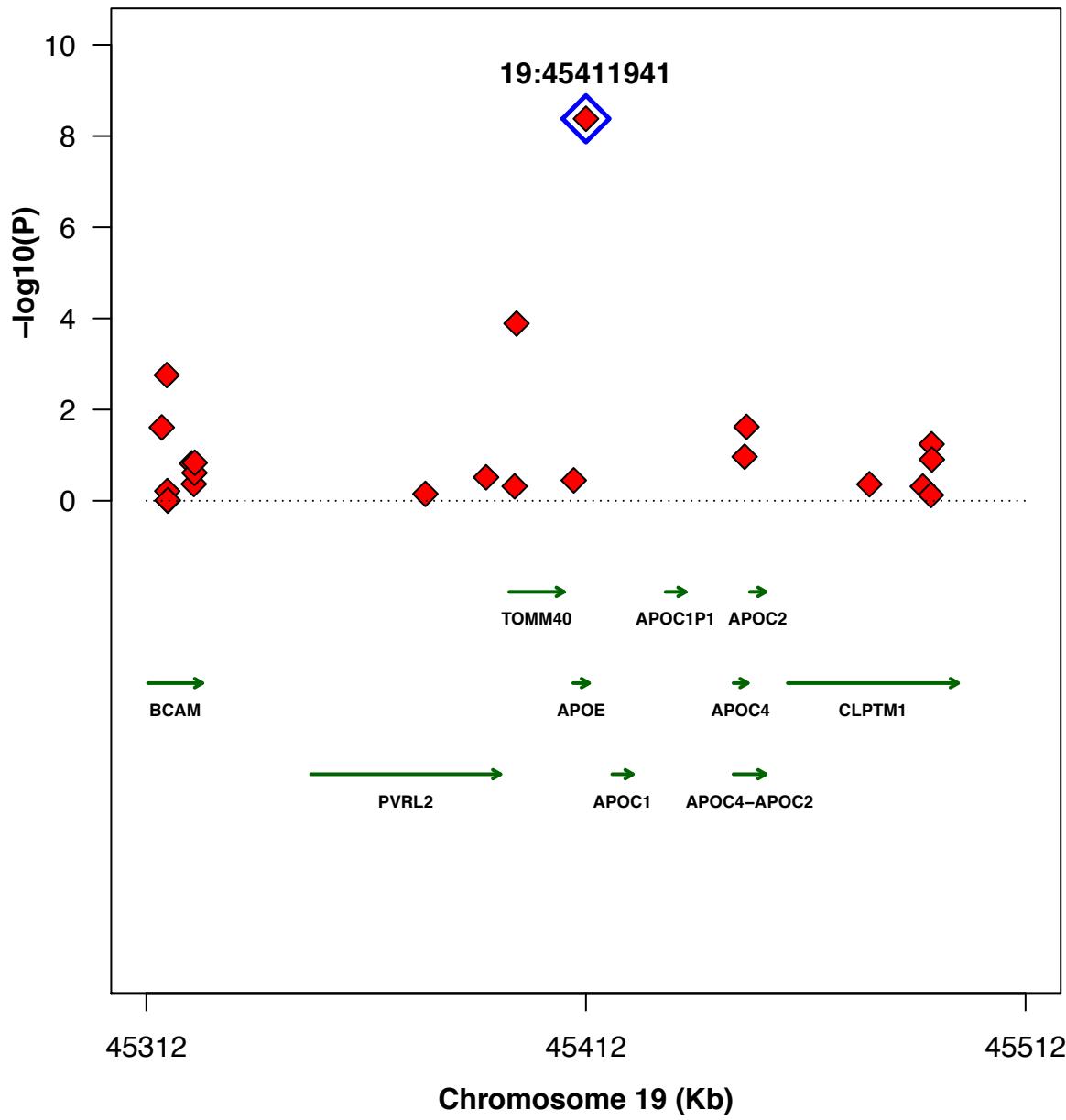


- PC were sufficient to correct the population stratifications.
- PC corrected more, some latent factors.

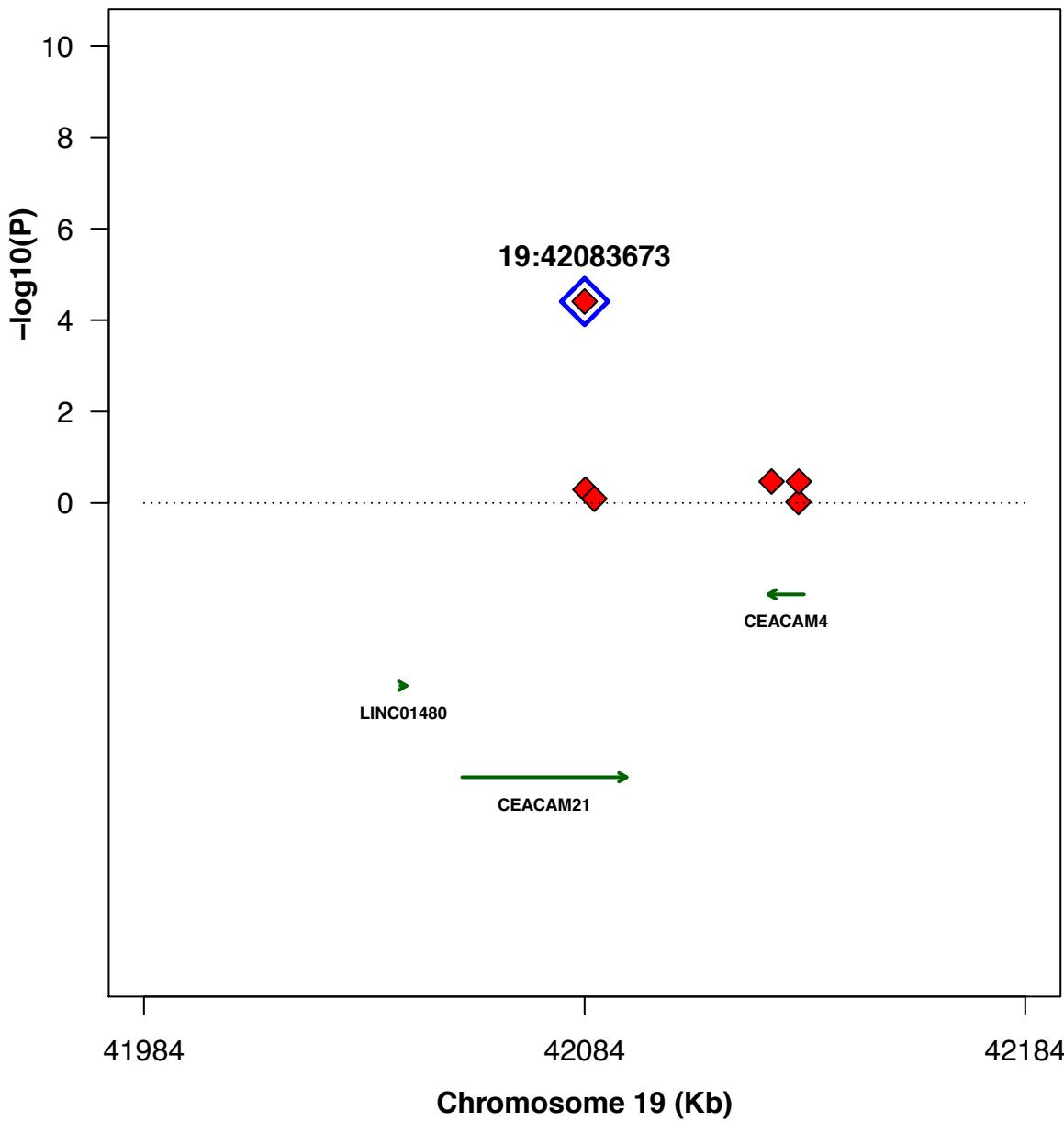
9 variants in 7 loci ($P < 1e-4$)



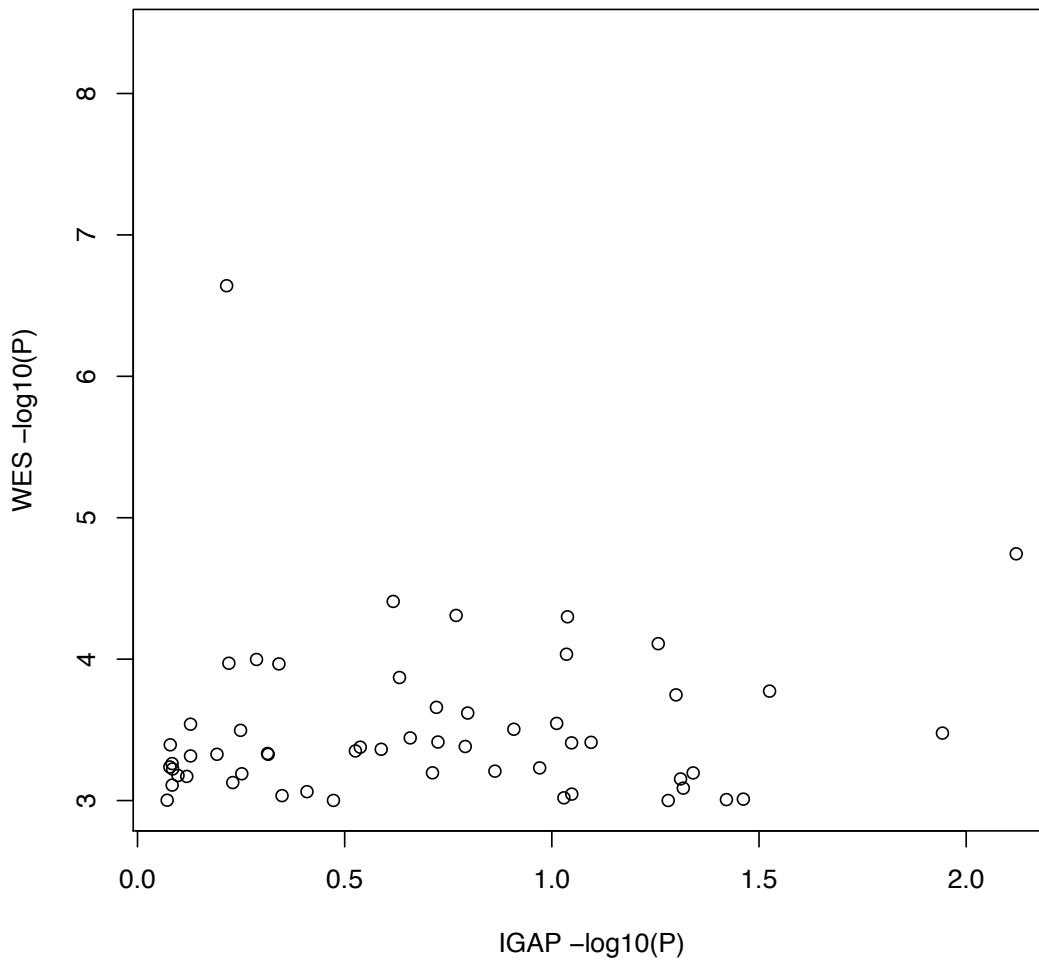
APOE locus



400 Kb away from APOE locus



67 variants in 45 loci ($P < 1e-3$)

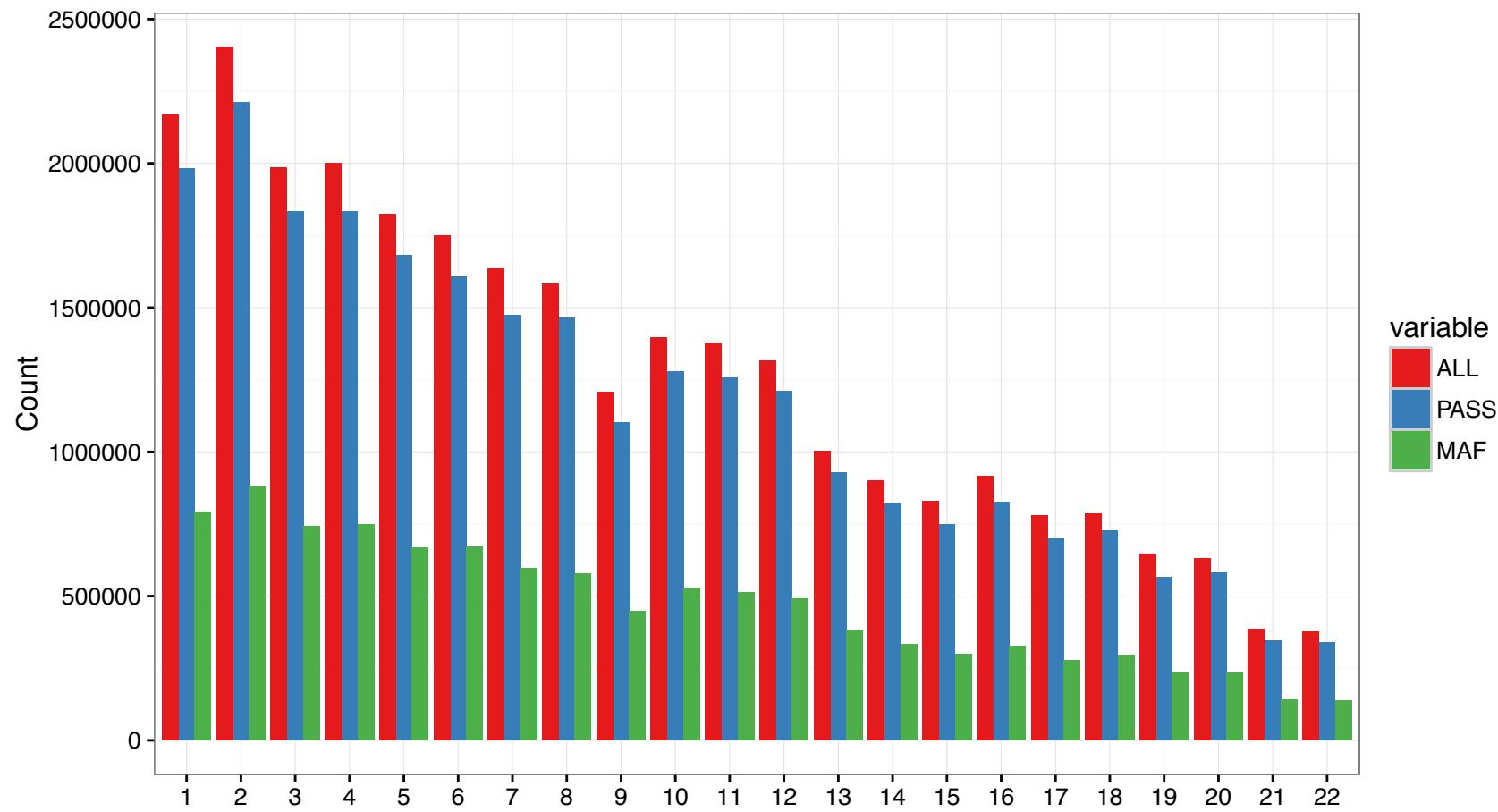


- 58 of the 67 had IGAP results, correlations were poor
- A table of the 67 variants available at the ADSP shared folder ([./Xulong/04May2016](#))

Variants by ADSP consortium: WGS also showed center-specific effects?

