

Genome-wide Association Approach for the Shanghai Breast Cancer Study

Chun Li

Department of Biostatistics
Center for Human Genetics Research
Vanderbilt University

August 15, 2008



Shanghai Breast Cancer Study (SBCS)

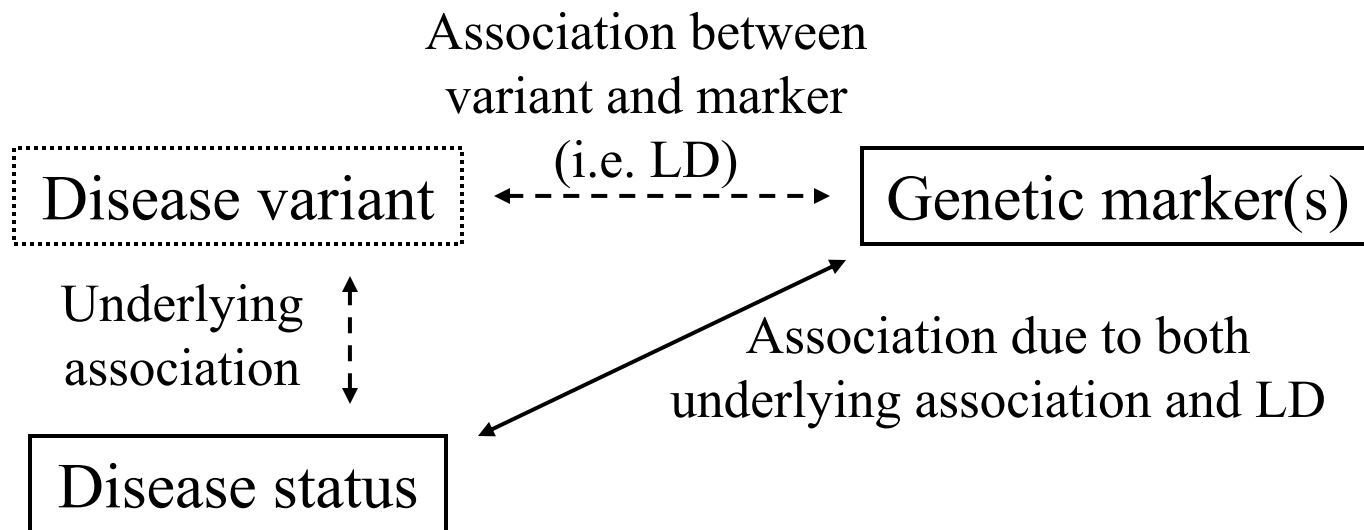
- Started in 1996 (PI: Dr. Wei Zheng)
- Already recruited ~3,500 cases and ~3,500 controls; still recruiting subjects
- Unique features:
 - East Asian population with low disease prevalence
 - Extensive survey and clinical data
 - 99% have tumor tissue collection

Genome-Wide Association Studies (GWAS)

Feasible due to success of the HapMap Project and technology improvement.

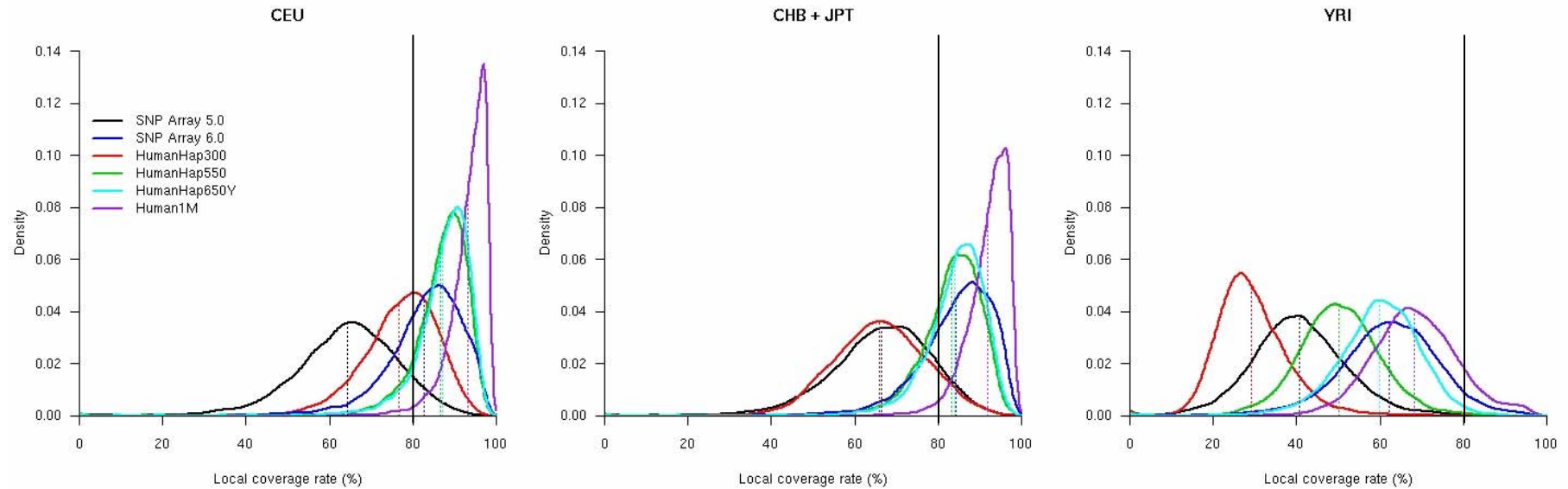
- Late 2005: ~ \$1,300 for Affymetrix 500K
- Late 2007: ~ \$300 for Affymetrix SNP Array 6.0 (900K SNPs + 900K non-polymorphic probes)

