Using SNPs to Characterize Genetic Effects in Clinical Trials

Bruce Weir
bsweir@uw.edu

Genome-wide association studies in pharmacogenomics

Ann K. Daly

Abstract | Genome-wide association (GWA) studies for pharmacogenomics-related traits are increasingly being performed to identify loci that affect either drug response or susceptibility to adverse drug reactions. Until now, only the largest effects have been detected, partly because of the challenges of obtaining large numbers of cases for pharmacogenomic studies. Since 2007, a range of pharmacogenomics GWA studies have been published that have identified several interesting and novel associations between drug responses or reactions and clinically relevant loci, showing the value of this approach.

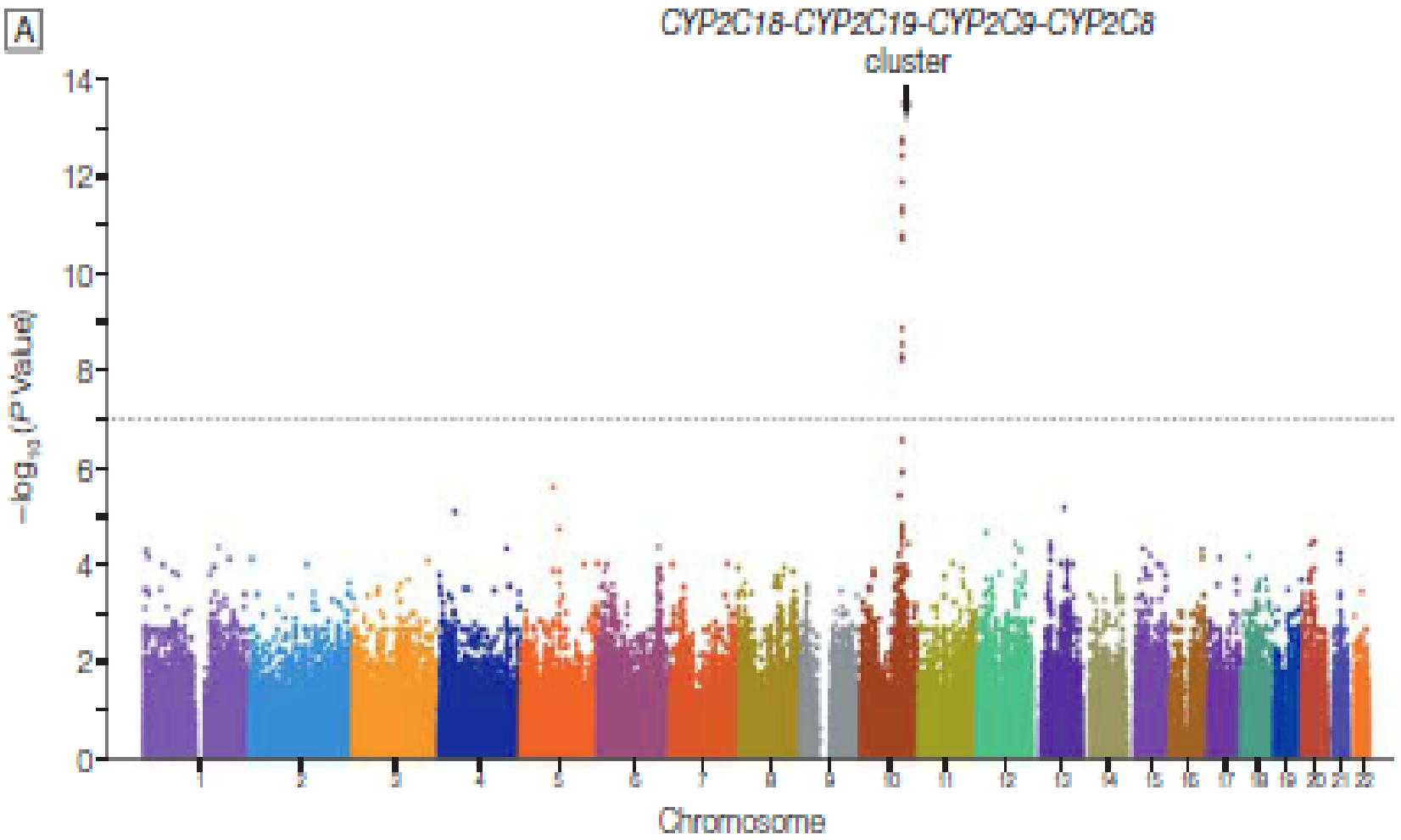
Clopidogrel Therapy

"Association of Cytochrome P450 2C19 Genotype With the Antiplatelet Effect and Clinical Efficacy of Clopidogrel Therapy." JAMA. 2009;302(8):849-857.
Alan R. Shuldiner; Jeffrey R. OConnell; Kevin P. Bliden; et al.

Method We administered clopidogrel for 7 days to 429 healthy Amish persons and measured response by ex vivo platelet aggregometry. A genome-wide association study was performed followed by genotyping the loss-of function cytochrome P450 (CYP) 2C19*2 variant (rs4244285).

Conclusion CYP2C19*2 genotype was associated with diminished platelet response to clopidogrel treatment and poorer cardiovascular outcomes.

Figure 2. Genome-Wide Association Study of Adenosine Diphosphate–Stimulated Platelet Aggregation in Response to Clopidogrel



Clopidogrel Therapy

"Effects of CYP2C19 Genotype on Outcomes of Clopidogrel Treatment." N Engl J Med 2010;363:1704-14.
Guillaume Par, Shamir R. Mehta, Salim Yusuf; et al.

Method We genotyped patients from two large, randomized trials that showed that clopidogrel, as compared with placebo, reduced the rate of cardiovascular events. Patients were genotyped for three single-nucleotide polymorphisms (*2, *3, *17) that define the major CYP2C19 alleles.

Conclusion Among patients with acute coronary syndromes or atrial fibrillation, the effect of clopidogrel as compared with placebo is consistent, irrespective of CYP2C19 loss of function carrier status.

GARNET Study

GARNET is a series of genome-wide association studies of treatment response in randomized clinical trials that looks to identify genetic variants associated with response to treatments for conditions of clinical or public health significance.

GARNET aims to utilize existing clinical trial data and sample resources to:

1. Identify genetic variants that influence an individual's response to treatment.
2. Determine whether specific treatments are more or less effective in groups defined by genotype.
3. Develop and disseminate innovative methods for adding genome-wide technologies to randomized clinical trials and interpreting the results in the context of a randomized treatment assignment.

GARNET Study

- Alex Reiner, Women's Health Initiative, GWAS of Hormone Treatment and CVD and Metabolic Outcomes in the WHI.

- Michele Sale, University of Virginia , Pharmacogenomic studies in VISP (Vitamin Intervention for Stroke Prevention) results & implications for clinical trial design

- Richard Weinshilboum, Mayo Clinic, A genome-wide association study in breast cancer patients from the prospectively randomized SUCCESS trial.

SNP Data

- 2005 HapMap I: 1 million SNPs in 269 individuals from four populations.
- 2007 HapMap II: 3.1 million SNPs in 270 individuals from four populations.
- 2010 HapMap III: 1.6 million SNPs in 1184 individuals from 11 populations.
- 2010 1000 Genomes: 15 miilion SNPs in 697 individuals from 7 populations.

1000 Genomes Project

"The aim of the 1000 Genomes Project is to discover, genotype and provide accurate haplotype information on all forms of human DNA polymorphism in multiple human populations."

Pilot project had three components:

- Whole-genome high coverage (average 42X) of two trios.
- Whole-genome low coverage (2-6X) of 179 individuals.
- Targeted 8,140 exons with high coverage ($>$ 50X) of 697 individuals.

Implications for GARNET

- Direct testing of genetic variants as opposed to testing of array-based markers in LD with causal variants: no immediate plans for GARNET studies, but ...

- Imputation with much denser panel of markers than possible with HapMap II or III: an activity planned by the Coordinating Center.

Direct Association Testing

The paper reports an experiment of seeking associations with markers on 142 low-coverage samples with expression levels determined from the corresponding cell lines. The numbers of significant eQTLs discovered was higher than with the 1M Illumina chip array:

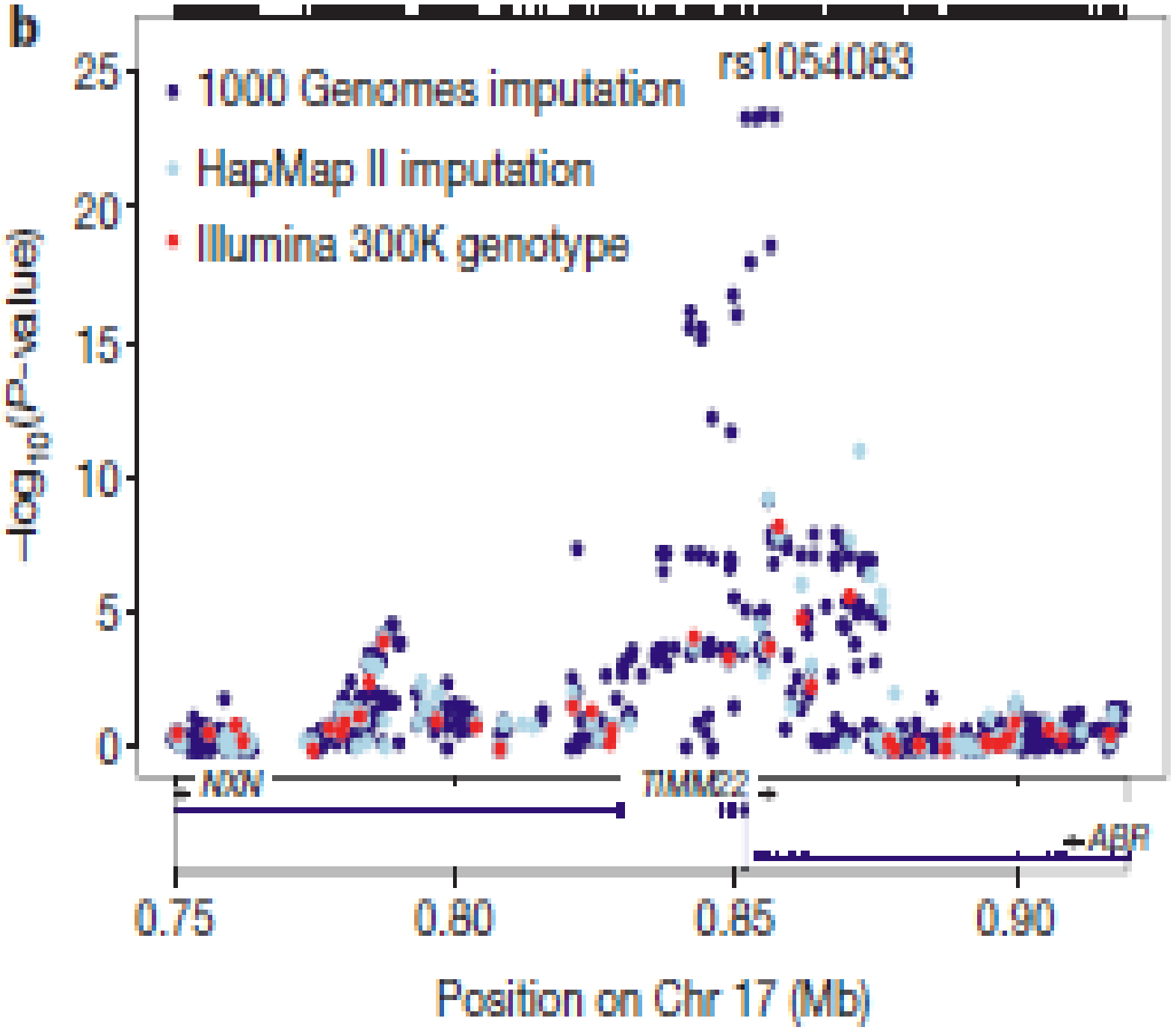
Population	Sample size	1M	1000G
CEU	43	420	522
YRI	42	345	518
CHB+JPT	59	968	1154

(Is this a good omen for understanding genetic mechanisms?)