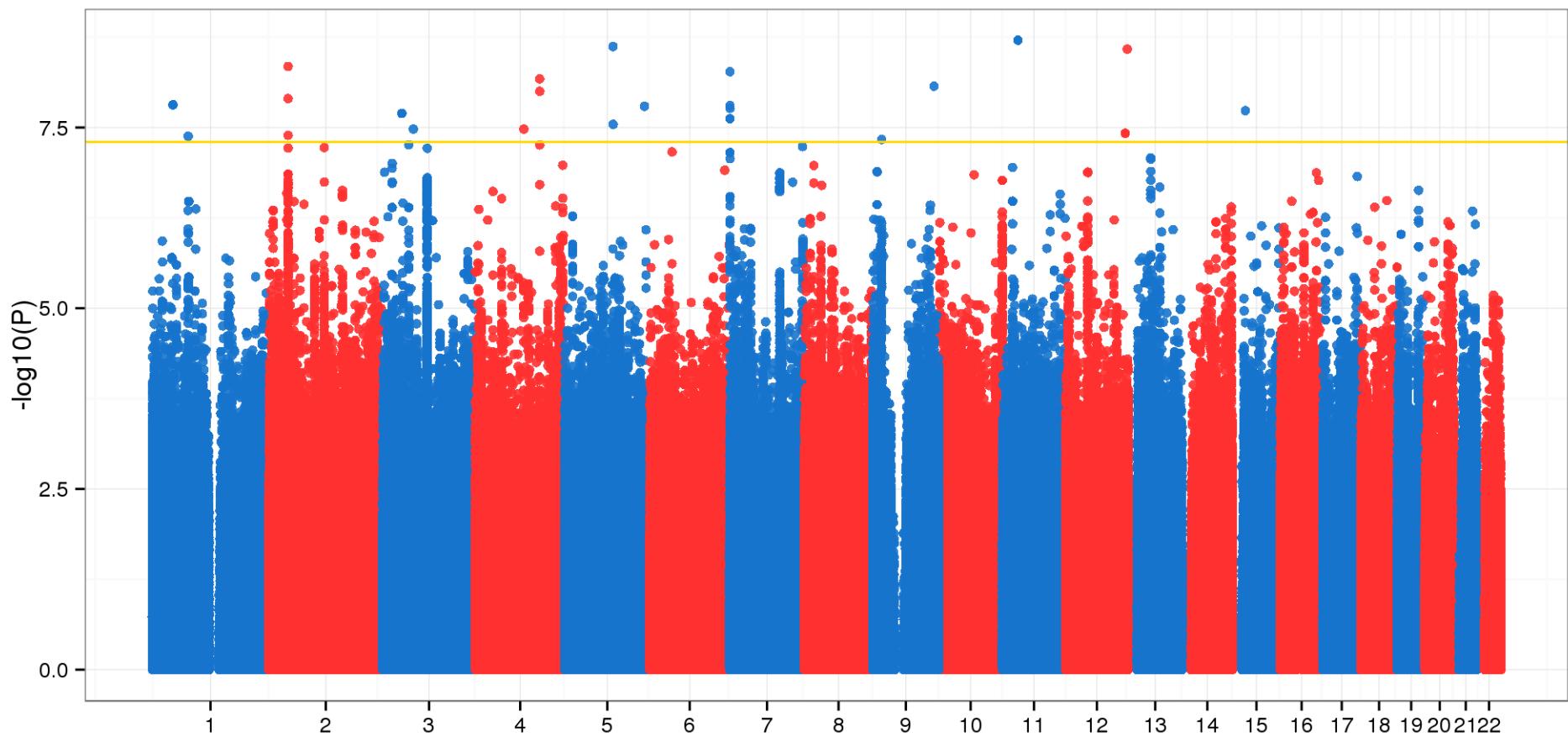
```
##FILTER=<ID=DISTINCT_REF_ALT,Description="FAIL: present in both Broad and Baylor, but alleles differ">
##FILTER=<ID=BOTH_failed,Description="FAIL: failed in both Broad and Baylor">
##FILTER=<ID=BOTH_passed,Description="PASS: passed in both Broad and Baylor">
##FILTER=<ID=BAYLOR_failed_BROAD_uncalled,Description="PASS: present in Broad, passed; uncalled in Baylor">
##FILTER=<ID=BAYLOR_passed_BROAD_uncalled,Description="PASS: present in Baylor, passed; uncalled in Broad">
##FILTER=<ID=BROAD_failed_BAYLOR_passed,Description="PASS: present in both, failed in Broad, passed in Baylor">
##FILTER=<ID=BROAD_passed_BAYLOR_failed,Description="PASS: present in both, passed in Broad, failed in Baylor">
##FILTER=<ID=BROAD_passed_BAYLOR_uncalled,Description="PASS: present in Broad, passed; uncalled in Baylor">
##FILTER=<ID=BROAD_failed_BAYLOR_uncalled,Description="FAIL: present in Broad, failed; uncalled in Baylor">
```

Variants by ADSP consortium

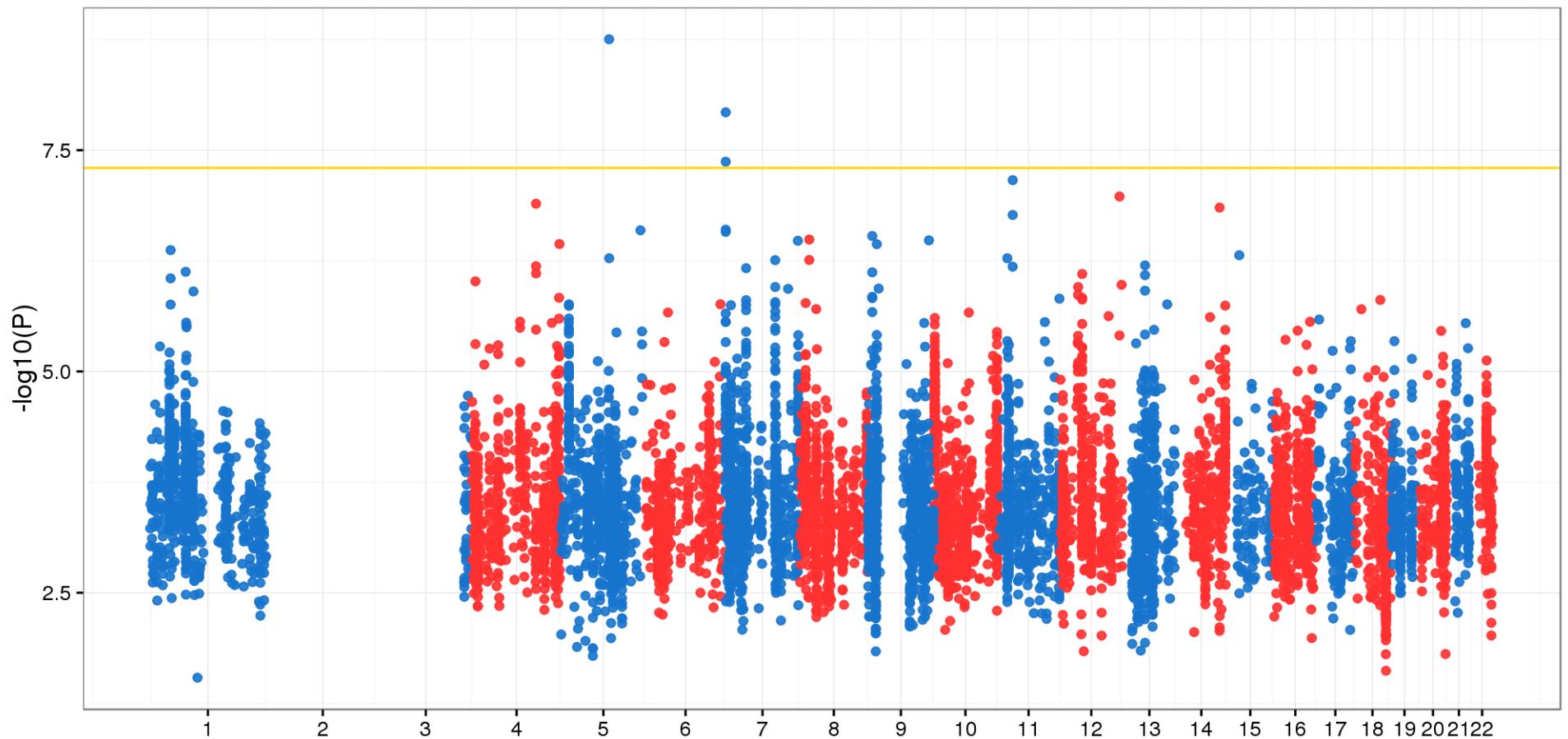


- ALL: 27.9 million; PASS: 25.5 million; MAF: 10.3 million
- ADSP released VCF for 578 out of 584 samples, 570 out of the 578 had disease status

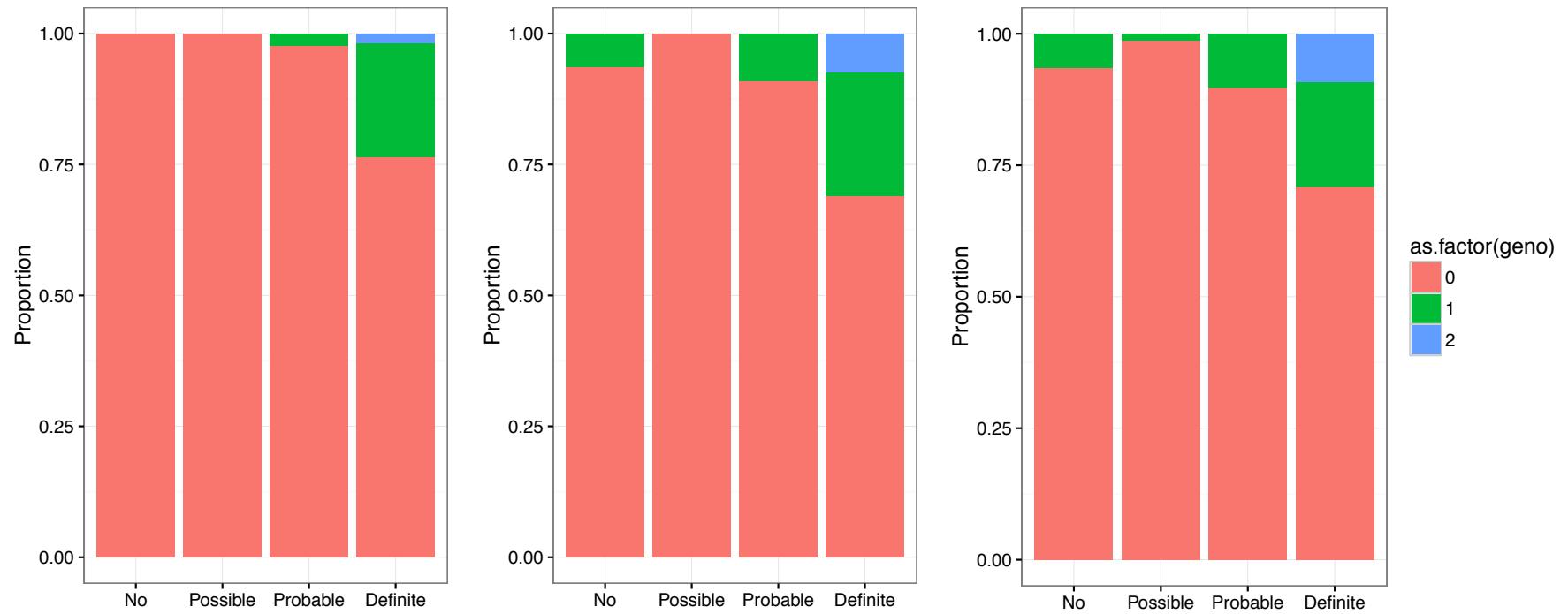
GWAS of ADSP WGS using ADSP's variants
Method: BAYES.GLMM without kinship correction
Covariates: Age, Sex



GWAS of ADSP WGS using CS's variants
Method: BAYES.GLMM with kinship correction
Covariates: Age, Sex

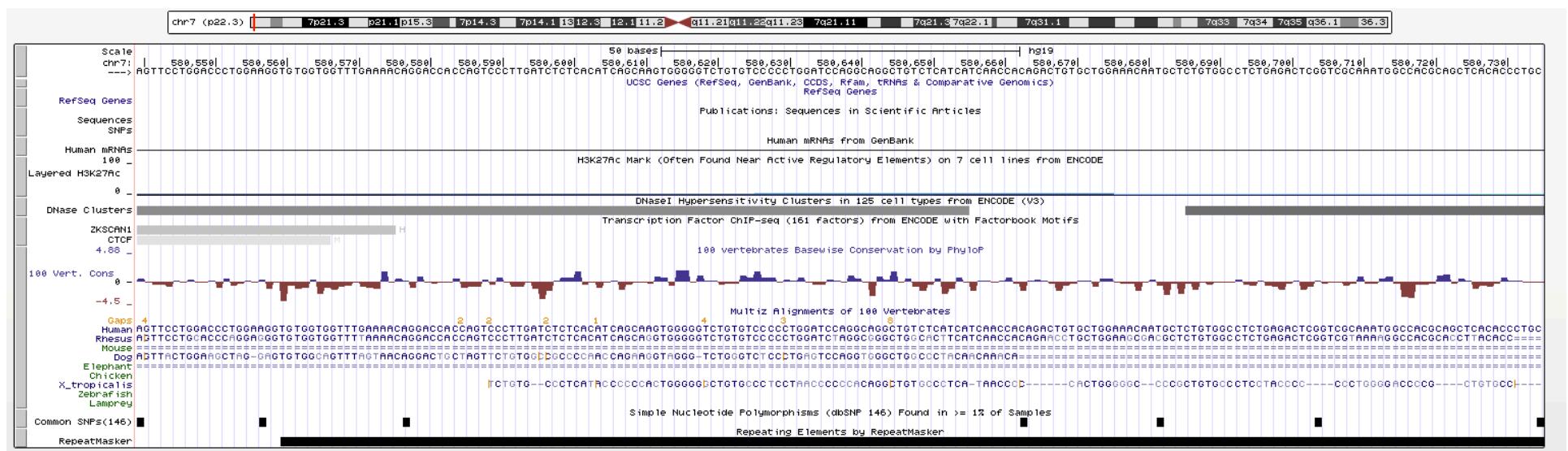


3 significant variants in 2 loci, all intergenic

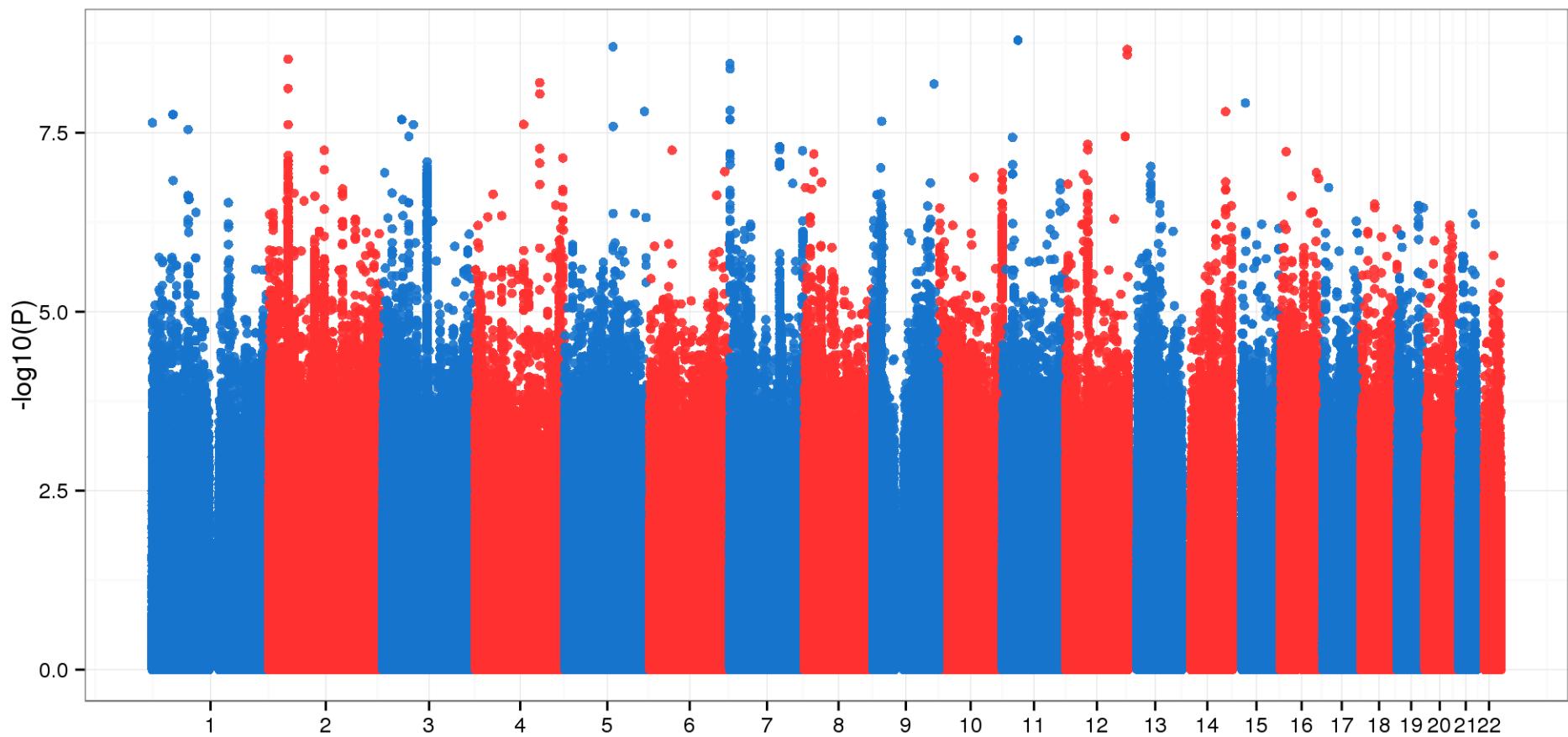


- MAF 0.02 (5:102726073_C/T) and 0.05 (7:580540_A/G, 7:580735_C/A)

