

Genome-Wide Association for Late-Onset Alzheimer Disease (LOAD) Confirms Risk Locus on Chromosome 12

Gary W Beecham, PhD

Miami Institute of Human Genomics

University of Miami Miller School of Medicine

Miami, Florida

**Reference: Beecham GW, Martin ER, Li YJ,
Slifer MA, Gilbert JR, Haines JL, Pericak-
Vance MA. *Am J Hum Genet.* 2009
Jan;84(1):35-43.**



Outline

- Background and Genetics of LOAD
- Genome-Wide Association
- Imputation and Meta-Analysis
- Imputation Results
- Conclusions

Alzheimer Disease (AD)

- AD is the most common cause of dementia
 - Gradual decline in memory and cognitive ability, including language, problem solving and performance of everyday tasks
- Lifetime risk is 10-15% in the general population
- Prevalence doubles every five years after age 60
 - Approaching 50% in those age 85 & older
- Most (>90%) AD is late onset Alzheimer disease (LOAD; > 60 years)
- Occurs in multiple racial/ethnic groups

Risk Factors for AD

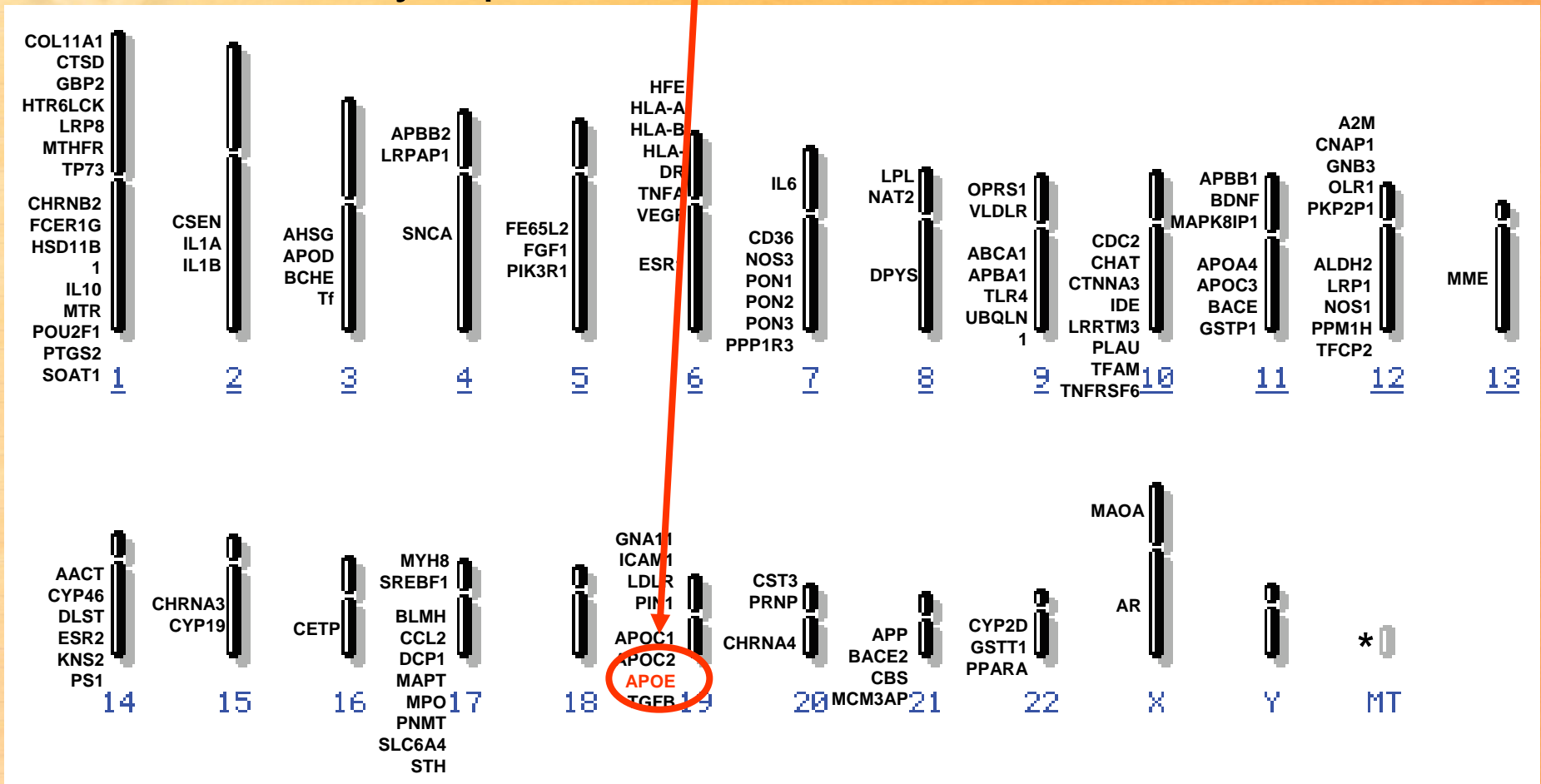
- Family history of AD
 - Strongest association in first degree relatives
 - 2-3 fold increase in risk
 - Higher with each additional relative
 - Heritability: 50%-80%
- Genetics
 - Four AD genes with major effects identified so far
 - *APP*, *PS1*, and *PS2* are all associated with early-onset AD
 - Apolipoprotein E (*APOE*) is the only consistently associated LOAD risk gene
- Environmental (e.g., Education level, head trauma, etc.)

