Understanding the contribution of copy number polymorphisms to multigenic traits

Rob Scharpf

October 25, 2013

Serum uric acid concentrations

- serum uric acid concentrations are highly heritable
- high levels of serum uric acid (hyperuricemia) is associated with gout, cardiovascular disease, and renal complications
- SNPs in *SLC2A9* and *ABCG2* increase the risk for hyperuricemia and gout

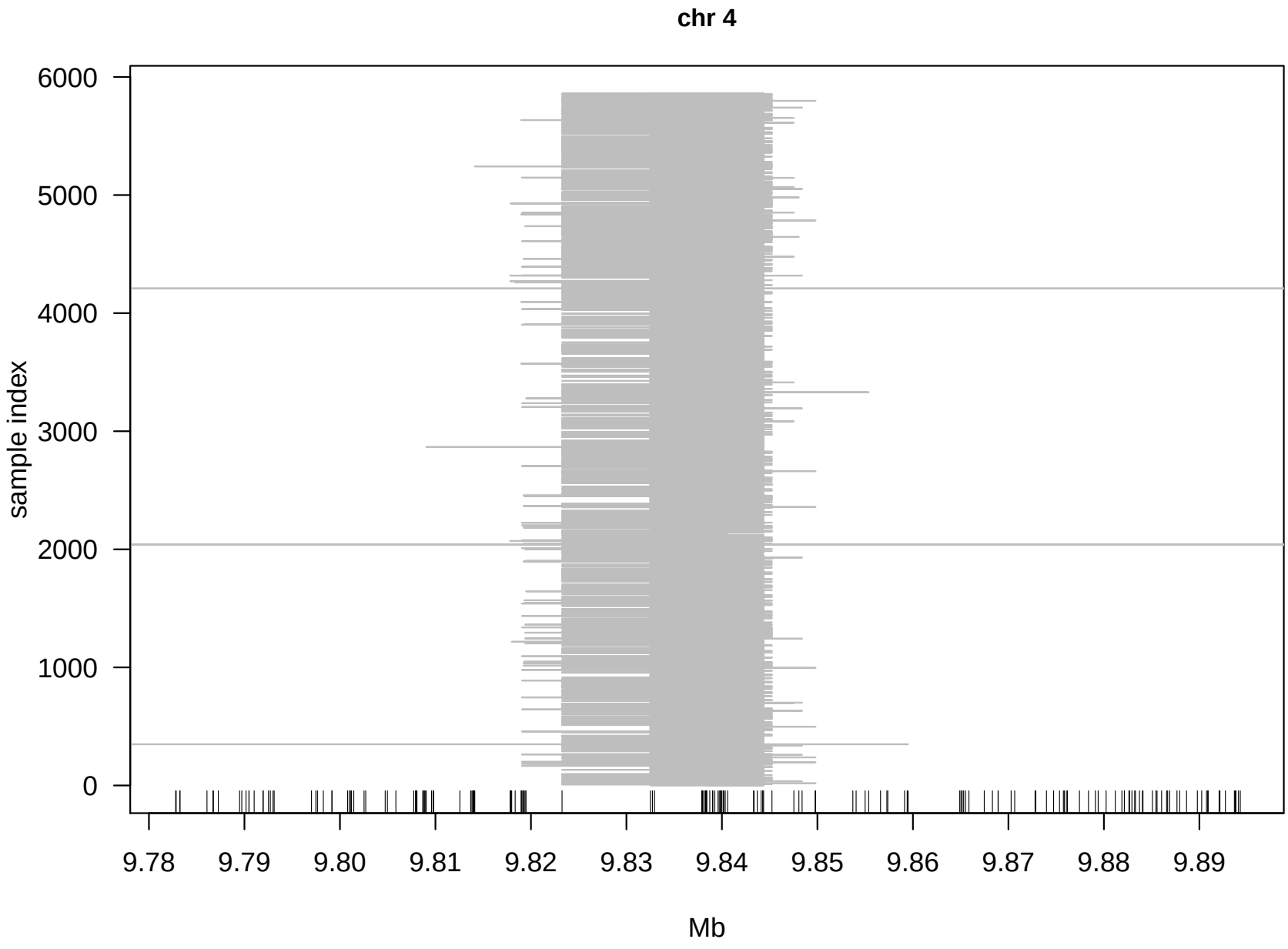
Atherosclerosis Risk in Communities Study

- 8,598 European ancestry (EA), 3,392 African Americans (AA) passing QC
- IQR age: 49-59 (median 54)
- 47% of EA and 38% of AA participants are male

Copy number estimation

1. Preprocess Affy 6.0 CEL files (R package `crlmm`)
2. Fit 6-state HMM to each sample (R package `VanillaICE`)
   $\rightarrow$ genomic intervals with estimated copy number

Copy number polymorphism (CNP)