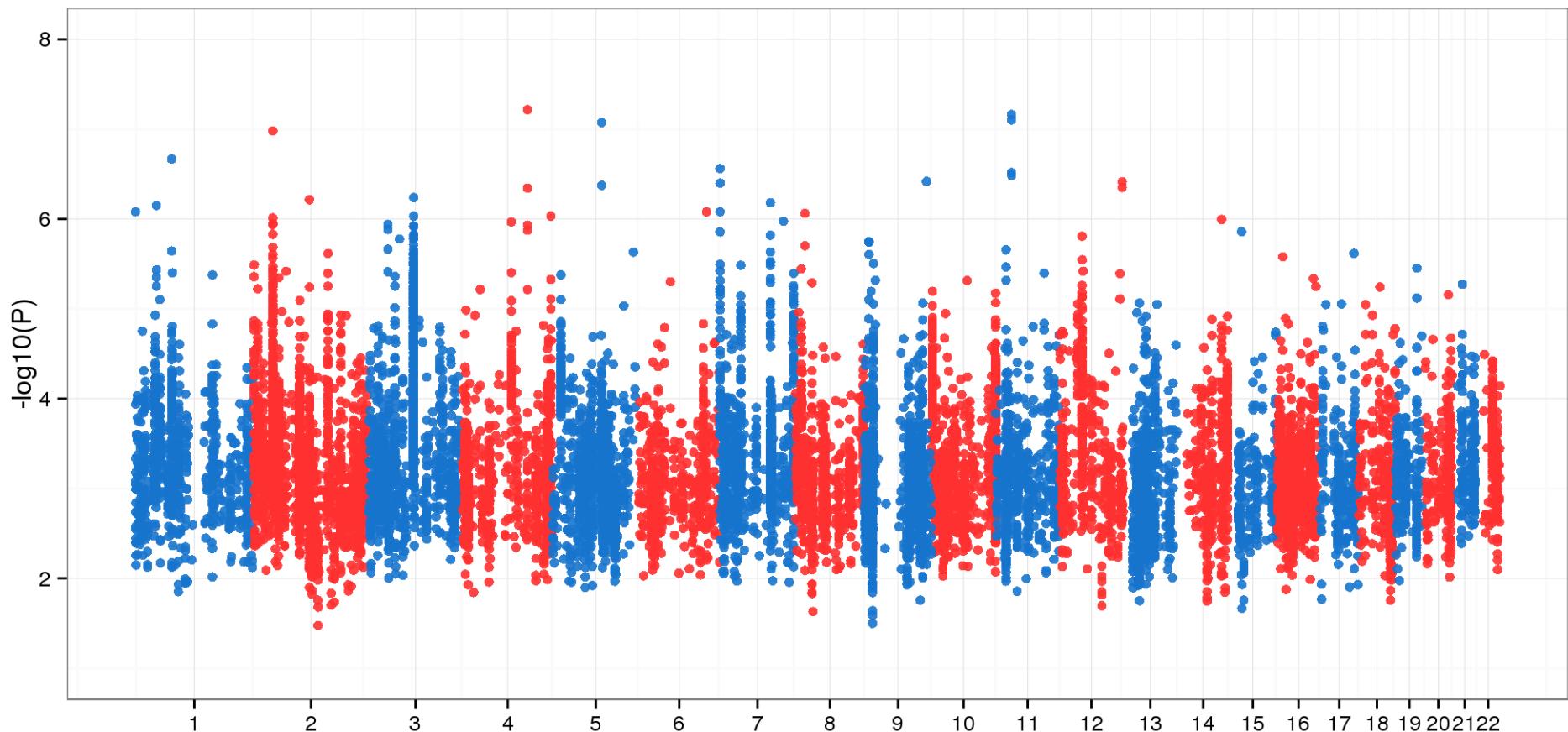
chr7:580540 and chr7:580735



GWAS of ADSP WGS using CS's variants
Method: BAYES.GLMM without kinship correction
Covariates: Age, Sex



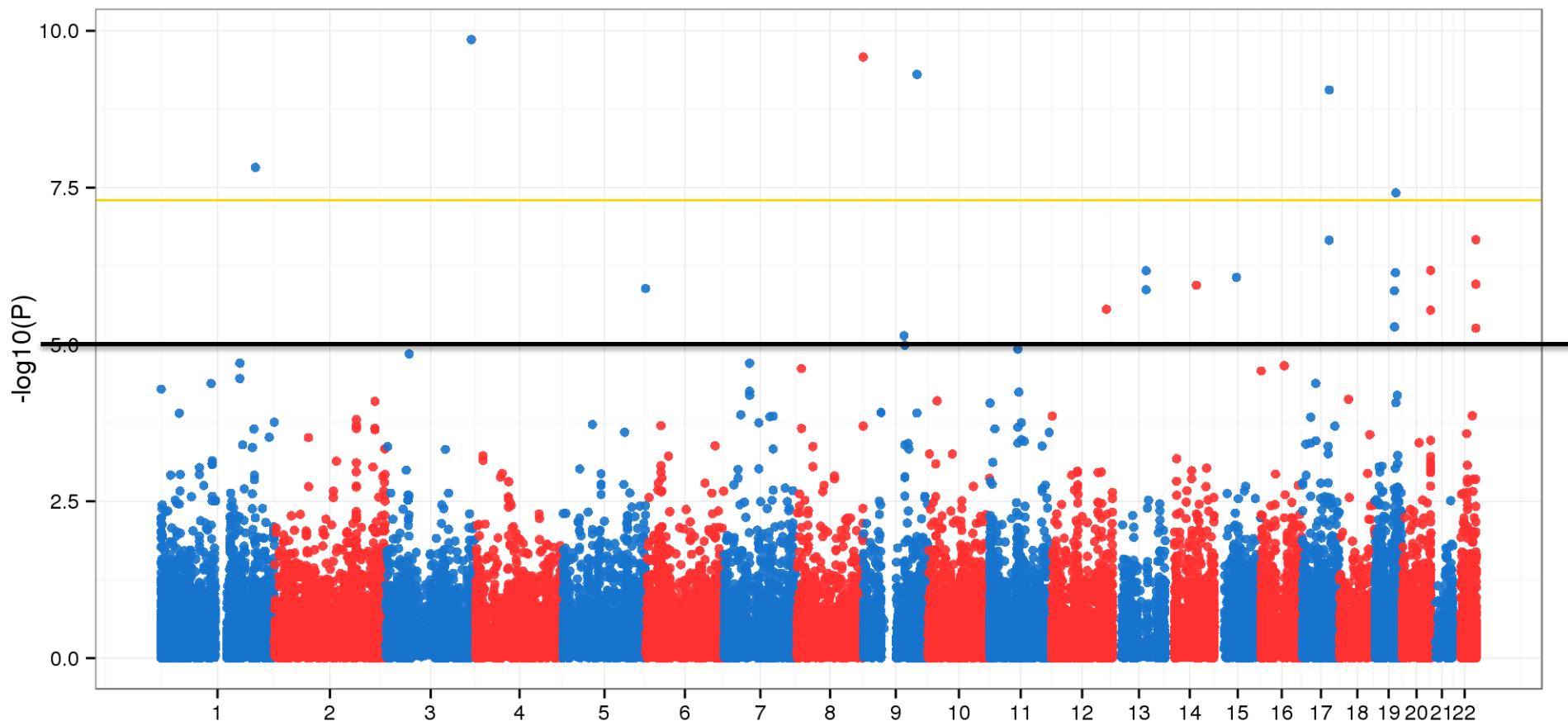
GWAS of ADSP WGS using ADSP's variants
Method: BAYES.GLMM with kinship correction
Covariates: Age, Sex



- 570 vs 576 samples
- Kinship matrices were different

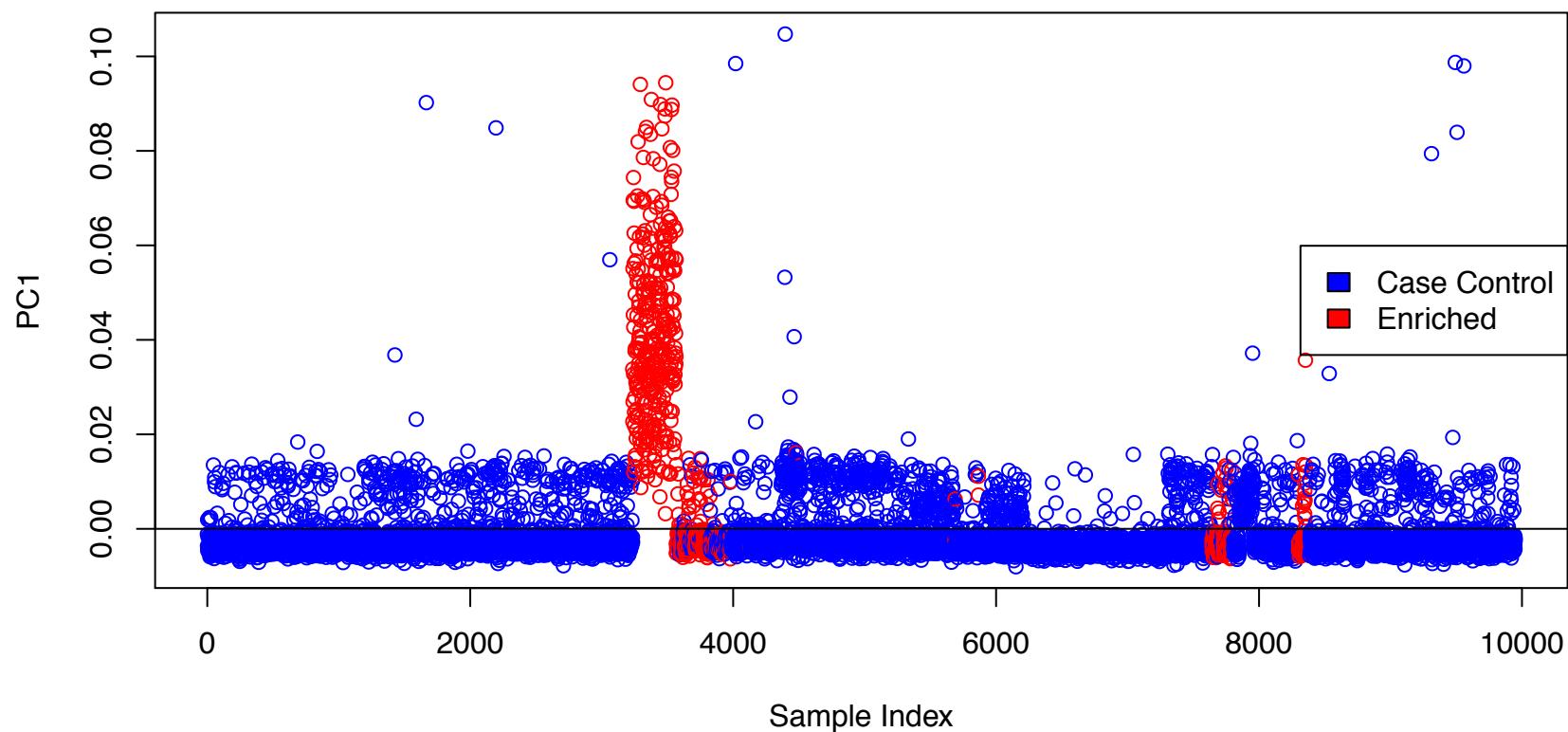
GWAS with all samples

Method: GLMM in GMMAT with covariates: Age, Sex

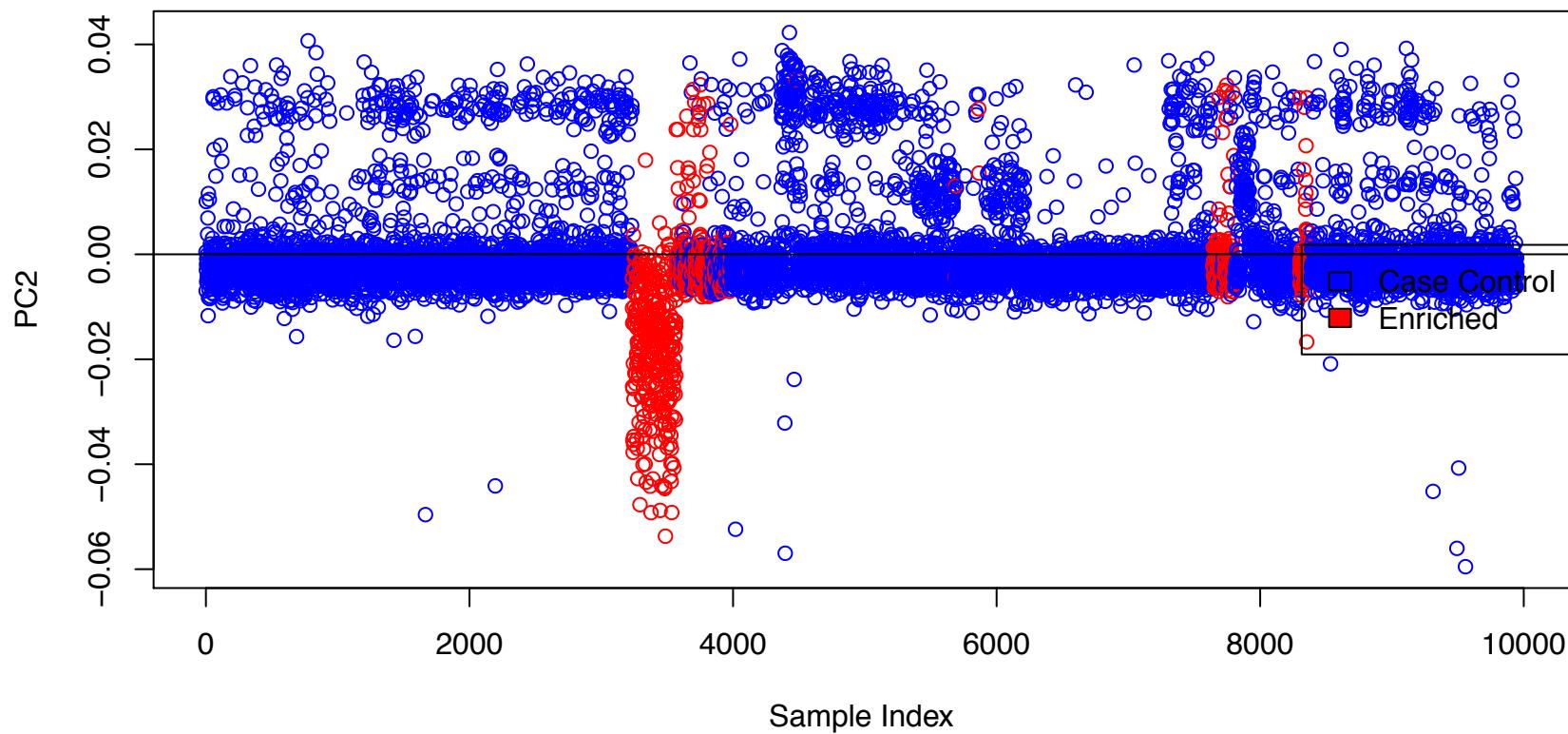


- Chen et al., AJHG 2016 used penalized quasi-likelihood method as proposed by Breslow and Clayton 1993
- 22 above 5

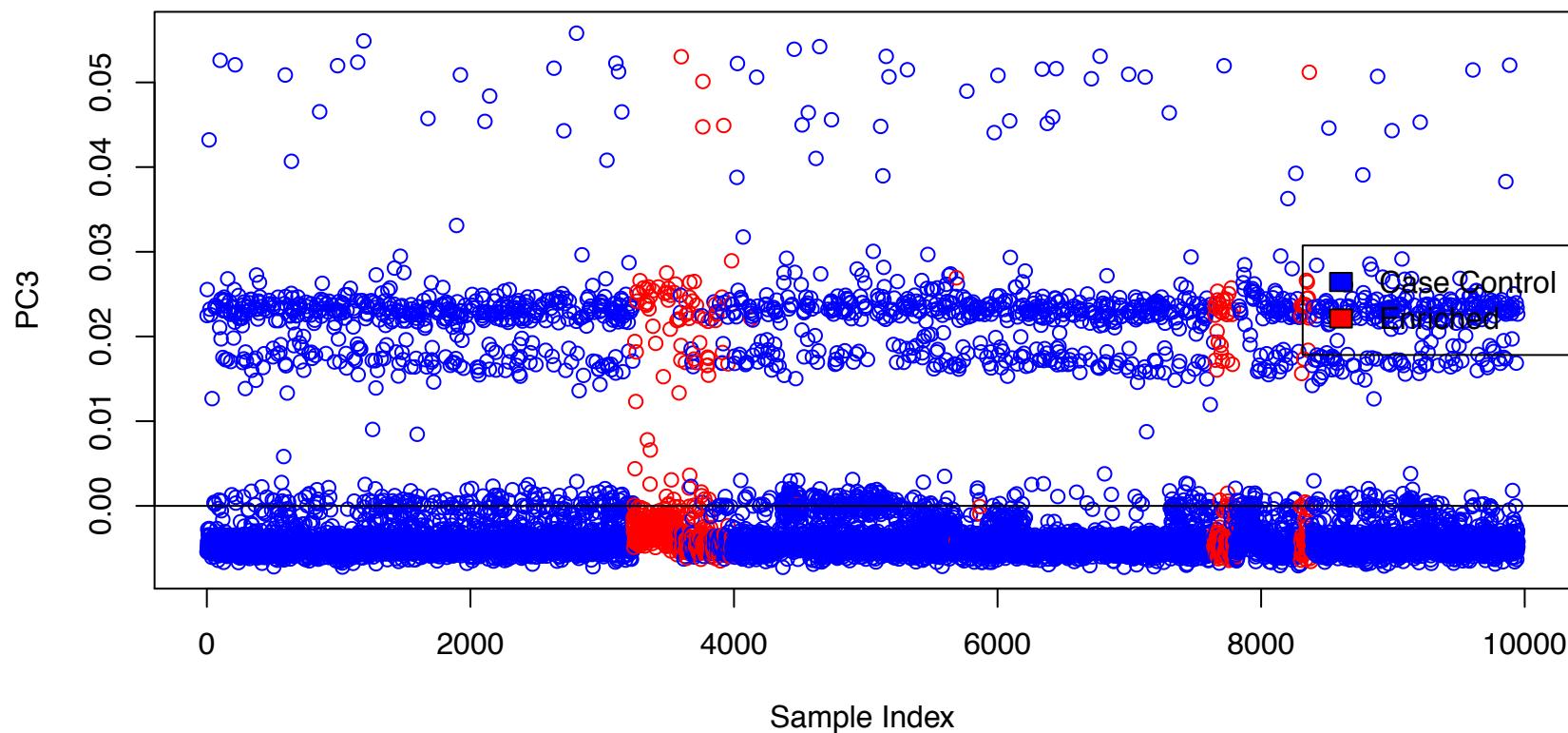
PCA on the 49,235 variants and 9949 samples