***APOE* and Alzheimer Disease**

- Apolipoprotein E (*APOE*) is the only LOAD risk gene with consistently observed associations
- *APOE* accounts for at most 50% of the genetic effect
- E4 carries greater risk of LOAD and earlier age of onset
 - 0 copies of E4: ~20% risk of AD, Average age-of-onset: 86
 - 1 copy of E4: ~40% risk of AD, Average age-of-onset: 78
 - 2 copies of E4: ~85% risk of AD, Average age-of-onset: 74

Many Candidate Genes in AD

- All shown have at least one positive association
- Apolipoprotein E (APOE) on chromosome 19 only one consistently replicated



Outline

- Background and Genetics of LOAD
- **Genome-Wide Association**
- Imputation and Meta-Analysis
- Imputation Results
- Conclusions

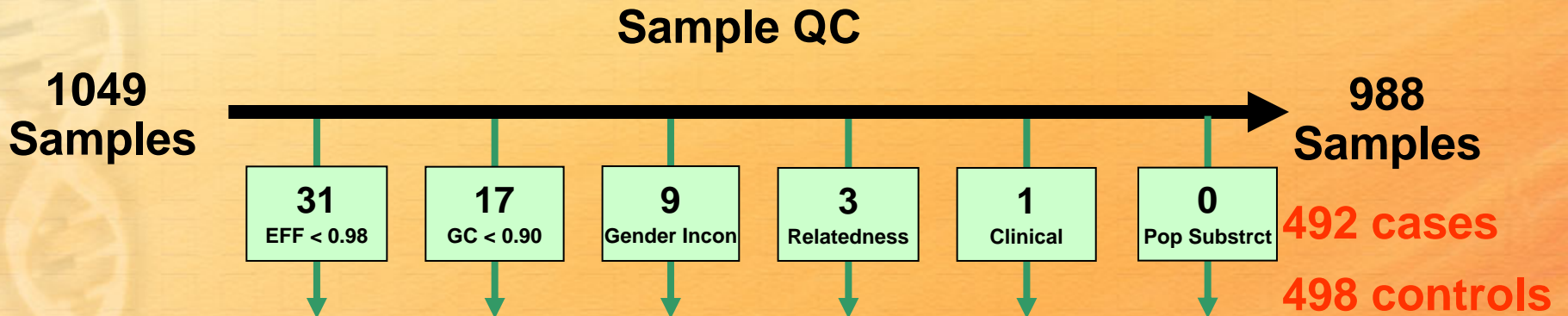
Genome-Wide Association Study

- We performed a GWAS to discover disease risk genes for LOAD
 - Increased coverage relative to candidate gene approach
 - Increased resolution and power relative to linkage studies

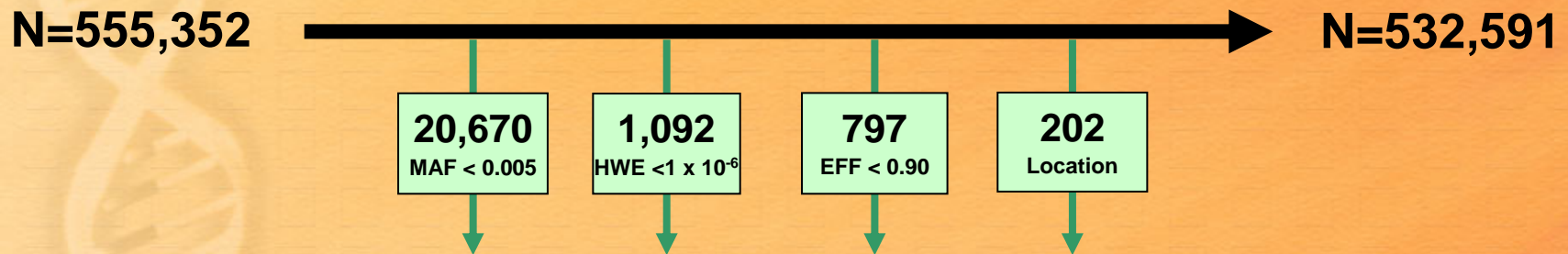
Methods- Discovery GWAS

- Case Definition/Diagnosis
 - LOAD affected samples meet NINCDS/ADRDA criteria for probable or definite AD, with AAO > 60
 - Controls had no signs of dementia by history and upon interview (MMSE > 26 or 3MS > 86)
- Ascertainment
 - Samples from the Collaborative Alzheimer Project (MIHG and CHGR)
 - 1049 samples (518 case, 531 control) were genotyped
- Genotyping: Illumina Infinium II Assay
 - Using Illumina Beadstation, 550HH Beadchip, & BeadStudio
 - Genotyping efficiency > 99%
 - CEPH controls genotyped multiple times

Quality Control- Discovery GWAS



Marker QC
(Genotype Quality) (Genotype Intensity) (Genotype Heterozygosity) (Genotype Missingness) (Genotype Missingness by Site)