Association Analysis



LD: Linkage disequilibrium

Coverage of Euchromatic Regions



SNP Chip	Overall coverage (%)			Fraction with coverage <80%		
	CEU	CHB+JPT	YRI	CEU	CHB+JPT	YRI
SNP Array 5.0	64	66	41	.91	.88	1.00
SNP Array 6.0	83	84	62	.31	.26	.95
HumanHap300	77	66	29	.59	.89	1.00
HumanHap550	87	83	50	.12	.28	1.00
HumanHap650Y	87	84	60	.10	.23	.99
Human1M	93	92	68	.02	.03	.88

Association Analysis in Case-Control Studies

Rationale: Cases are more likely to carry disease-predisposing variants than controls. In other words, the fraction of disease variants is expected to be higher in cases than that in controls.

This is the basis of association analysis, in which we look for *association* between case/control status and variant/normal allele state:

	Variants	Normal alleles	Total
Cases	50	50	100
Controls	25	75	100

Multiple comparison is a big issue in GWAS.

Association Analysis of SNPs in Case-Control Studies

Commonly used tests:

Allele-based: Pearson's χ^2 test on
 2×2 table (additive model).

	<i>A</i>	<i>a</i>
Case	125	75
Control	116	84

Trend test (additive model).

“Dominant” model: Collapse *AA/Aa*
and test on resulting 2×2 table.

“Recessive” model: Collapse *Aa/aa*
and test on resulting 2×2 table.

Pearson's χ^2 test on 2×3 table.

	<i>AA</i>	<i>Aa</i>	<i>aa</i>
Case	40	45	15
Control	36	44	20

GWAS for the SBCS

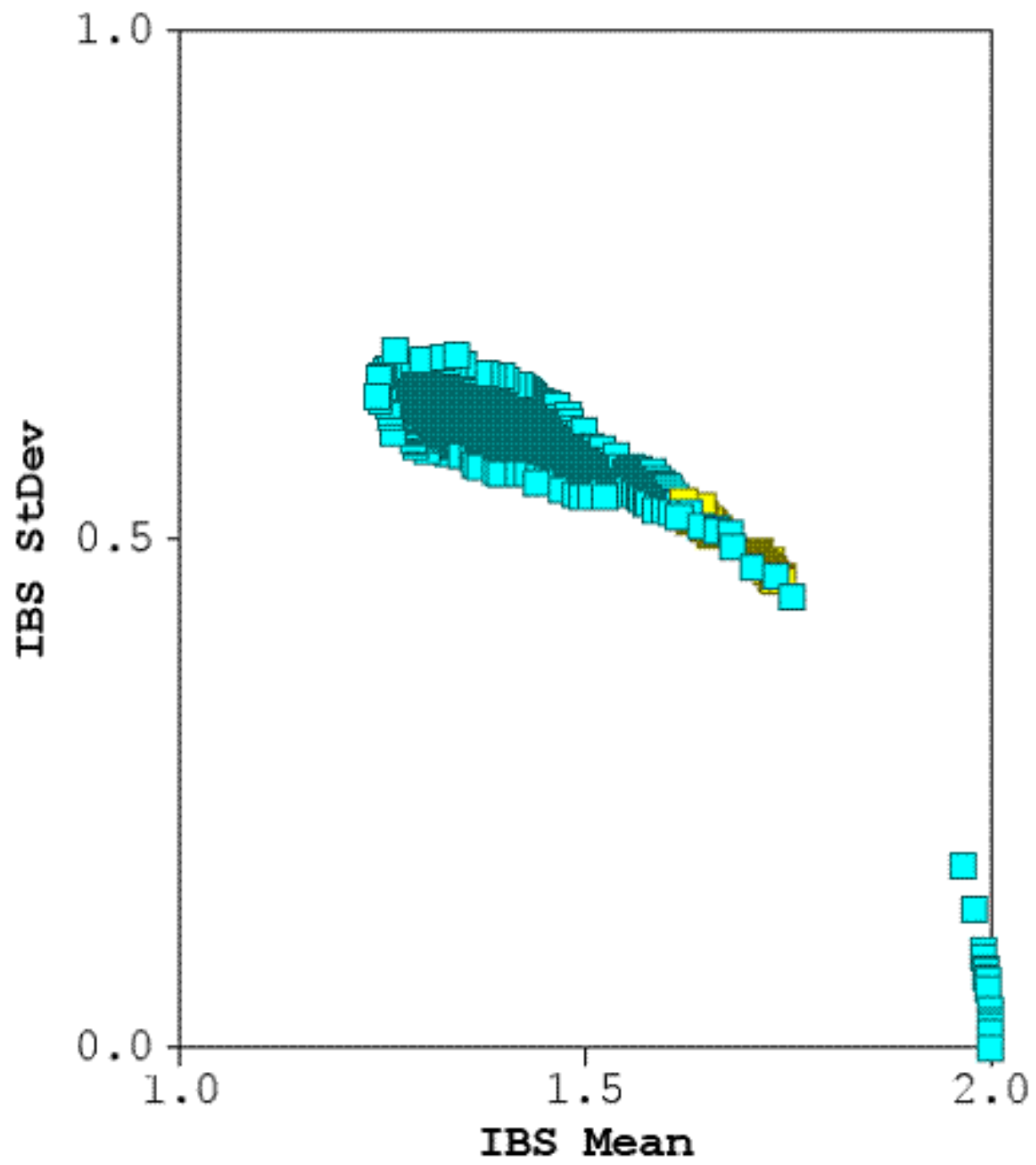
- Initial plan:
 - Genome-wide scan in 1,000 cases/1,000 controls
 - Validate 10,400 promising SNPs in 1,500 cases/1,500 controls
 - Validate 300 promising SNPs in 1,000 cases/2,000 controls
 - Evaluate gene-gene and gene-environment interactions in 3,500 cases/4,500 controls
- Current plan:
 - Genome-wide scan in ~2,000 cases/~2,000 controls
 - Validate ~500 promising SNPs in ~2,000 cases/~2,000 controls