PCA on the 49,235 variants and 9949 samples

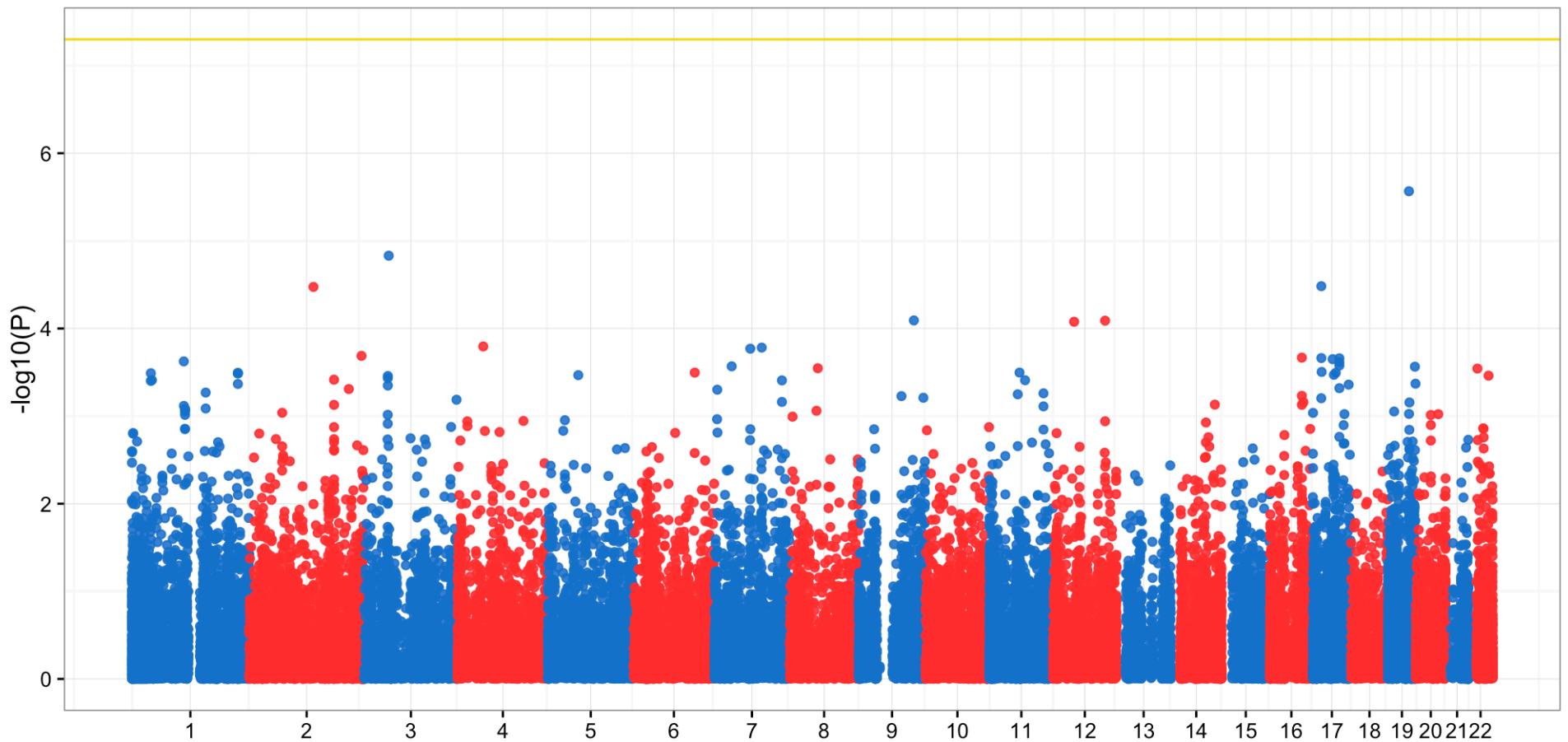


PCA on the 49,235 variants and 9949 samples



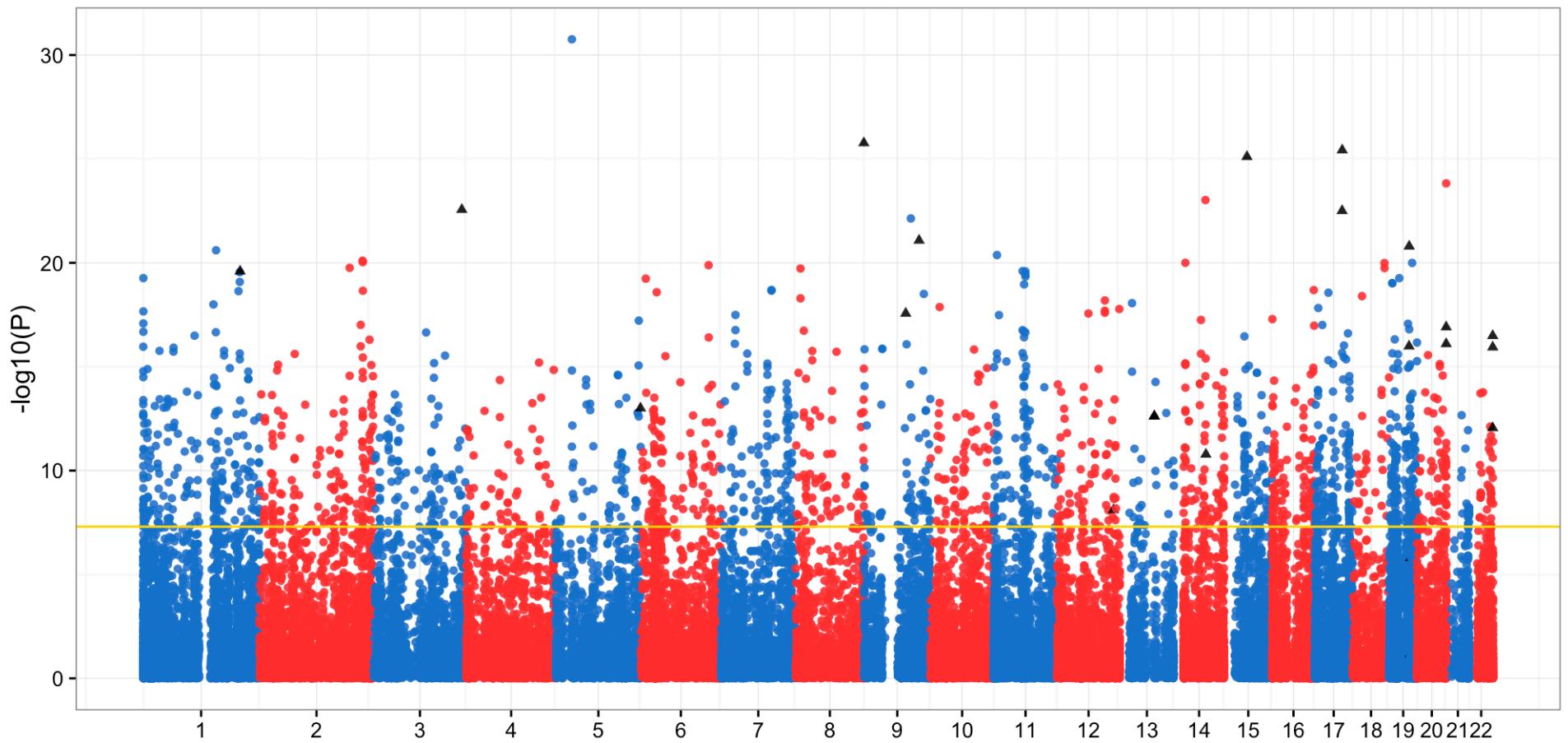
GWAS without the “enriched” samples

Method: Logistic regression in Plink with covariates: Age, sex



GWAS with only the “enriched” samples

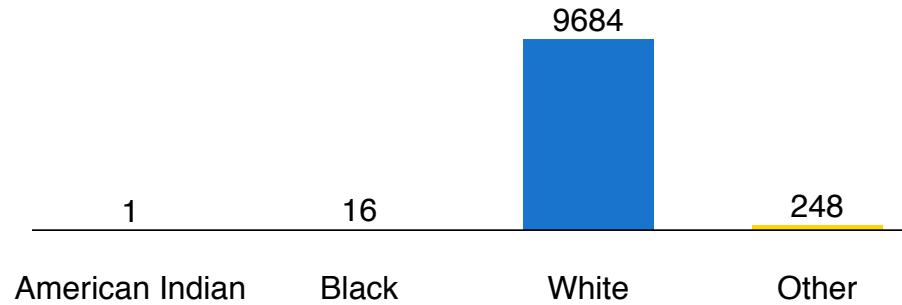
Method: Logistic regression in Plink with covariates: Age, sex



To do

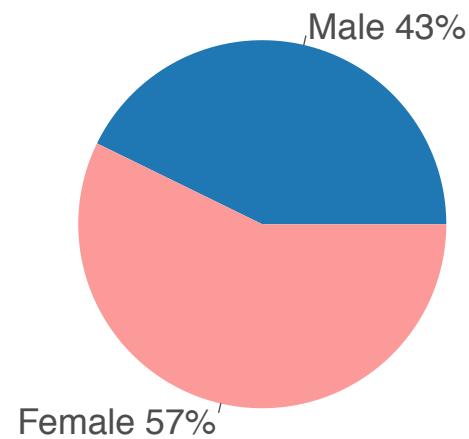
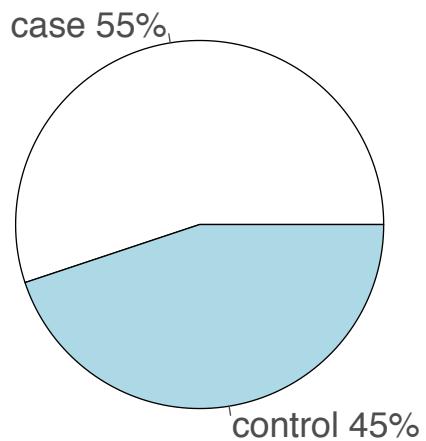
1. Functional annotations of top variants
2. WGS: GLMM by BAYES.GLMM in established pipeline (21/22 done)

ADSP WES data: 9949 samples by 3 sequencing centers



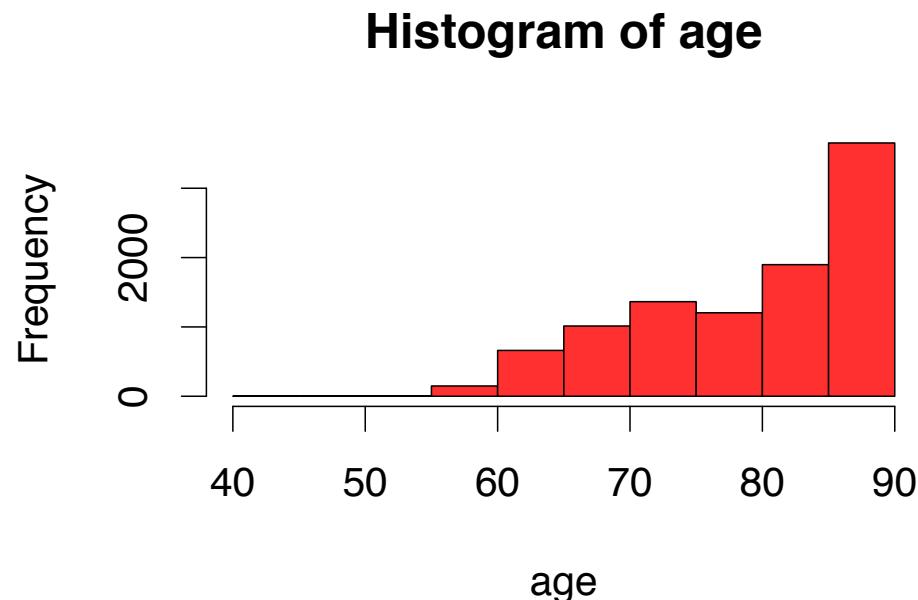
- 10 of the 9949 samples were sequenced independently by the 3 centers. This led to 20 redundant sequencing results, which were removed.
- The 9949 samples were each diagnosed as case or control in Alzheimer's disease status.
- The 9949 samples came from two groups: case control, and enriched. The 816 enriched samples had 645 case and 171 controls.

5484 cases and 4465 controls, 4254 male and 5695 females



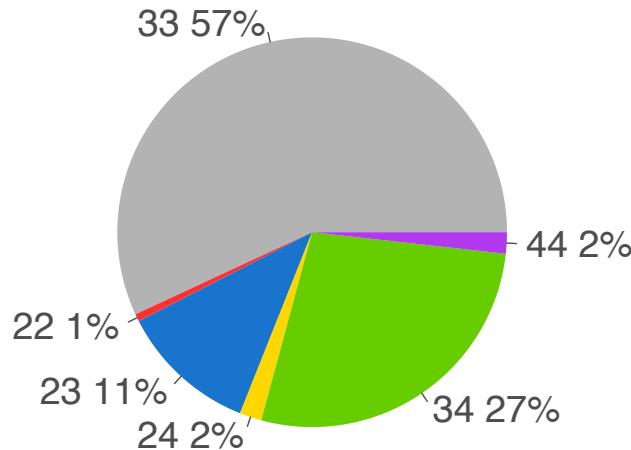
- 10% more case than control, and 14% more female than male

Anyone older than 90 were annotated as “90+”



- 1307 people and 13.1% of the total were annotated as “90+”.
- We viewed these people as 90 years old in the association analysis.