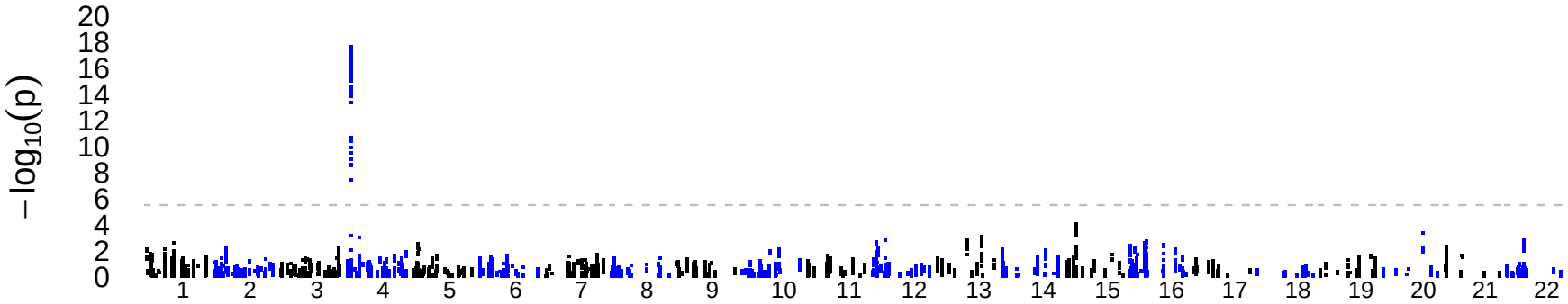
Copy number estimation

3. Translate list of CNVs to a rectangular matrix where rows are disjoint genomic intervals and columns are samples

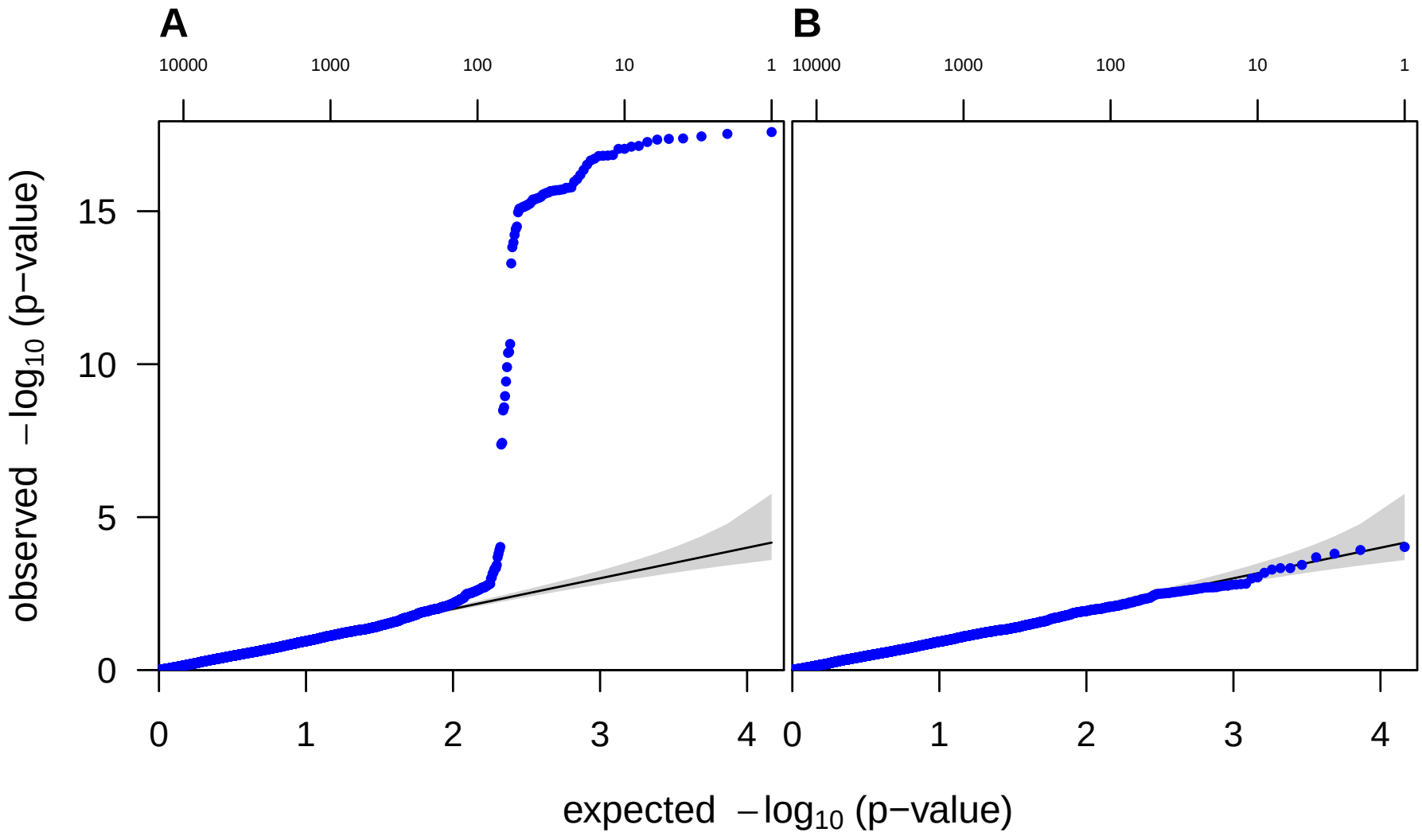
4. Define copy number polymorphisms (CNPs) as intervals for which at least 1% of ARIC participants have a CNV
    - 14,678 such intervals in ARIC corresponding to approximately 434 regions