Analysis Approaches

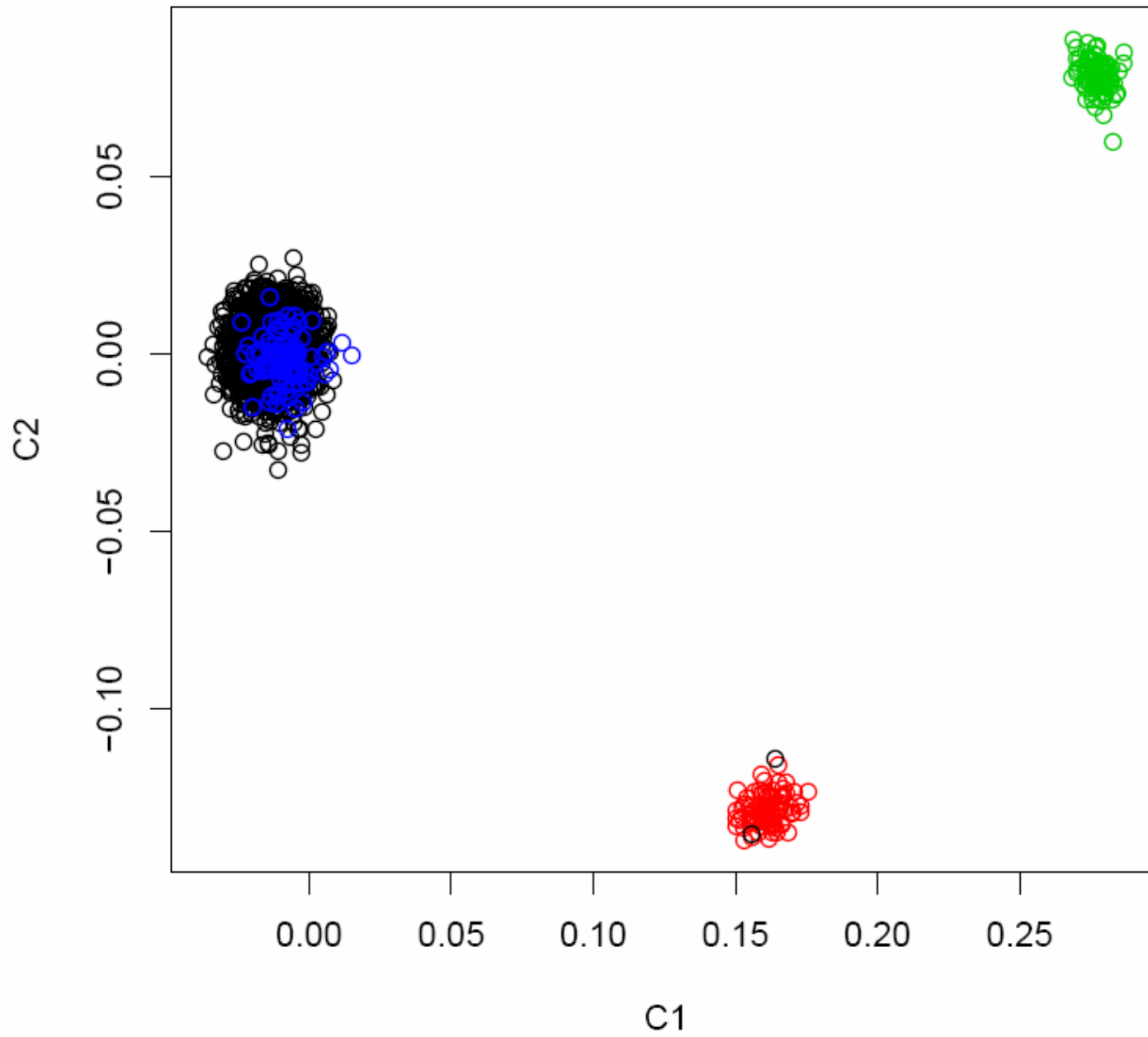
- Single SNP association analysis
- Multiple-SNP association analysis
- Imputation for
 - analysis of untyped markers
 - combined analysis of data genotyped with different SNP chips
- Copy-number variation (CNV) detection and analysis