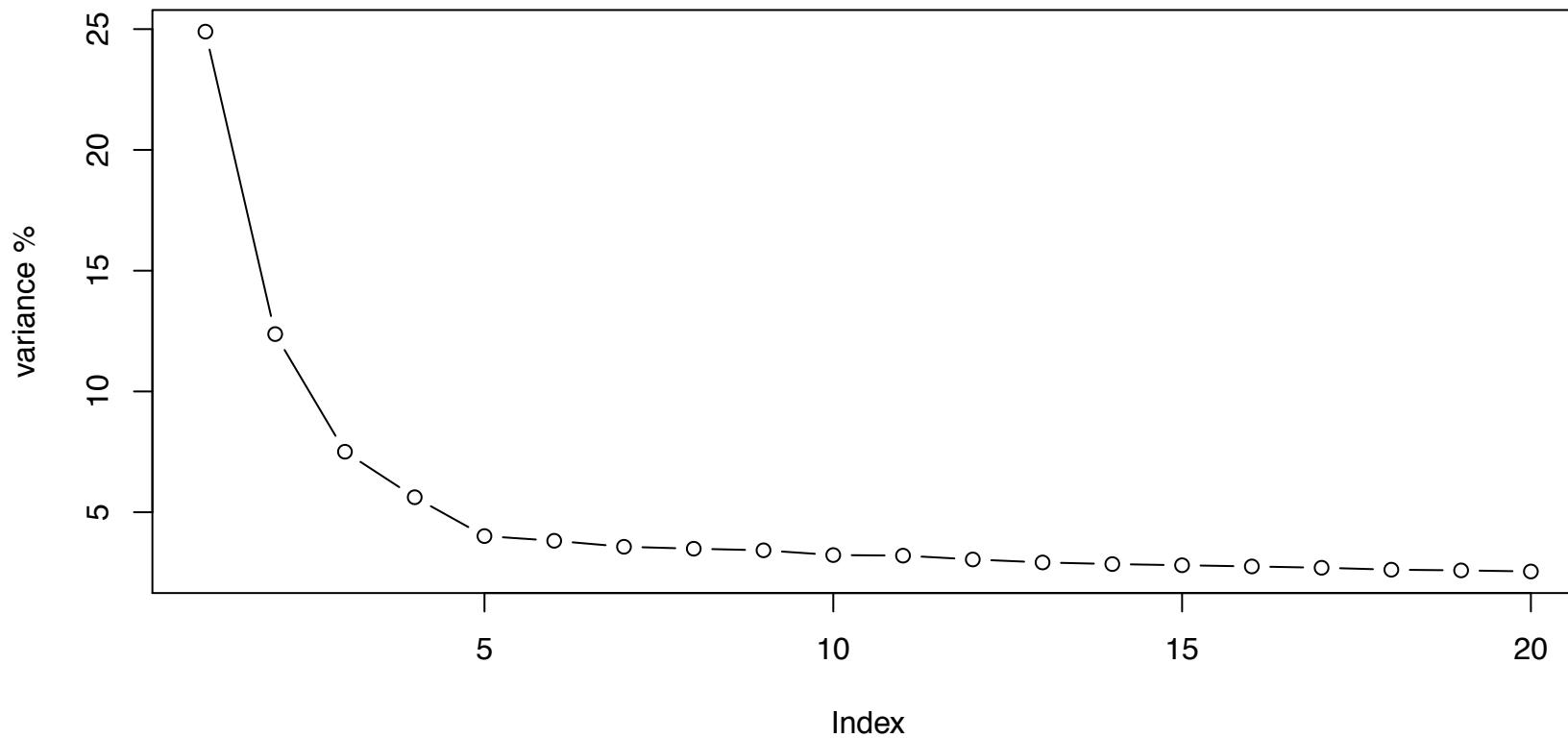
APOE genotypes



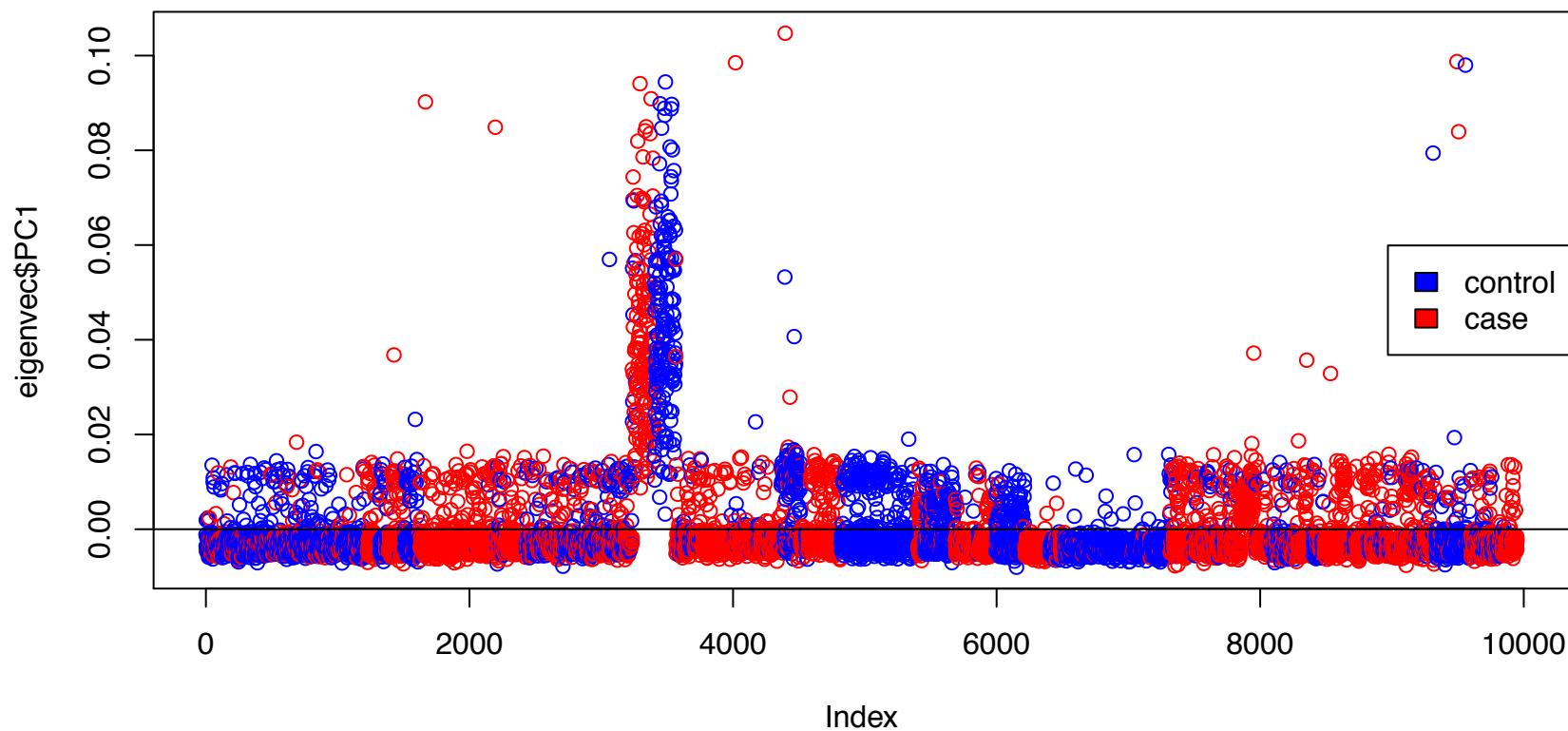
	33	22	23	24	34	44
case	2709	20	334	115	2142	164
control	2954	42	802	67	588	12

- We had 6 unique APOE genotypes, which constitute all the possible bi-allelic combinations of 3 APOE allele types.
- Homozygous e3 was predominantly the majority, while homozygous e2 and e4 were rare.
- Homozygous e4 carriers showed a 15-fold increased risk for developing AD comparing to the homozygous e3 carriers

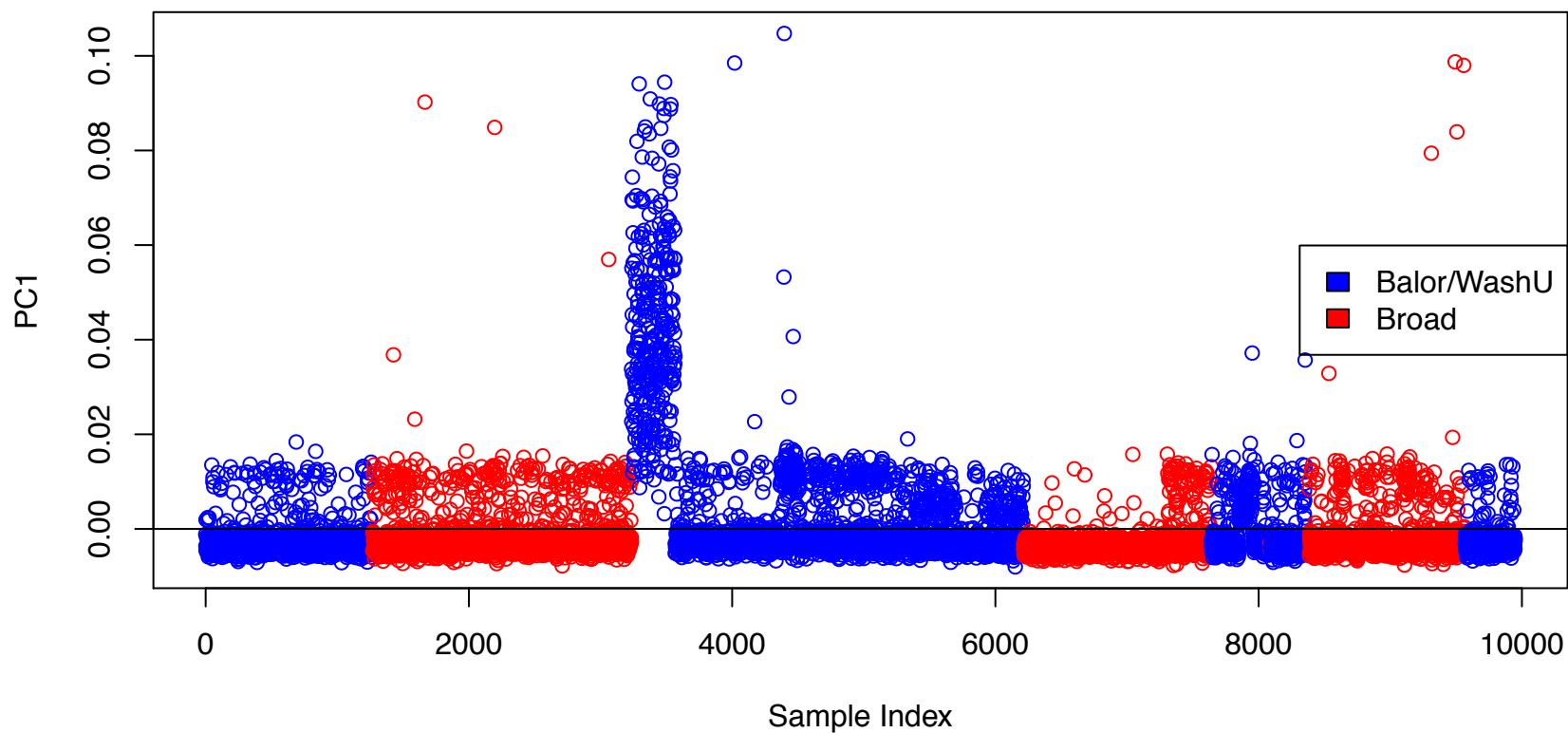
PCA on the 49,235 variants and 9949 samples



PCA on the 49,235 variants and 9949 samples



PCA on the 49,235 variants and 9949 samples





[HOME](#) [ABOUT](#) ▾ [DATA ACCESS](#) ▾ [STUDY INFO](#) ▾ [CONTACT US](#) ▾ [LINKS](#) ▾ [AD](#)

Welcome to the **Alzheimer's Disease Sequencing Project**

The overarching goals of the ADSP are to:

1. Identify new genomic variants contributing to increased risk of developing Late-Onset Alzheimer's Disease (LOAD)
2. Identify new genomic variants contributing to protection against developing Alzheimer's Disease (AD)
3. Provide insight as to why individuals with known risk factor variants escape from developing AD
4. Examine these factors in multi-ethnic populations as applicable in order to identify new pathways for disease prevention

Study Design

Learn about study design, sample selection, and data generation procedures

Apply for Data

Instructions on how to apply for ADSP data

Access Data

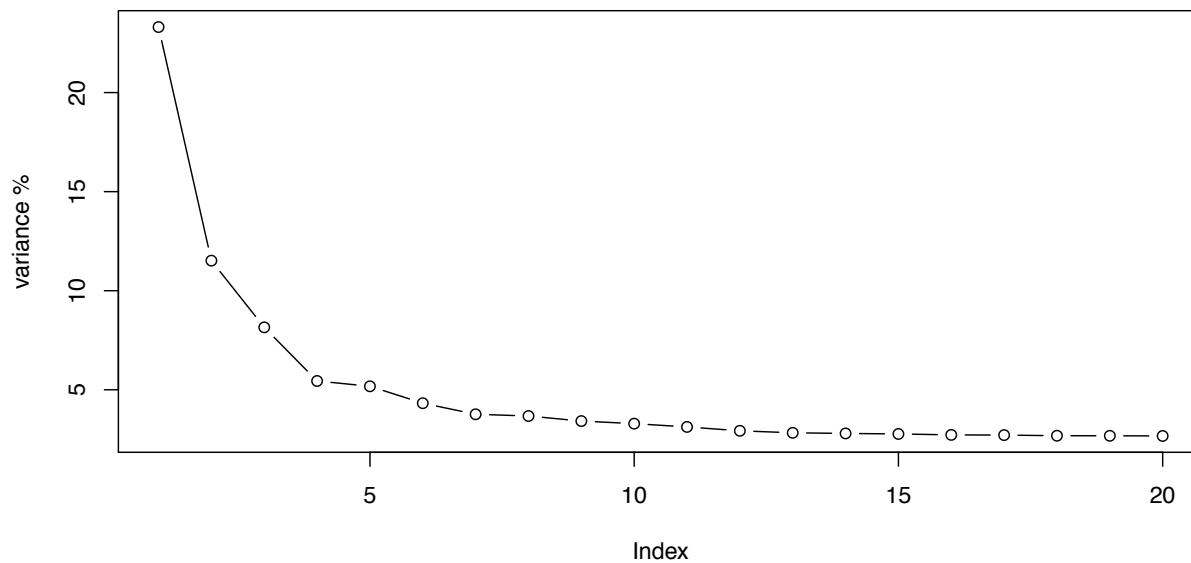
Go to the ADSP data portal

Variants and association analysis

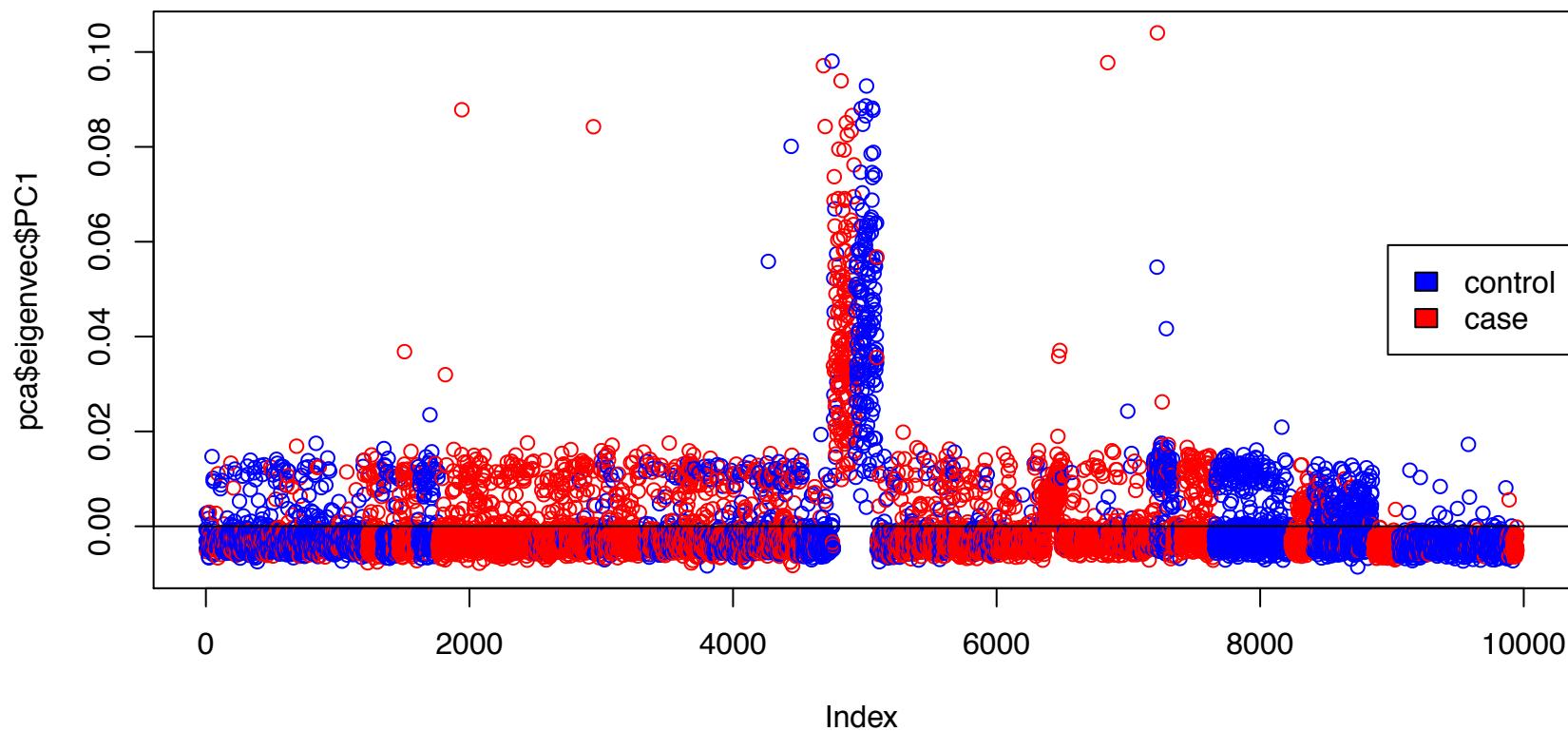
Variants were filtered out by 3 criteria, below, which left 30959 variants

- (1) outside of the shared probe regions,
- (2) carried a non-"PASS" flag in any sample,
- (3) MAF threshold 0.01. 30959 variants were kept for the association analysis.