5. Regress uric acid on copy number estimates

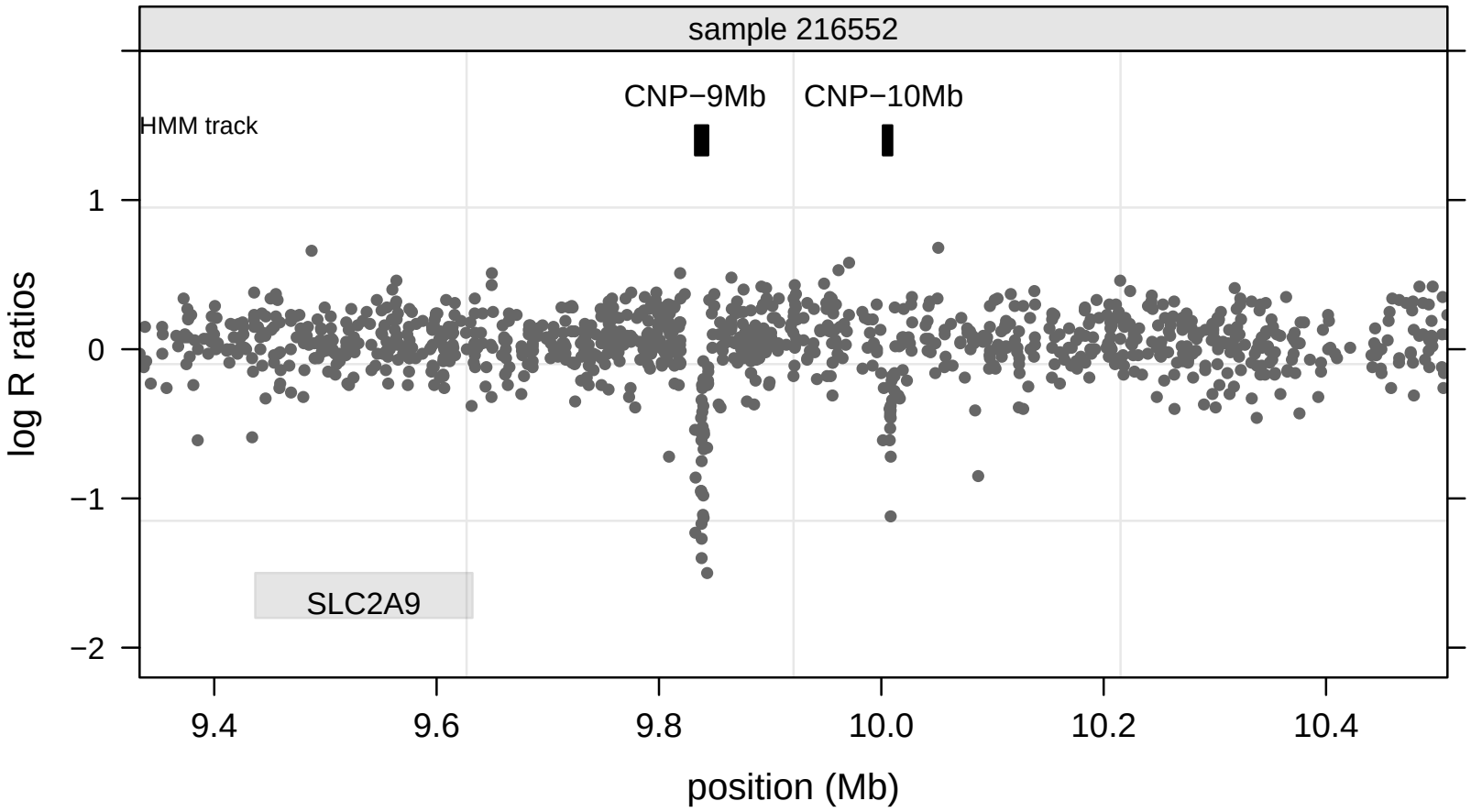
Copy number GWAS



Copy number GWAS

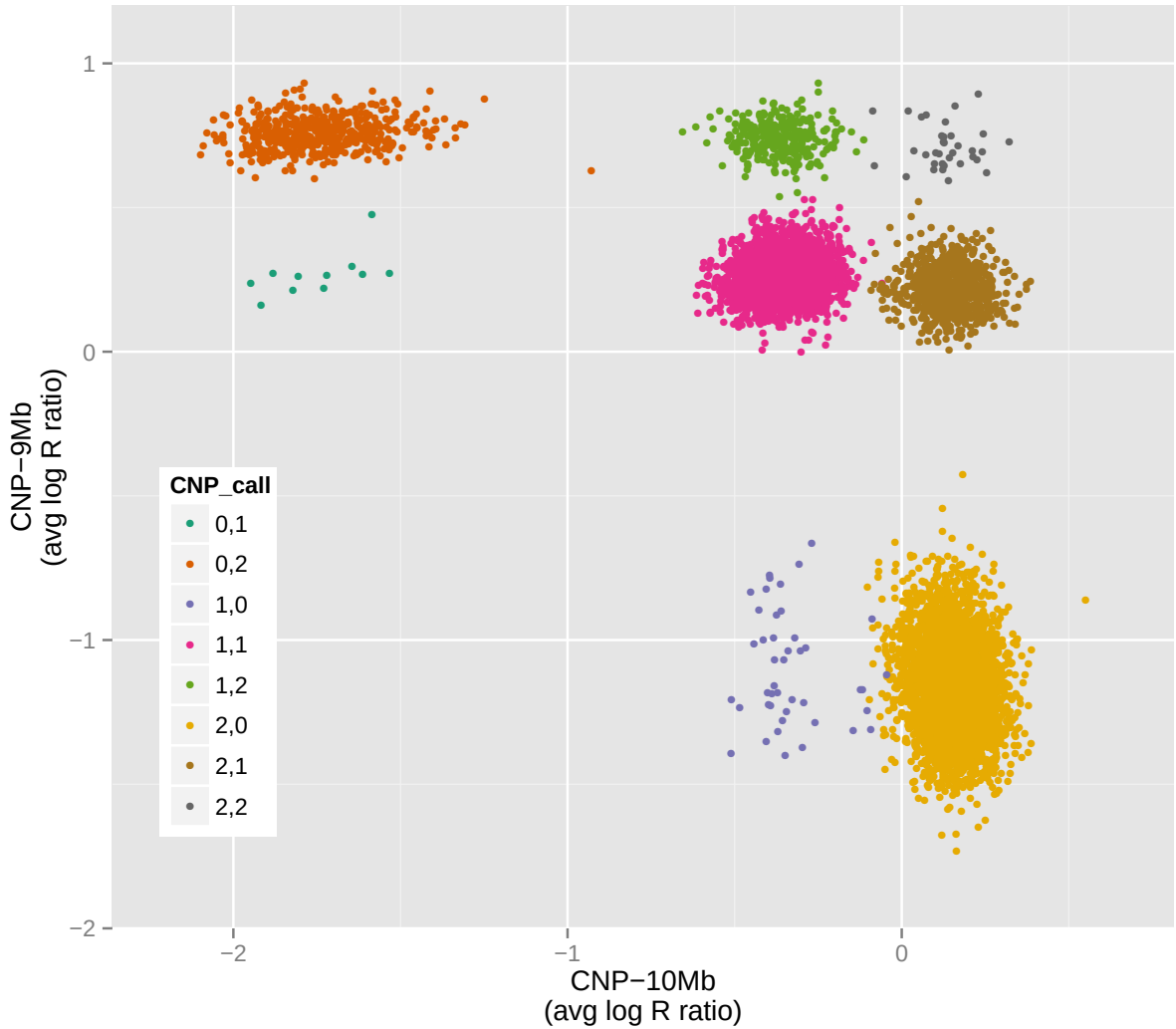


Signal is from two non-overlapping CNPs

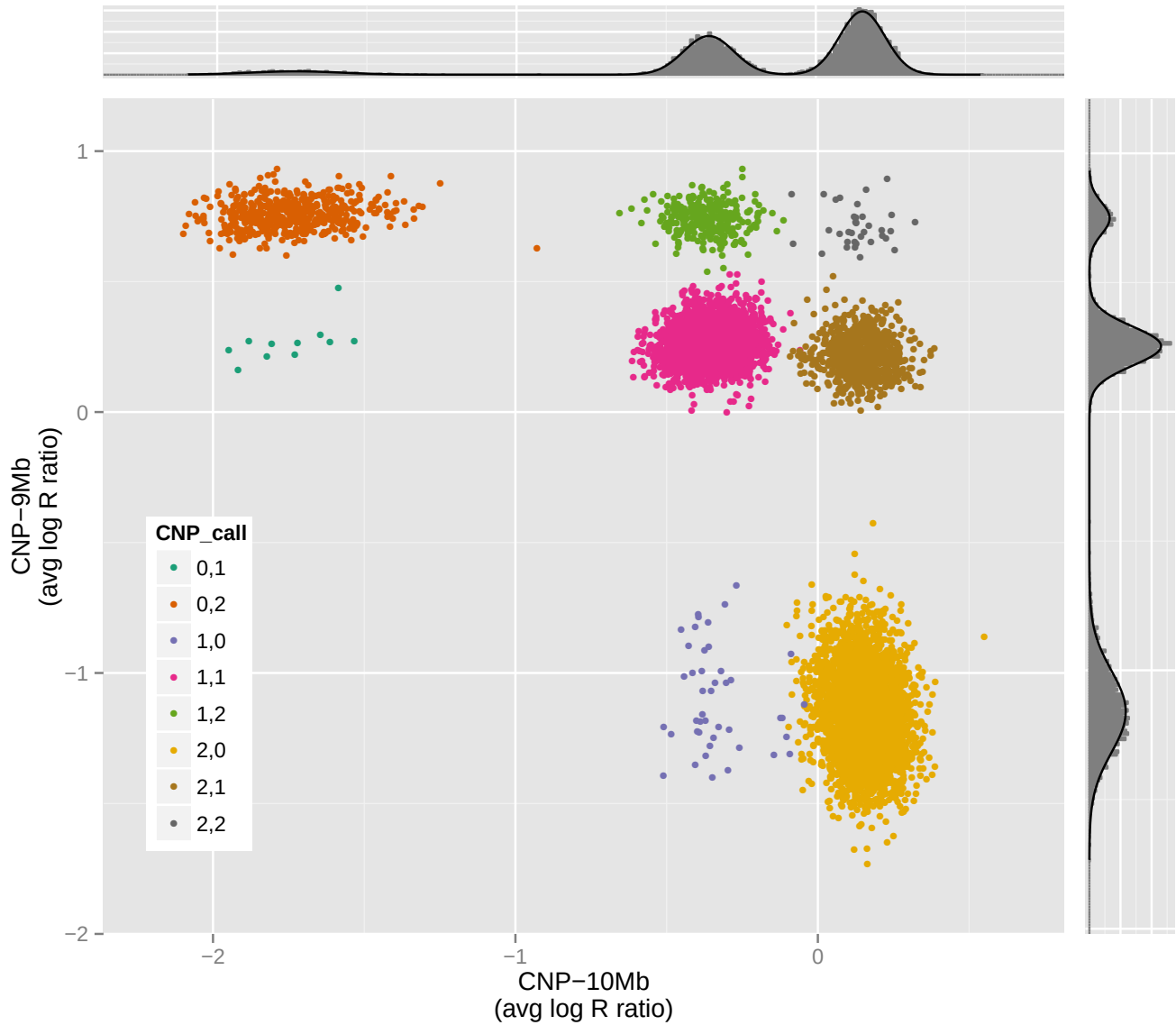


- CNP regions are small and data is noisy, but CNPs are near *SLC2A9*
- *SLC2A9* is transcribed in the reverse direction