Average MAF:
0.246

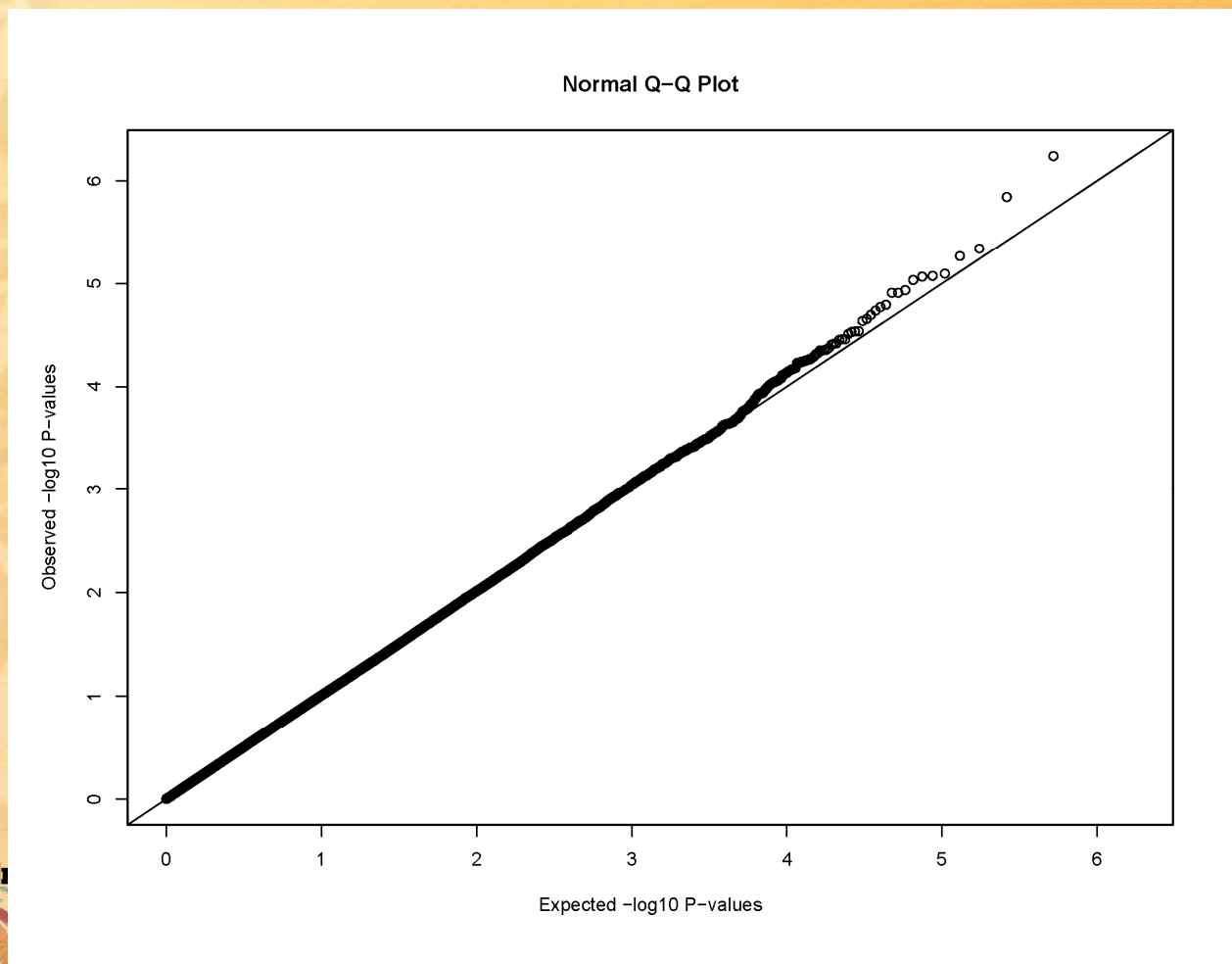
Average SNP Efficiency:
99.83%

Association- Discovery GWAS

- Cochran-Armitage trend test ($df = 1$)
- Logistic regression, adjusting for:
 - APOE status (# of APOE e4 alleles)
 - Age at Onset (AAO)/Age at Exam (AAE)
 - Gender
- Top results compared with consensus linkage regions

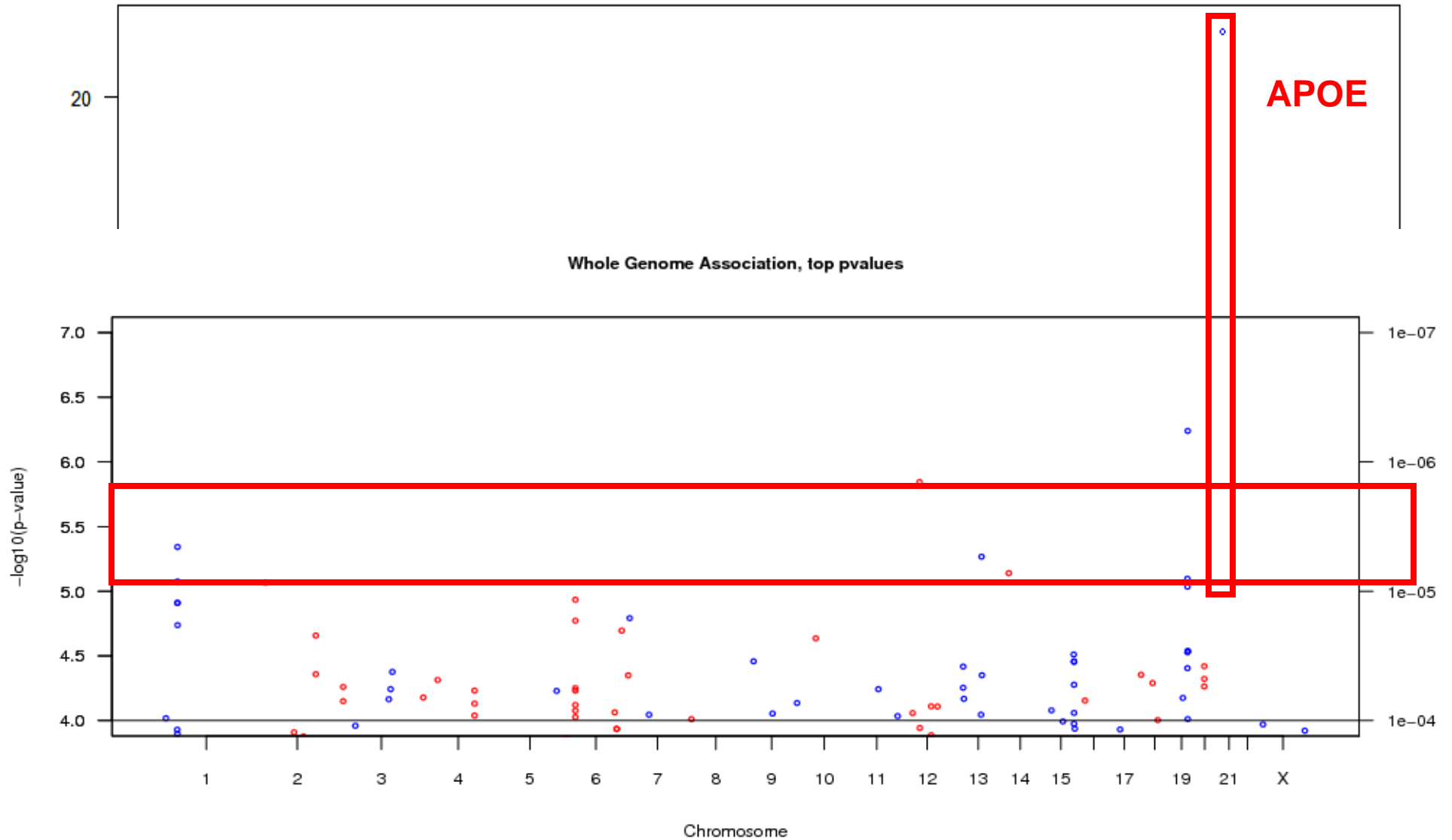
Results- Discovery GWAS

- Departure of P-values from Expectation (QQ Plot)