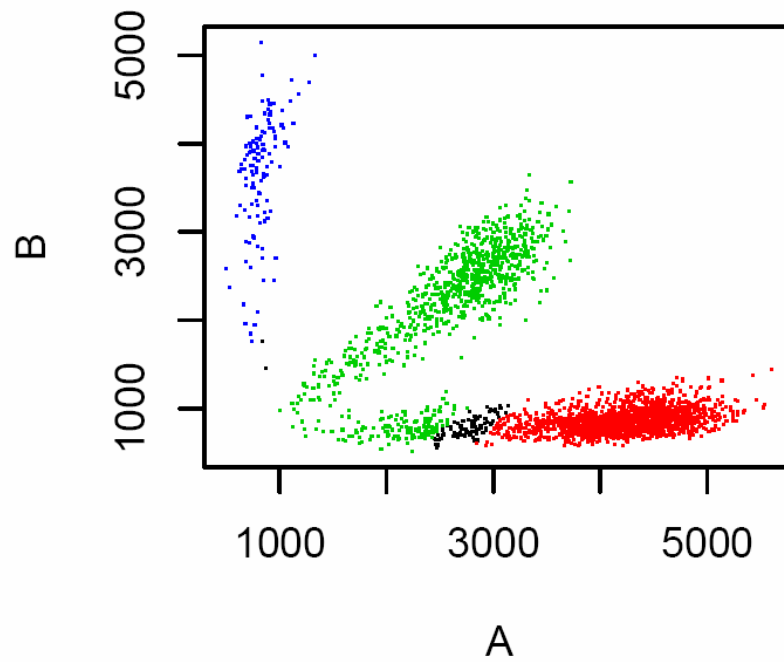
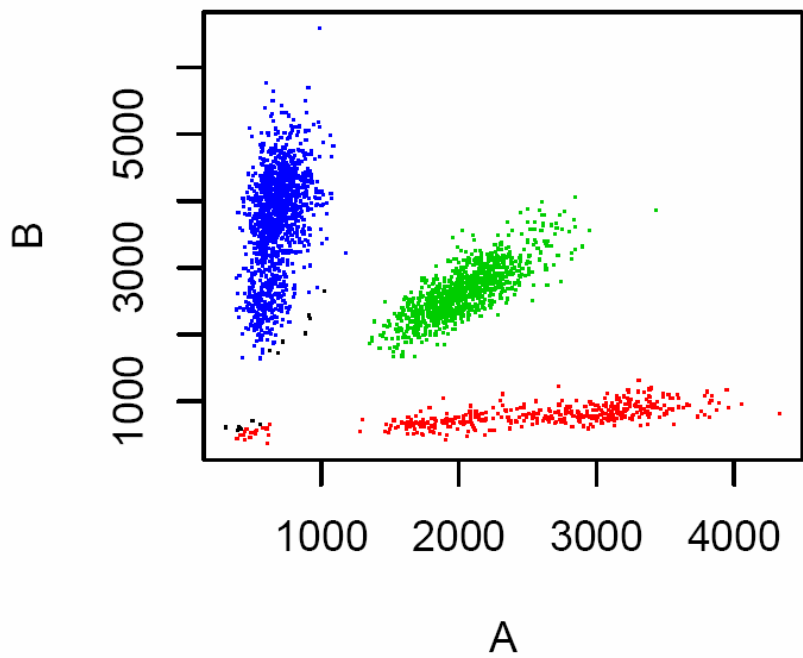
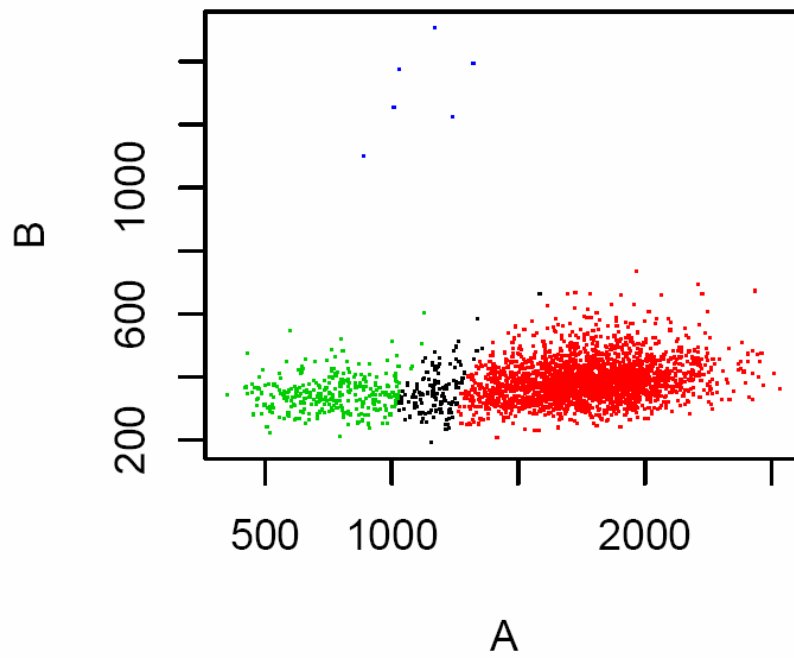
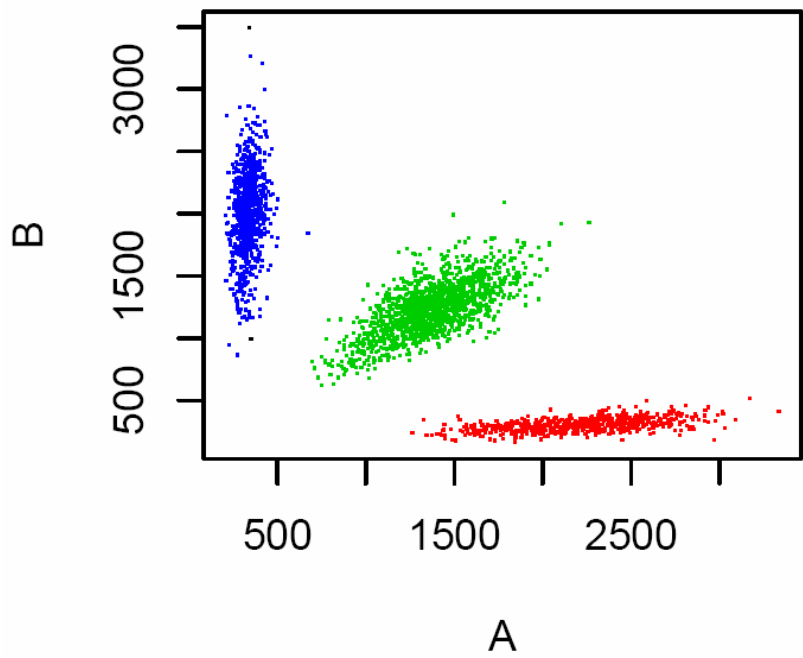
Data Cleaning

