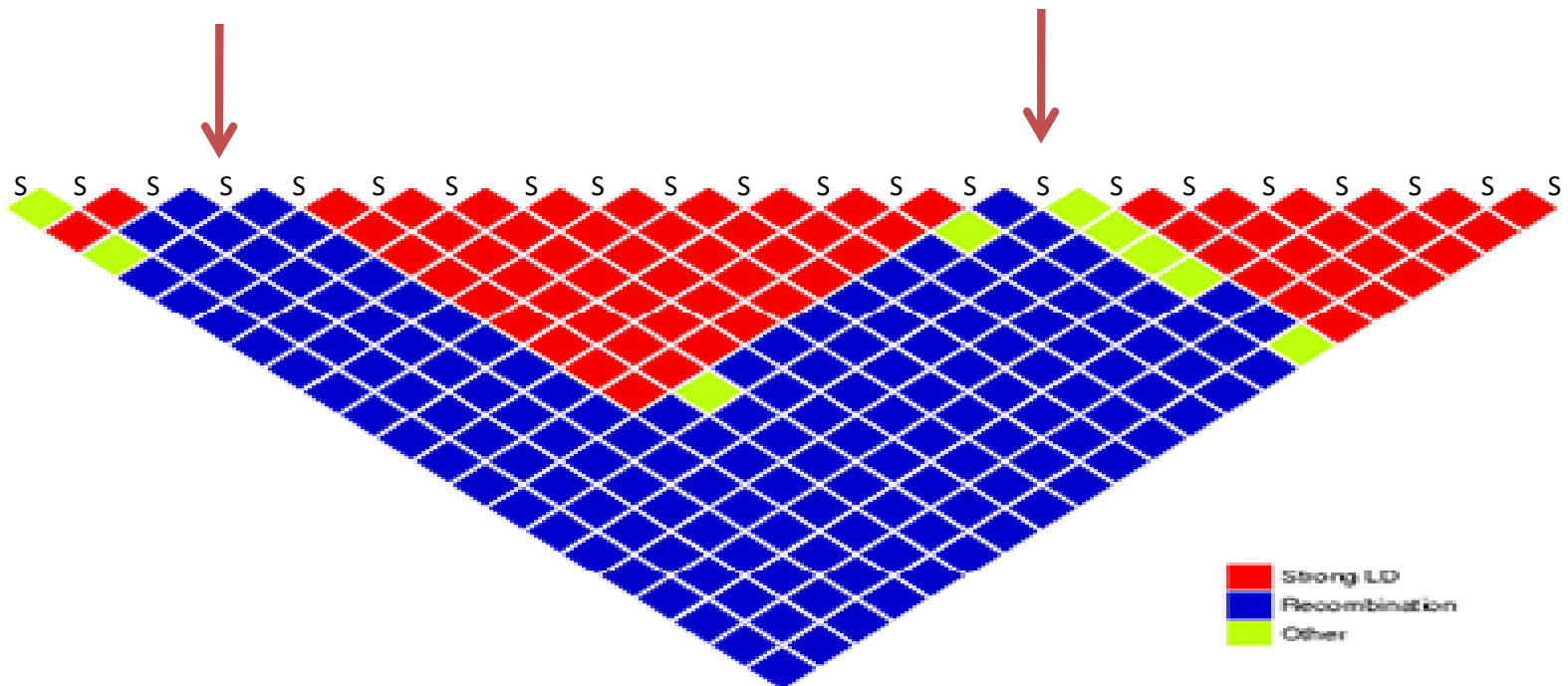


Singleton SNPs in the human
genome and
implications for genome-wide
association studies

Bioinformaatika 2008

Singleton SNP

SNPs that only tag themselves and dont contribute power to the rest of the region