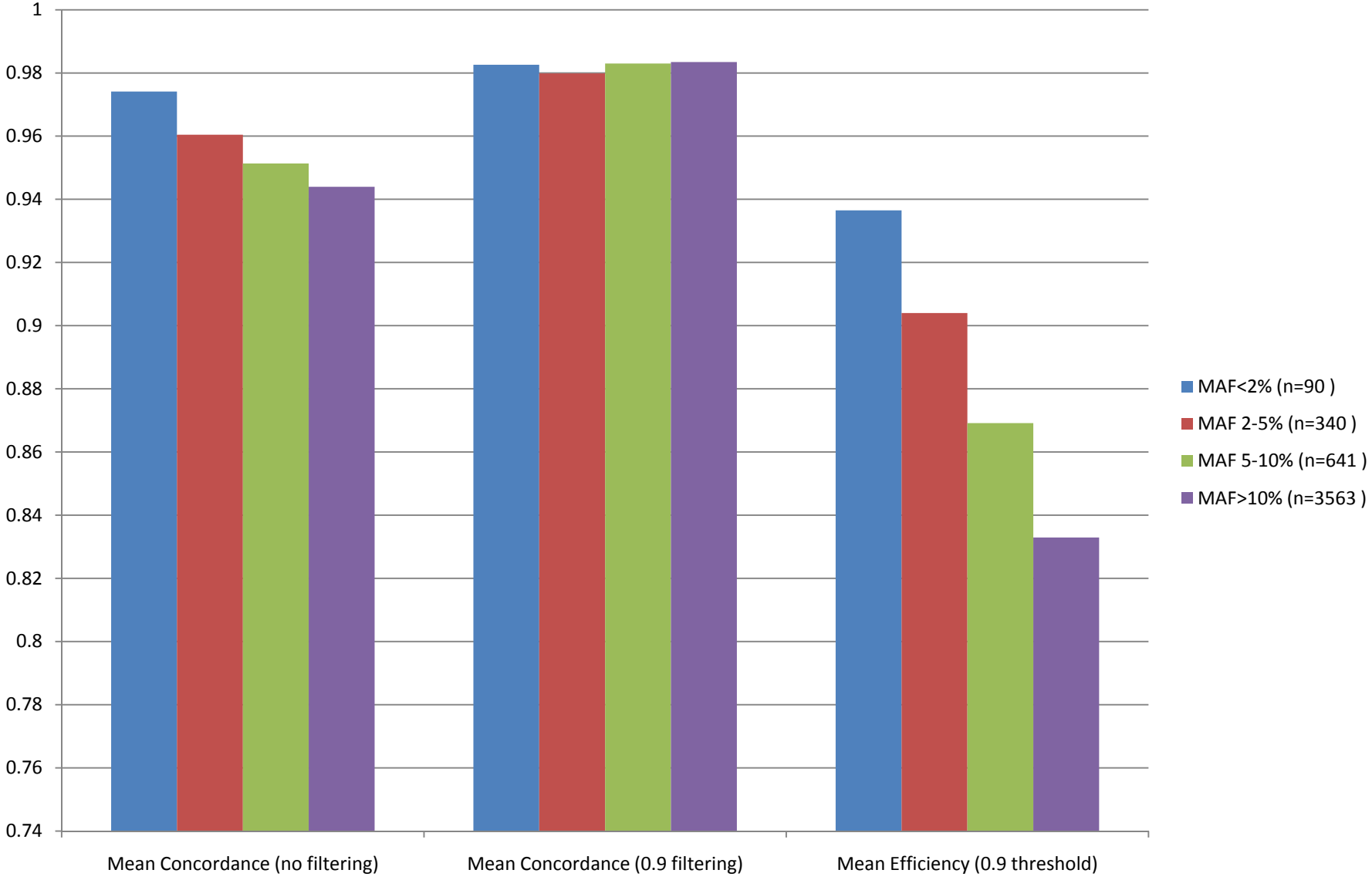
Association Testing After Imputation

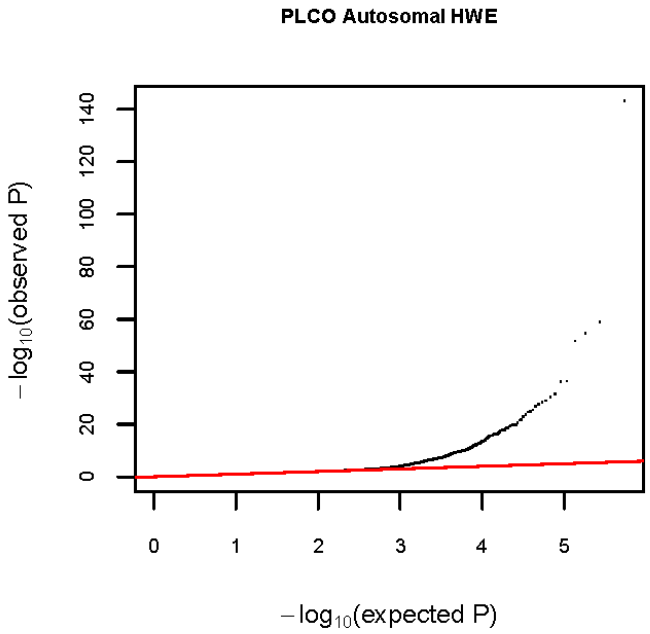
Instead of whole-genome sequencing, could use 1000 Genomes data to impute large numbers of SNPs and conduct association tests with those markers. Paper reports comparison of results of imputing with HapMap II and the low-coverage 1000 Genomes.

The number of detected cis-eQTLs increased much more with 1000 Genomes data.


b
rs1054083
1000 Genomes imputation
HapMap II imputation
Illumina 300K genotype
−log10(P-value)
25
20
15
10
5
0
NDN
TIMM22
ABR
0.75
0.80
0.85
0.90
Position on Chr 17 (Mb)

1000 Genomes imputation (March 2010 release): Quality Metrics for masked Chr22 SNPs in Bierut Eur. American dataset




PLCO Autosomal HWE
−log10(observed P)
0
20
40
60
80
100
120
140
0
1
2
3
4
5
−log10(expected P)

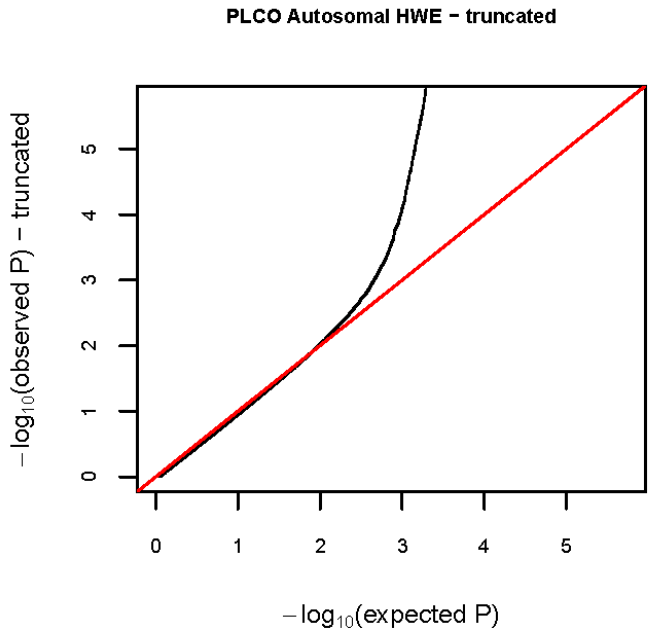
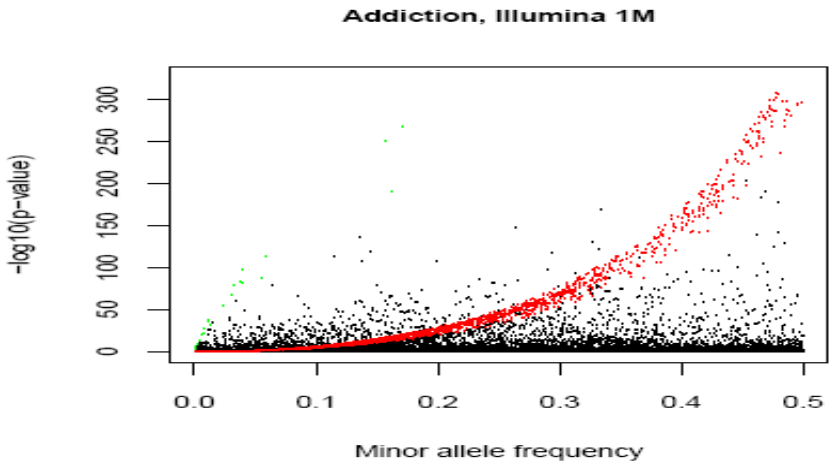

PLCO Autosomal HWE − truncated
−log10(observed P) − truncated
0
1
2
3
4
5
0
1
2
3
4
5
−log10(expected P)

Figure 4


Addiction, Illumina 1M
-log10(p-value)
0
50
100
150
200
250
300
0.0
0.1
0.2
0.3
0.4
0.5
Minor allele frequency

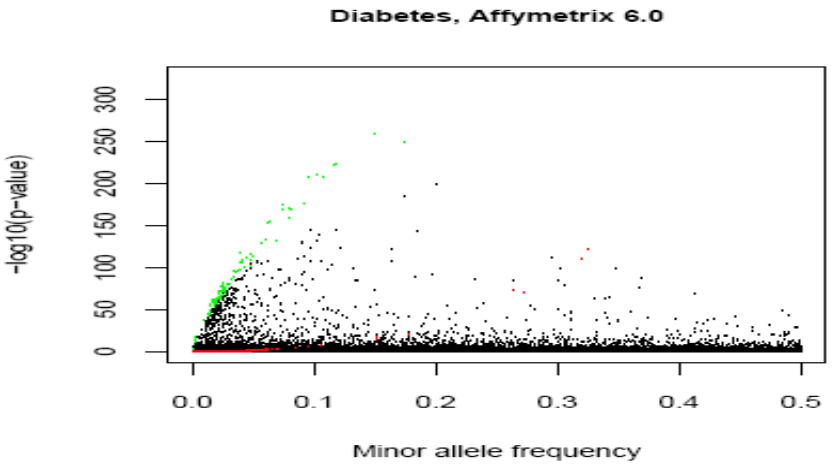
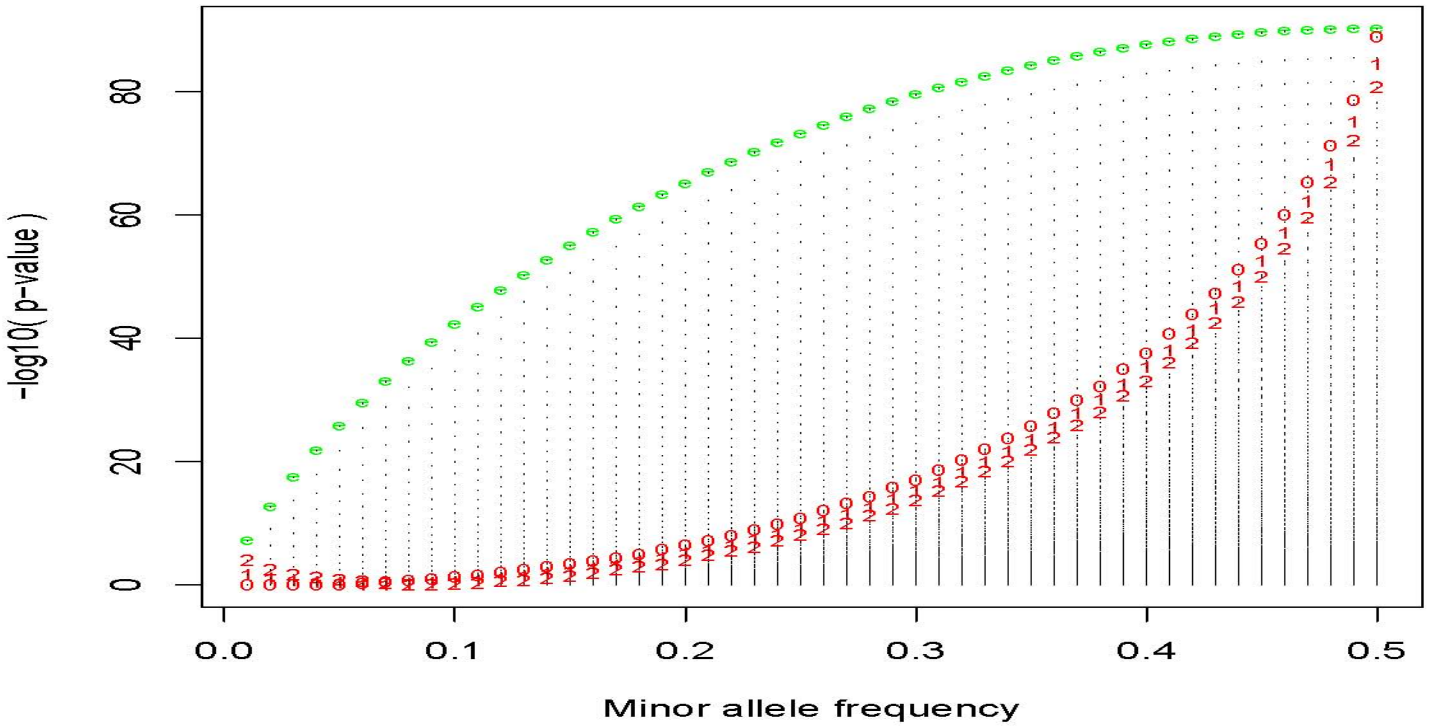
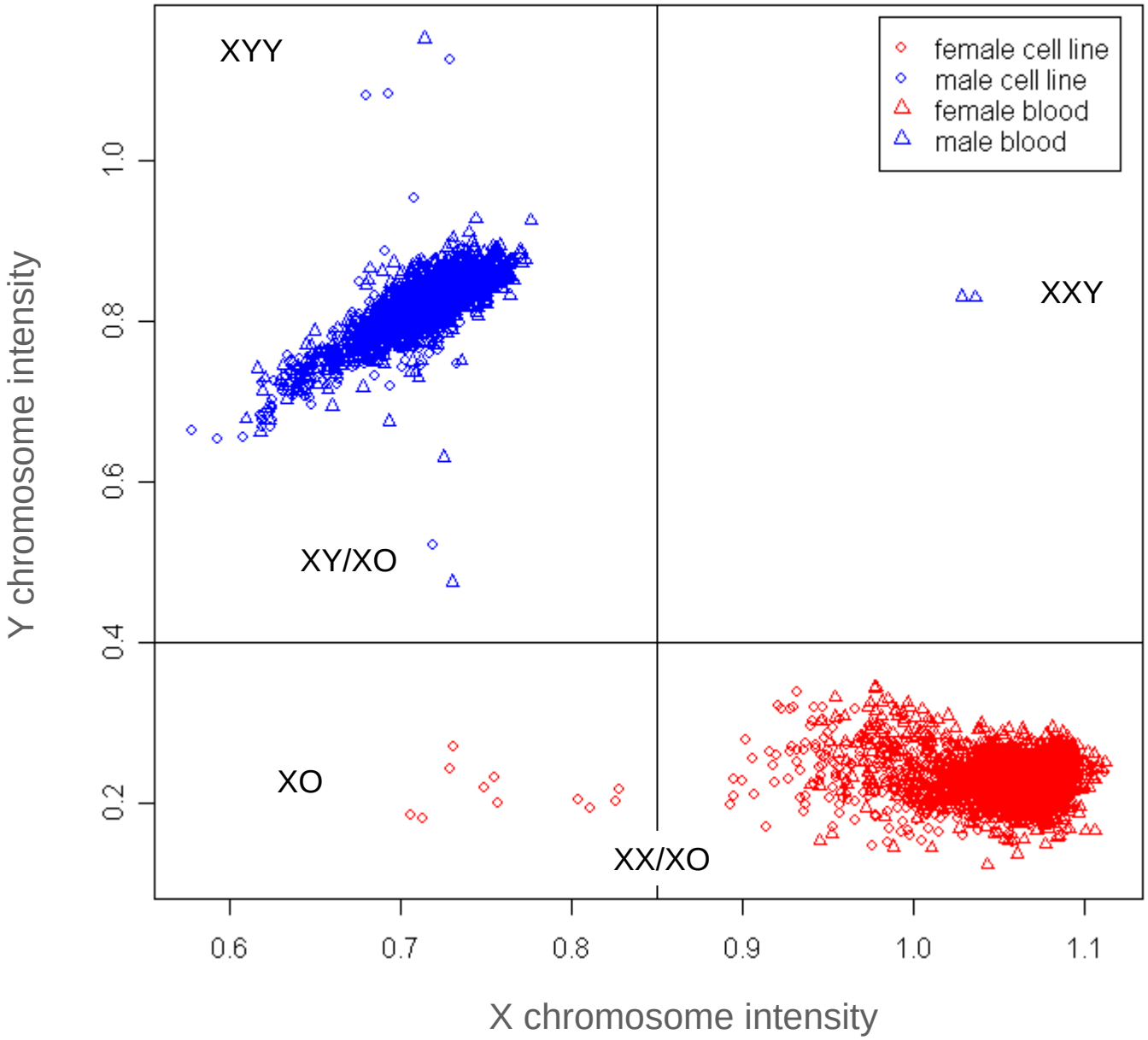