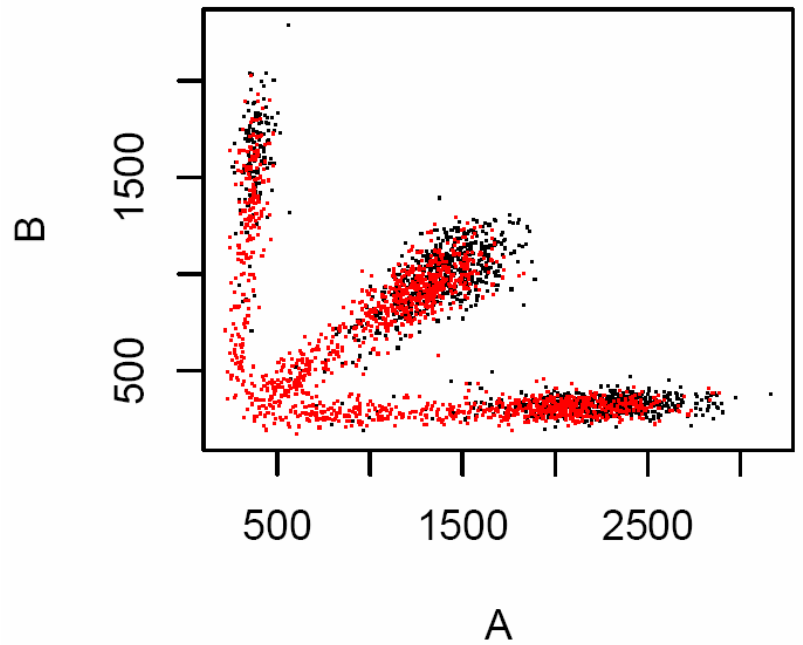
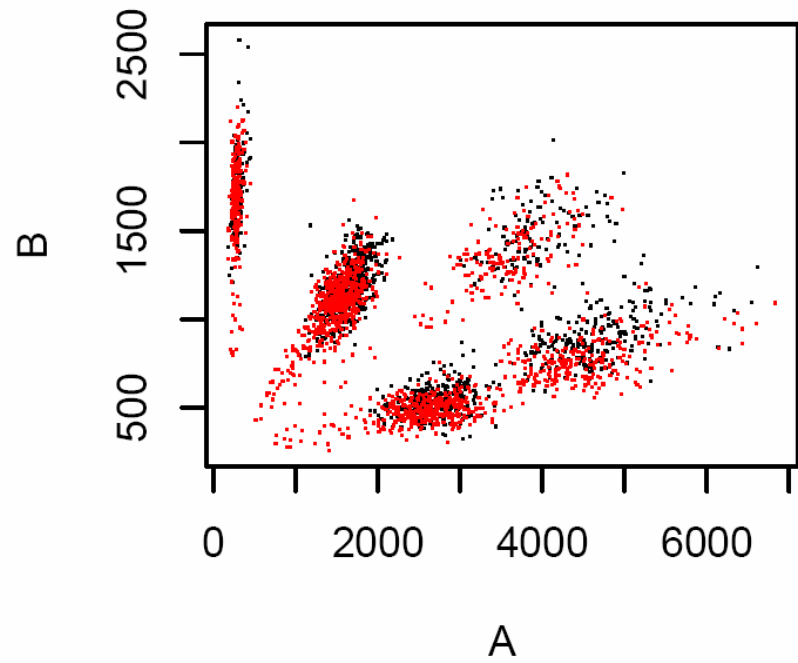
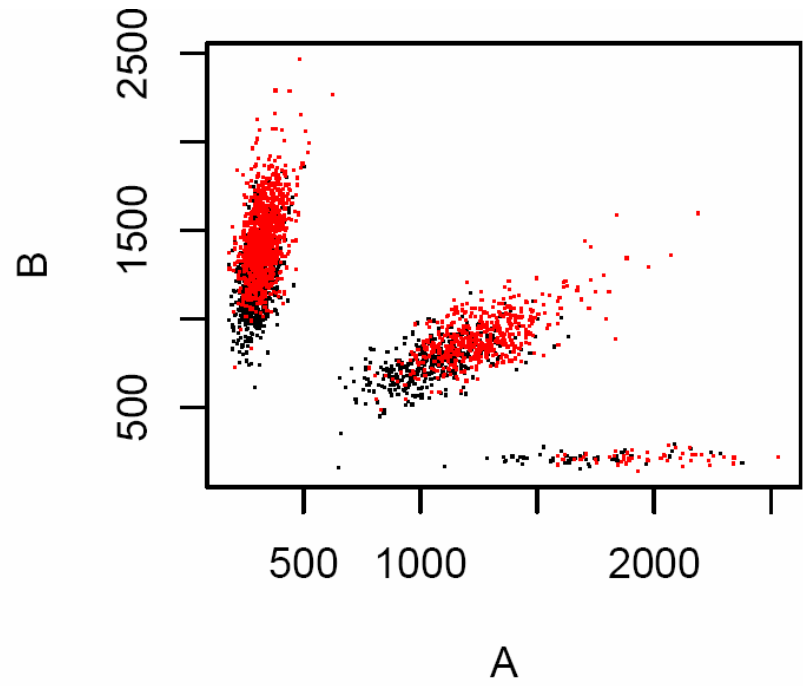
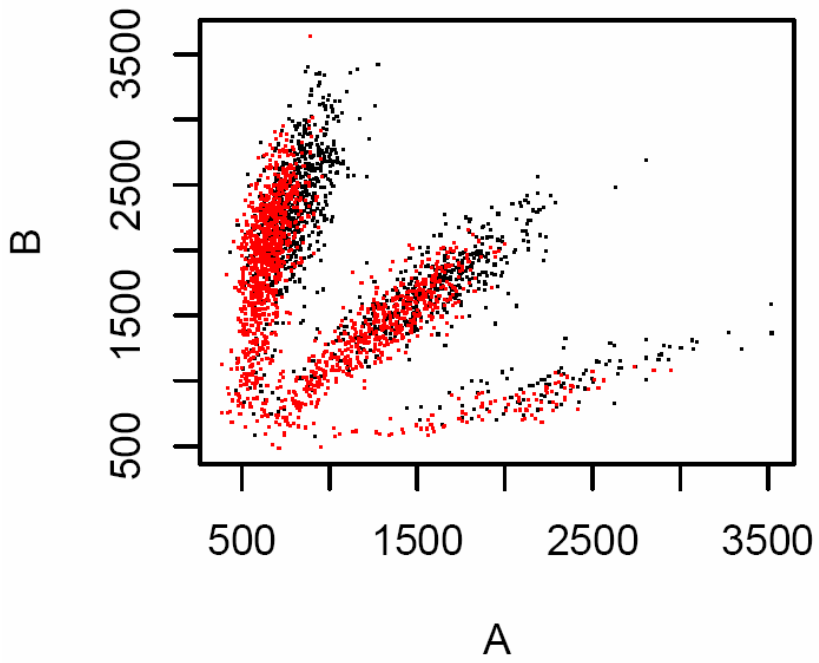
- Relationship checking
- Population structure
- Genotype call performance
- Exclusion of SNPs (low call rate or low minor allele frequency)