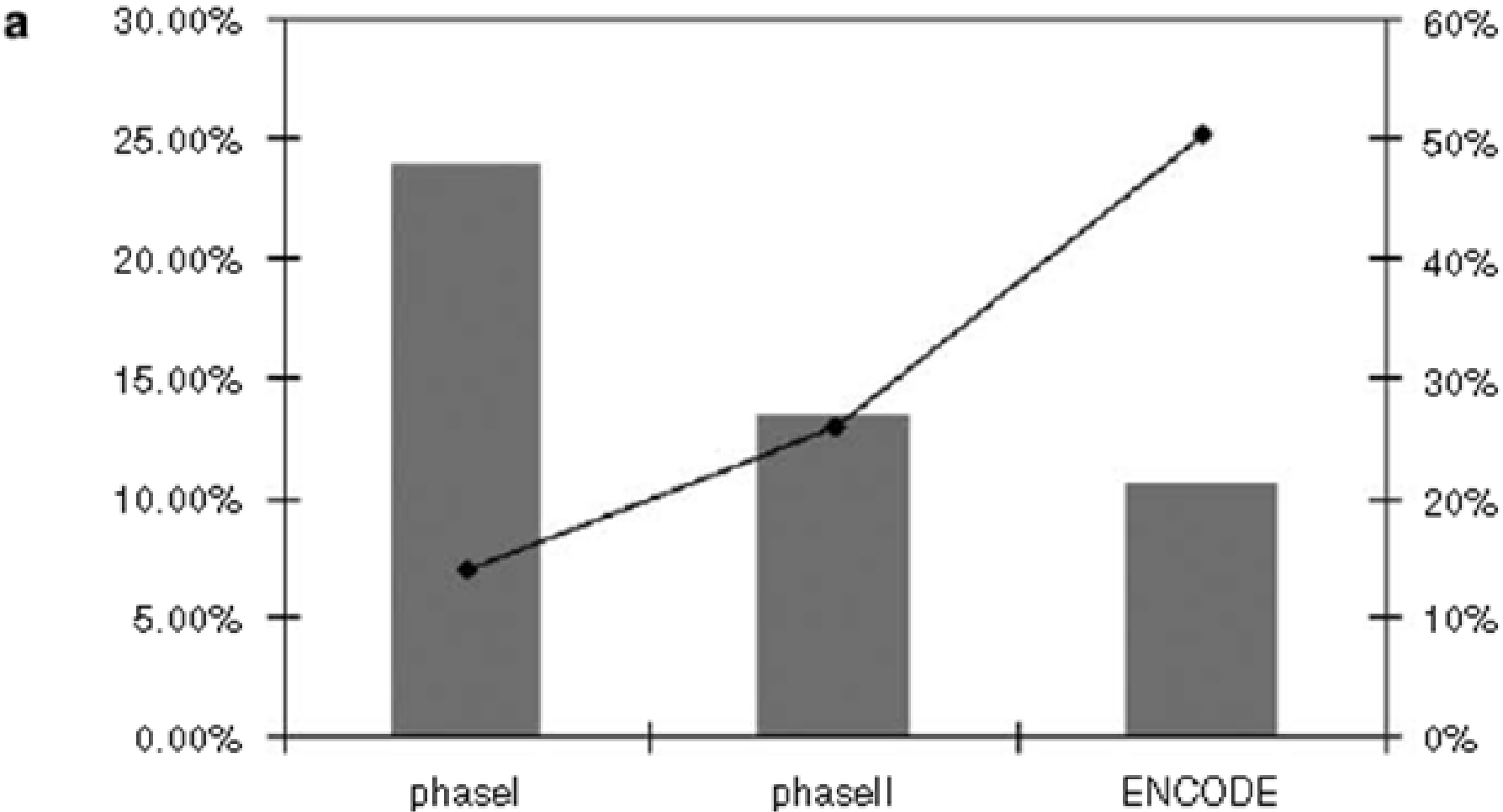


SNP genotyping and LD

- In the Human Genome ~ one SNP per 300 bp
- LD between can be used to save \$\$\$
- Large proportion of SNPs turn out to be singletons i.e. not in LD with any neighbouring SNP
- These are the first ones to be removed to save more \$\$\$
- Is it wise strategy?

Topics of interest:

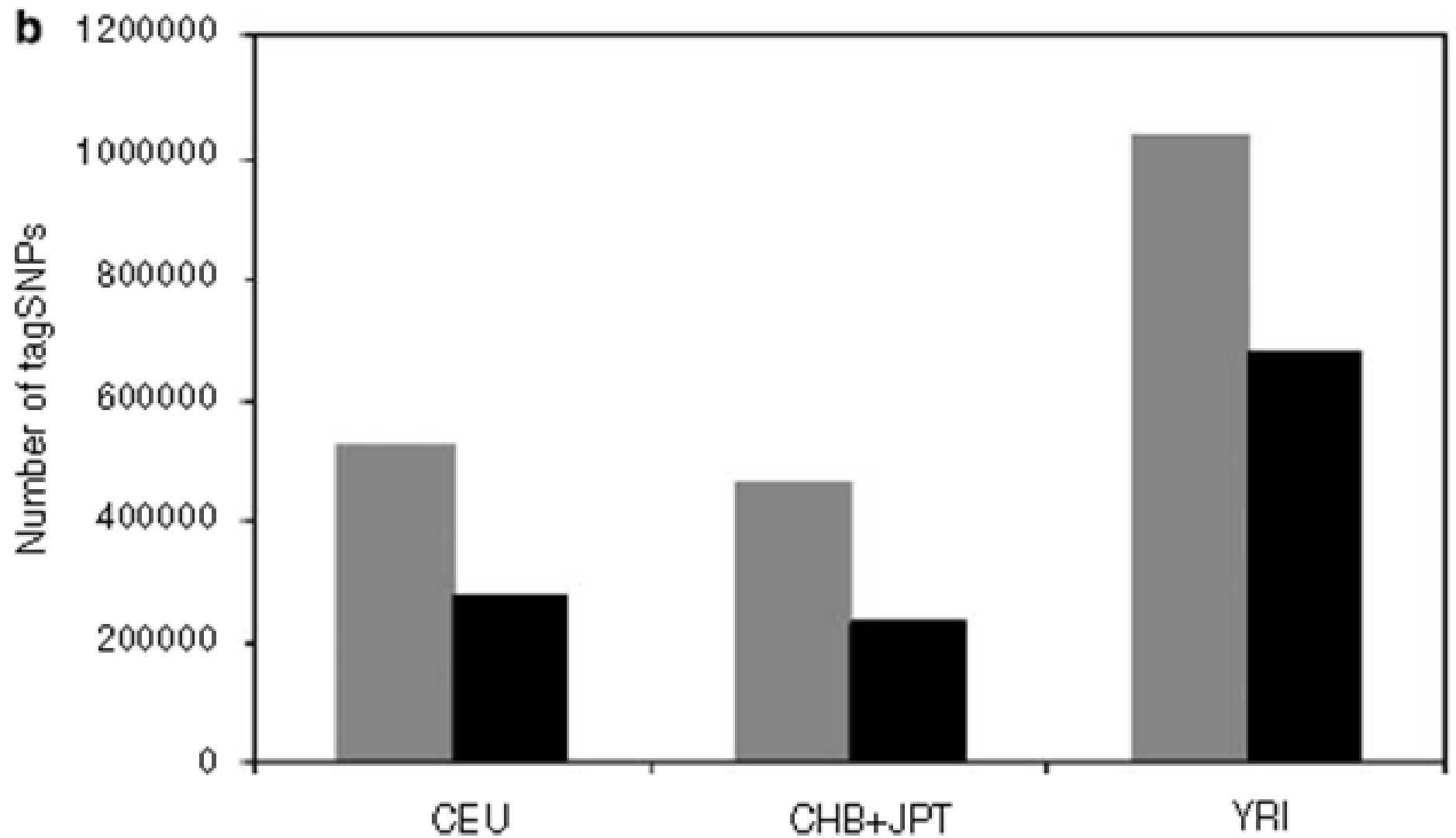
- Count of singleton SNPs
- Functional implications of singleton SNPs
- How many singleton SNPs can be tagged by currently popular genome-wide chips



Singleton SNPs in the human autosomal genome.