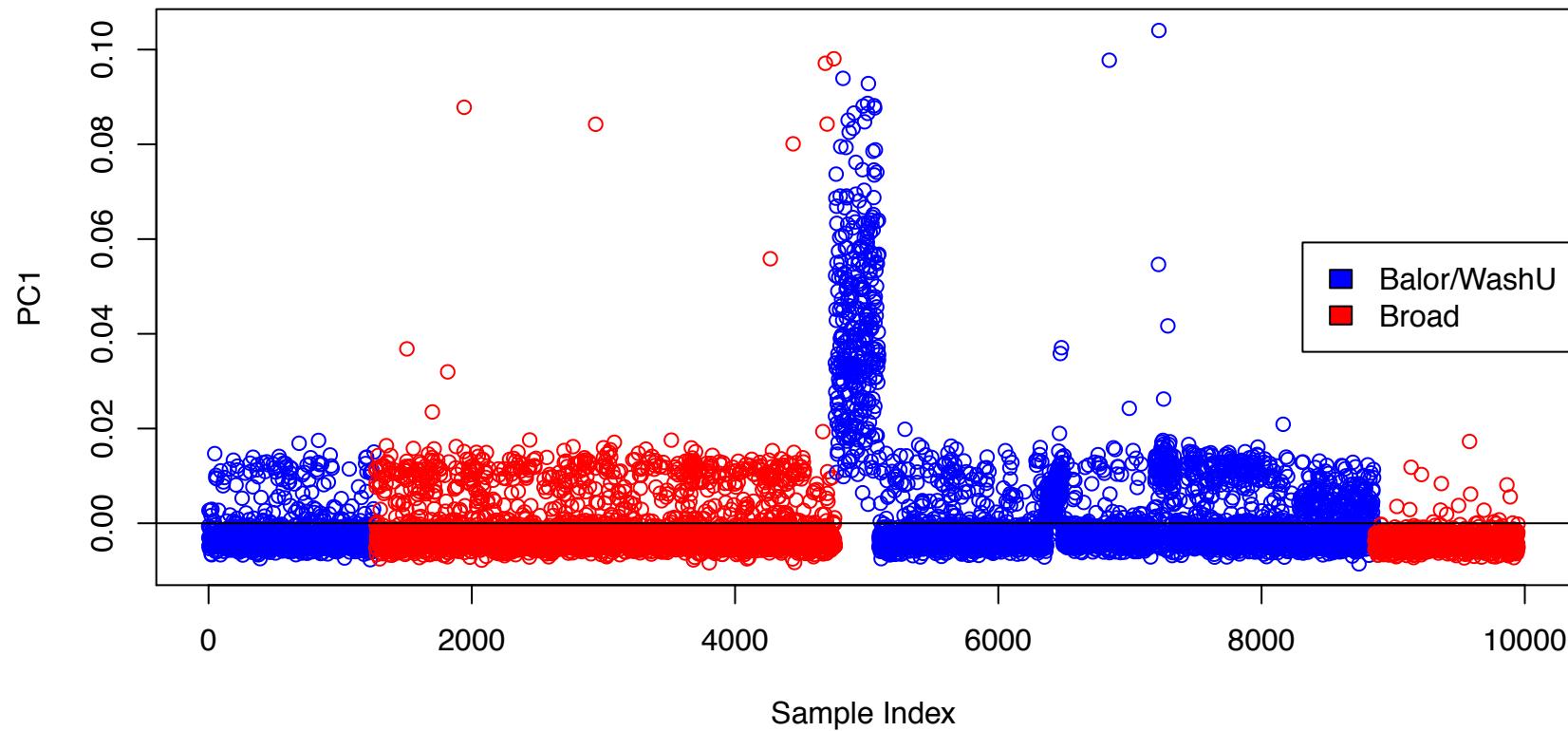
PCA on the 30959 variants and 9949 samples



PCA on the 30959 variants and 9949 samples

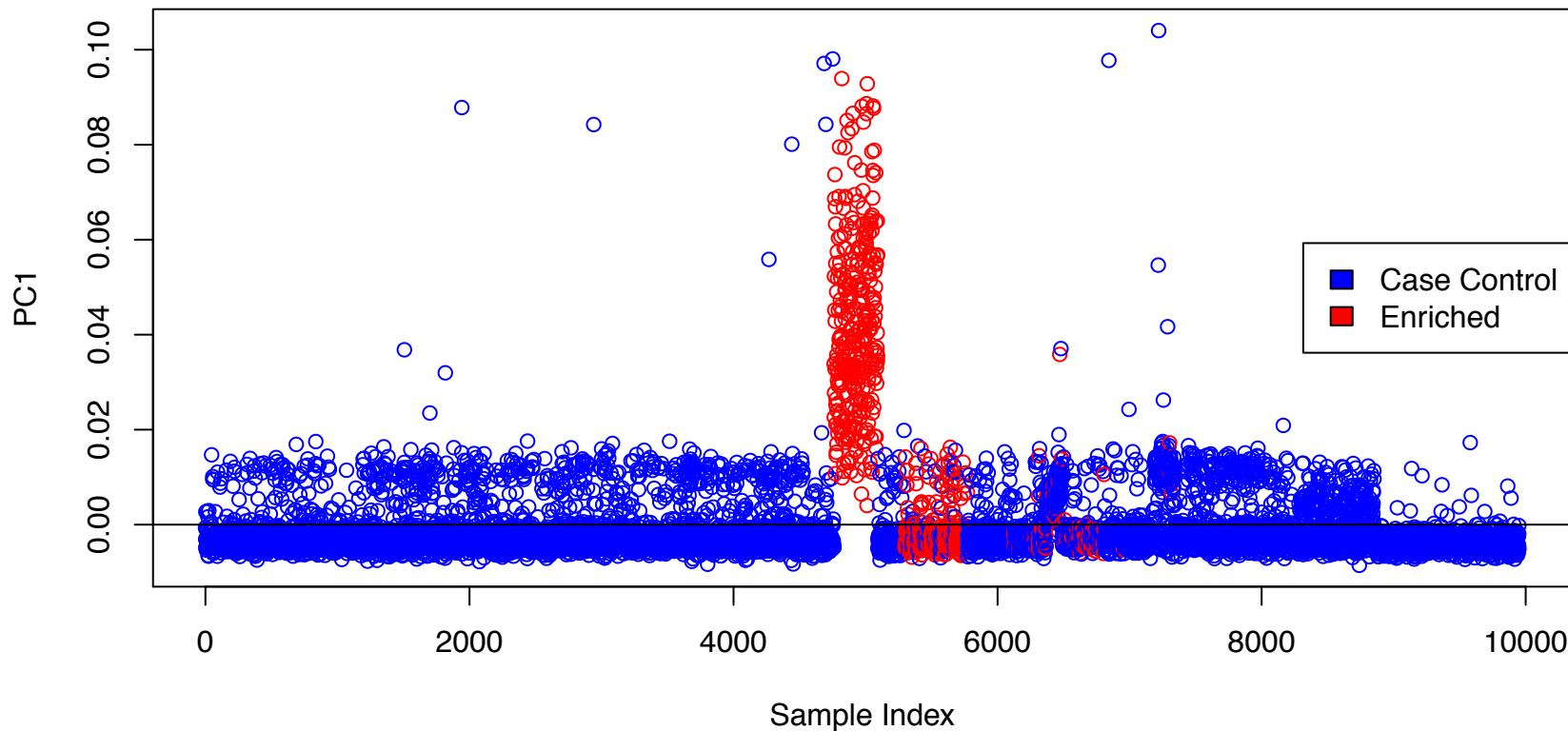


PCA on the 30959 variants and 9949 samples



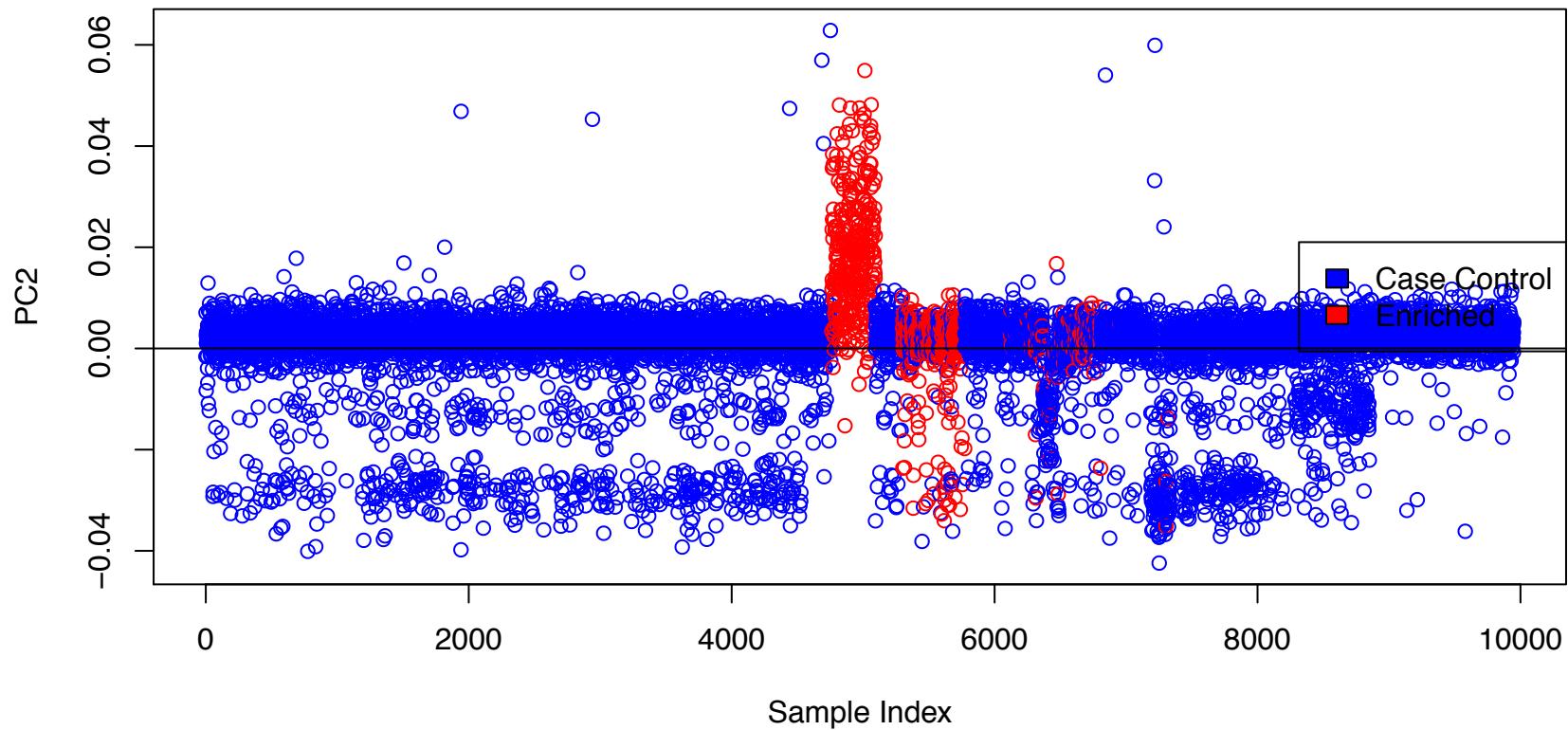
- No more sequencing center-specific effects

PCA on the 30959 variants and 9949 samples



- Enriched samples were different from the others in genotypes

PCA on the 30959 variants and 9949 samples



- Enriched samples were different from the others in genotypes

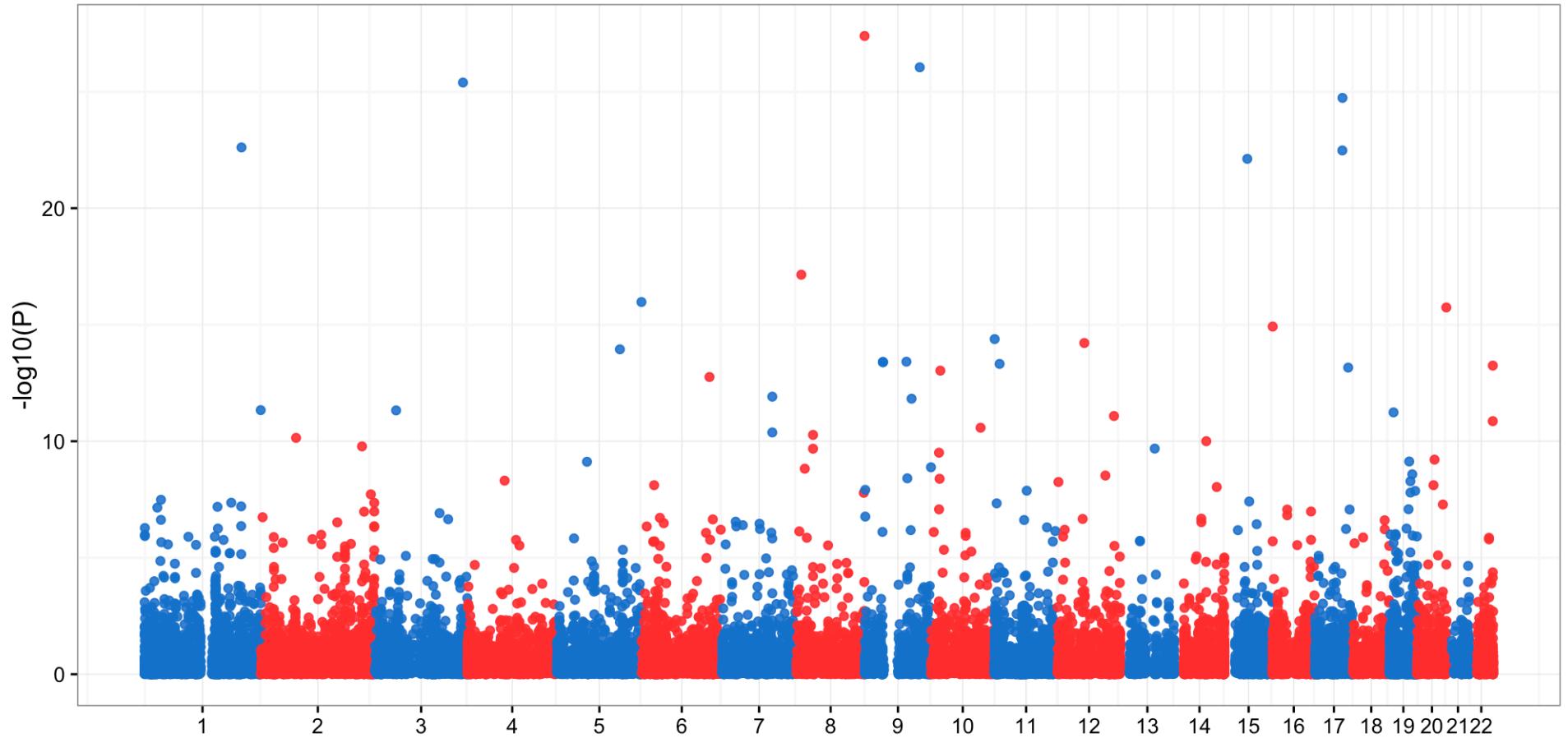
Enriched case

2. **Case-Control sequencing:** whole exome capture sequencing (WES) of 5,000 cases / 5,000 controls for both risk raising and protective loci.

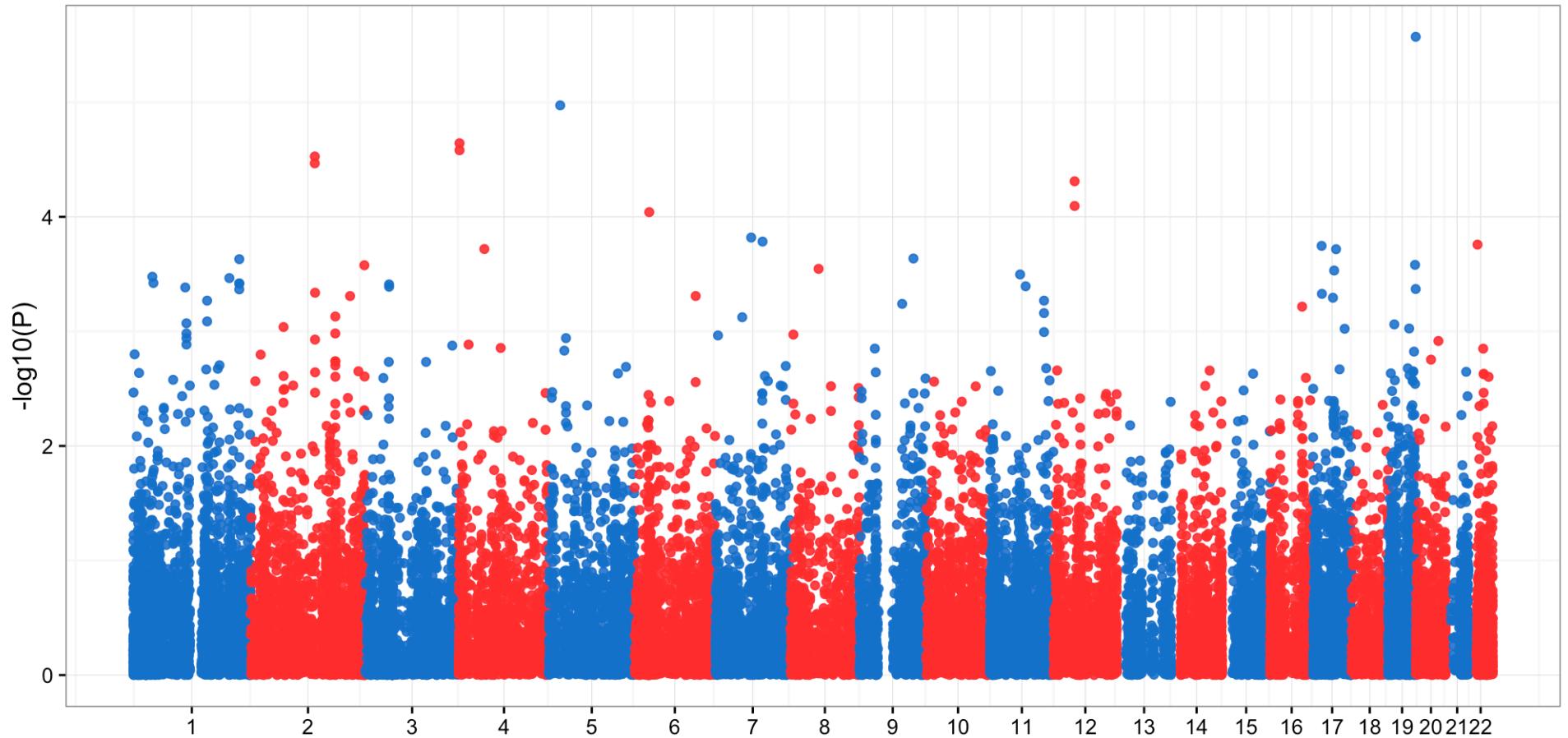
Enriched case sequencing: in addition to the cases above, WES from an additional case group made-up of one individual from 1,000 additional AD families to identify regions associated with increased risk or protection from AD.