Diabetes, Affymetrix 6.0
-log10(p-value)
0
50
100
150
200
250
300
0.0
0.1
0.2
0.3
0.4
0.5
Minor allele frequency

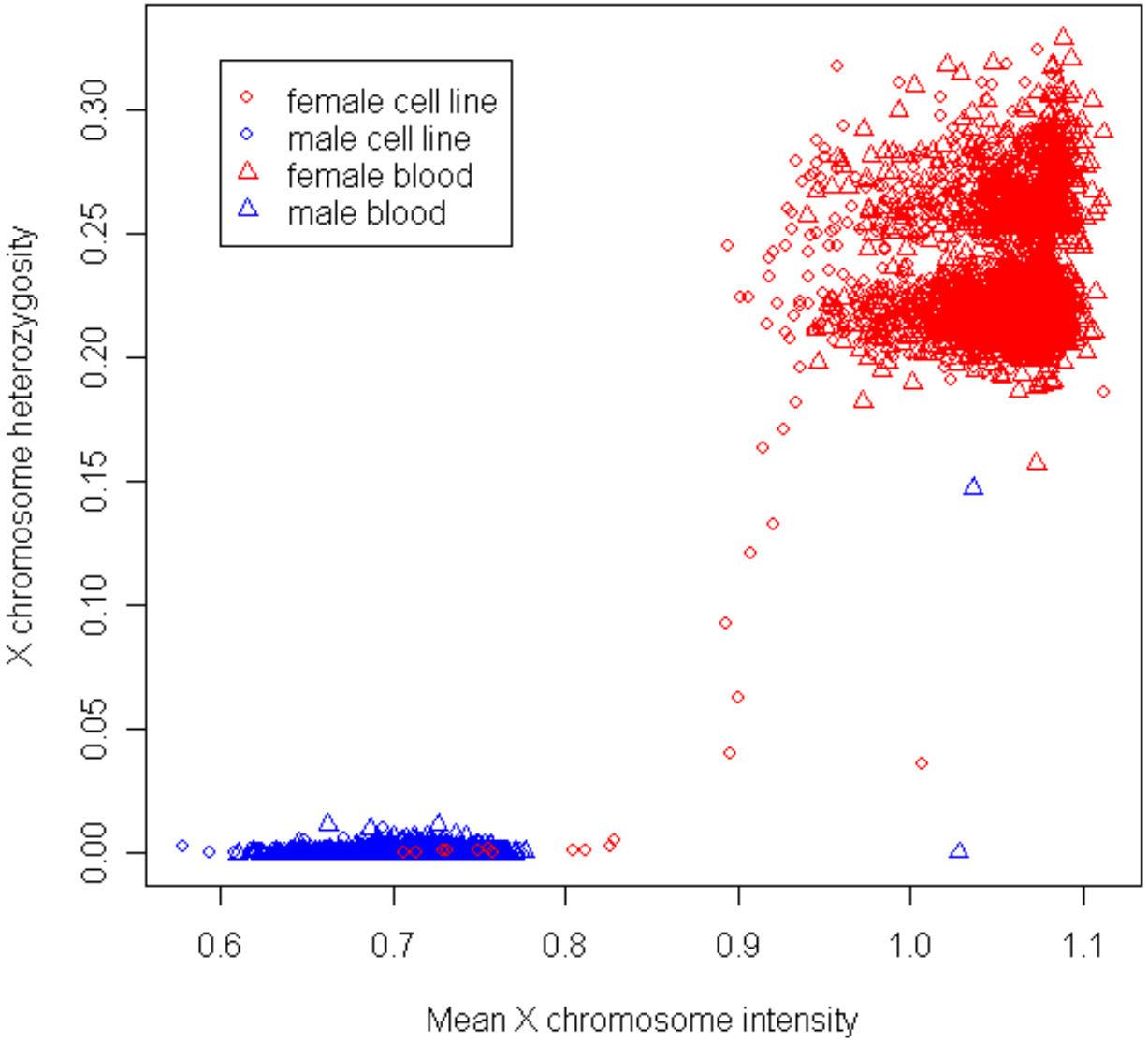
Figure S12


-log10(p-value)
0
20
40
60
80
0.0
0.1
0.2
0.3
0.4
0.5
Minor allele frequency

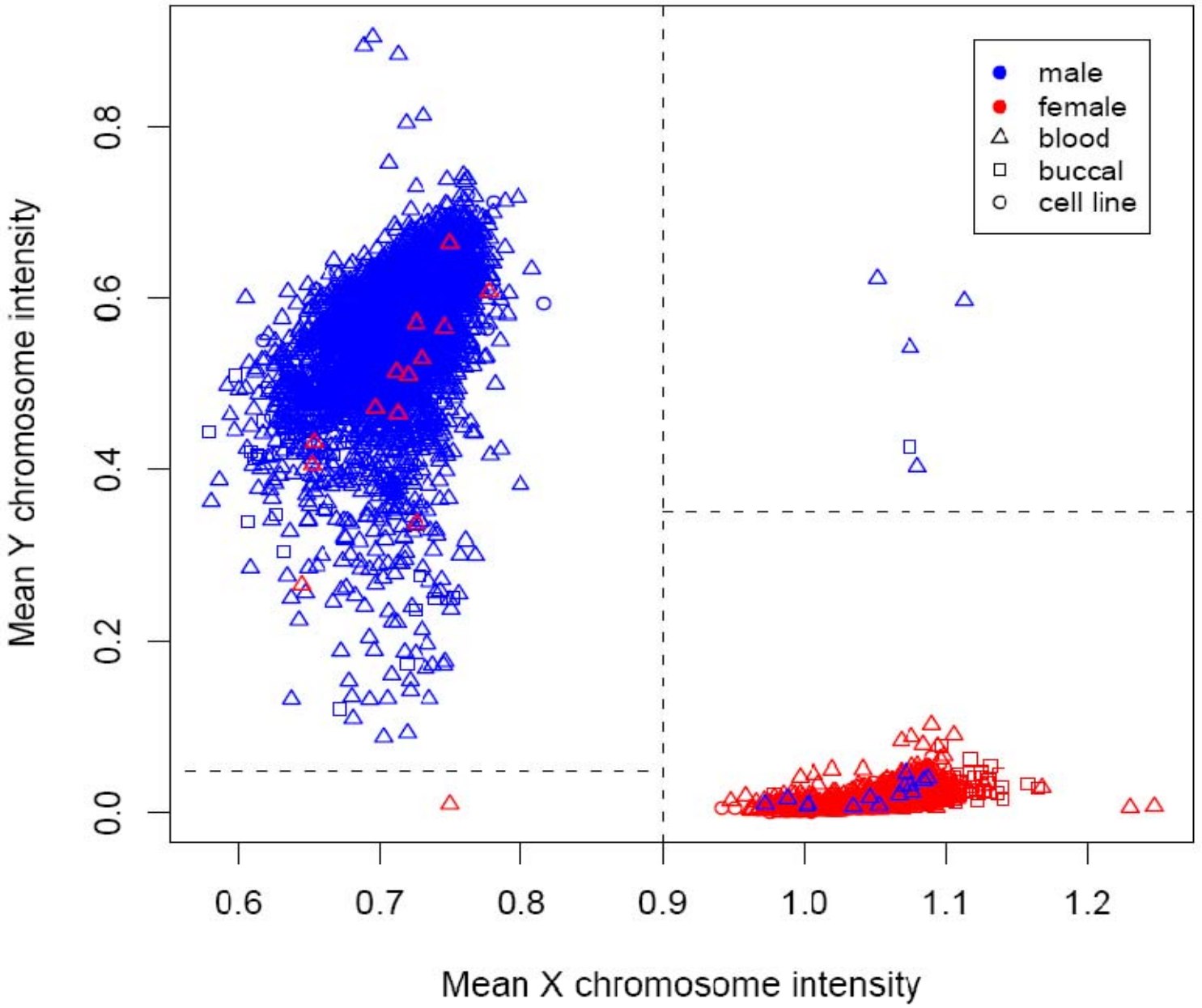
Gender and sex chromosome aneuploidy

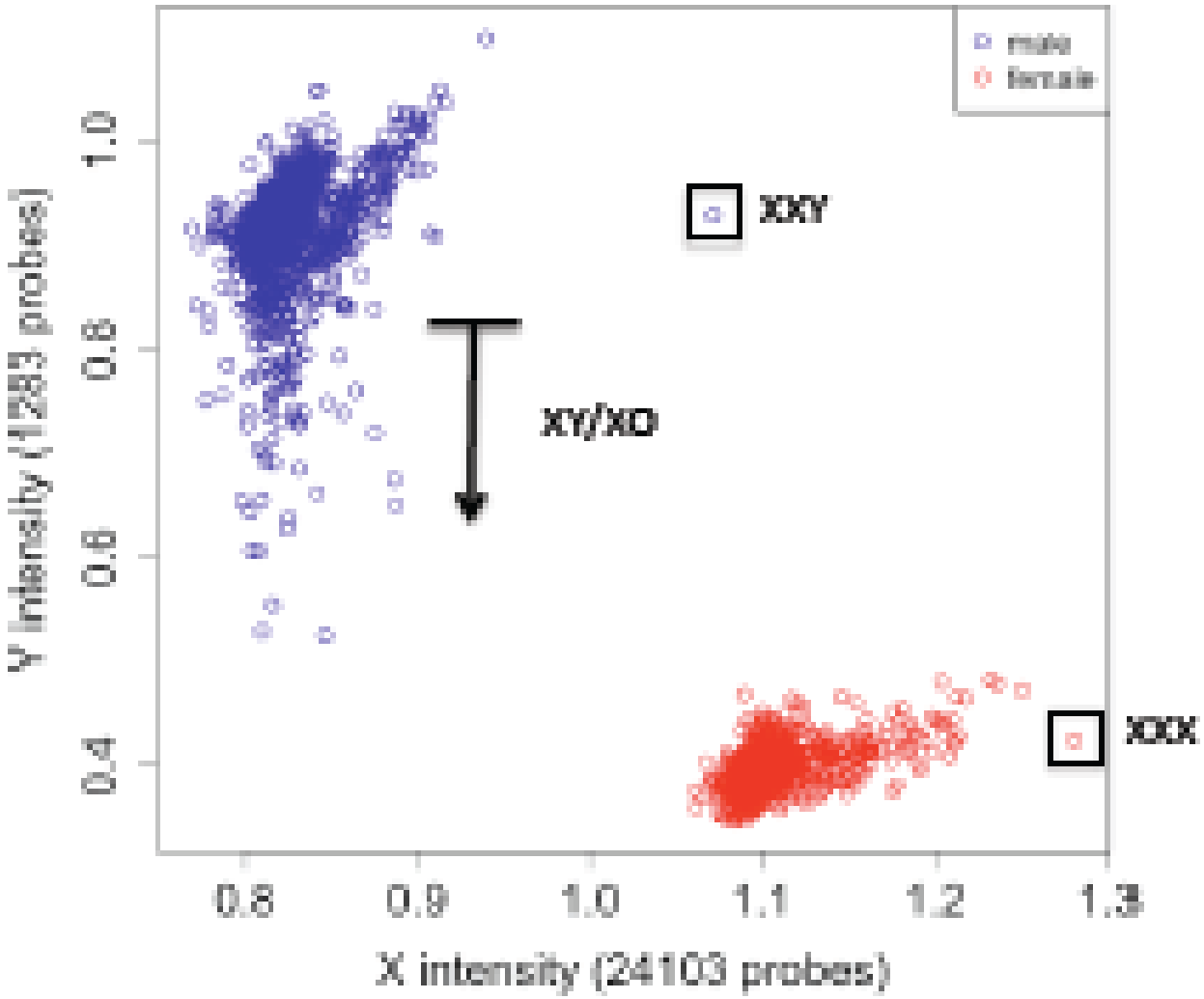


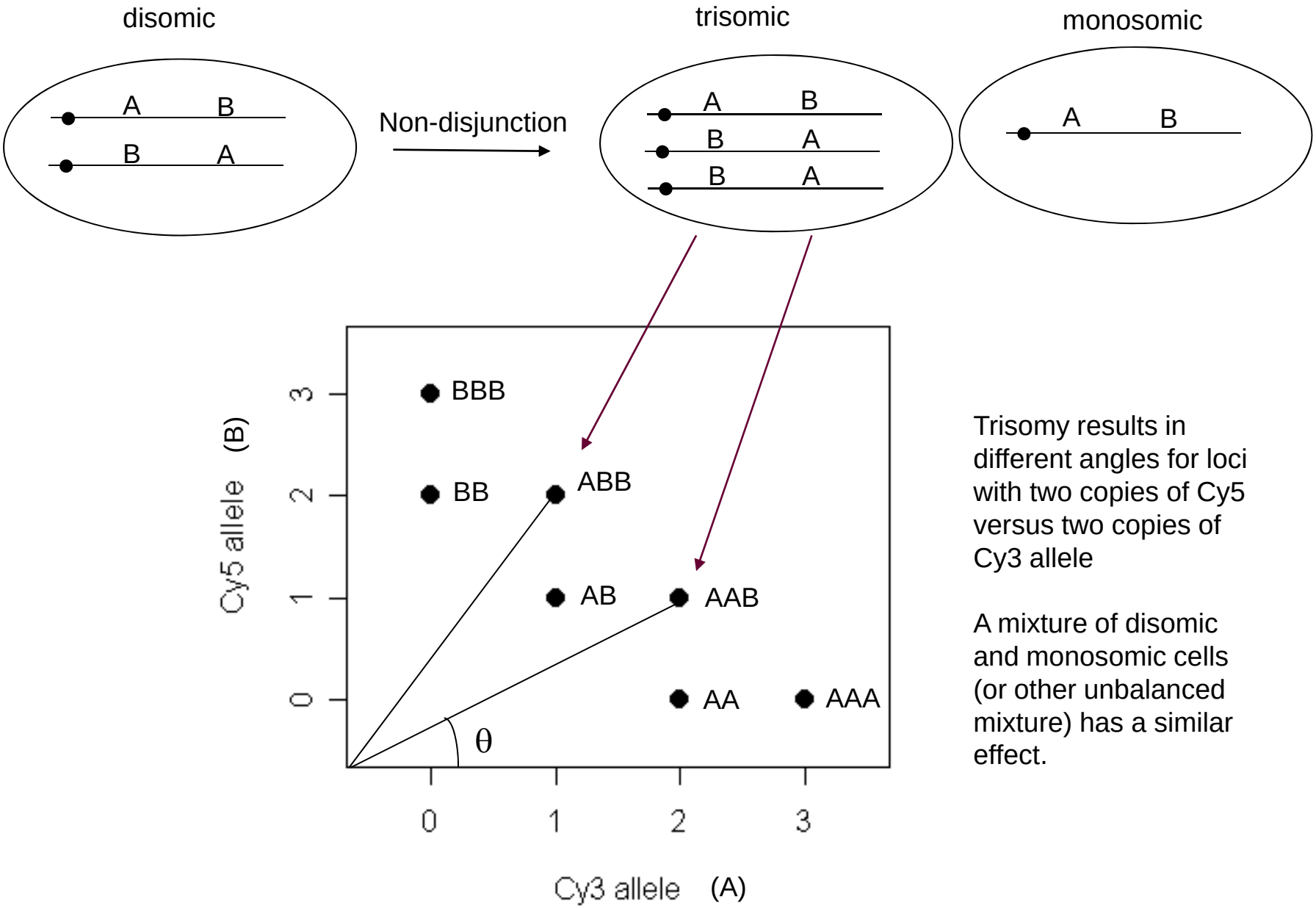
X Chromosome Heterozygosity