CNP-10Mb vs CNP-9Mb log R ratios

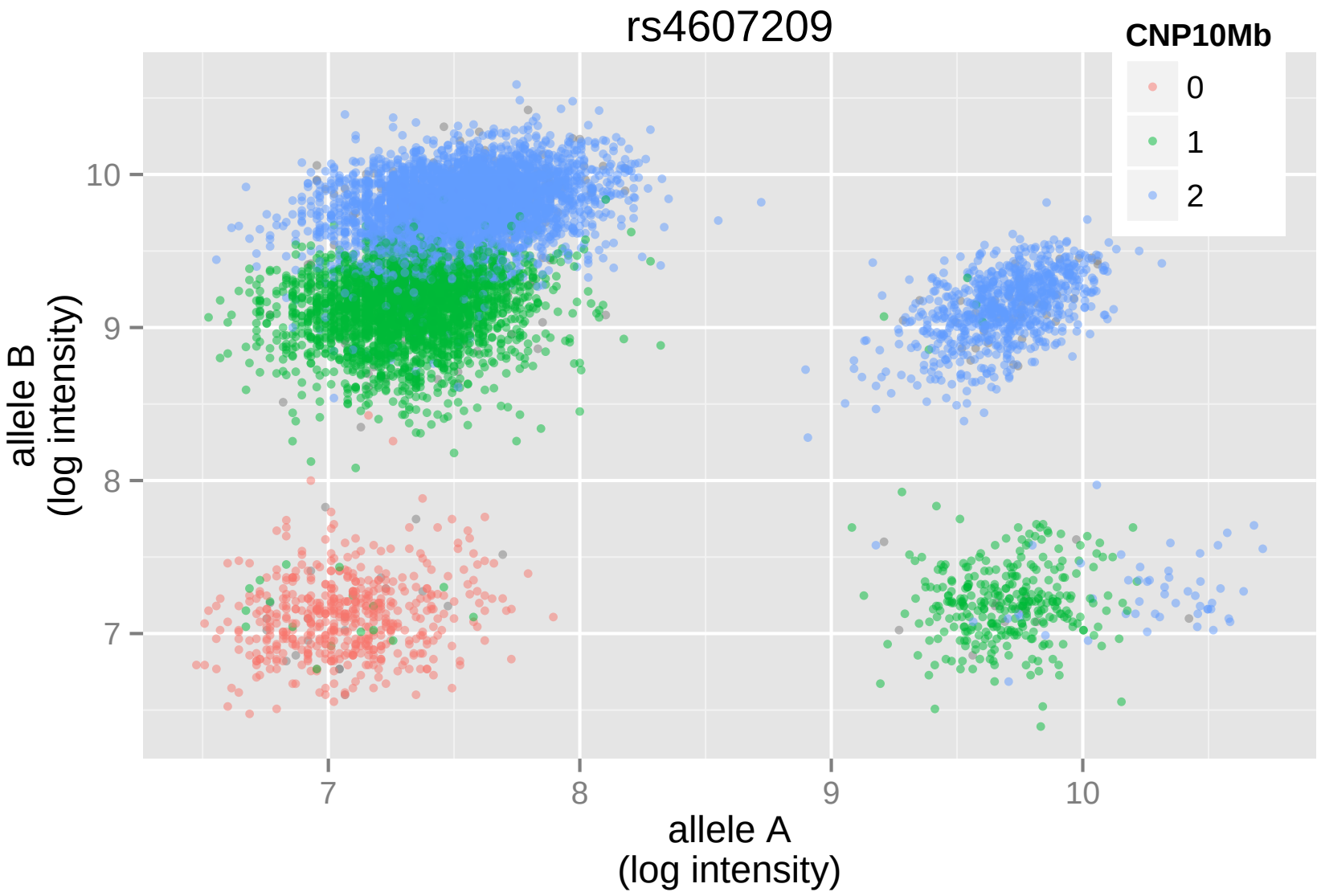


Fit a Bayesian mixture model to the marginals

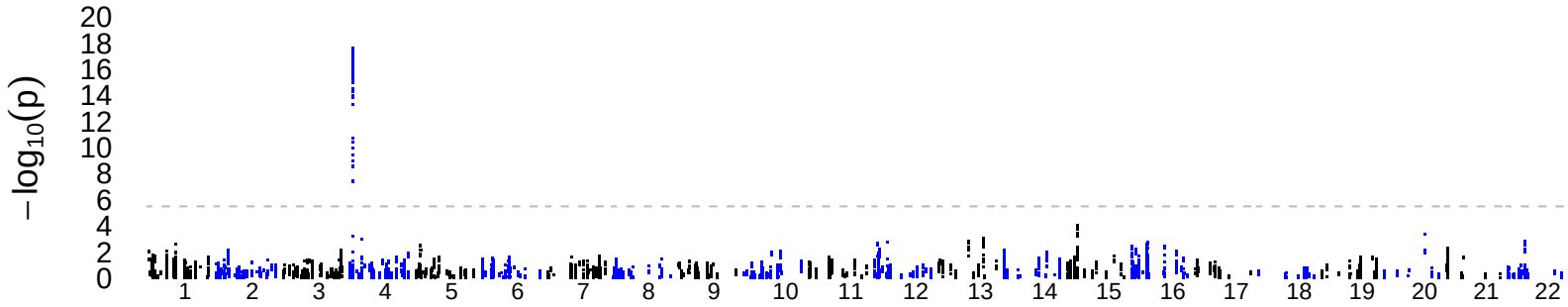


- EM implementation (Korn *et al.*, NG 2008)
- Bayesian hierarchical MM (Cardin *et al.*, Gen Epi 2011)