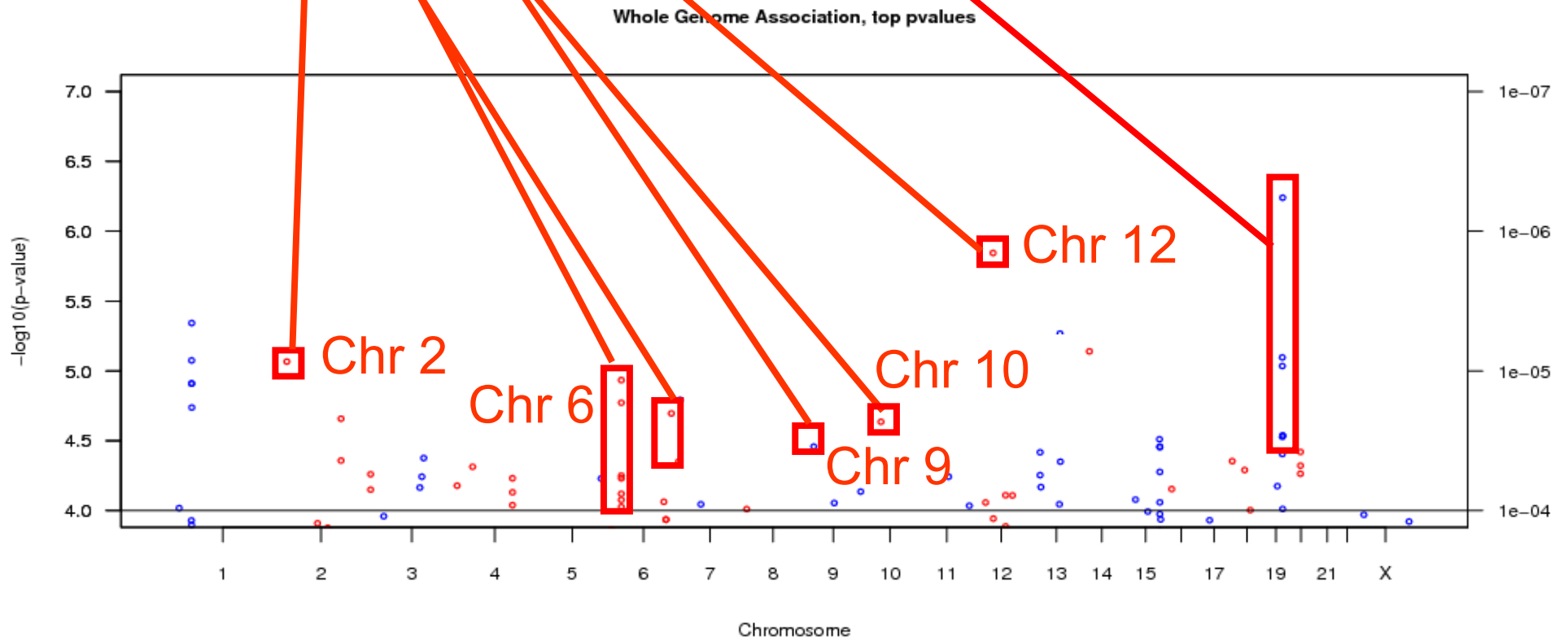
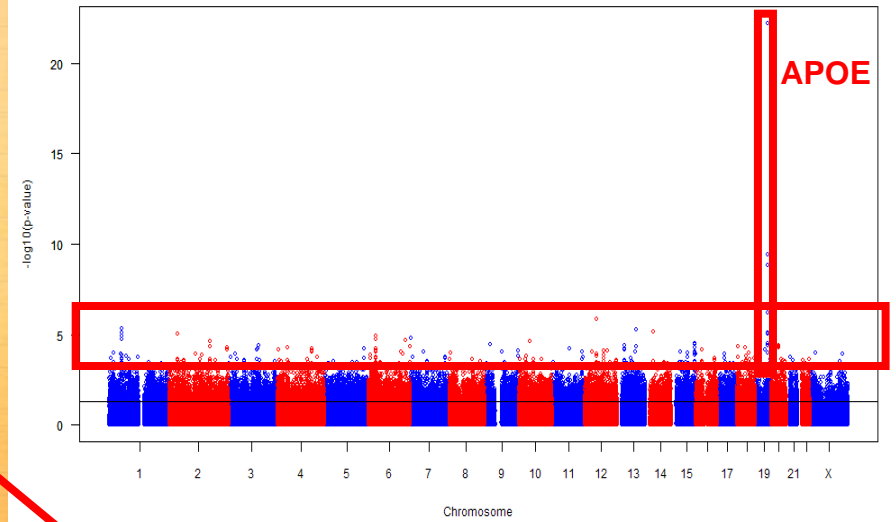
Excluding top 3 associations in/near APOE on Chr 19

Results- Discovery GWAS



Results- Discovery GWAS

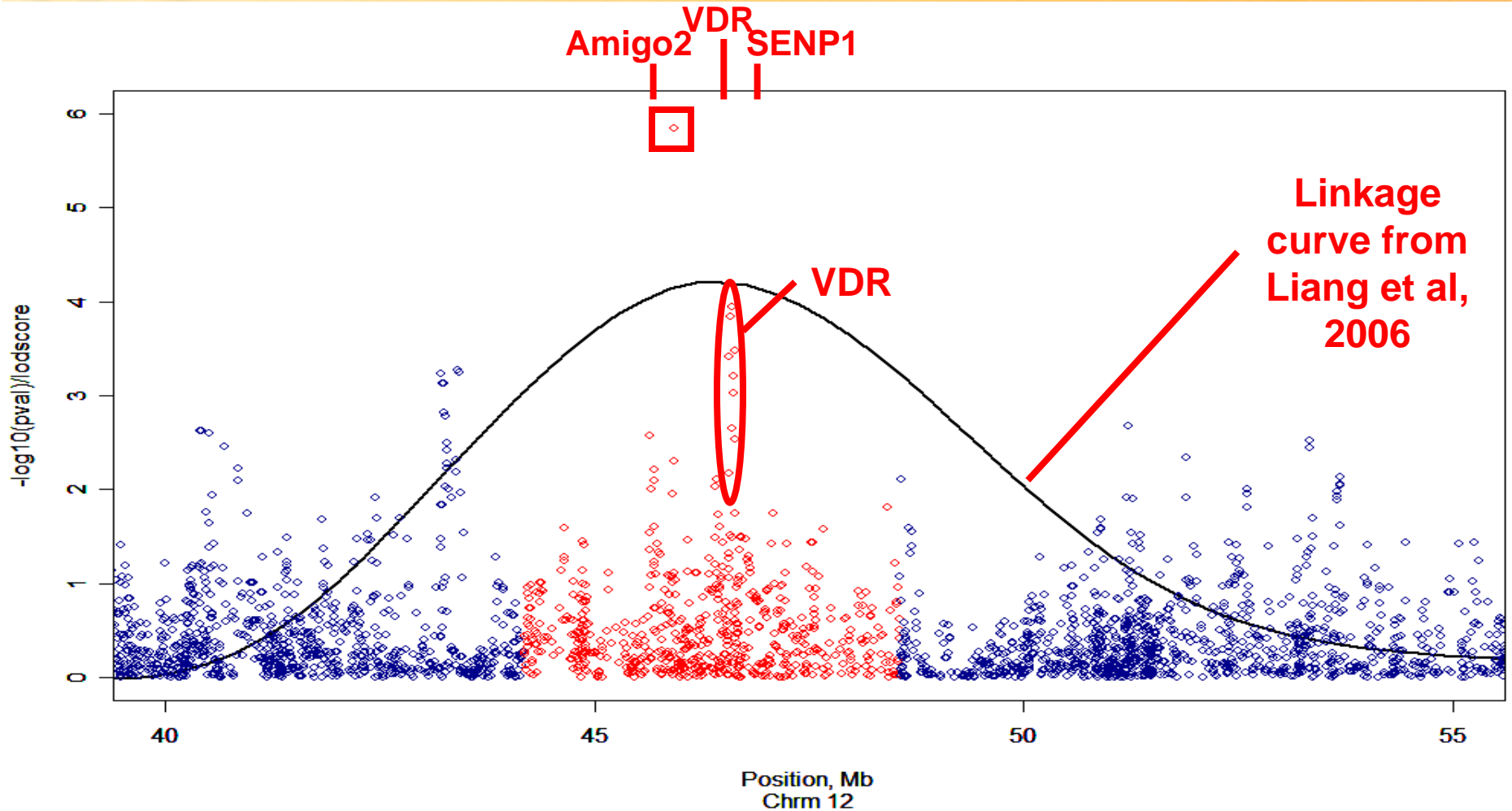
- **Chr 19: APOE Confirmed**
- Associations observed under previously observed linkage peaks