Gender check and sex chromosome anomalies




male
female
XXY
XY/XO
XXX
Y intensity (1283 probes)
X intensity (24103 probes)
0.4
0.6
0.8
1.0
0.8
0.9
1.0
1.1
1.2
1.3

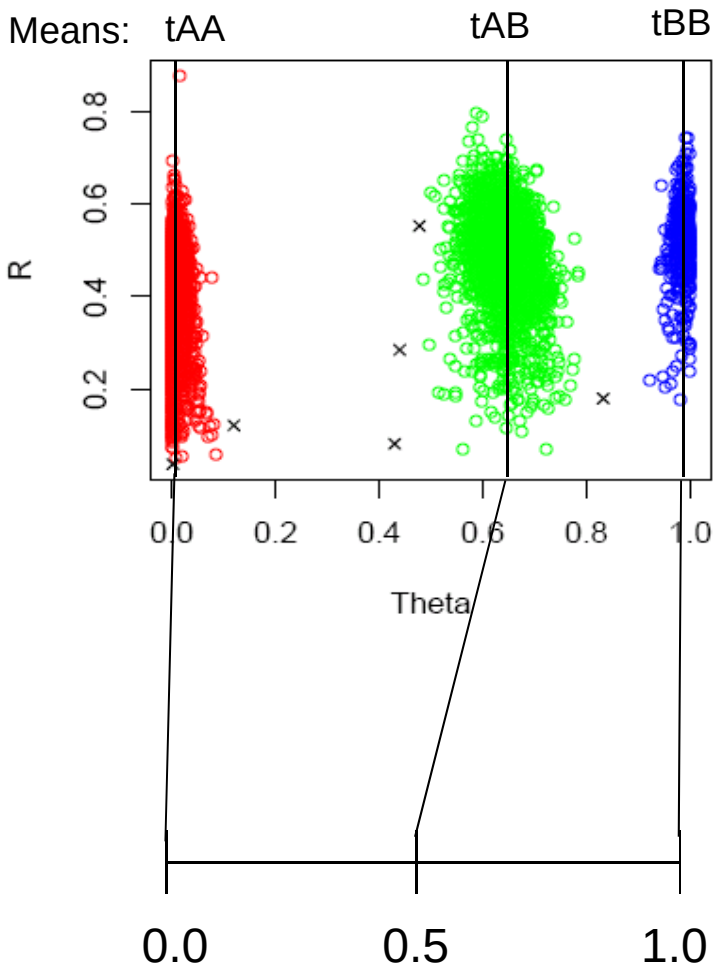


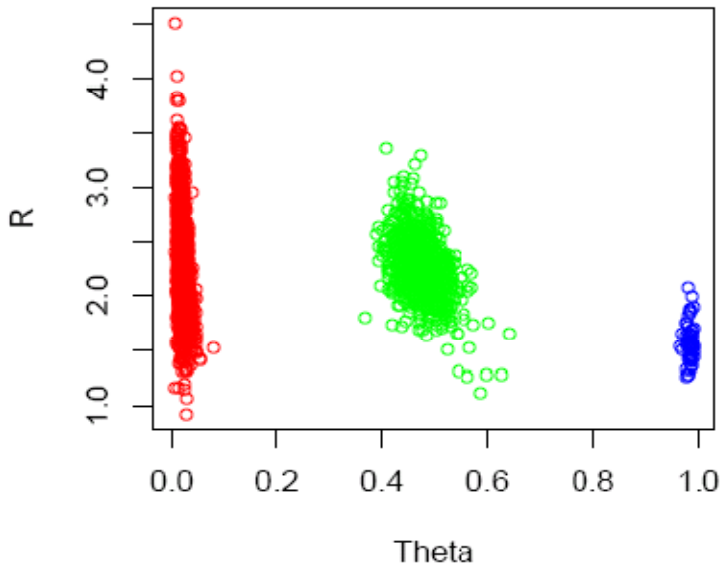
Trisomy results in different angles for loci with two copies of Cy5 versus two copies of Cy3 allele

A mixture of disomic and monosomic cells (or other unbalanced mixture) has a similar effect.