PRIMARY:
 $R^2 \geq 0.8$
CEPH population

SECONDARY:
rare SNPs
 $MAF \leq 5\%$



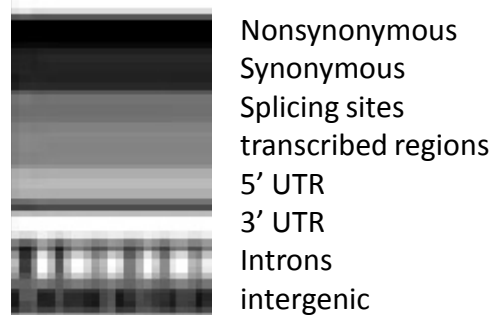
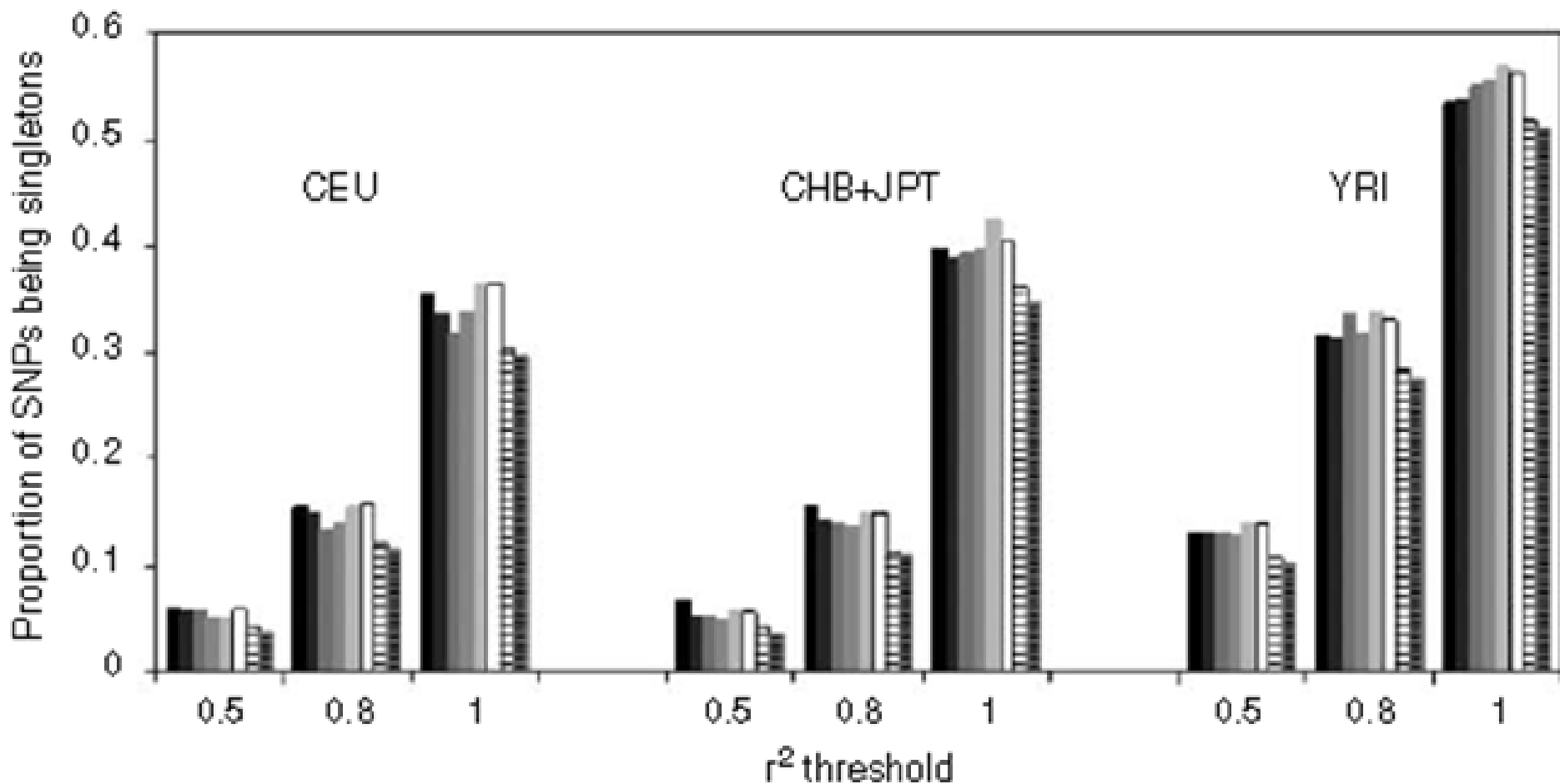
TagSNP selection: **common** SNP counts