Top Non-APOE Hit- rs11610206

- SNP rs11610206 is top hit outside of APOE SNPs
 - $P = 1.43 \times 10^{-6}$ (genome-wide significance)
 - MAF = 0.085
 - Efficiency 100%
 - HWD p-value = 0.64
 - Located at 45.9 Mb, Chromosome 12
- Why is it of interest?
 - Lies under the linkage signals from several studies (independent datasets)
 - Near biological candidate *VDR* (Vitamin D Receptor) with positive associations in multiple association studies
 - Significant differences in expression of *VDR* between AD cases and controls (Xu, 2007)

Chromosome 12 Linkage Region



Liang et al., *Mol Psychiatry* Mar;11(3):280-5. 2006

Outline

- Background and Genetics of LOAD
- Genome-Wide Association
- Imputation and Meta-Analysis
- Imputation Results
- Conclusions

Imputation of SNPs in GWAS

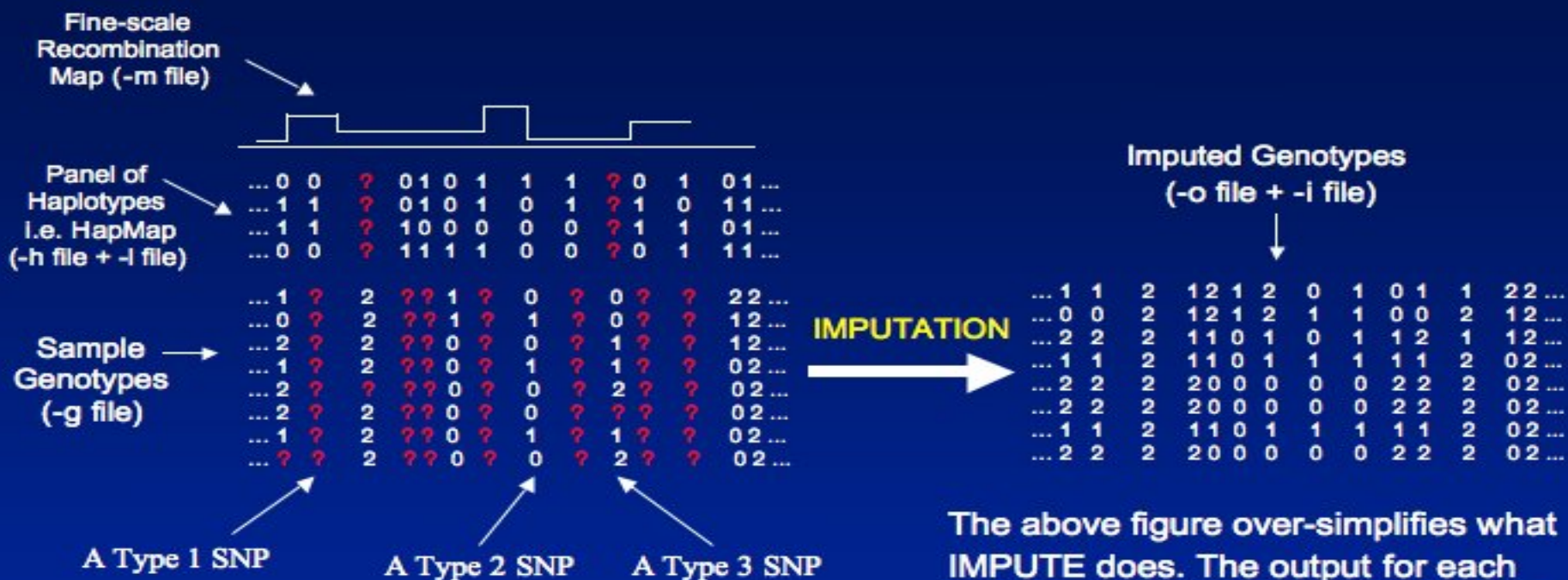
- What is SNP imputation?
 - Use recombination pattern information in a dense genotype map to impute genotypes in untyped SNPs in a less dense genotype map
 - E.g., use HapMap genotypes and known LD patterns to infer genotypes for SNPs that are not in the Illumina 550k HumanHap
- Several Advantages
 - More powerful than tagging
 - Tagging approaches test only single SNPs or small haplotypes of SNPs on a genotyping chip,
 - Provides increased resolution
 - Can be used as a multi-point test
 - **Facilitates meta-analysis**
 - **Allows datasets collected with different genotyping chips to be combined for increased power**

How Does Imputation Work?