BAlleleFreq (BeadStudio) is a transformation of Theta

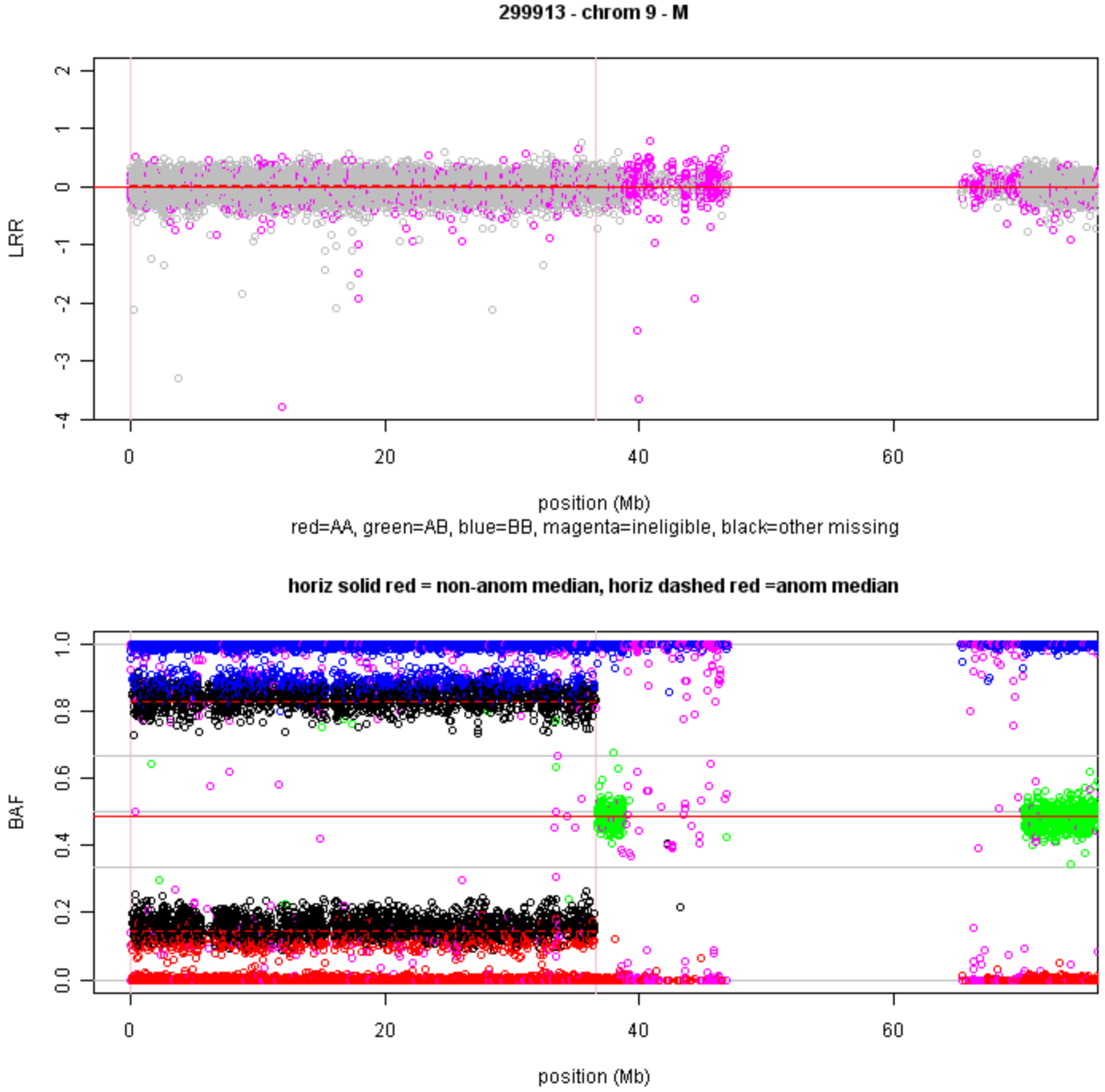
(to standardize mean positions of AA, AB and BB to 0, 0.5 and 1.0)

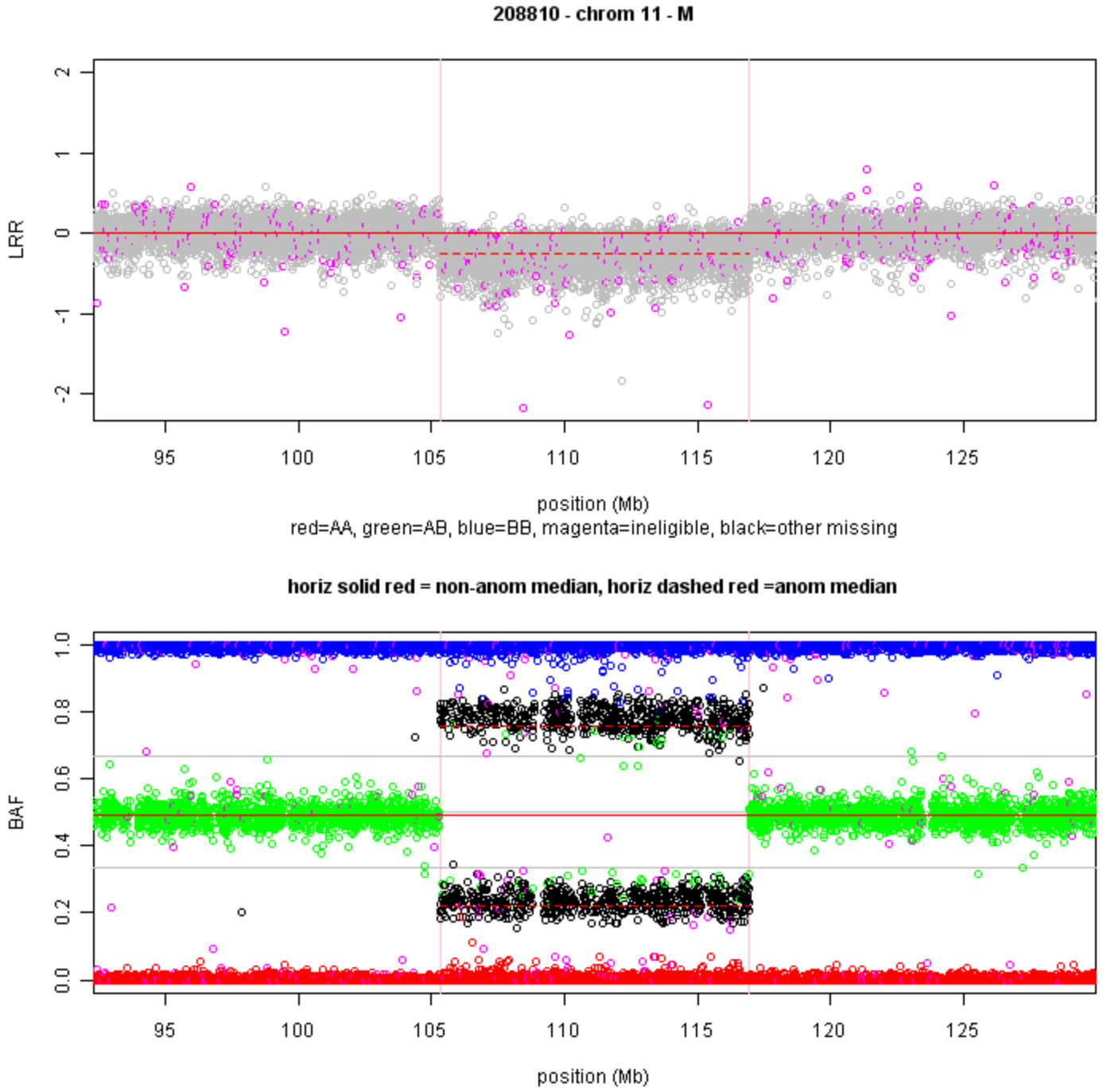


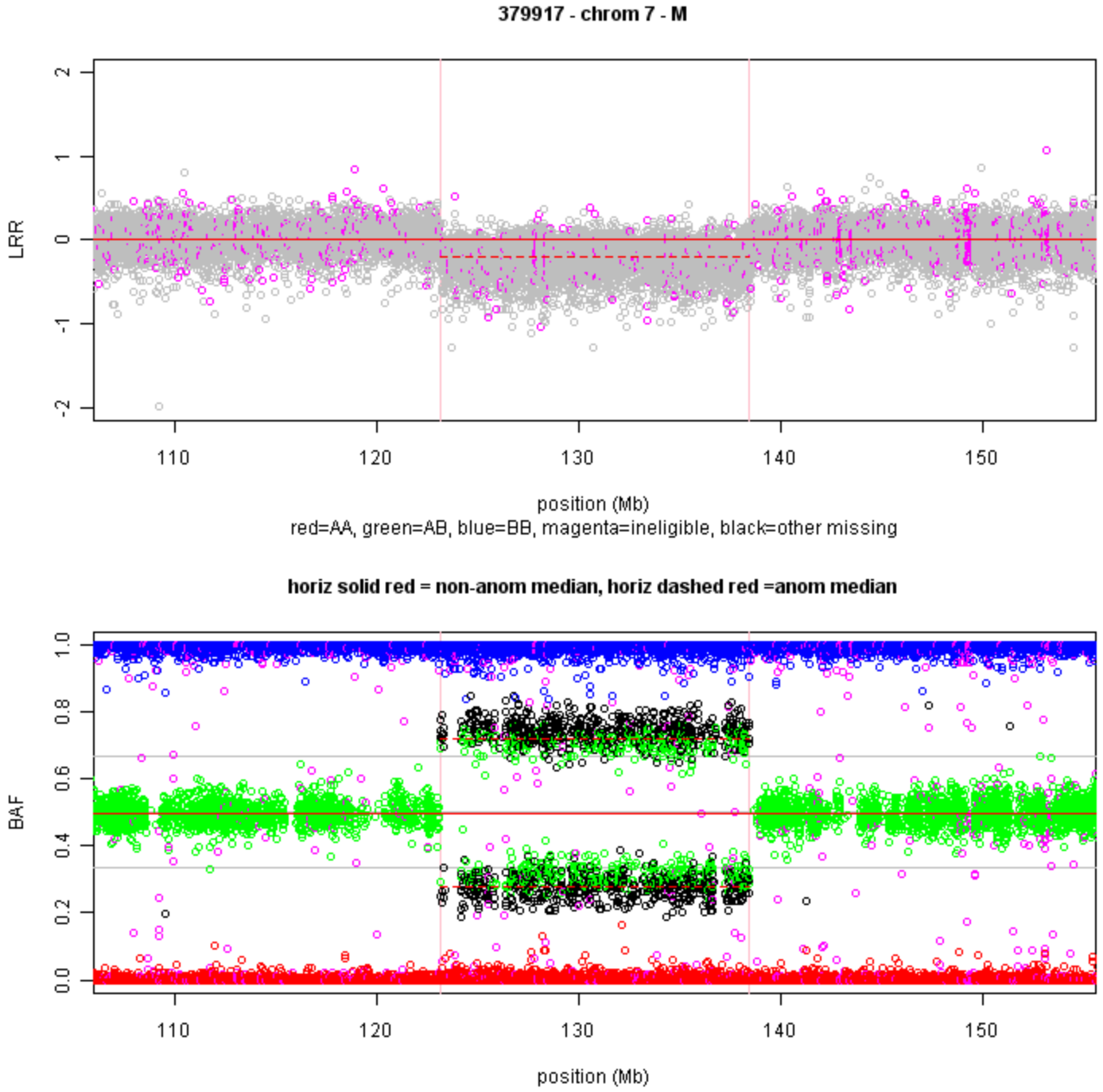


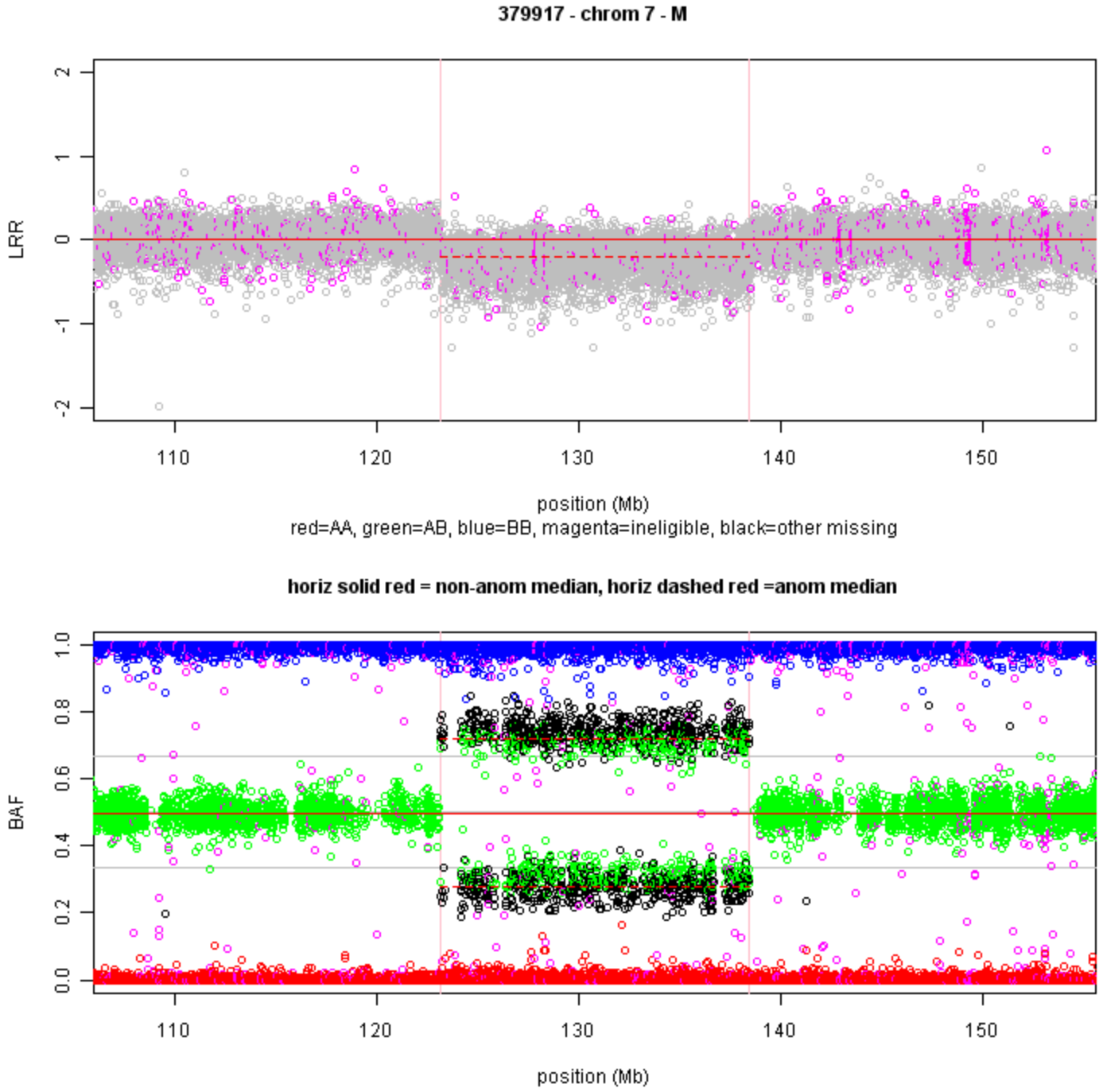
R = intensity of allele A + intensity of allele B
Theta = polar coordinate angle