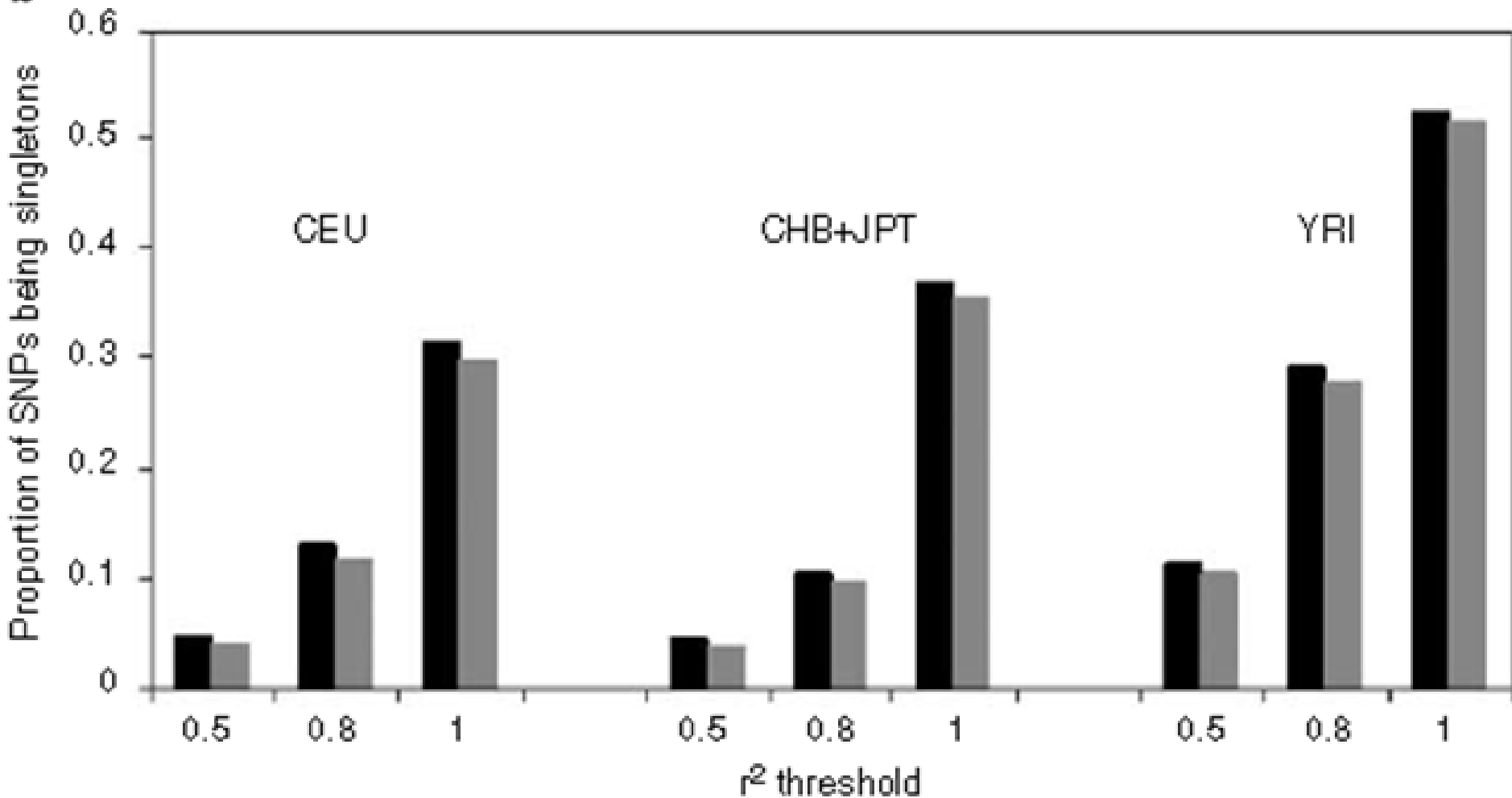
Phase II data, CEPH pop. MAF>5%

Grey – all; black - singletons

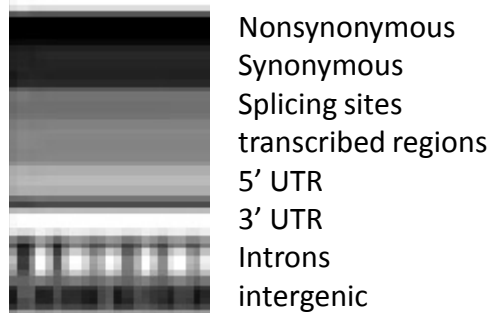
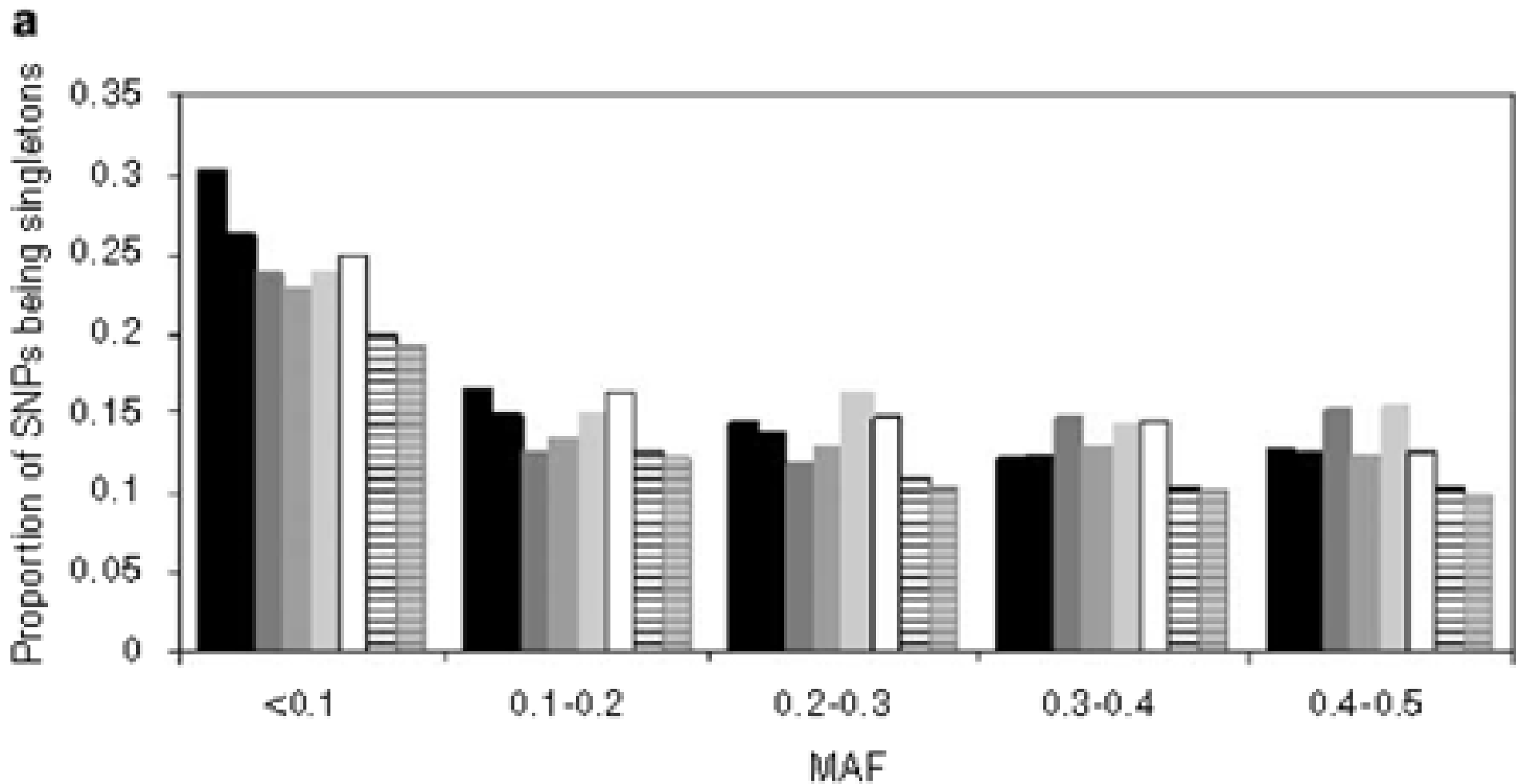
Functional implications of singleton SNPs

- Among 8 876 160 SNPs:
 - 0.51% nonsynonymous
 - 0.46% synonymous
 - 0.04% in splicing sites
 - 0.2% in 5' UTR
 - 0.84% in 3' UTR
 - 49.78% introns
 - 46.83% intragenic