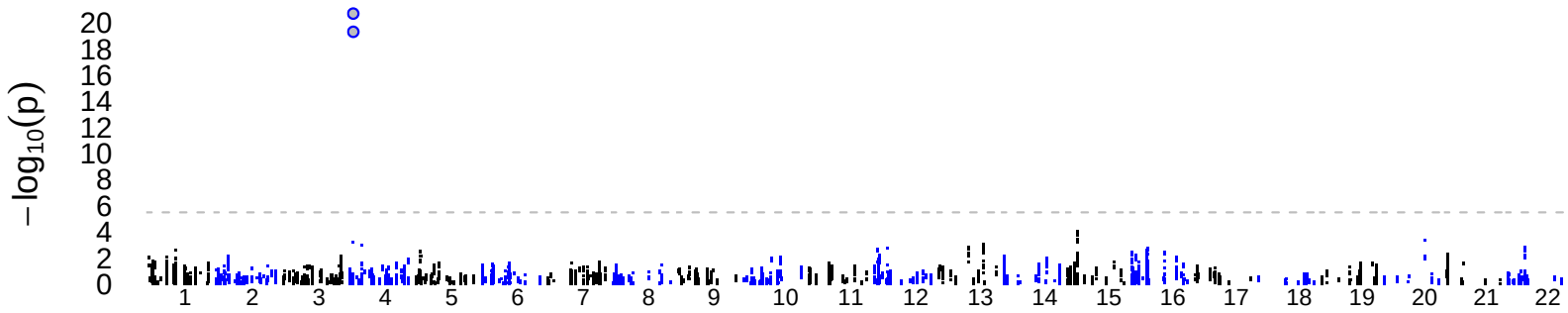
SNP in CNP-10Mb locus



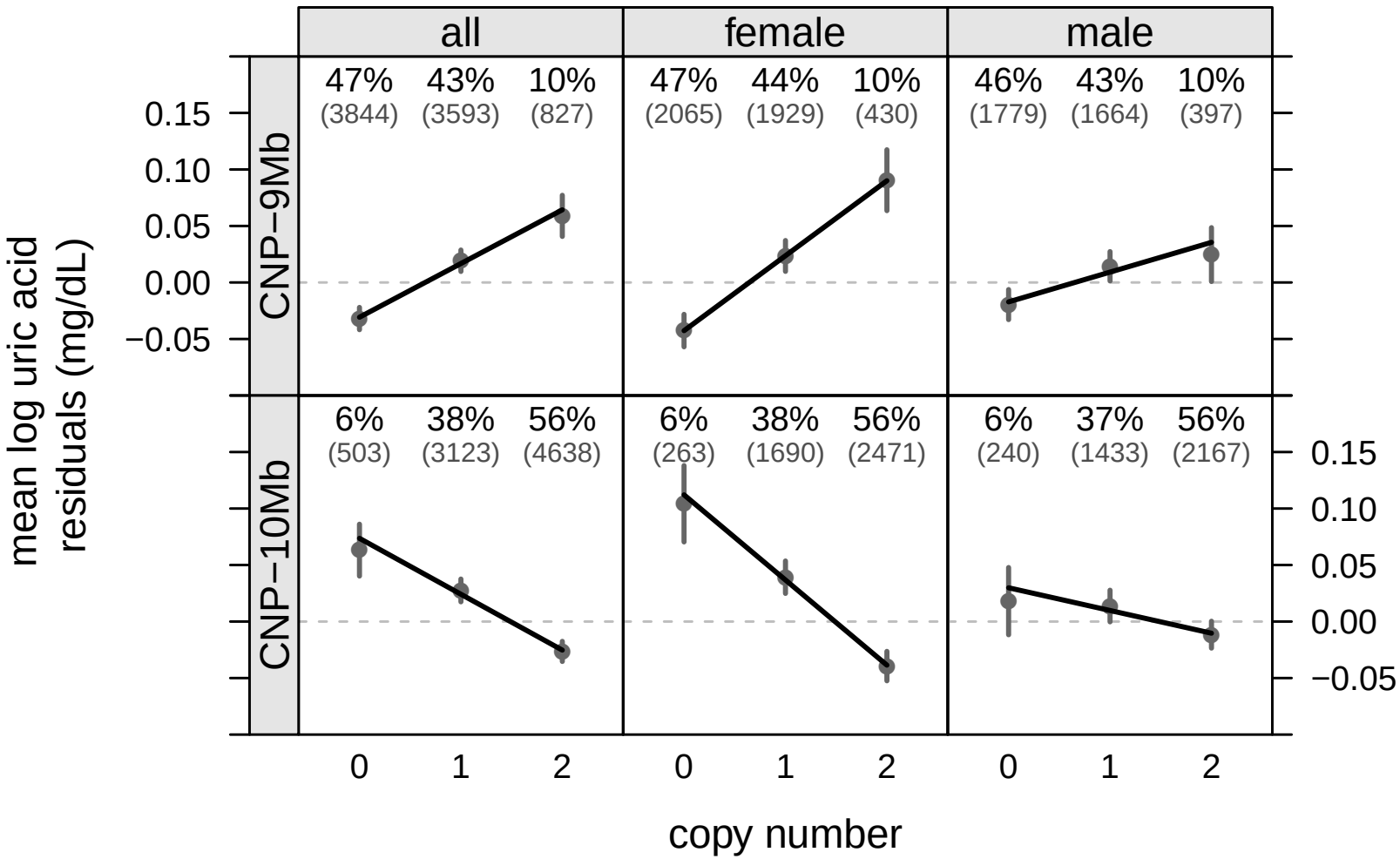
Association from HMM estimates



Association from mixture model



Slopes are gender-specific



- Copy number estimates are negatively correlated
- Relationship between log (uric acid) and copy number is approximately linear

Next steps

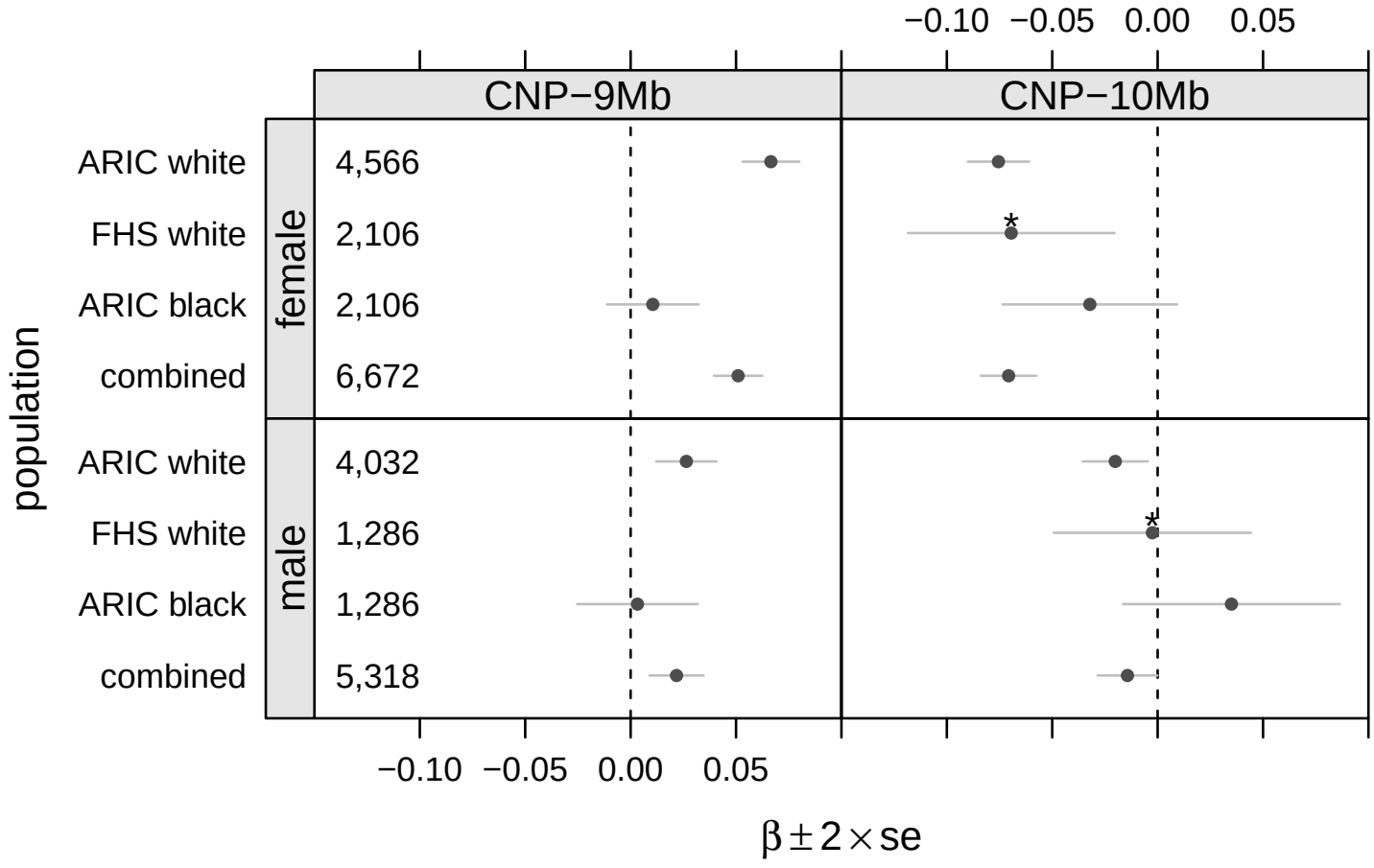
1. Does the CNP association replicate in other European ancestry cohorts?