- Estimate the probability of an unobserved genotype given the observed genotypes and a genetic model
 - Model provided by the reference haplotypes and recombination map
 - Genotypes consistent with local LD patterns are considered more likely
 - Marker information decreases with genetic distance (no need to set sliding windows, select SNP# etc to define haplotypes)
 - Estimated using a Hidden Markov Model (HMM)

Using IMPUTE for Imputation

Overview of Imputation using IMPUTE



Type 1 SNP : in haplotypes file only
 Type 2 SNP : in haplotypes and genotypes file
 Type 3 SNP : in genotypes file only

The above figure over-simplifies what IMPUTE does. The output for each genotype is a probability distribution on genotypes i.e.

| | | |
|------|------|------|
| 0 | 1 | 2 |
| 0.01 | 0.18 | 0.81 |

This captures the uncertainty in the prediction.

Benefits of Imputation

- Accurate (based on imputation of known genotypes from WTCCC data)
- Can fill in missing data
- Can improve coverage in genomic regions

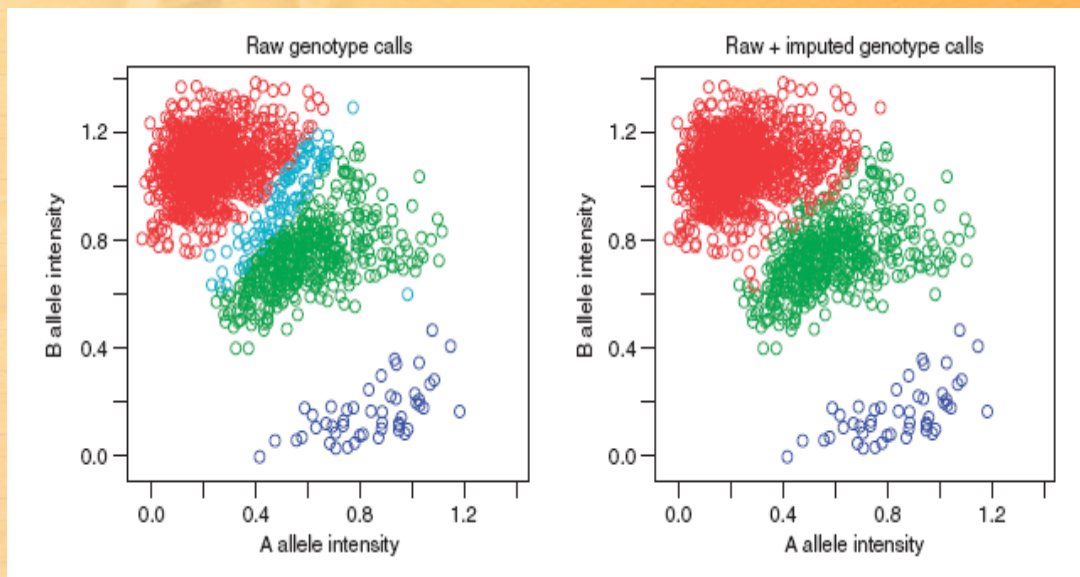


Table 1. Estimates of Genomic Coverage for Currently Available Genome-wide SNP Platforms Alone and after Imputation

| | Percentage of Genomic Coverage at $r^2 \geq 0.8$ | Percentage of Genomic Coverage at $r^2 = 1$ |
|--|--|---|
| Affymetrix SNP Array 5.0 | 65 | 43 |
| Affymetrix SNP Array 5.0 plus imputed SNPs | 73 | 54 |
| Affymetrix SNP Array 6.0 | 80 | 59 |
| Illumina HumanHap 300 | 77 | 42 |
| Illumina HumanHap 300 plus imputed SNPs | 81 | 50 |
| Illumina HumanHap 550 | 87 | 57 |
| Illumina HumanHap 1M | 91 | 68 |

Limitations to Imputation

- Potential Problems
 - Assumes LD structure is the same between reference and sample
 - Potential sampling error for reference
 - Studies suggest it is robust against misspecification
 - Cases **SHOULD** be different in a region of association
 - Assumes uniform mutation rate, no insertions/deletions, similar recombination rates between reference and sample
 - Potential bias in missing data problems
- Multiple Testing- generating >2 million SNP genotypes
- Coverage
 - When combining datasets, if SNPs on two different platforms are too far apart, none of the intermediate SNPs will be imputed with confidence
 - Any signal in the low coverage area is not found

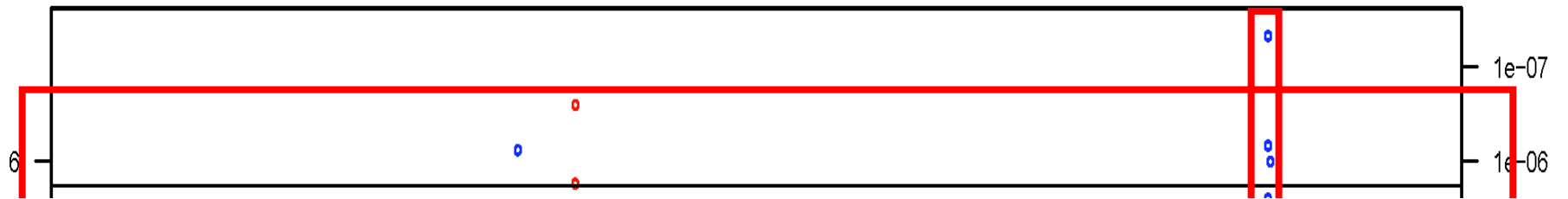
Imputation in the LOAD GWAS

- Data was acquired from a previous GWAS on LOAD cases and controls that used Affy 500K (Reiman et al. 2007)
 - Partial overlap with Illumina 550K panel
 - For all cases and controls, imputed 2.5 million SNPs identified in the HapMap CEU trios
- Implemented two strategies:
 - 1) Examining/comparing the strongest associations from both studies ($P < 0.0001$)
 - 2) Examining nominally associated markers within candidate genes ($P < 0.05$)