a

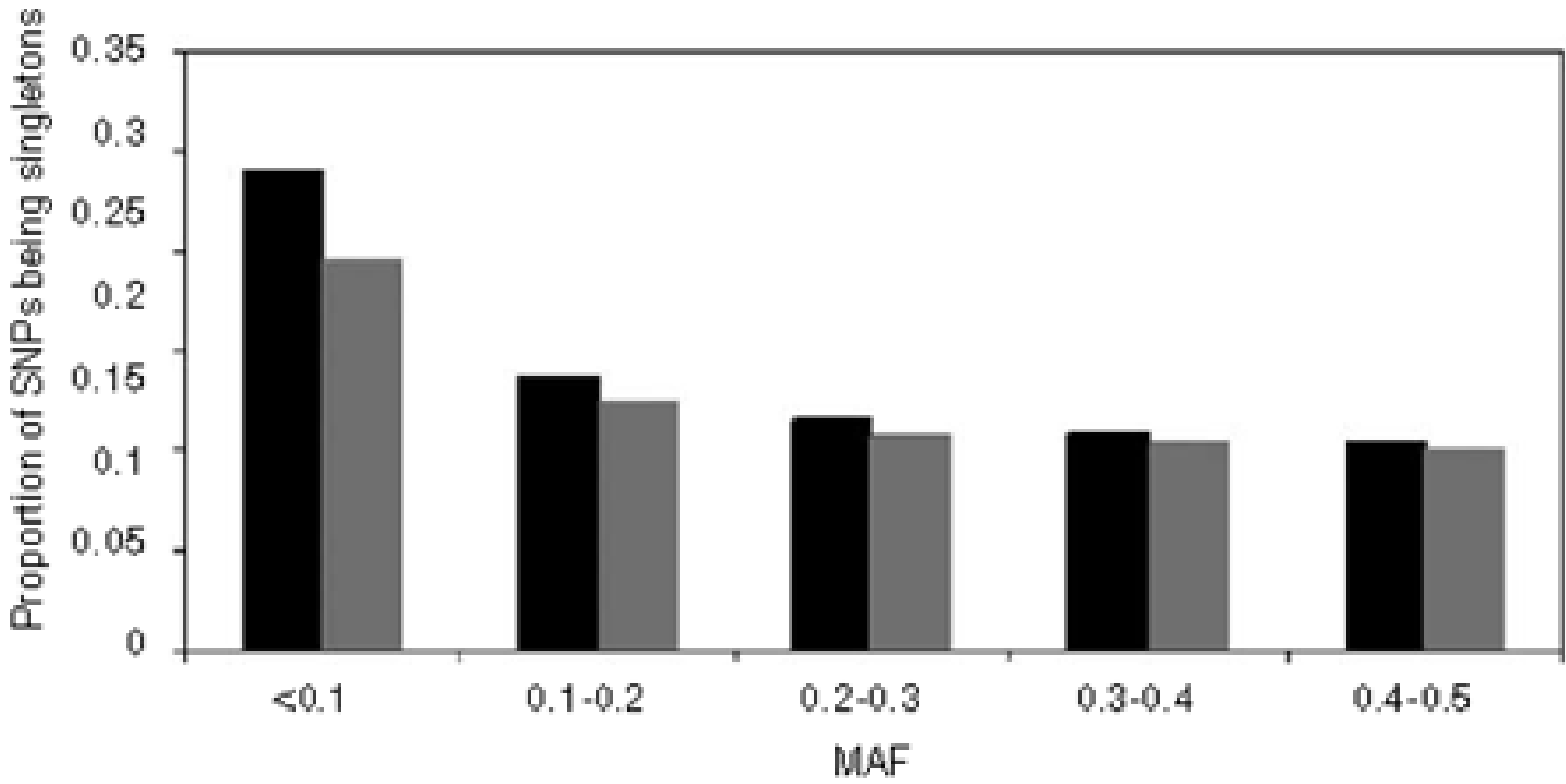
b

Proportion of SNPs being singleton SNPs at pairwise r^2 threshold 0.50, 0.80 and 1.0 in conserved (black bars) and non-conserved (grey bars) regions.

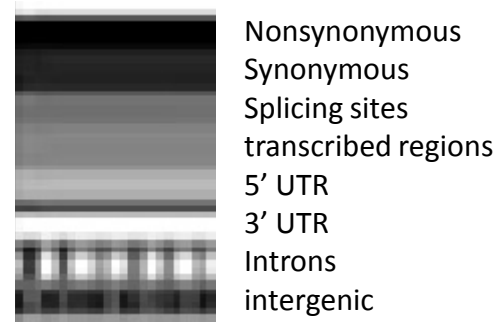
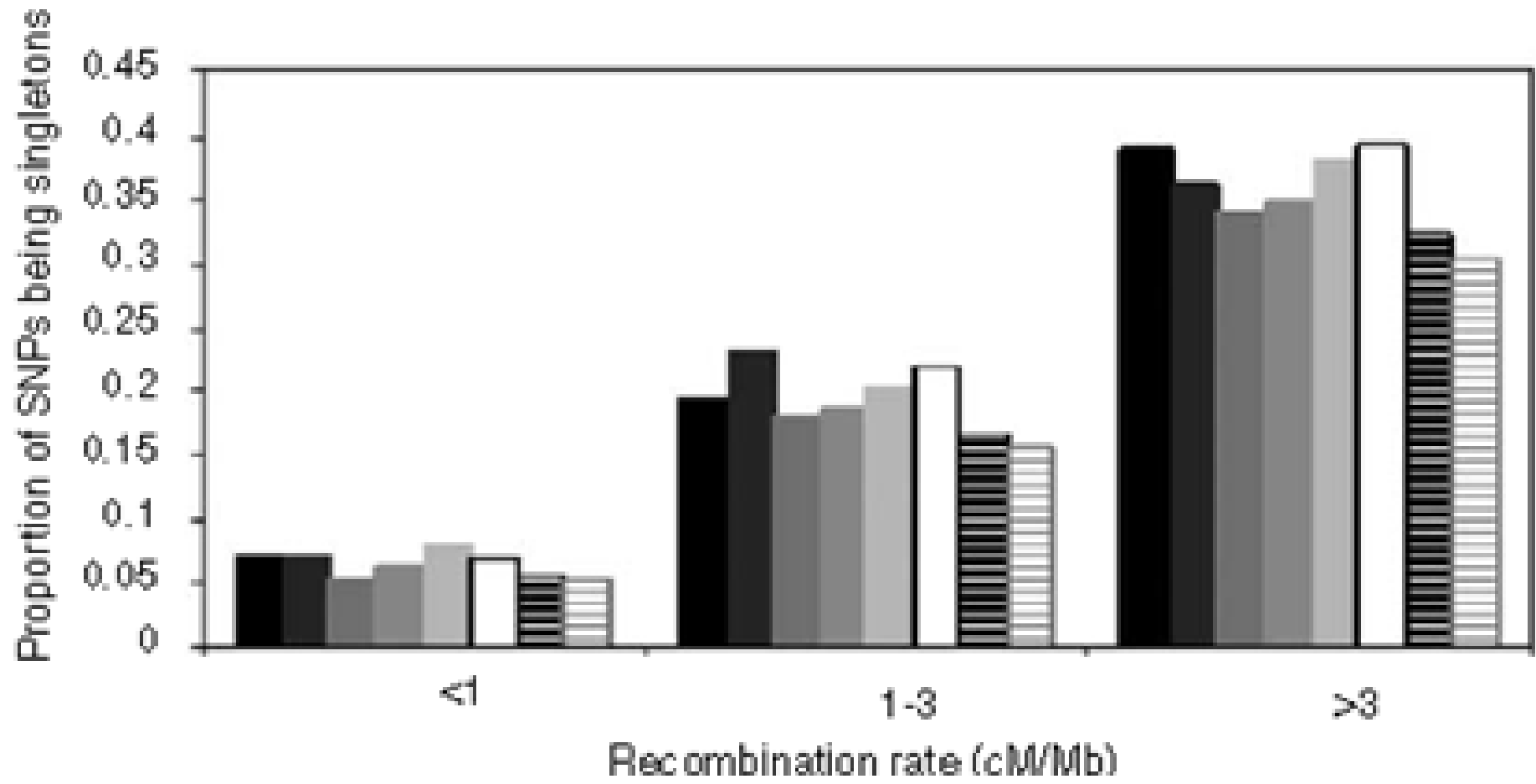