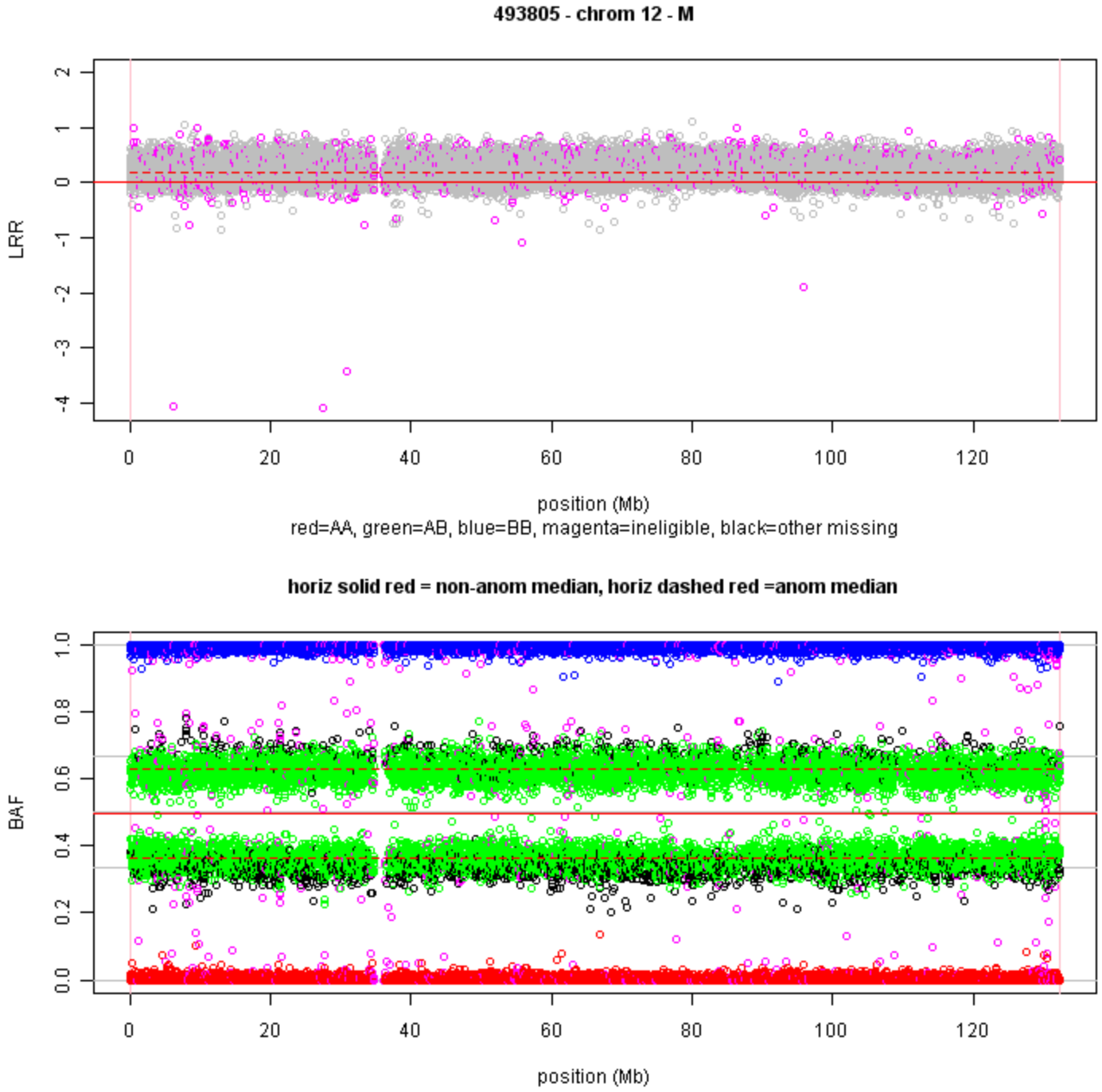
BAlleleFreq = continuous estimate of allele frequency in a single individual

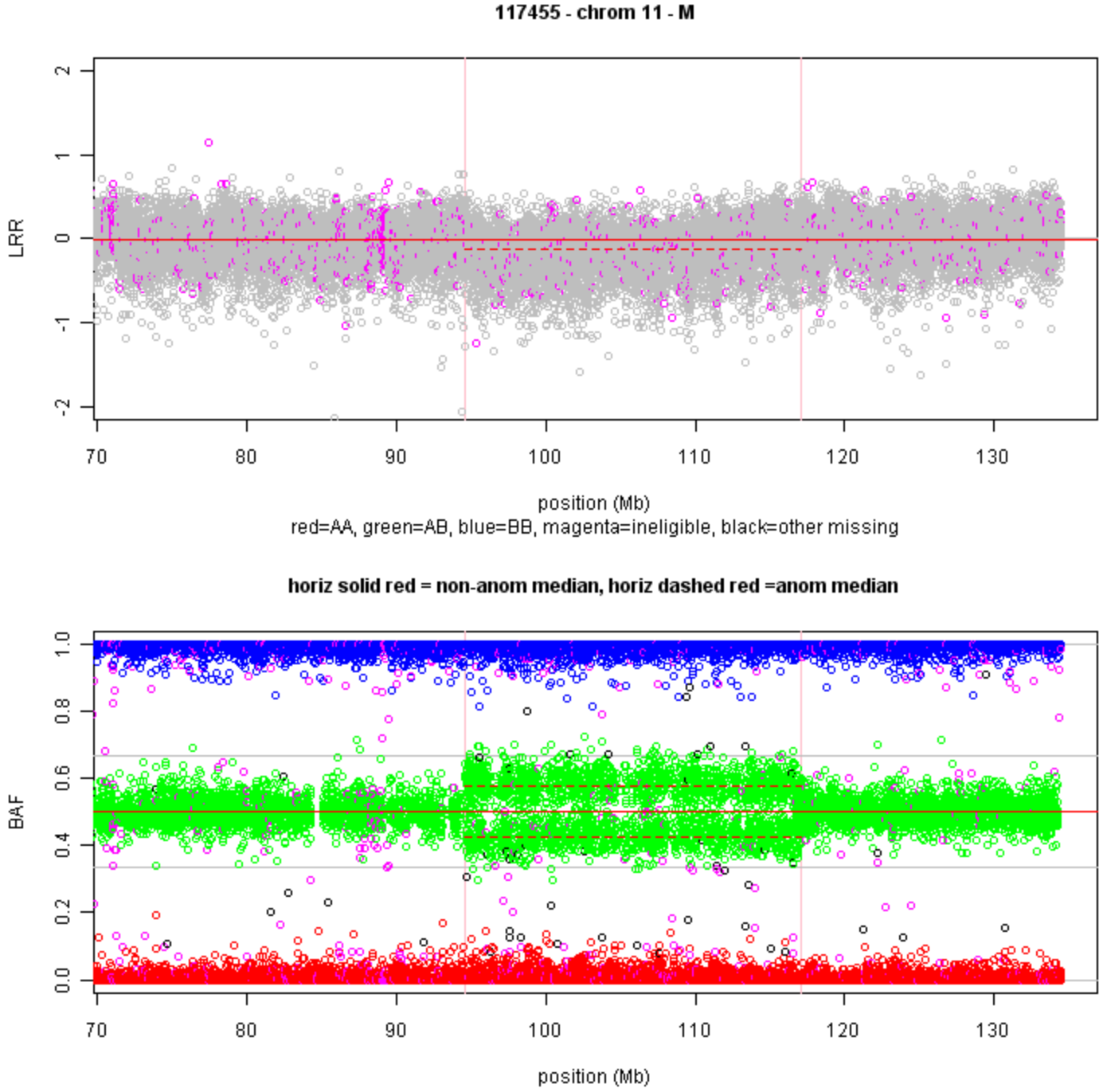

299913 - chrom 9 - M
LRR
2
1
0
-1
-2
-3
-4
0
20
40
60
position (Mb)
red=AA, green=AB, blue=BB, magenta=ineligible, black=other missing
horiz solid red = non-anom median, horiz dashed red =anom median
BAF
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
position (Mb)

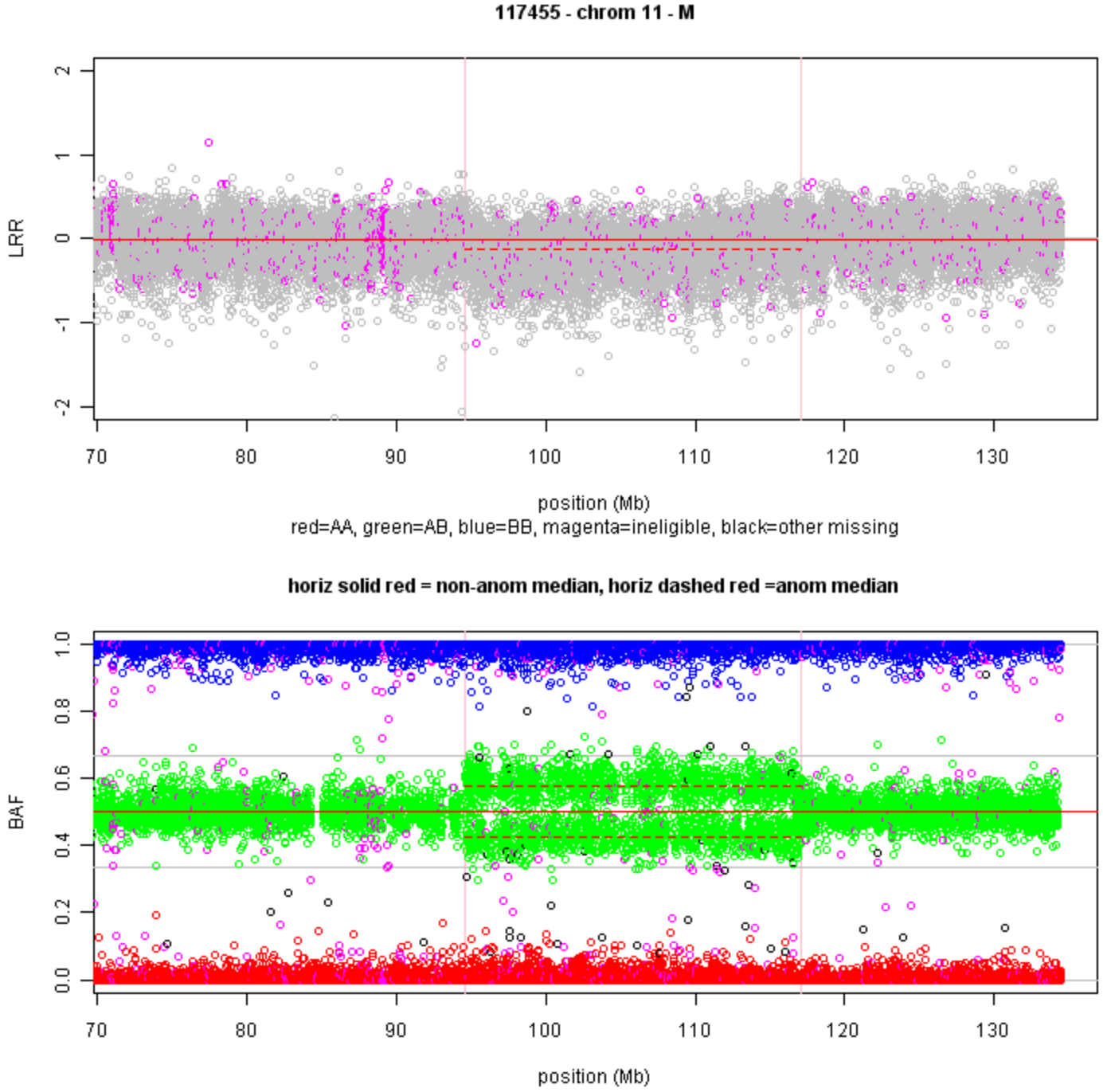

208810 - chrom 11 - M
LRR
2
1
0
-1
-2
95
100
105
110
115
120
125
position (Mb)
red=AA, green=AB, blue=BB, magenta=ineligible, black=other missing
horiz solid red = non-anom median, horiz dashed red =anom median
BAF
1.0
0.8
0.6
0.4
0.2
0.0
95
100
105
110
115
120
125
position (Mb)