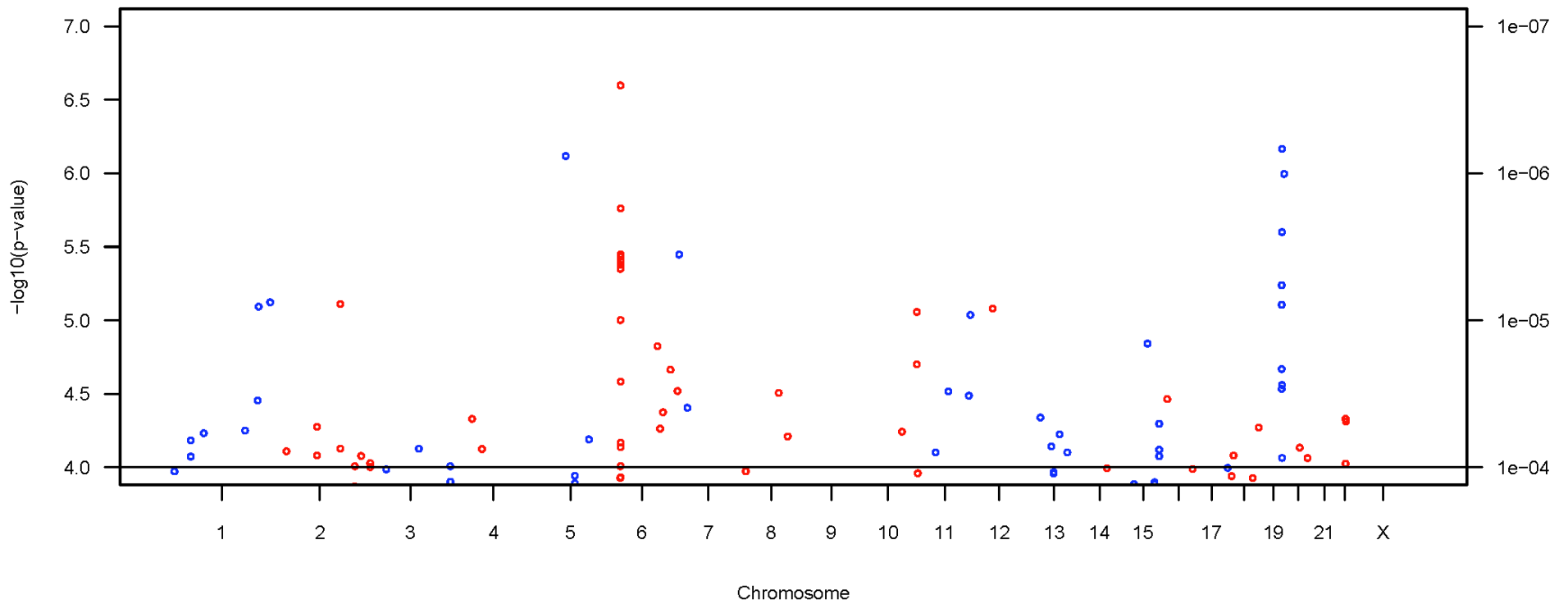
Outline

- Background and Genetics of LOAD
- Genome-Wide Association
- Imputation and Meta-Analysis
- Imputation Results
- Conclusions