Proportion of SNPs being singleton SNPs
at different allele frequencies

Phase II data, CEPH pop. MAF>5%

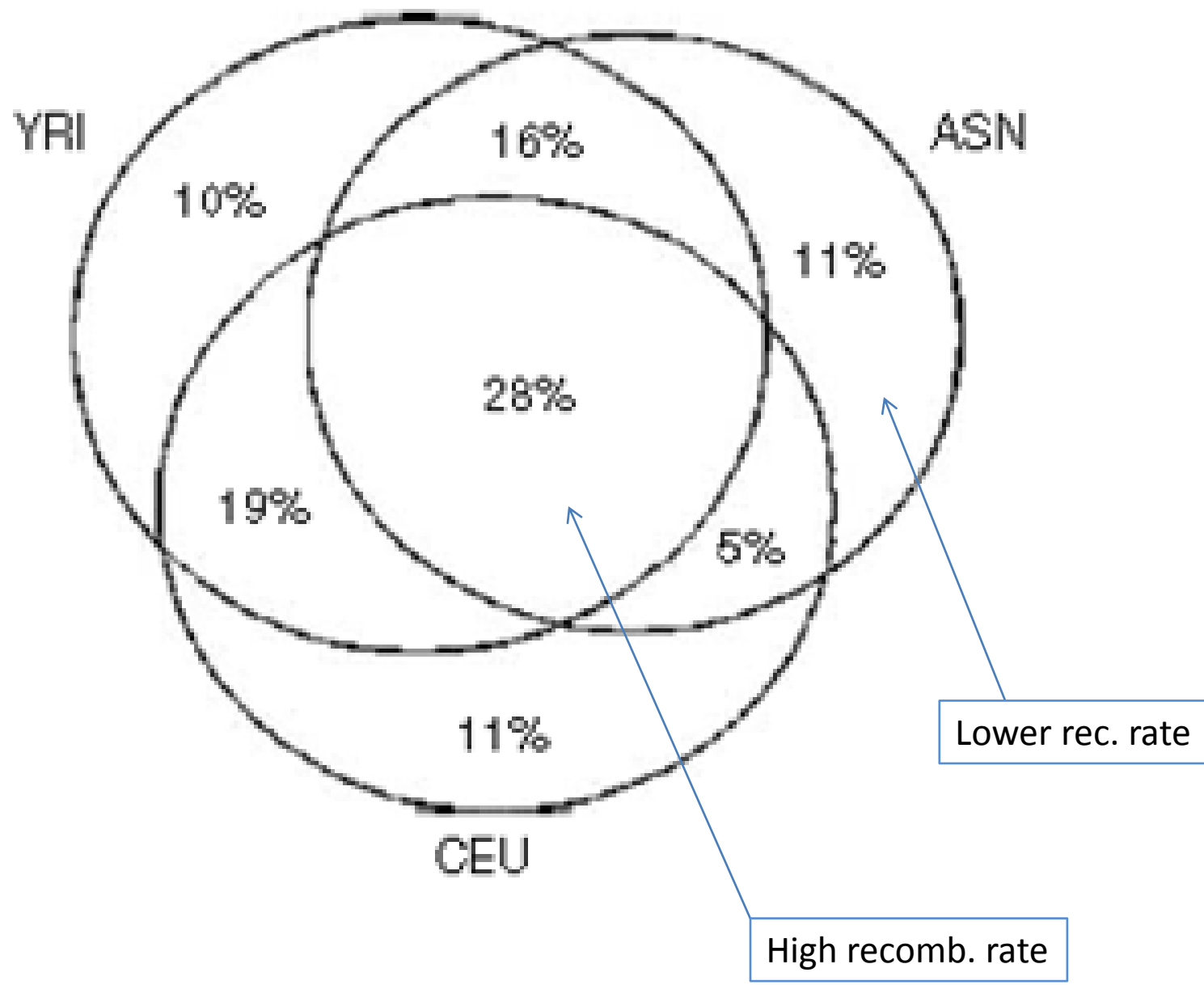
b

Distribution of singleton SNPs in conserved vs non-conserved regions at different MAF spectrum.