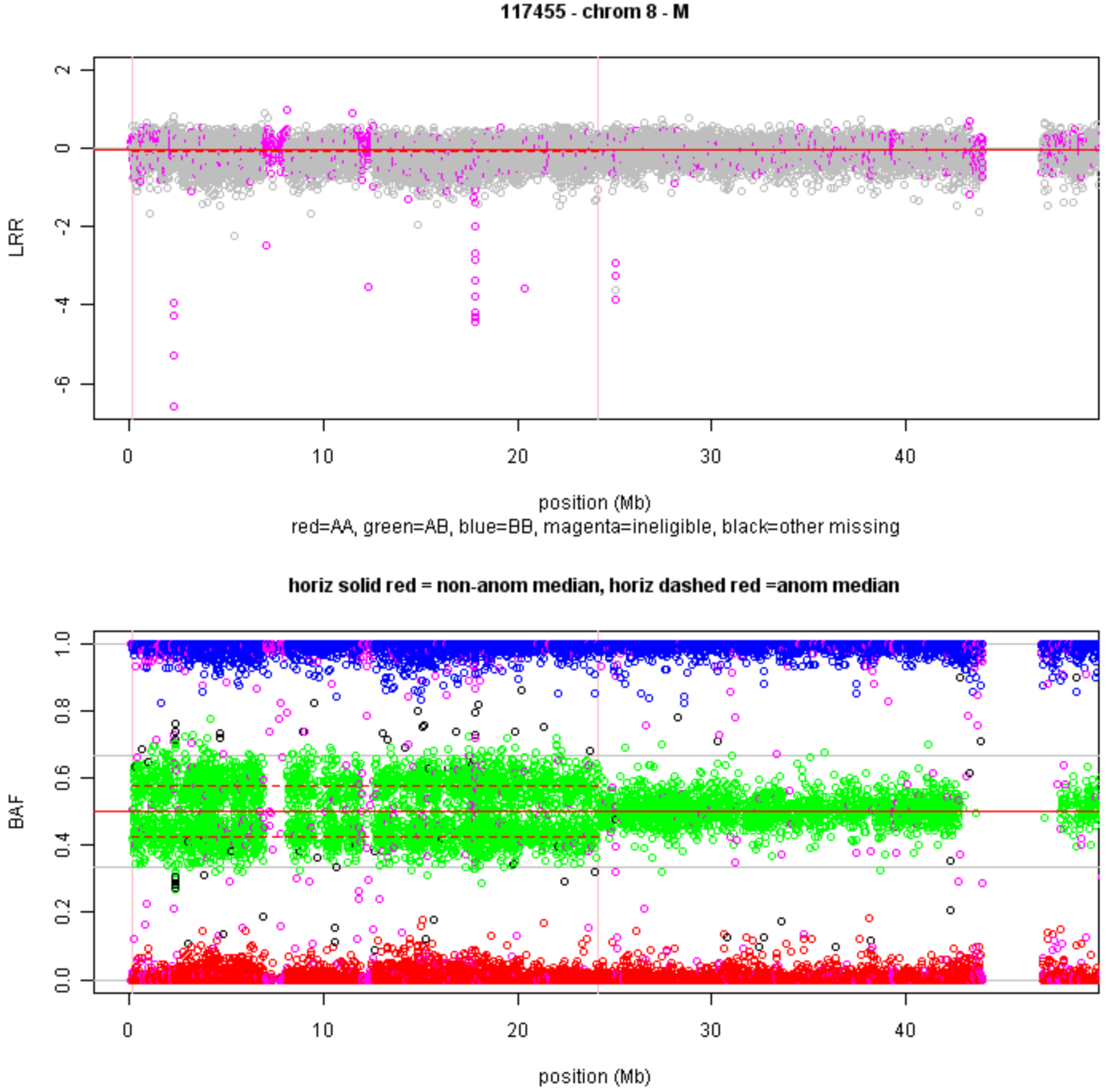

379917 - chrom 7 - M
LRR
2
1
0
-1
-2
110
120
130
140
150
position (Mb)
red=AA, green=AB, blue=BB, magenta=ineligible, black=other missing
horiz solid red = non-anom median, horiz dashed red =anom median
BAF
1.0
0.8
0.6
0.4
0.2
0.0
110
120
130
140
150
position (Mb)


379917 - chrom 7 - M
LRR
2
1
0
-1
-2
110
120
130
140
150
position (Mb)
red=AA, green=AB, blue=BB, magenta=ineligible, black=other missing
horiz solid red = non-anom median, horiz dashed red =anom median
BAF
1.0
0.8
0.6
0.4
0.2
0.0
110
120
130
140
150
position (Mb)

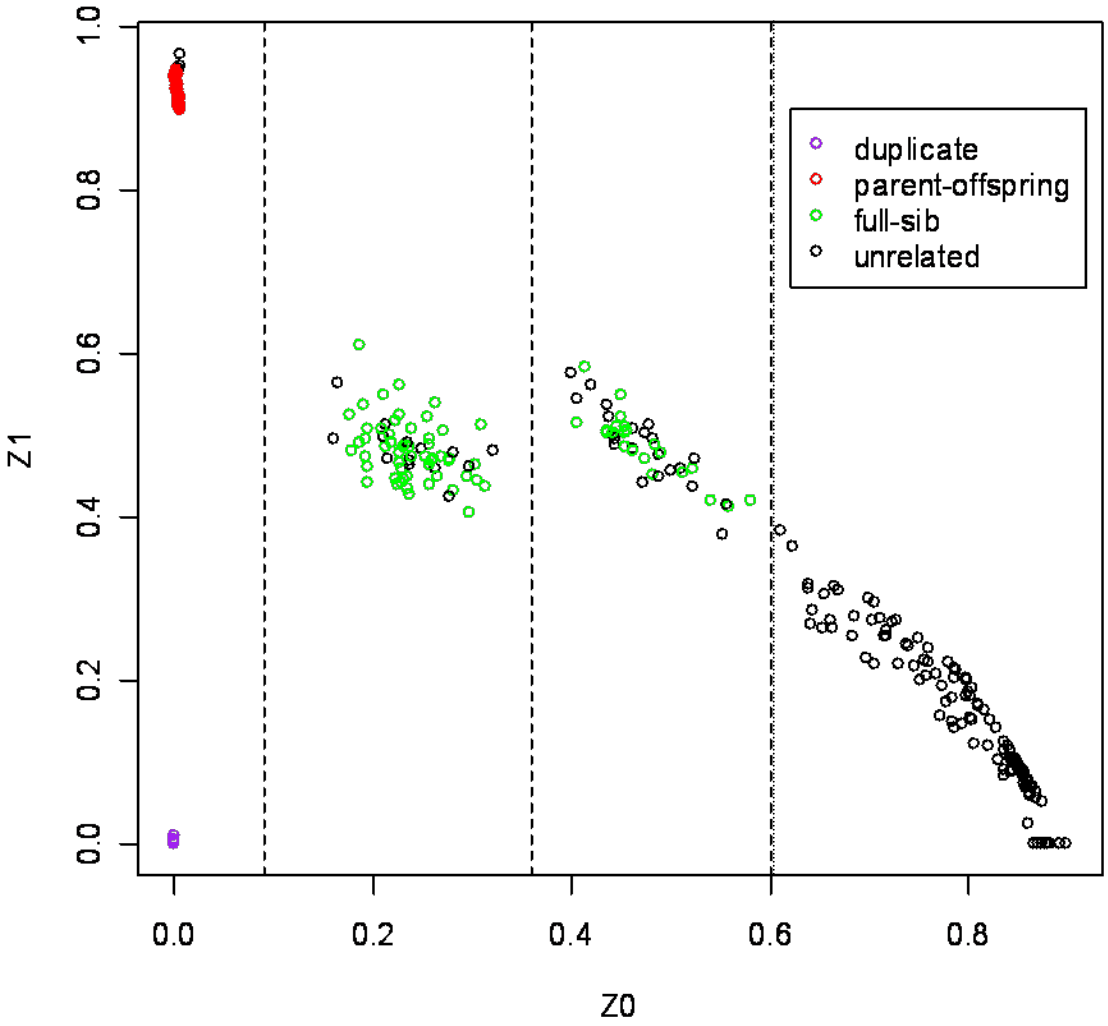

493805 - chrom 12 - M
LRR
2
1
0
-1
-2
-3
-4
0
20
40
60
80
100
120
position (Mb)
red=AA, green=AB, blue=BB, magenta=ineligible, black=other missing
horiz solid red = non-anom median, horiz dashed red =anom median
BAF
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
120
position (Mb)


117455 - chrom 11 - M
LRR
2
1
0
-1
-2
70
80
90
100
110
120
130
position (Mb)
red=AA, green=AB, blue=BB, magenta=ineligible, black=other missing
horiz solid red = non-anom median, horiz dashed red =anom median
BAF
1.0
0.8
0.6
0.4
0.2
0.0
70
80
90
100
110
120
130
position (Mb)


117455 - chrom 11 - M
LRR
2
1
0
-1
-2
70
80
90
100
110
120
130
position (Mb)
red=AA, green=AB, blue=BB, magenta=ineligible, black=other missing
horiz solid red = non-anom median, horiz dashed red =anom median
BAF
1.0
0.8
0.6
0.4
0.2
0.0
70
80
90
100
110
120
130
position (Mb)

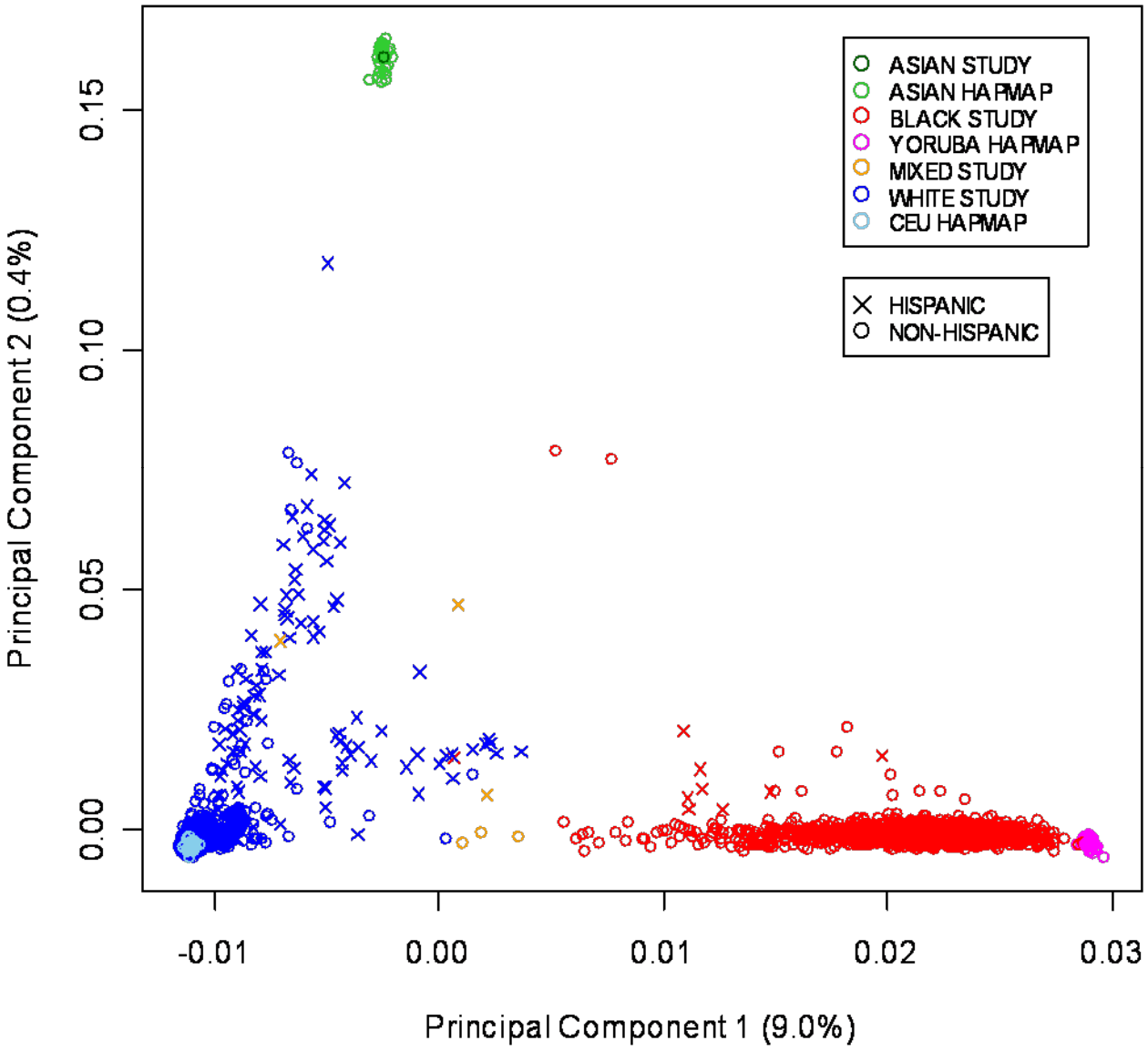

117455 - chrom 8 - M
LRR
2
0
-2
-4
-6
0
10
20
30
40
position (Mb)
red=AA, green=AB, blue=BB, magenta=ineligible, black=other missing
horiz solid red = non-anom median, horiz dashed red =anom median
BAF
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
position (Mb)