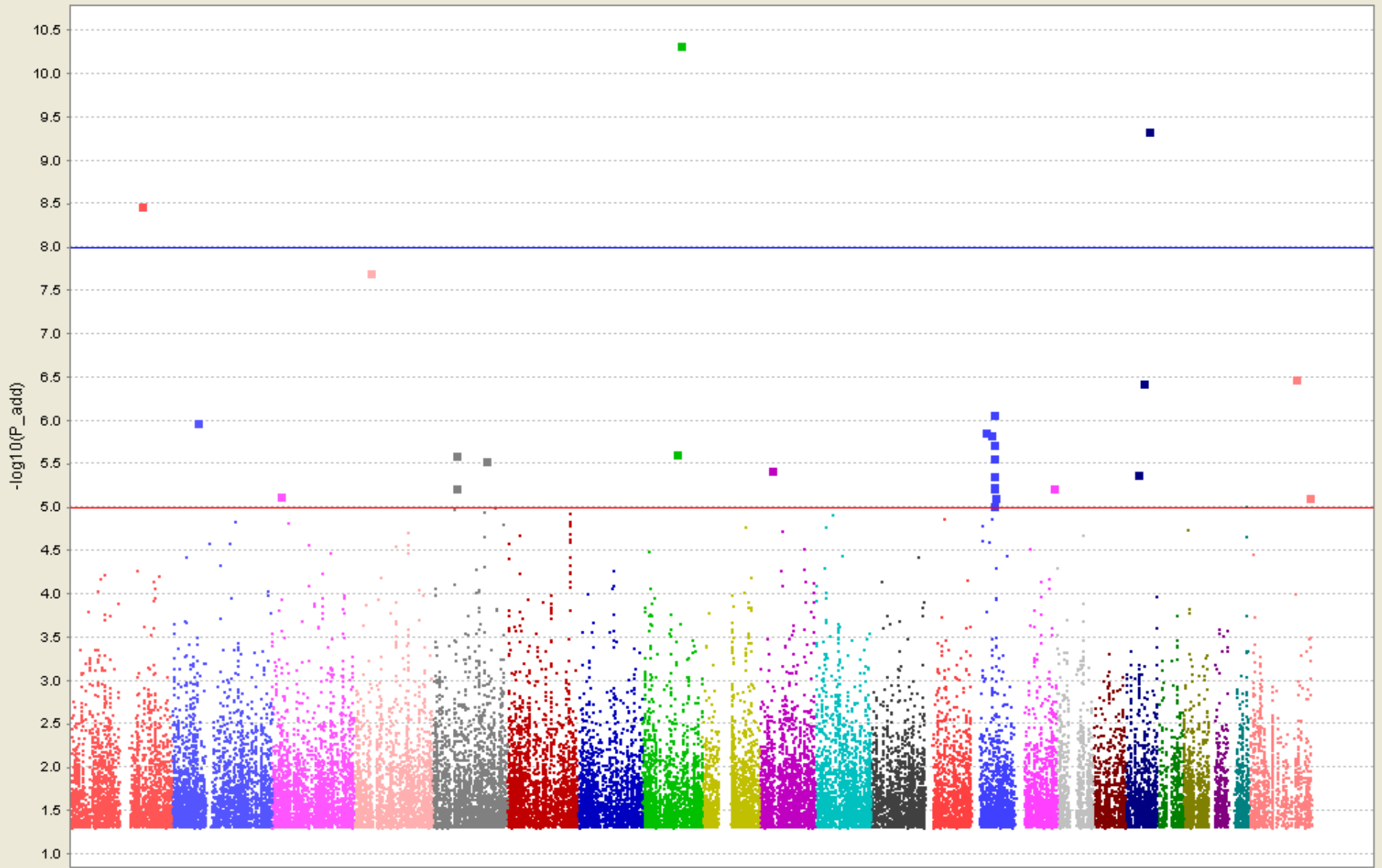
SBCS2994-HapMap-MDS







P_trend_0.05_maf_0.05_genos_0.95



Chr1 Chr2 Chr3 Chr4 Chr5 Chr6 Chr7 Chr8 Chr9 Chr10 Chr11 Chr12 Chr13 Chr14 Chr15 Chr16
Chr17 Chr18 Chr19 Chr20 Chr21 Chr22 ChrX

Genome-wide Analysis of CNV

Stage I: Conduct a CNV genome wide association scan using the intensity data of the GWAS