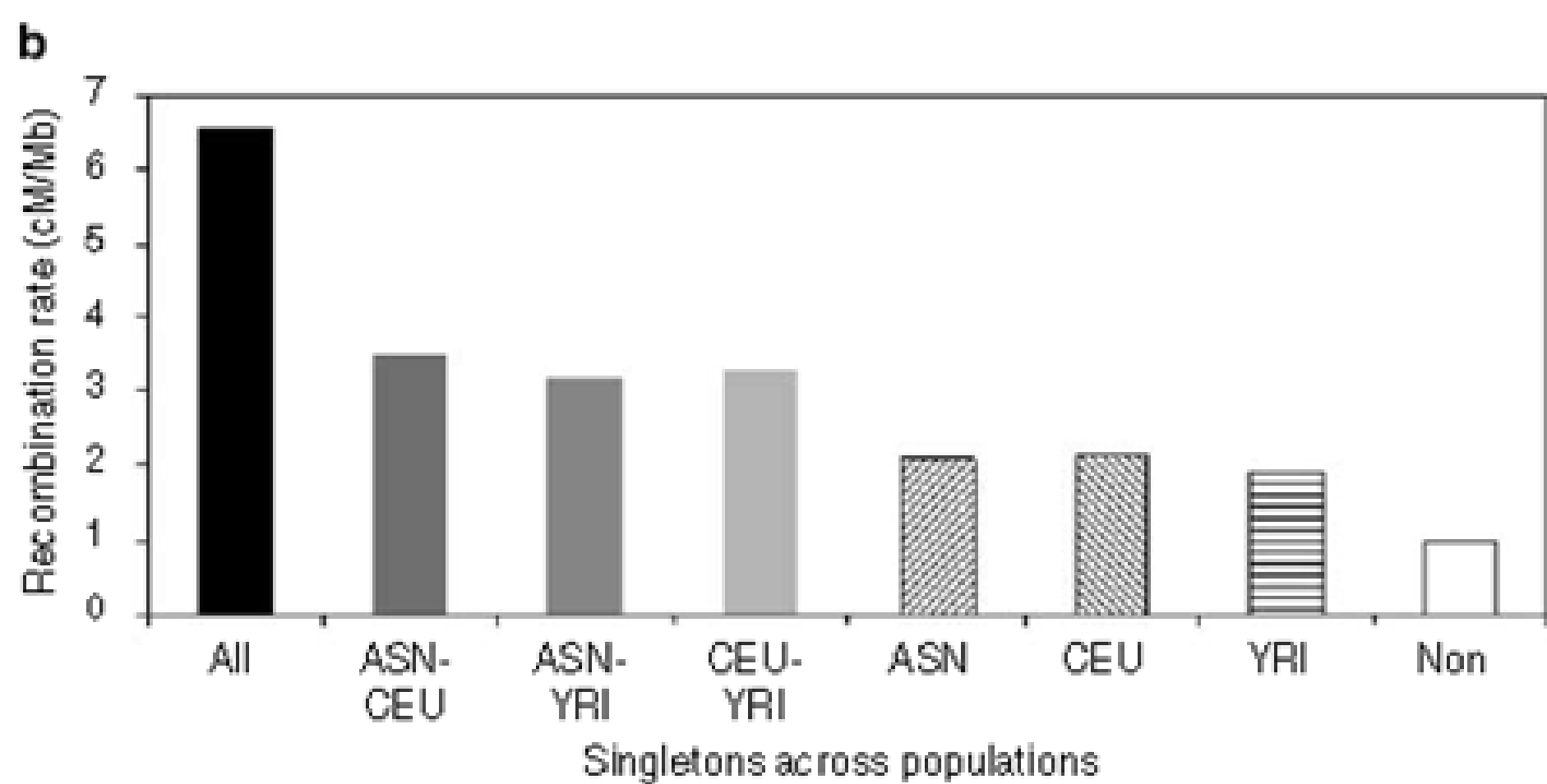
c

Distribution of singleton SNPs of different functional groups at regions of high (>3 cM/Mb), intermediate (1 –3 cM/Mb) and (<1 cM/Mb) recombination rates

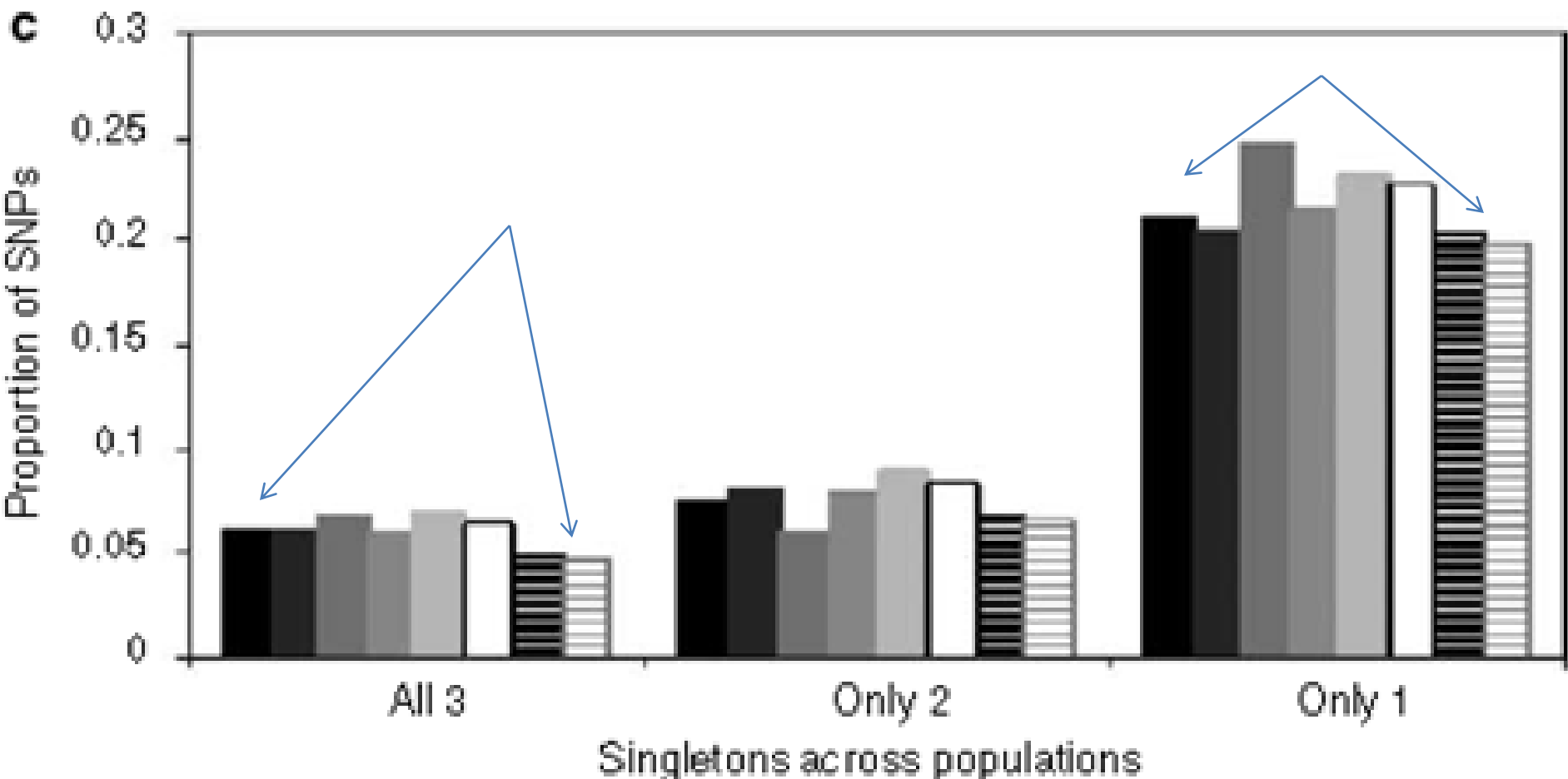
a



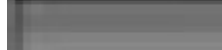
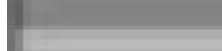