Timeline for data production: June 2013- August 2014

Manhattan with all samples



Manhattan without the “enriched” samples

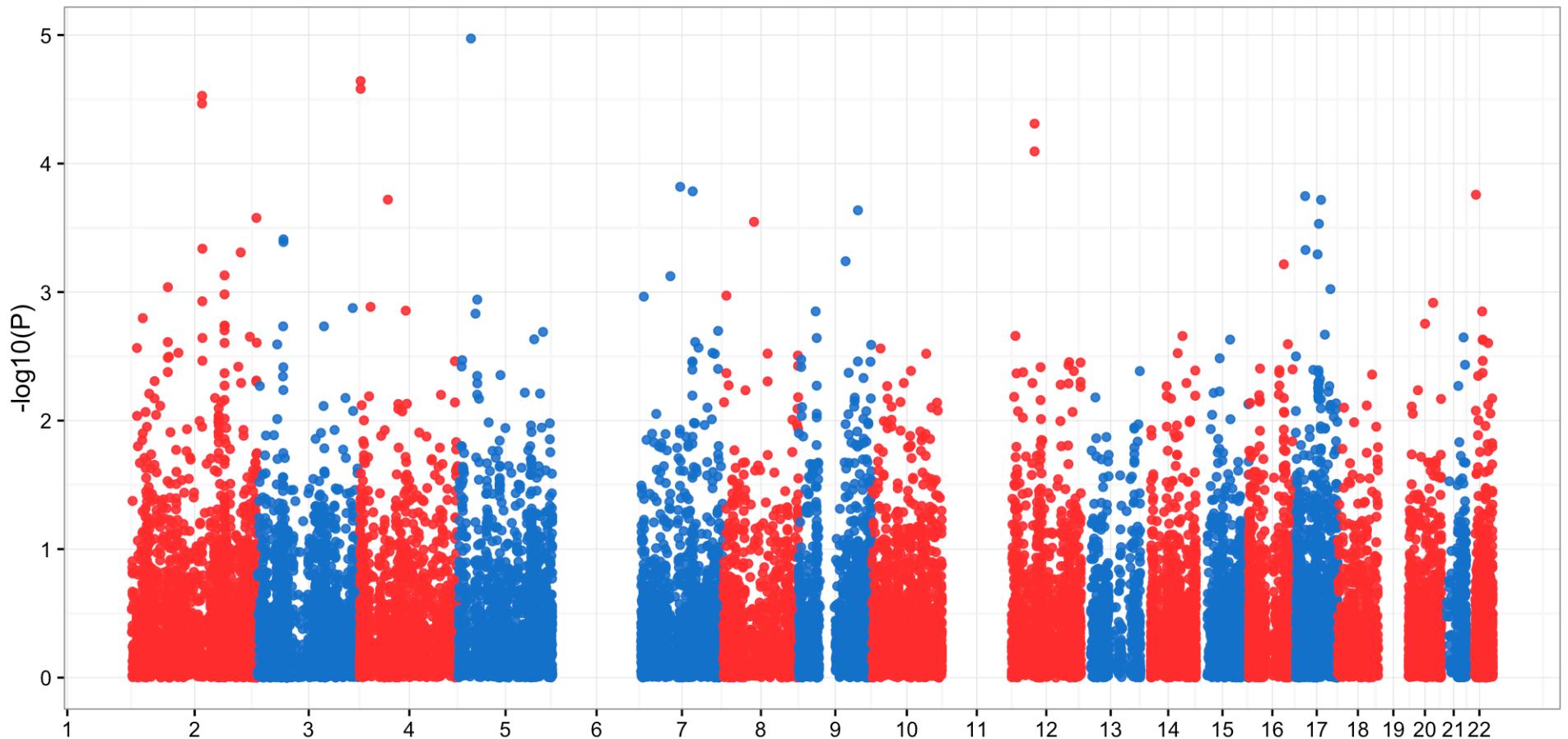


- APOE SNP: rs7412 and rs429358, were all flagged as “LowQual” and “LowCoverage”

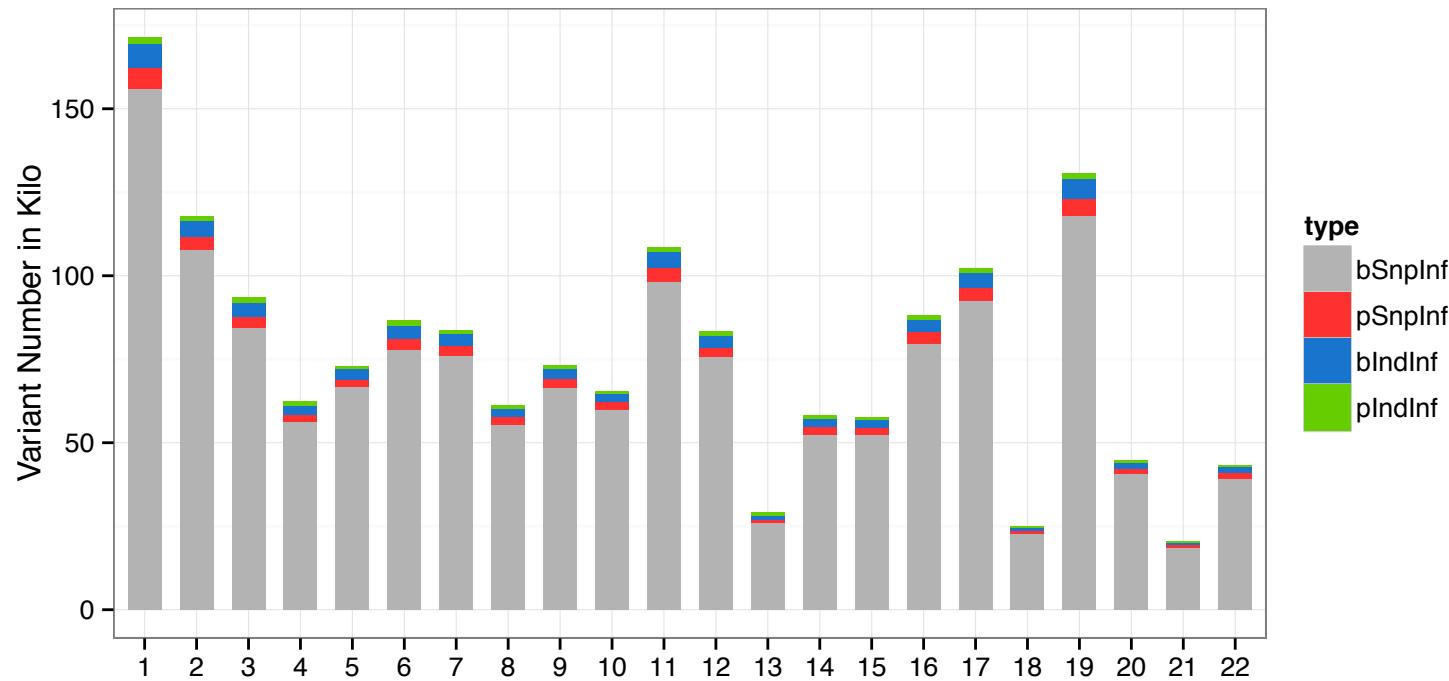
Variants with small proportions of non-”PASS” flag ought to be tolerated

- (1) No chromosome: 1, 6, 11, 19
- (2) 48091 variants MAF > 0.01
- (3) 46139 of 48091 showed “PASS” in at least 80% of carriers

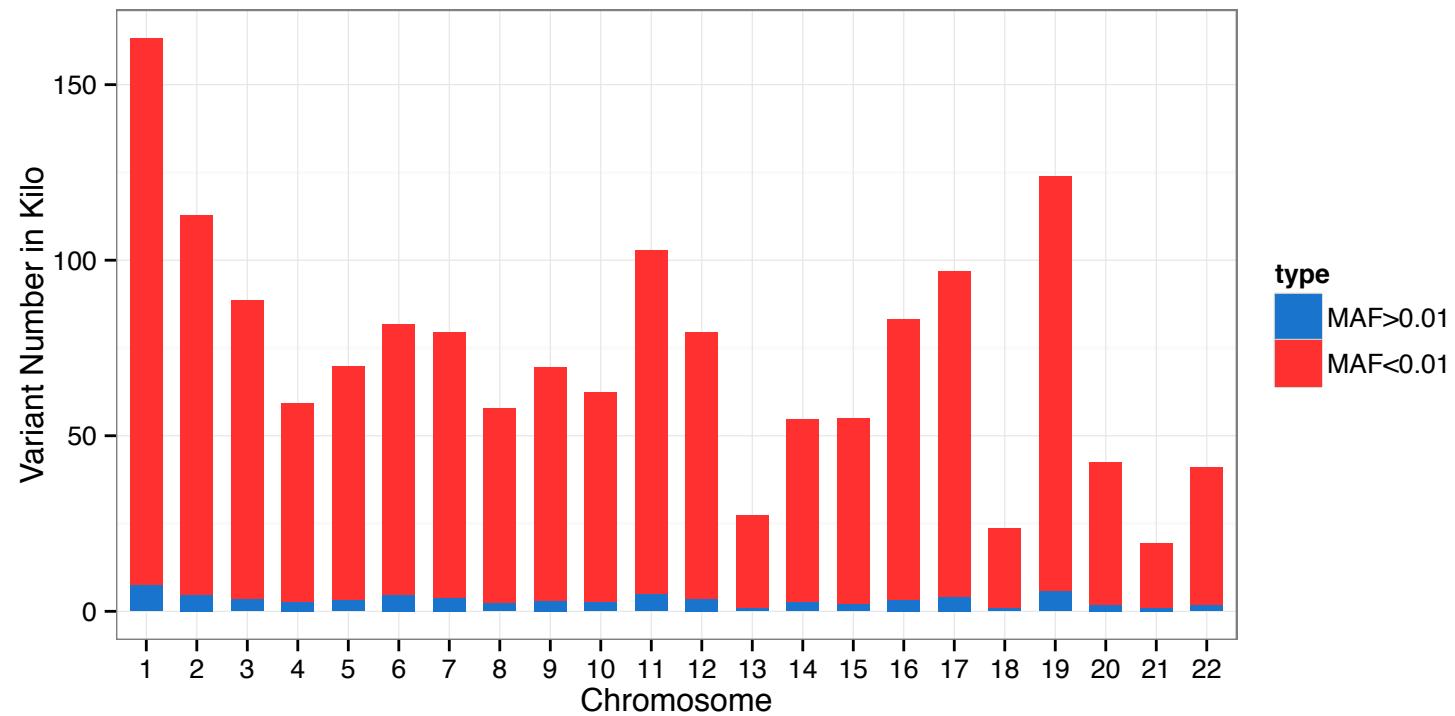
Manhattan with expanded variants



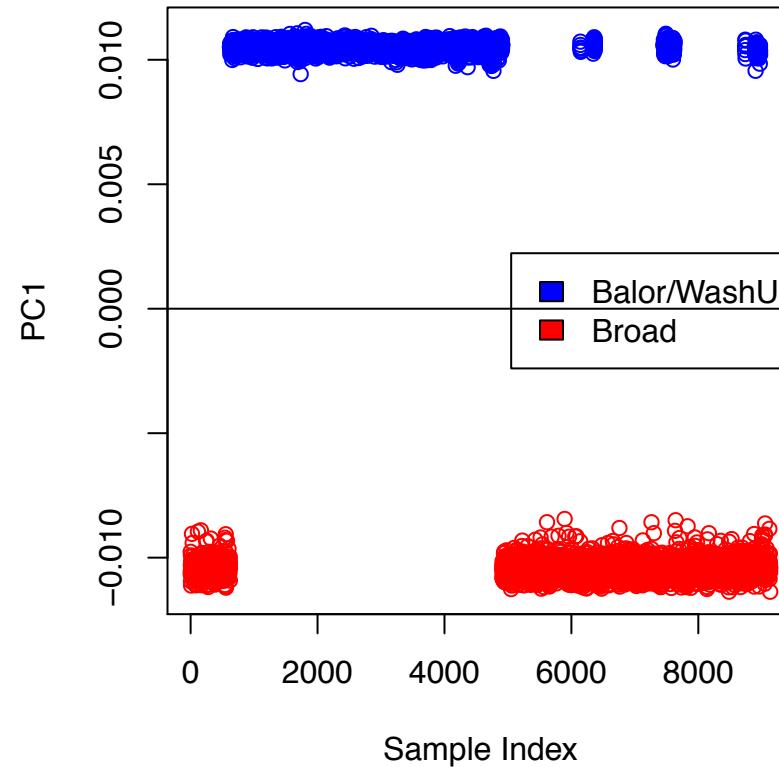
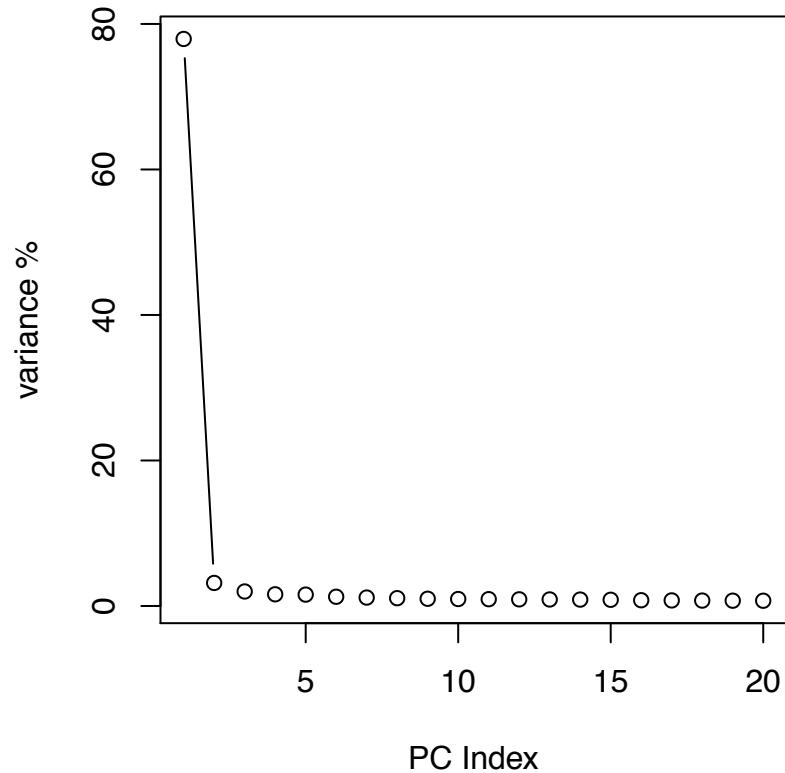
Genomic Variants by GATK