Results- Imputed Data

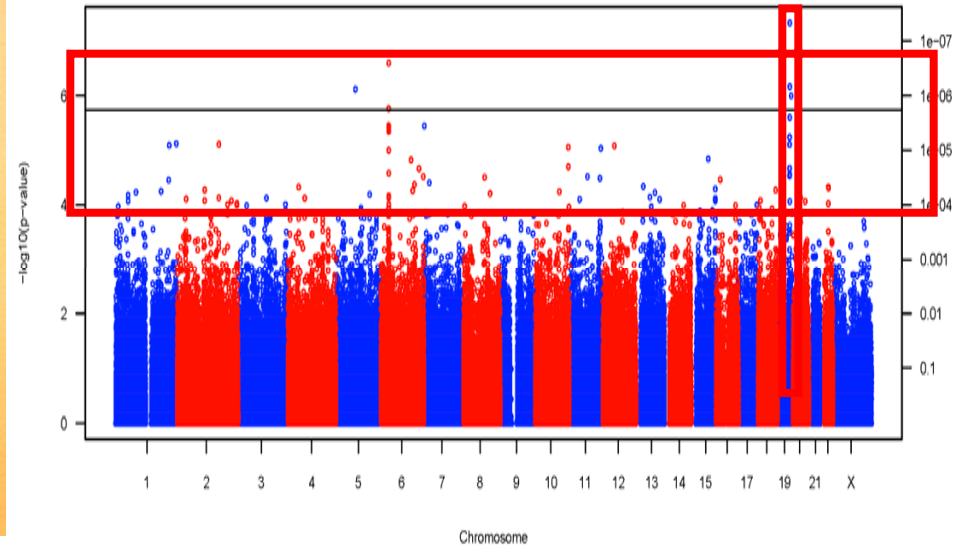


Whole Genome Association, top p values



Results- Imputed Data

- Chr 19 hits
- Chr 1 hit in *DISC1*

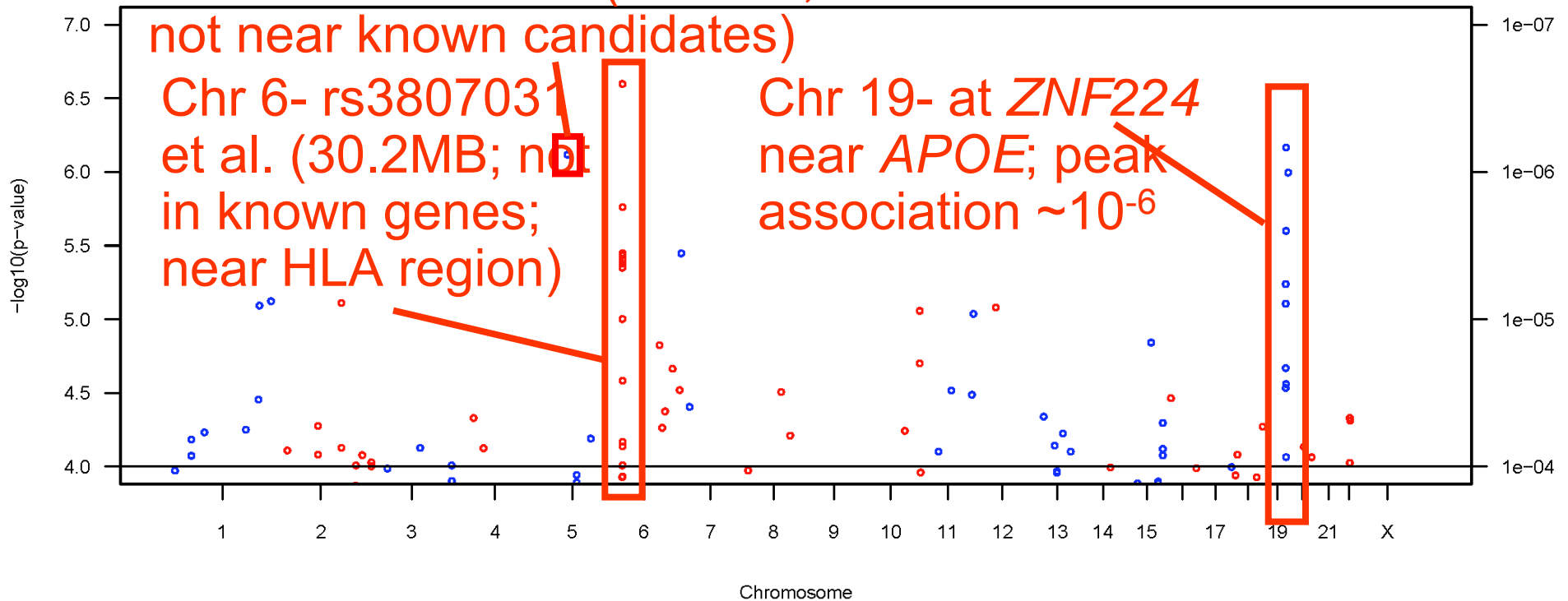


Whole Genome Association, top p-values

Chr 5- rs10474567 (78.1MB;
not near known candidates)

Chr 6- rs3807031
et al. (30.2MB; not
in known genes;
near HLA region)

Chr 19- at *ZNF224*
near *APOE*; peak
association $\sim 10^{-6}$



Joint Analysis with Reiman GWAS

- Jointly analyzed discovery dataset with 550K (Affy) GWAS dataset examined in Reiman et al. (2007)
- Several associations observed in both studies & joint analyses

| SNPs | CHR:BP | Type | Original | Reiman | Joint |
|------------|-------------|------|----------|----------|----------|
| rs12044355 | 1:229910970 | R | 3.90E-05 | 0.008216 | 9.20E-06 |
| rs1425967 | 4:138508340 | R | 3.90E-05 | 0.01052 | 1.25E-05 |
| rs4416533 | 6:815 | | 3.61E-05 | 0.0101 | 1.13E-05 |
| rs13213247 | 6:815 | | | | |
| rs4508518 | 19:49303260 | I | 0.02627 | 4.37E-05 | 1.95E-06 |
| rs3746319 | 19:49304071 | B | 6.05E-05 | 0.02326 | 3.01E-06 |
| rs2061332 | 19:49305501 | B | 6.19E-05 | 0.03786 | 4.91E-06 |
| rs2061333 | 19:49306048 | I | 0.01745 | 2.51E-05 | 1.51E-06 |

SNPs in/near *ZNF224*

Intronic SNP of AD candidate *DISC1*



Associations at prior candidates

- Identified candidate genes with prior associations from the Alzheimer Research Forum AlzGene database*
- Nine candidate gene demonstrated nominally significant associations ($P < 0.05$) of SNPs in each GWAS and in joint analyses
 - P between 0.003-0.05 in individual analysis
 - P between 0.0001-0.01 in joint analysis

Joint Analysis with Reiman GWAS

- Nine AD candidate genes in AlzGene Compendium* showed associations in both datasets and joint analyses

– *ADAM12*

– *CSF1*

– *GBP2*

– *KCNMA1*

– *NOS2A*

– *SORCS2*

– *SORCS3*

– *SORL1*

– *WWC1*

Related to *SORL1*:
GBP2 (guanylate-binding protein 2)
SORCS2 (sortilin-related VPS10 domain containing receptor 2)
upregulated in hippocampus; has shown previous association

SORCS3 (sortilin-related VPS10 domain containing receptor 3)

WWC1 (WW and C2 containing domain 1)
previous association in Spanish population; associated with memory performance based on a task

SORL1 (Sortilin-related receptor)
associated in multiple LOAD populations

Reiman et al., *Neuron*
2007 Jun 7;54(5):713-20



* <http://www.alzforum.org/res/com/gen/alzgene/>

GWAS Imputation Findings

- Strong associations were observed on Chr 19 for SNPs in *ZNF224* (P from Joint = 1.51×10^{-6} – 4.91×10^{-6})
- Nine candidate genes previously associated with AD were observed to have SNPs with nominal associations ($P < 0.05$) in both GWAS datasets and in the joint analysis

Outline

- Background and Genetics of LOAD
- Genome-Wide Association
- Imputation and Meta-Analysis
- Imputation Results
- Conclusions

Conclusions

- *APOE* associations observed in discovery GWAS dataset
- Most significant SNP association outside of *APOE*: rs11610206 on 12q13
 - Under Liang et al (2006) linkage signal
 - Near the *VDR* gene, a biological candidate gene with multiple positive associations and supporting expression data
- These results provide evidence for an Alzheimer disease risk locus on 12q13

Conclusions

- Imputation is an efficient approach to using existing genotype data and patterns of linkage disequilibrium on each chromosome to infer genotypes at untyped SNPs in GWAS
- Application of imputation identified a strong associations on Chr 1, 4, 6, and 19.
- Nine candidate genes previously associated with AD were observed to have SNPs with nominal associations ($P < 0.05$) in both GWAS datasets and in the joint analysis

Acknowledgements

- AD patients and their families for participating in the study
- Clinical and Research staff of the Miami Institute for Human Genomics
 - Adam Naj
 - Margaret Pericak-Vance
 - Jonathan Haines
 - John Gilbert
 - Stephan Zuchner
 - Patrice Whitehead
 - Odette Gordon
 - Kenju Hara
 - Eden Martin
 - Paul Gallins
 - Monica Polk
 - Karen DeSombre
 - Robin Toft
 - Doris Caldwell
- This work was supported by grants from
 - The Alzheimer Association (ZEN-01-2935, IIRG-00-2006)
 - The National Institutes of Health (NIA and NINDS) (AG021547 and AG19757)
 - NCRAD is supported by a cooperative agreement grant (U24AG21886) from the National Institute of Aging