Estimated Relatedness

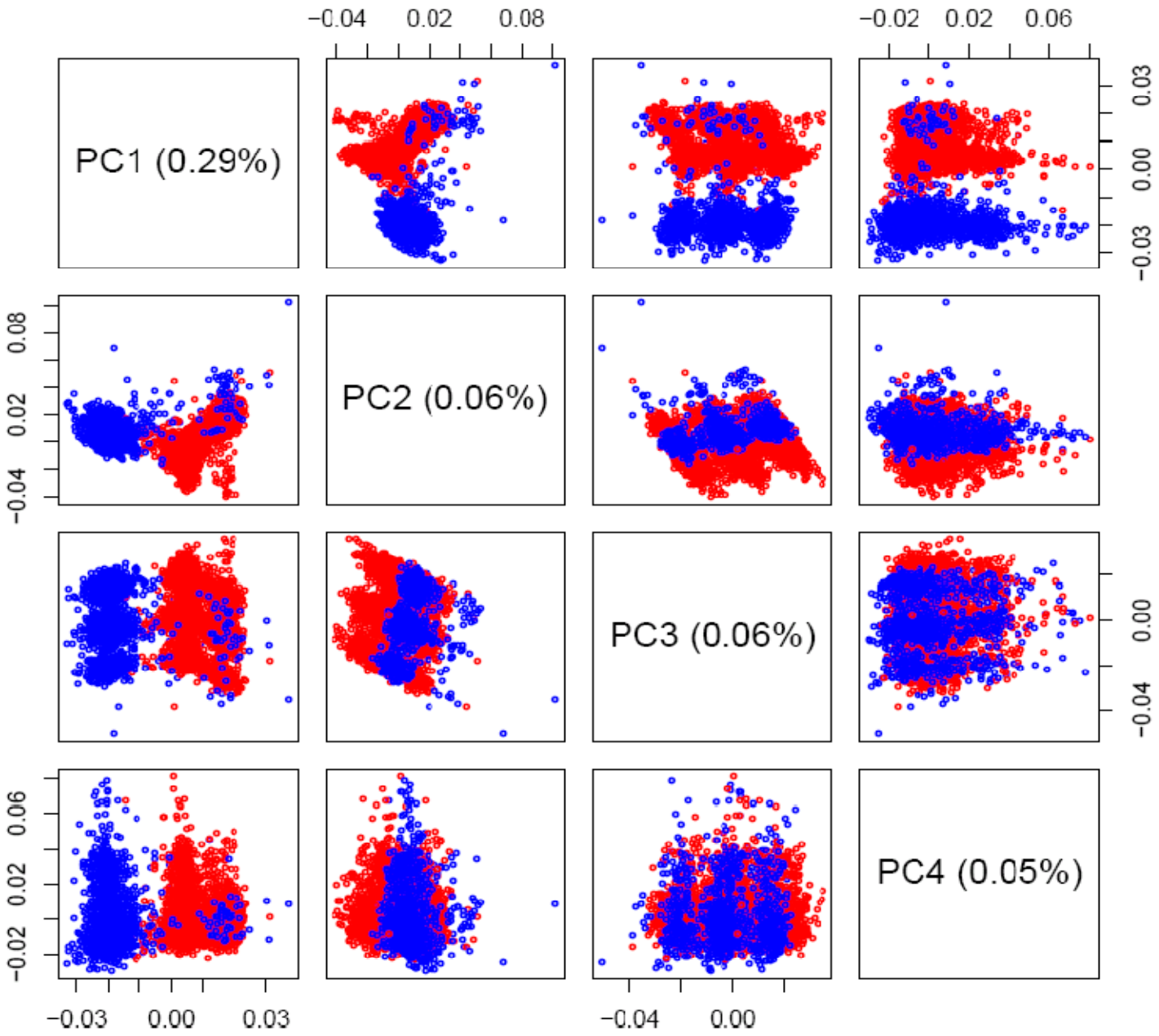


PLINK IBD estimates with kinship > 0.05

Population Structure

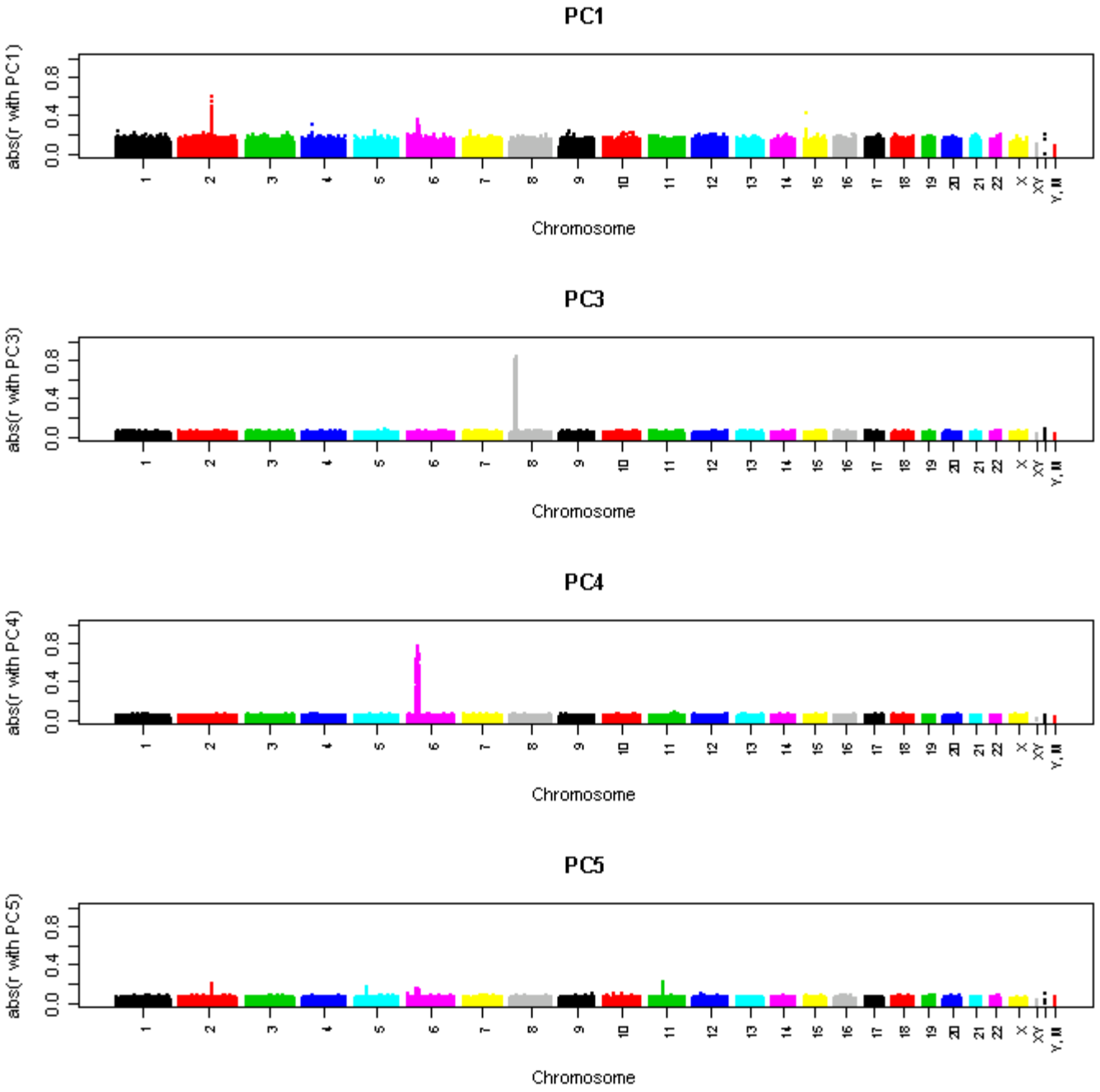


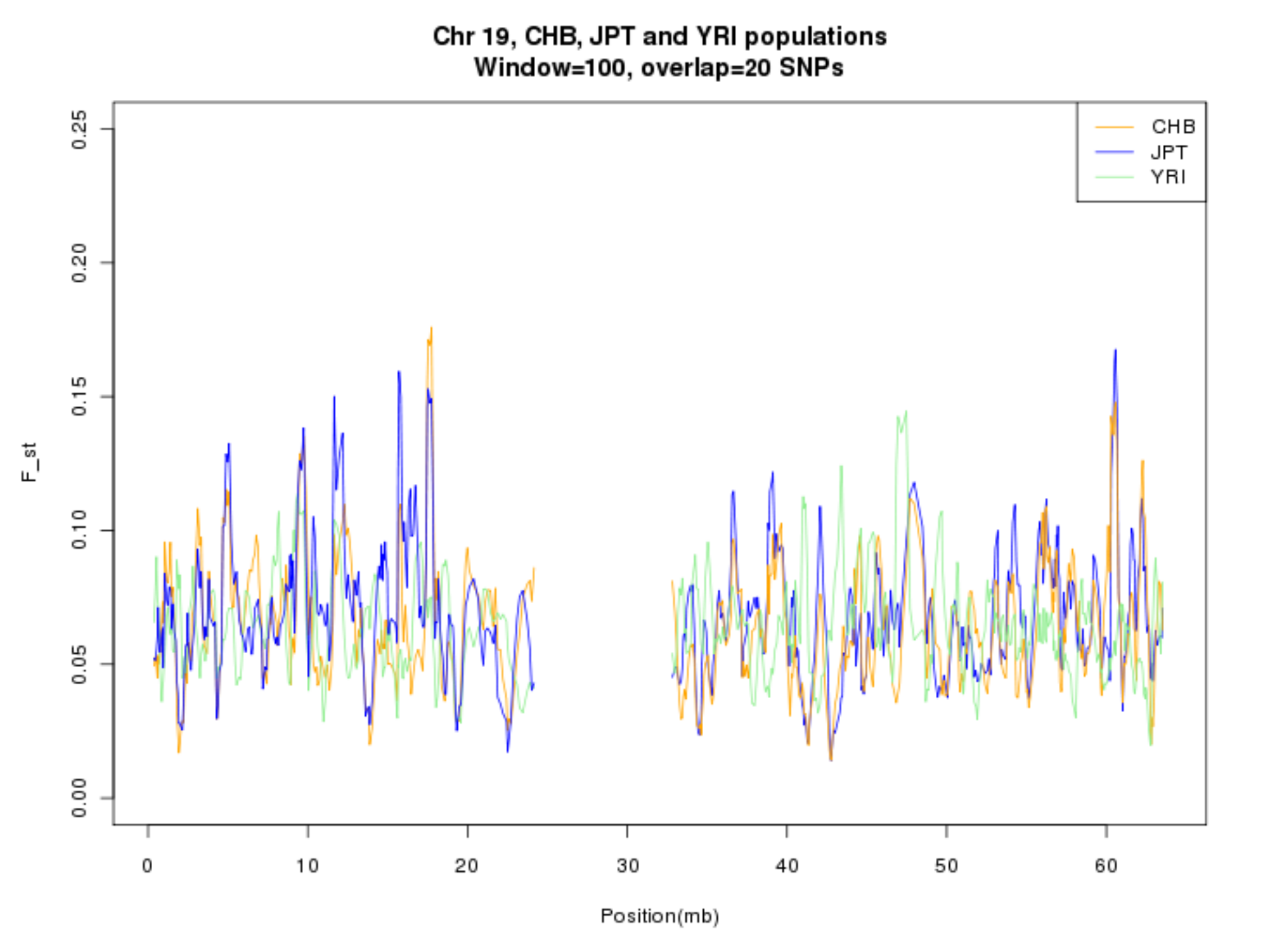
All unrelated samples (minus one outlier and minus HapMap)



SNP set A

All autosomal SNPs with missing rate < 5% used to calculate PCs




Chr 19, CHB, JPT and YRI populations
Window=100, overlap=20 SNPs
CHB
JPT
YRI
F_st
0.00
0.05
0.10
0.15
0.20
0.25
0
10
20
30
40
50
60
Position(mb)