Comparison of singleton SNPs between populations

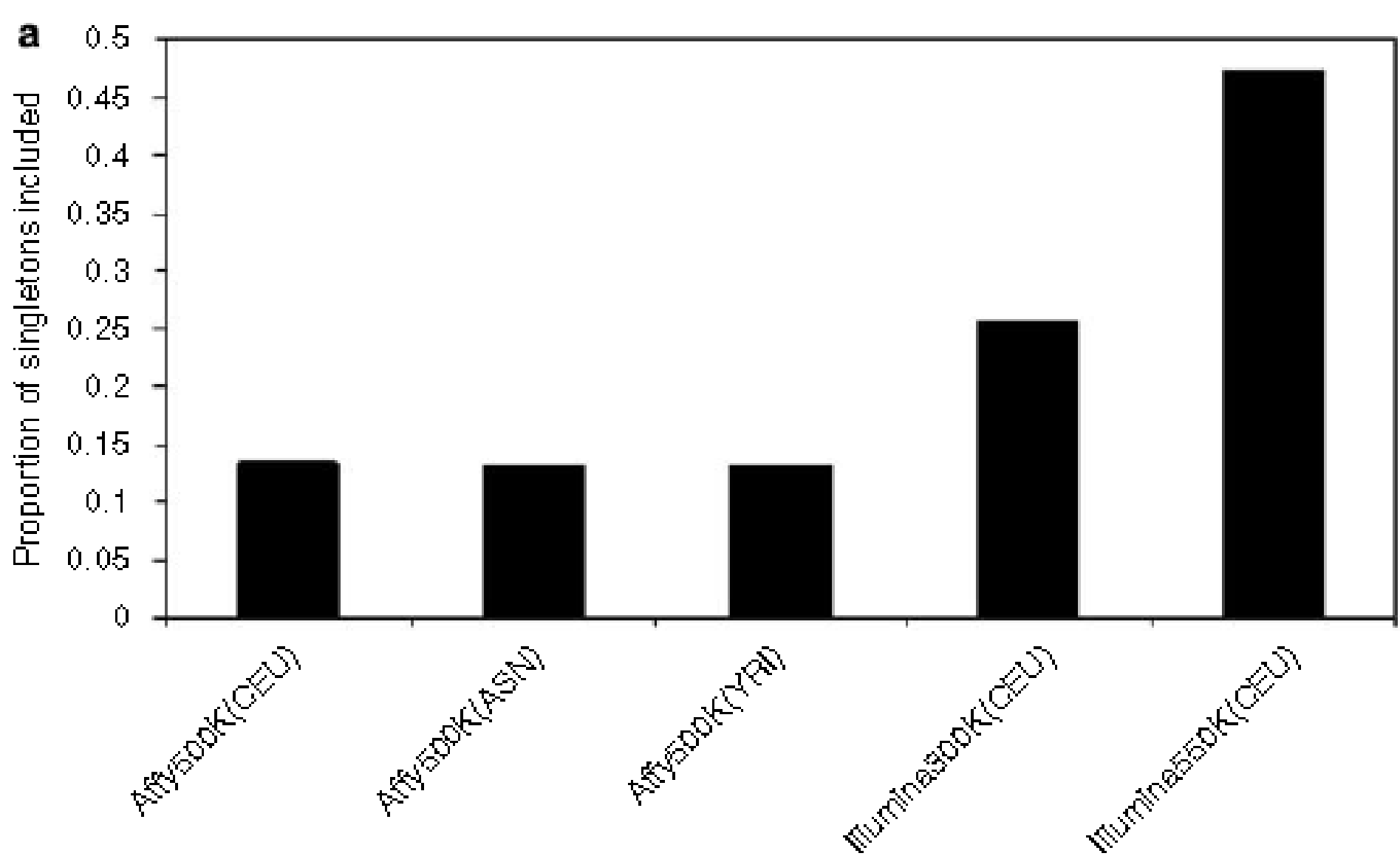


Average recombination rates of singleton SNPs shared between populations

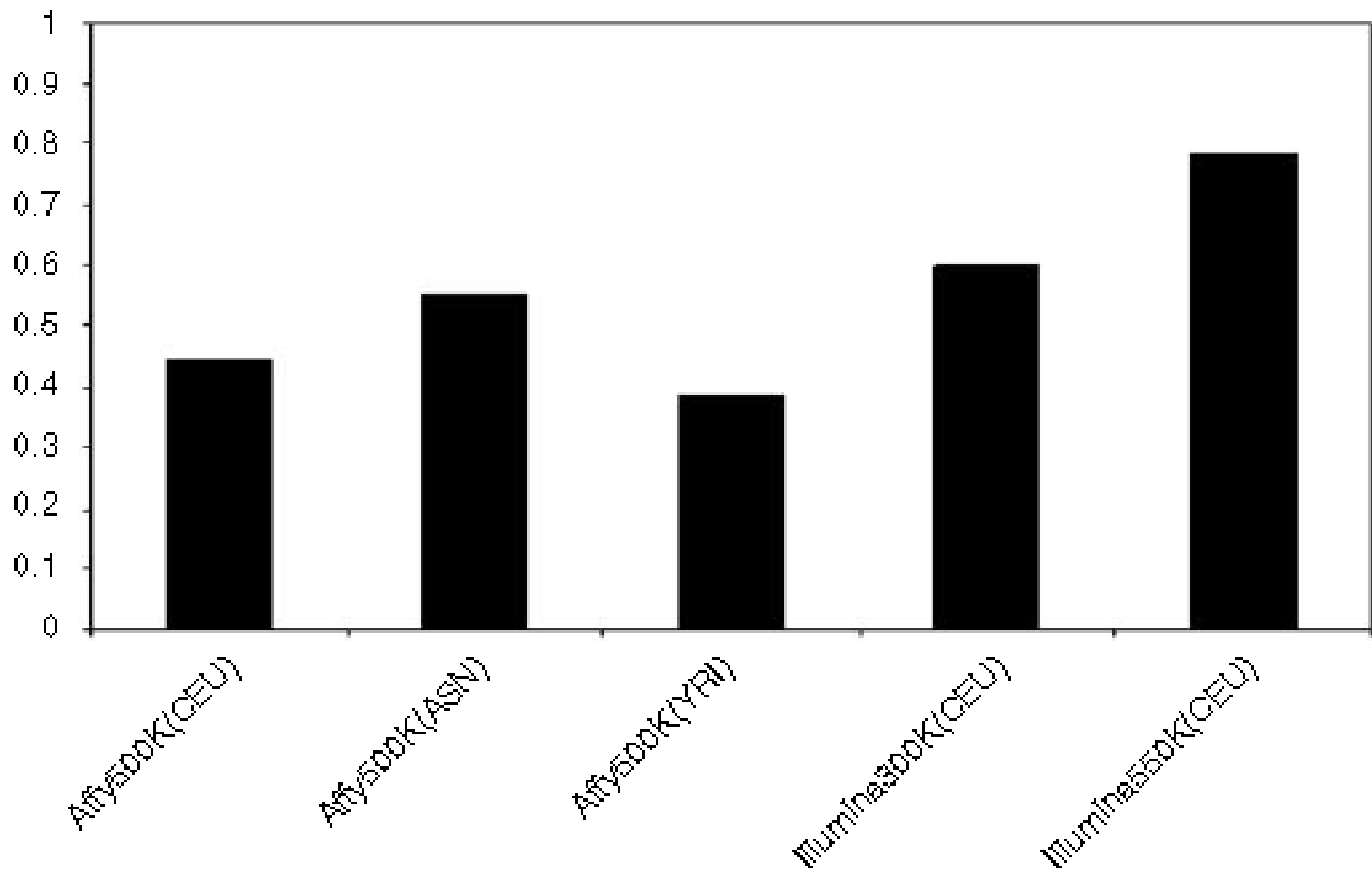


-  Nonsynonymous
-  Synonymous
-  Splicing sites
-  transcribed regions
-  5' UTR
-  3' UTR
-  Introns
-  intergenic

Distribution of singleton SNPs of different functional groups across populations



b
Proportion of singletons covered



Conclusions

- Rare singleton SNP has higher than average possibility to be functional
 - Some of them can be tagged by multimarker predictors
 - Still many of them will remain untaggable unless included directly
 - In marker selection, singleton SNPs should be treated as equally important ones
 - HapMap is biased towards more common SNPs
-
- If rare SNPs are so important then is it right to investigate common SNPs in GWAS?