Stage II: Evaluate 500 most promising CNVs identified in Stage I in an independent set of 1,500 cases and 1,500 controls

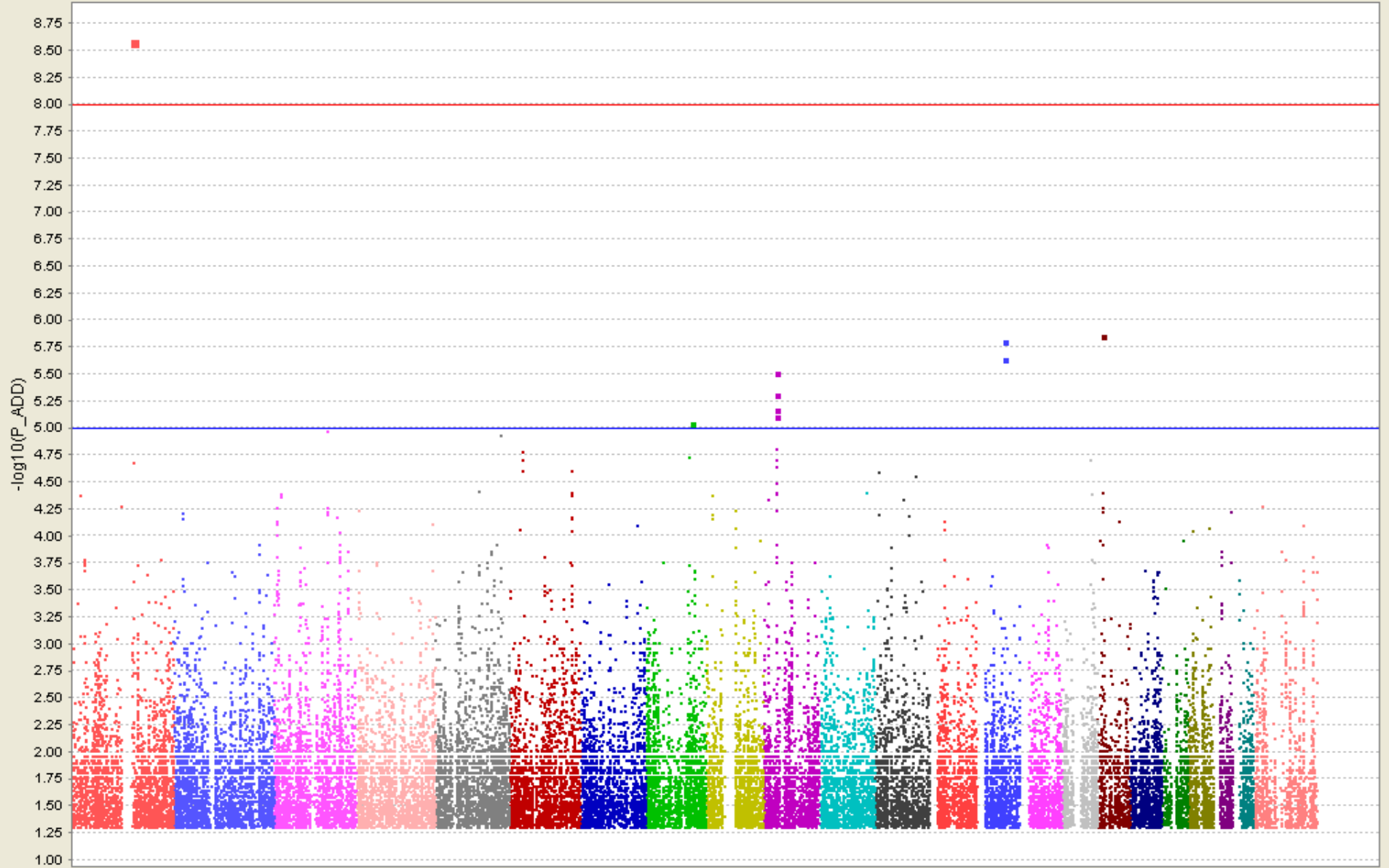
Stage III: Validate all promising CNVs (~30) identified in Stage II in a nested case-control study of 1,000 cases and 2,000 controls from the Shanghai Women's Health Study