2. Is the CNP association independent of previously known SNPs in *SLC2A9*?

Replication in Framingham Heart Study

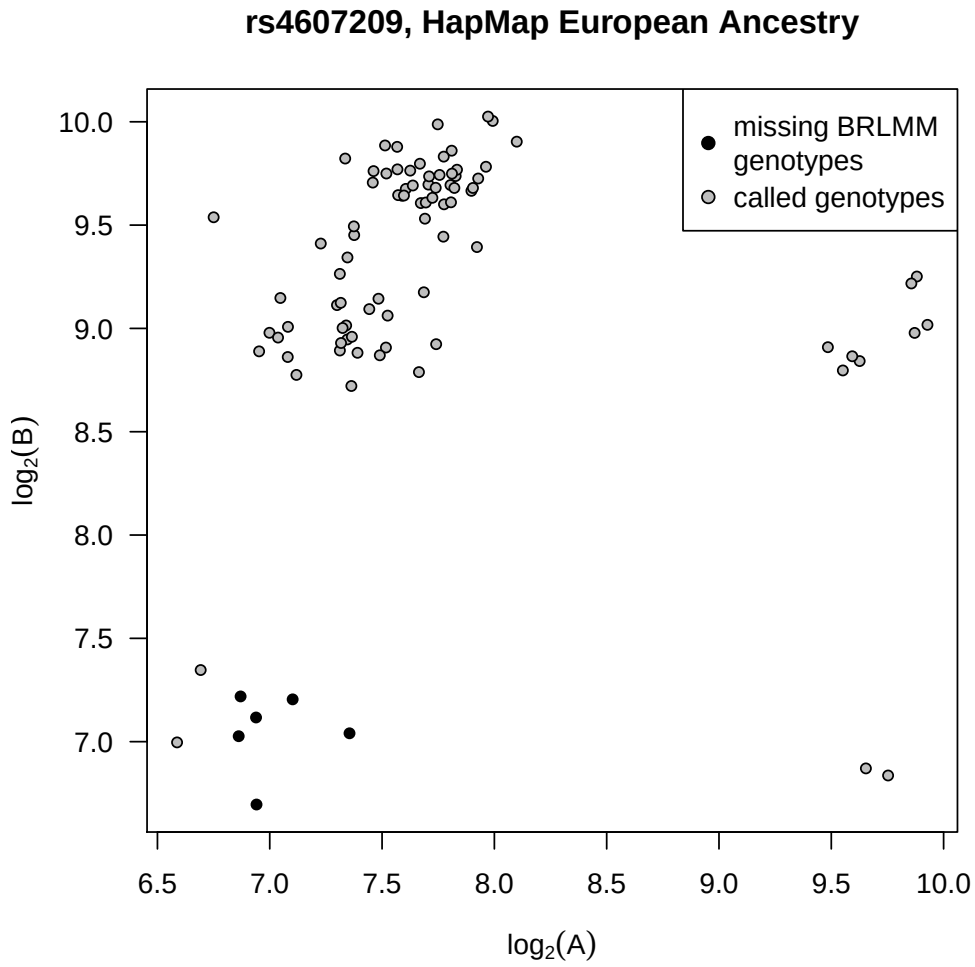
- Family-based study with 2,106 women and 1,286 men
- Older array technology (Affy 250k)
  - No markers in the CNP-9Mb locus
  - One SNP in the CNP-10Mb locus
- Only genotype calls available

Replication in Framingham Heart Study



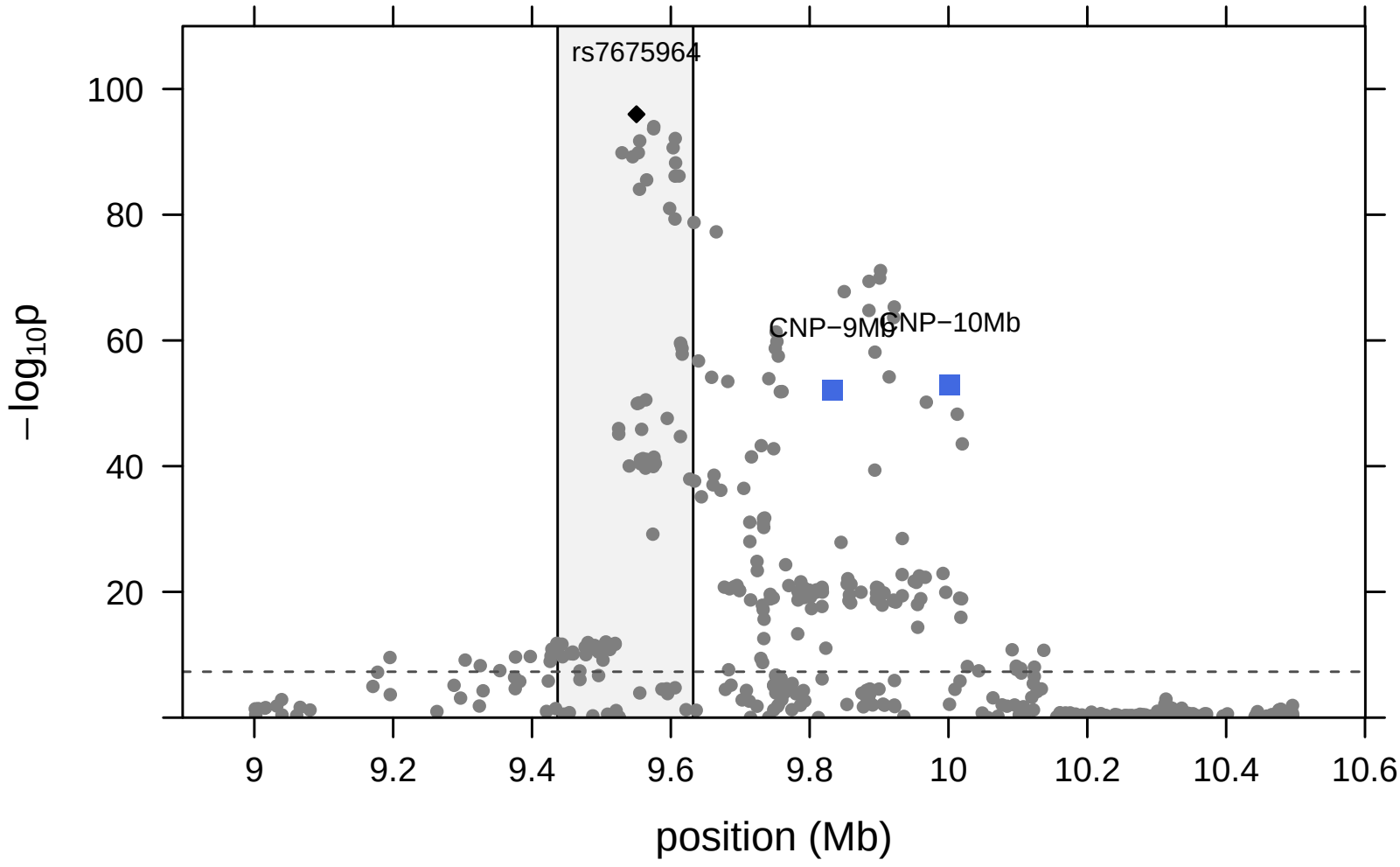
* Missing genotype calls used as a surrogate for deletion genotypes