72447 SNP and IND with MAF 0.01 and morec

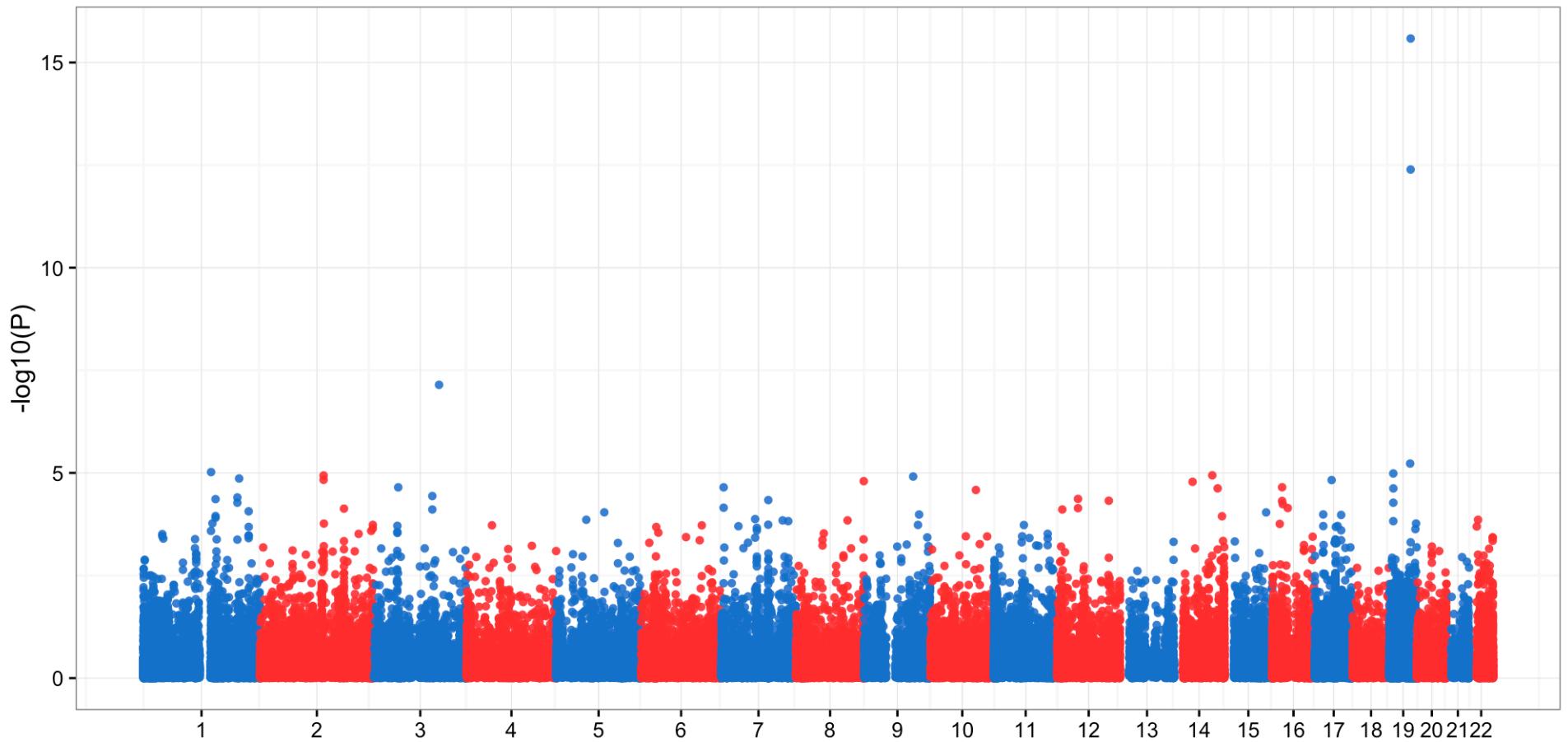


PCA on genotypes



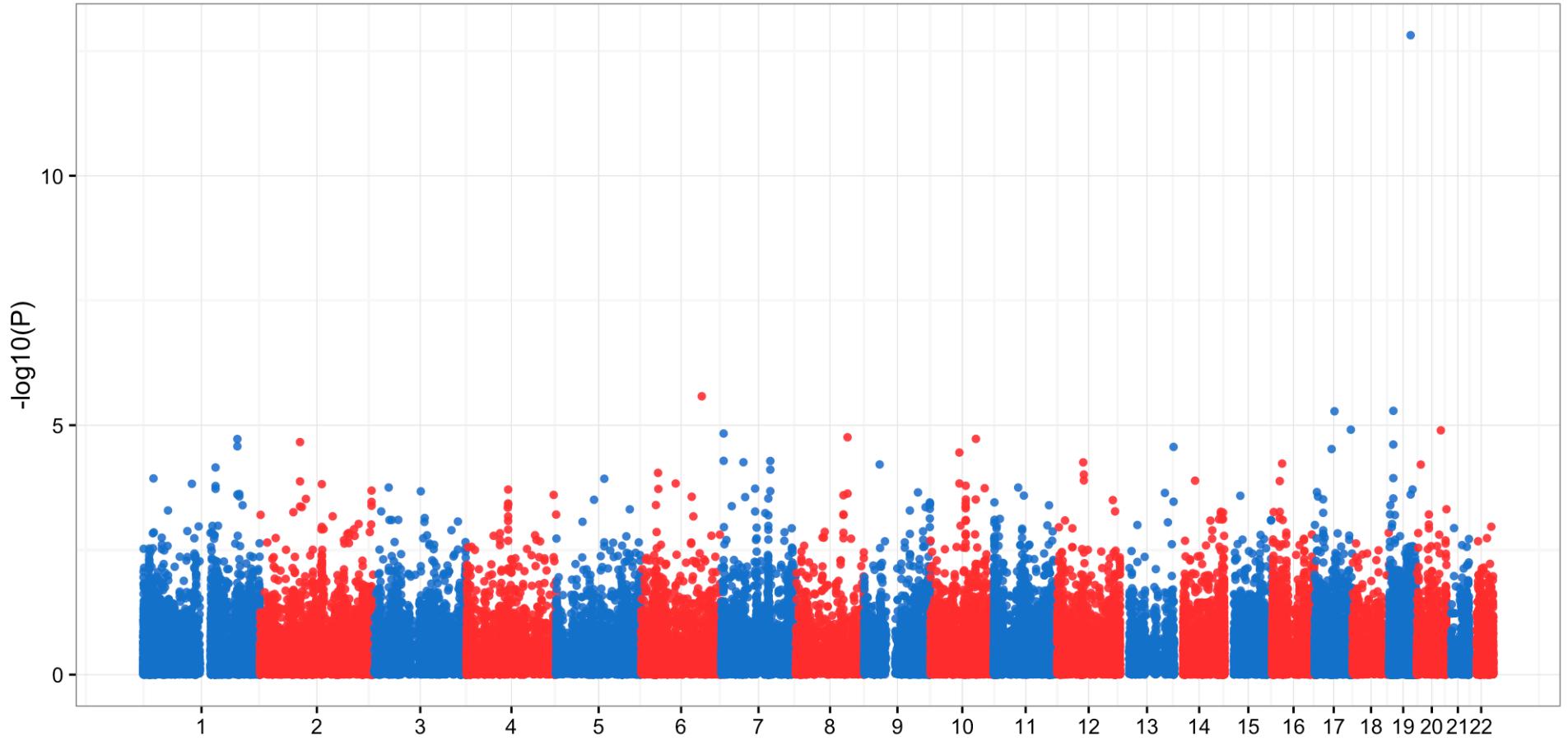
- Out of 72447, 9764 only called by Broad, and 3255 only called by Baylor/WashU

WES w/o priors



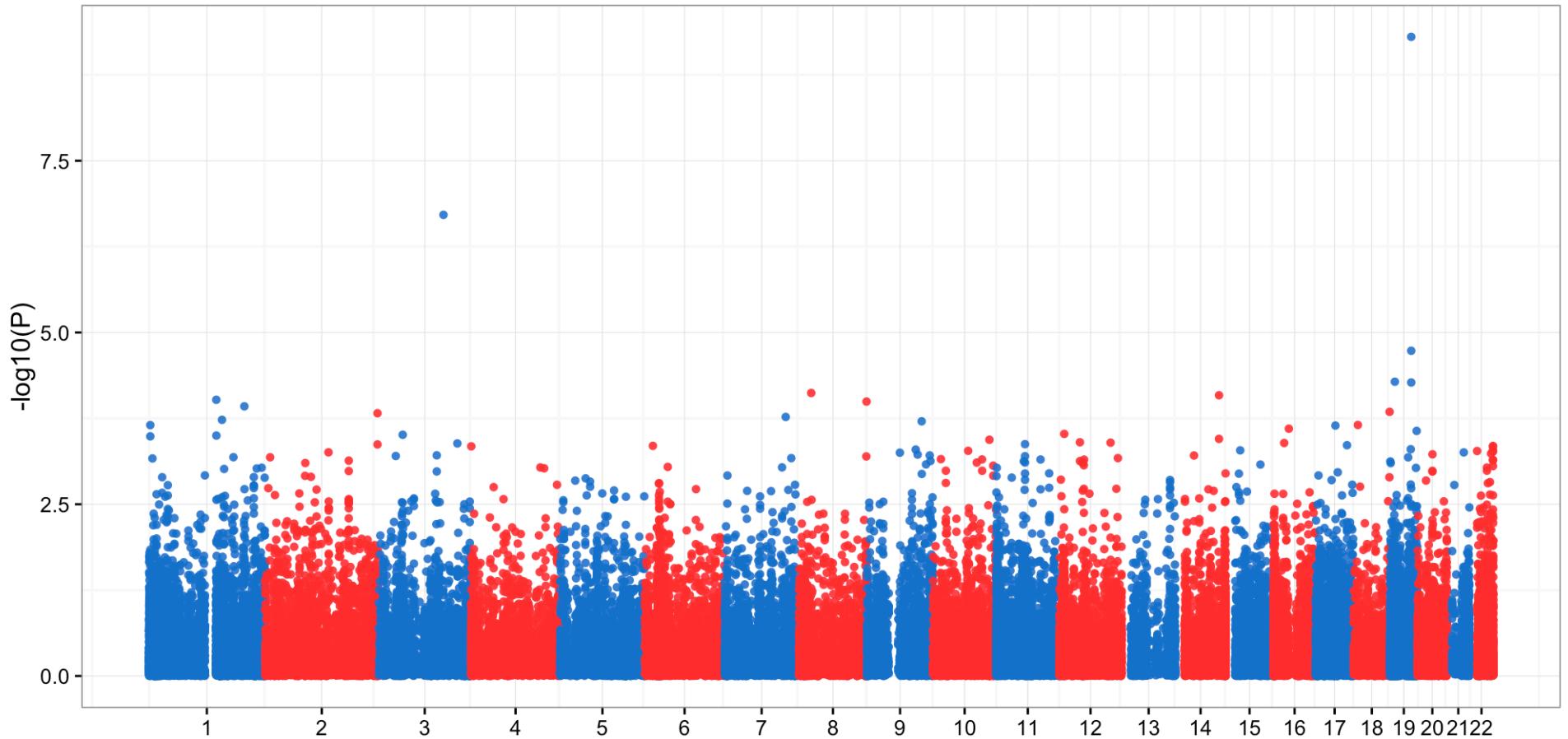
- Logistic method in Plink with covariates: age, sex, PC1

Broad only WES w/o priors



- Logistic method in Plink with covariates: age, sex

Bayor/WashU only WES w/o priors



- Logistic method in Plink with covariates: age, sex

Generalized linear mixed model in Bayesian framework (bayes-glmm)

$$\begin{aligned} P(Y_i \leq j) &= \pi_{i1} + \dots + \pi_{ij} \\ \text{logit}(P(Y_i \leq j)) &= \theta_j - \mathbf{X}\beta - g\beta_0 + u \quad j = 1, \dots, J-1 \end{aligned}$$

Ordered logistic model

$$\begin{aligned} u &= L * z \\ L &= \text{Chol}(K) \\ z &\sim mvN(0, \sigma I) \end{aligned}$$

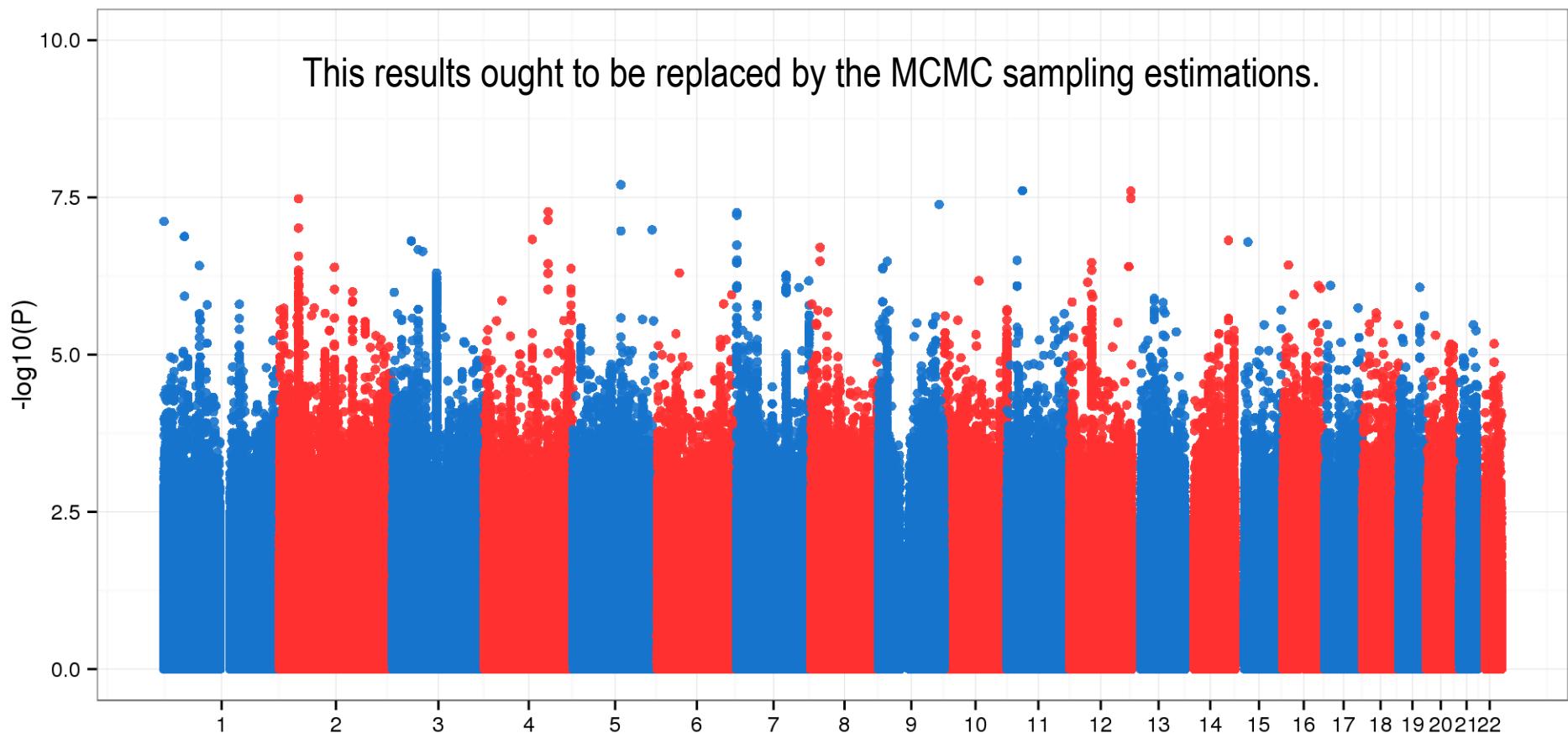
Random term by Cholesky factoring

$$\begin{aligned} \theta &= 10 * \text{cumsum}(\theta_0) \\ \theta_0 &\sim \text{dirichlet}(1) \\ \beta &\sim N(0, 1) \\ \beta_0 &\sim N(t * \sigma_0, \sigma_0) \\ t &\sim N(\text{prior}, 1) \\ \sigma_0 &\sim \text{inv-gamma}(2, 1) \\ \sigma &\sim \text{inv-gamma}(2, 1) \end{aligned}$$

Priors

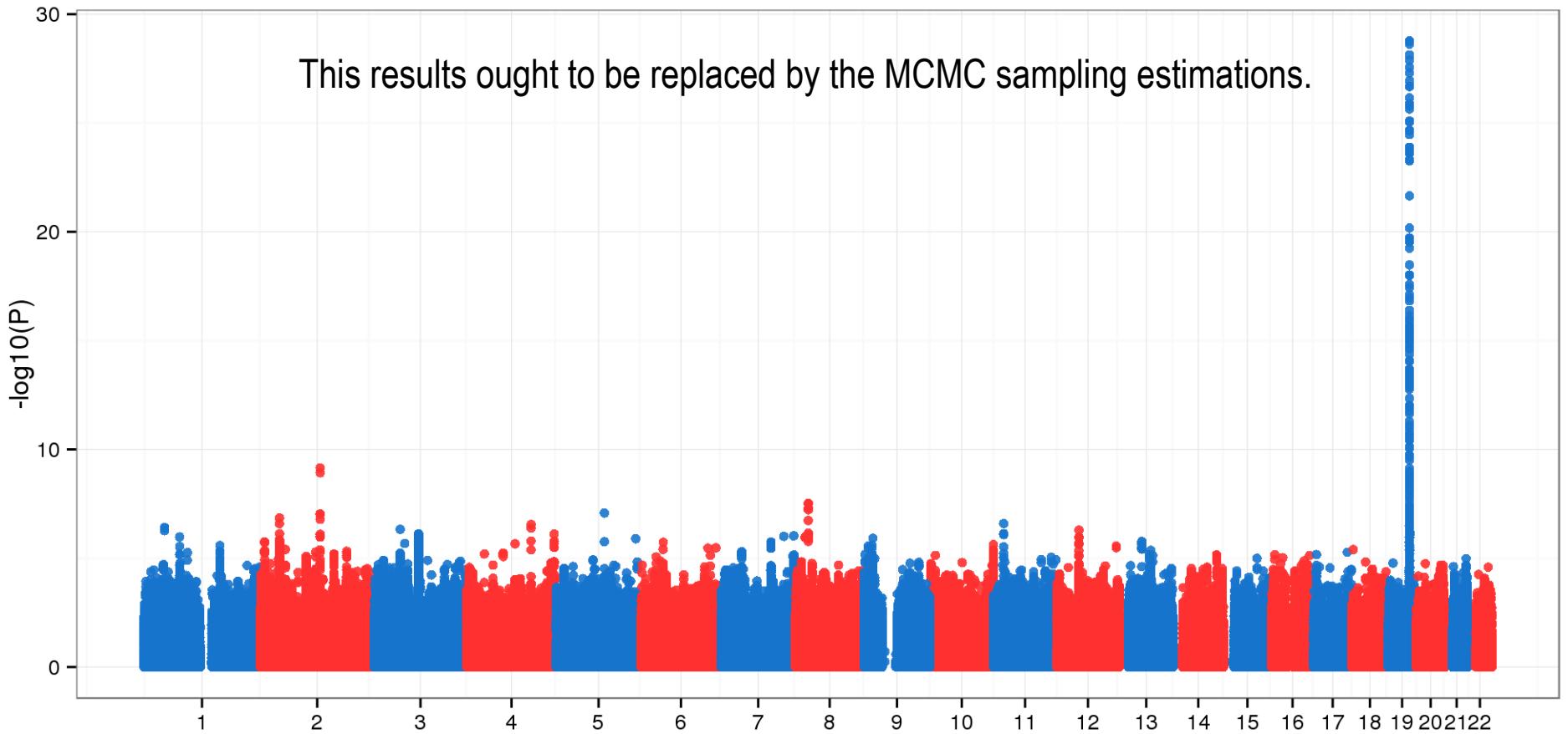
- AD status was modeled by an ordered categorical variable
- Population structure was modeled by a random term that followed a multivariate normal distribution with covariance matrix constrained by the IBS kinship relatedness
- Prior knowledge can be included as a standardized effect size
- Kinship computed by Plink; model implemented in Stan (<http://mc-stan.org/>)

WGS w/o priors



GLMM method with covariates: age, sex

WGS with priors



- GLMM method with covariates: age, sex
- APOE, BIN1, CLU loci genome-wide significant