Genome-wide Analysis of Overall and Disease-free Survival

Overall survival: 1,395 subjects; 255 events
(death from any cause)

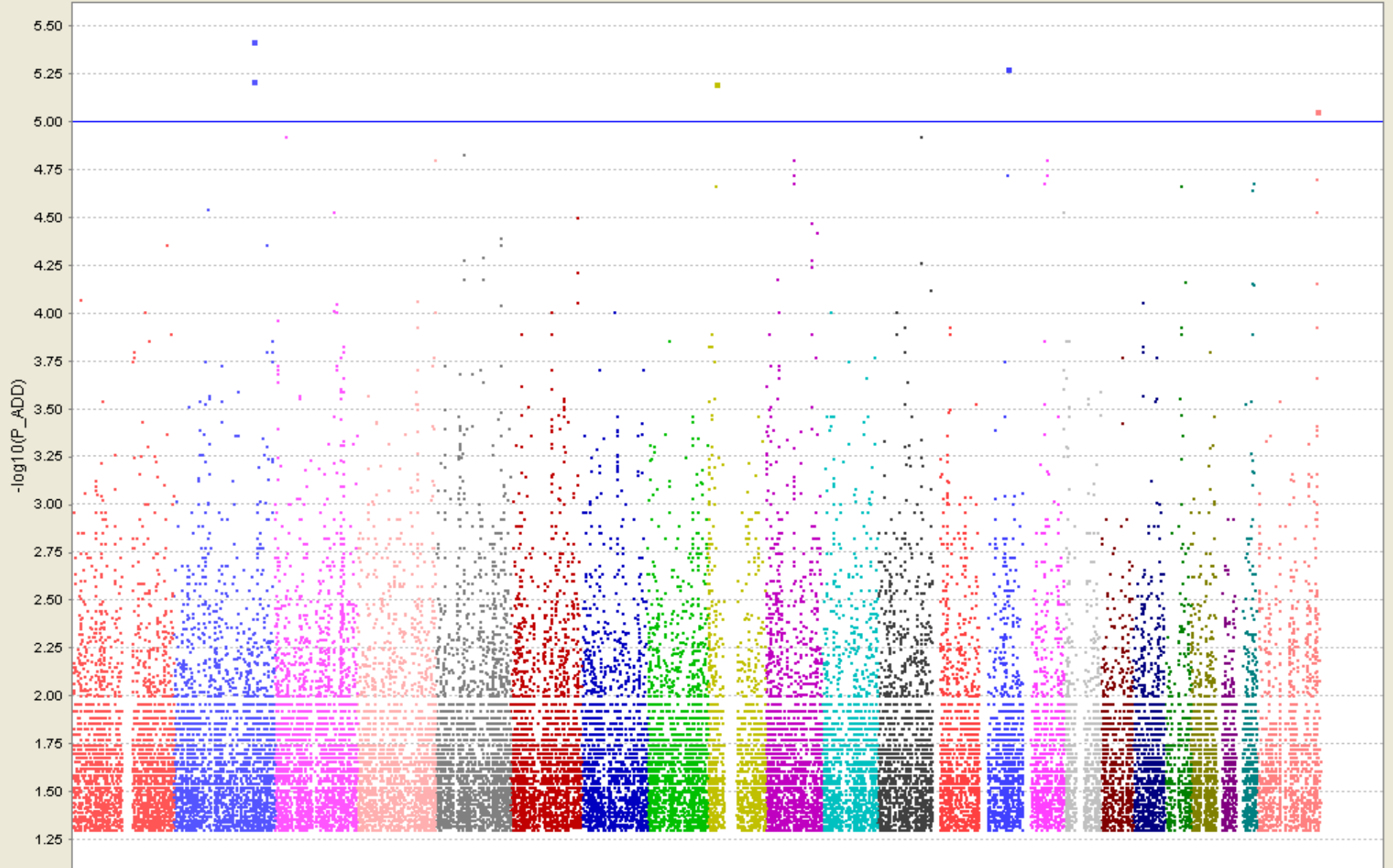
Disease-free survival: 1,225 “disease-free”
subjects; 253 events (cancer recurrence or
metastasis, death related to breast cancer)

Overall survival



Chr1 Chr2 Chr3 Chr4 Chr5 Chr6 Chr7 Chr8 Chr9 Chr10 Chr11 Chr12 Chr13 Chr14 Chr15 Chr16
Chr17 Chr18 Chr19 Chr20 Chr21 Chr22 ChrX

DF



Chr1 Chr2 Chr3 Chr4 Chr5 Chr6 Chr7 Chr8 Chr9 Chr10 Chr11 Chr12 Chr13 Chr14 Chr15 Chr16
Chr17 Chr18 Chr19 Chr20 Chr21 Chr22 ChrX