EA HapMap data for 250k chip

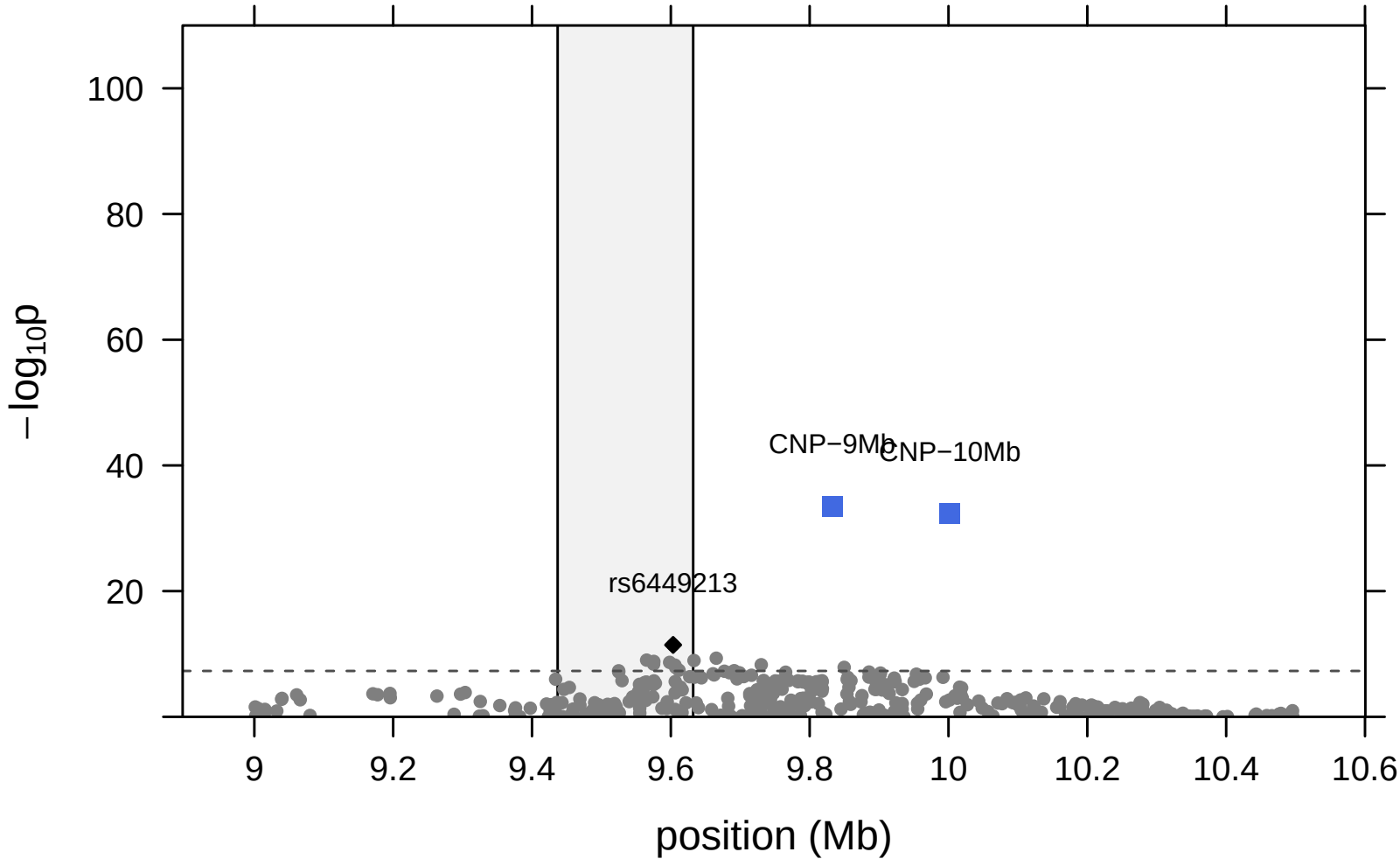


- Genotype calls in HapMap using the BRLMM algorithm that was used to genotype Framingham participants

SNP and CNP associations near the *SLC2A9* locus



Adjusted for rs7675964



Remarks regarding independence

- Reverse is also true: adjusted for CNP estimates, SNP rs7675964 remains genome-wide significant

Phasing SNP allelic haplotypes with copy number

- We phased copy number estimates at CNP-9Mb and CNP-10Mb with the rs7675964 and rs6449213 genotypes
- Example haplotype

Haplotype 1 : $--a--b--0--1--$
Haplotype 2: $--b--b--1--1--$

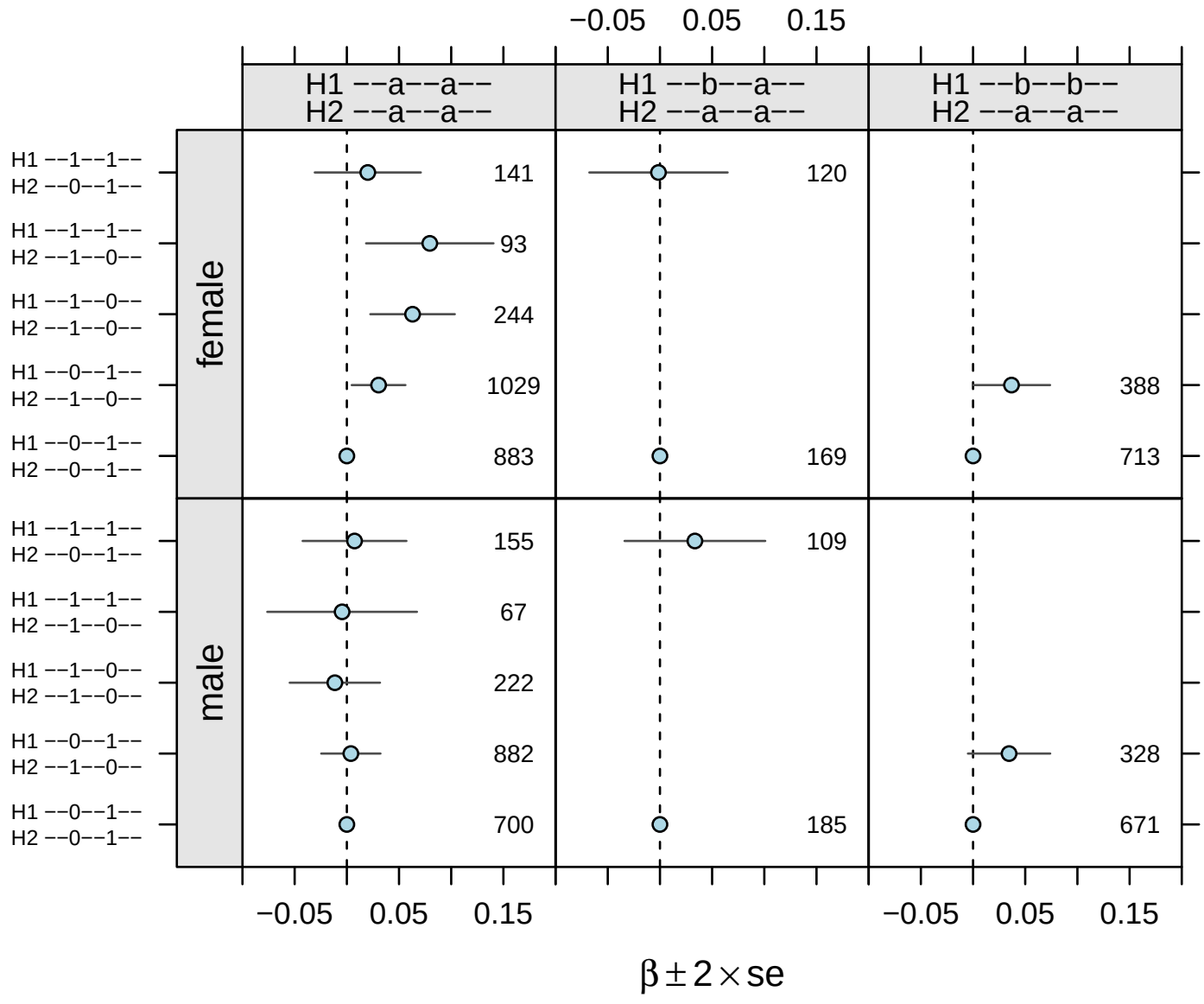
(Haplotype for subect that is heterozygous at SNP rs7675964, homozygous for the minor allele at SNP rs6449213, hemizygous deletion at CNP-9Mb, and diploid at CNP-10Mb)

- Phasing performed by fastPHASE: Scheet and Stephens, PLoS Genetics (2008)
- Thanks to Dan Arking for the suggestion of phasing

Phasing SNP allelic haplotypes with copy number

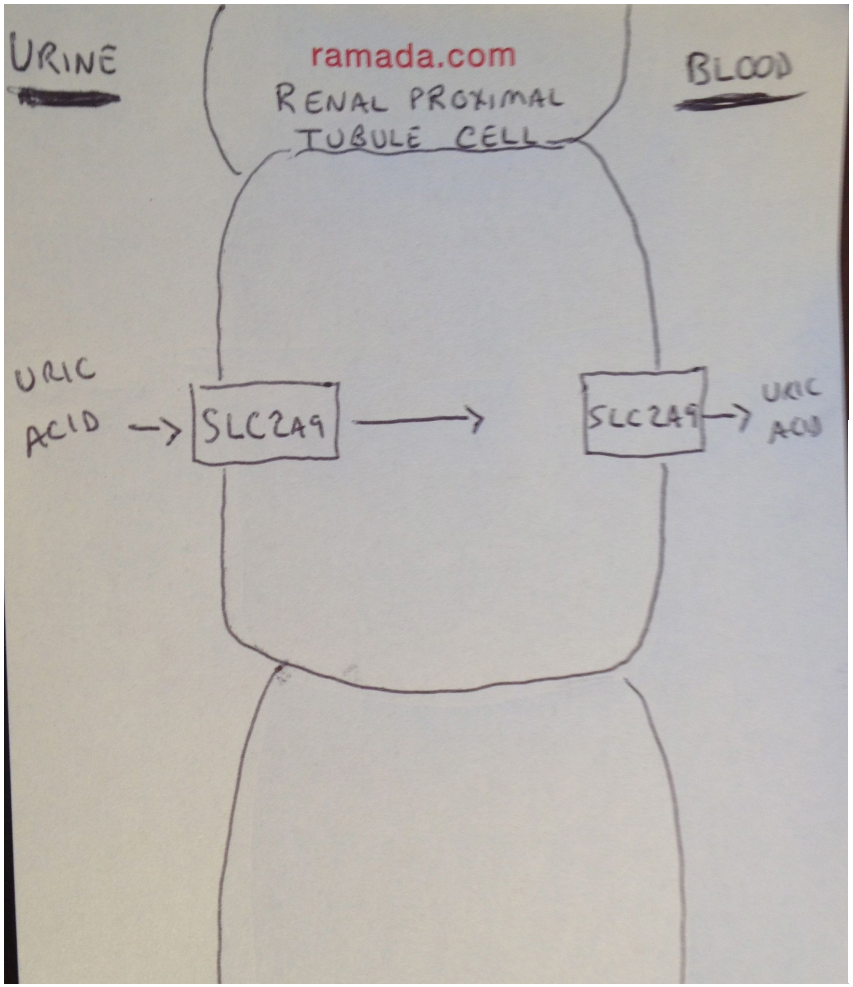
- Only three allelic haplotypes exhibit variation in the corresponding copy number haplotypes:
  1. H1:$--a--a--$
H2:$--a--a--$
  2. H1:$--b--a--$
H2:$--a--a--$
  3. H1:$--b--b--$
H2:$--a--a--$
- Among subjects with the same allelic haplotype, do we see variation in uric acid levels associated with copy number haplotypes?

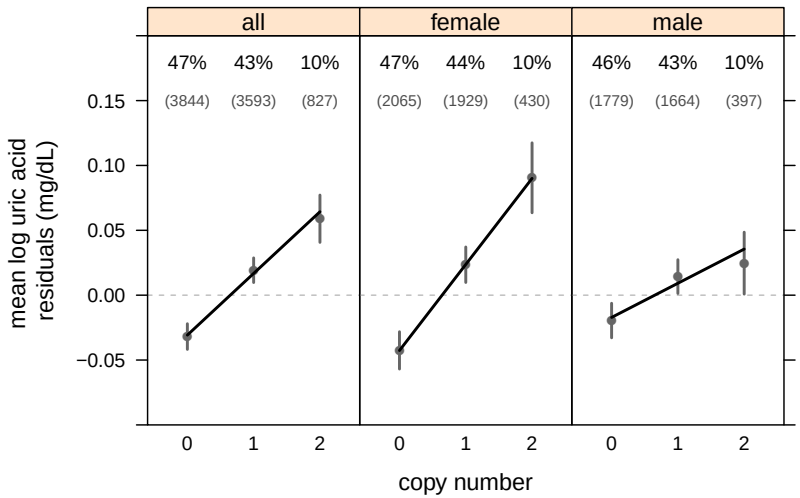
Association of CNP haplotypes



Discussion and Summary

- Deletion of an enhancer at CNP-9Mb is consistent with physiological role of *SLC2A9* in the kidney





Adapted from Woodward *et al.*, PNAS (2009)

Discussion and Summary

- Strong evidence that CNP-10Mb spans regulatory elements (DNase hypersensitivity/histone marks) in normal epithelial tissues (esophagus, mammary).
- No promoters have been reported at the CNP-9Mb locus and no regulatory elements have been published in the target tissues (kidney and liver) at either loci
- Copy number is needed to interpret ChIP-seq and DNAse hypersensitivity, particularly at CNP-9Mb
- Gene expression in target tissues and copy number is needed to evaluate whether deletions effect transcription of *SLC2A9* as hypothesized, and to evaluate gender differences in *SLC2A9* expression

Acknowledgements

- Linda Kao
- Lynn Mireles
- Adrienne Tin
- Stephen Cristiano
- Ingo Ruczinski
- Anna Köttgen
- Josef Coresh
- Aravinda Chakravarti
- Eric Boerwinkle
- Caroline Fox and Qiong Yang (Framingham Heart Study)
- Members of ARIC writing group

Contact: rscharpf@jhu.edu