SUPPLEMENT

Dissection of genetically complex traits with extremely large pools of yeast segregants

by

Ehrenreich, I.M. *et al.*

SUPPLEMENTAL TABLES

Chemical	Selective dose	Source of chemical
4-nitroquinoline	0.5 ug/ml	Sigma
Benomyl	600 mg/ml	Chem Service
Cadmium chloride	300 uM	Sigma
Chlorpromazine	20 uM	Fluka
Cobalt chloride	3 M	Sigma
Copper sulfate	12 mM	Sigma
Cycloheximide	50 ng/ml	Sigma
Diamide	1.5 mM	Sigma
E6-berbamine	100 uM	Enzo Life Sciences
Ethanol	8%	Sigma
Hydrogen peroxide	6 mM	Fisher
Manumycin	72.8 uM	Enzo Life Sciences
Neomycin	5 mg/ml	Sigma
Paraquat	3 mM	Sigma
SDS	0.15%	Sigma
Tunicamycin	3 uM	Biomol
Zeocin	100 ug/ml	Invitrogen

Table S1. The chemicals and doses used in our experiments.

Gene	Chromosome	Position (bp)	Estimated Peak Location (bp)	Distance from Causal Gene (bp)
RAD5	XII	204992-208501	203439	1553
MKT1	XIV	467133-469625	470929	1304

Table S2. Information regarding how precisely known loci for 4-NQO resistance were mapped using X-QTL.

SUPPLEMENTAL FIGURES

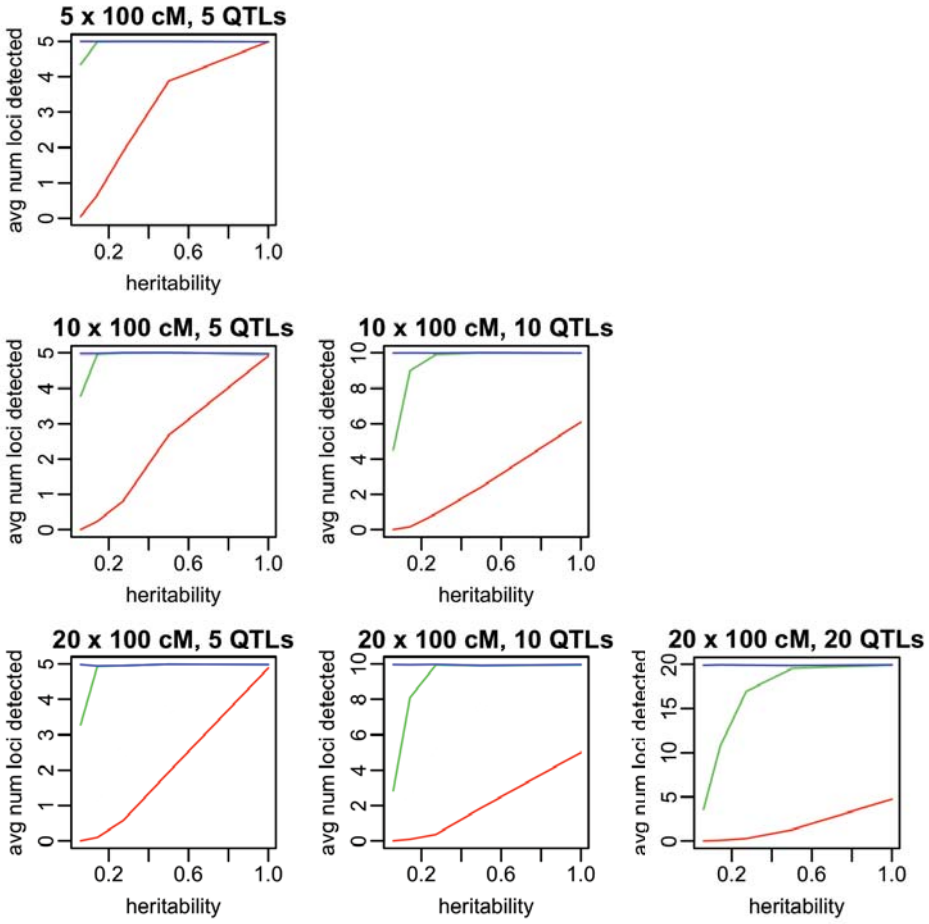

Figure S1A. Power of X-QTL to detect loci as a function of ploidy, heritability, genome size, segregating population size, and genetic architecture. Results for haploids with 100 cM chromosomes and populations of 10^3 (red), 10^4 (green), and 10^5 (blue) cross progeny are shown. Genomes were simulated with 5, 10, or 20 chromosomes and genetic architectures of 5, 10, or 20 loci with equal, additive effects. We required the number of loci in a simulation to be less than or equal to the number of chromosomes in order to minimize the effect of linkage on our power estimates. The average number of loci detected per simulation is plotted on the y axis, while the heritability is plotted on the x axis. These simulations show X-QTL should be powerful to simultaneously detect many loci with small effects. For example, effect sizes range from 1 to 20% variance explained in the case of 5 loci, 0.5 to 10% variance explained in the case of 10 loci, and 0.25 to 5% variance explained in the case of 20 loci.

In each simulation, genomic positions were randomly selected to be QTLs, with each QTL required to be at least 10 cM away from any other QTL. Phenotypes were assigned to the segregants based on their genotypes at each of the QTLs. The phenotypes of the segregants (**p**) were thus modeled as

$$\mathbf{p} = \mathbf{Ga} + \mathbf{e},$$

where **G** is a design matrix containing the genotype of each segregant at each additive QTL, **a** is a vector of equal and additive QTL effects that are scaled to the heritability, and **e** is a vector of environmental noise generated as random deviates sampled from a normal distribution of mean = 0 and standard deviation = square root(1 - heritability). In haploids, QTL effects were set to equal 2 × square root(heritability/(number of QTLs)). In diploids, a factor of square root(heritability/(number of QTLs)) was added to the phenotype of individuals heterozygous for a QTL, while homozygous individuals had 2 × square root(heritability/(number of QTLs)) added. All high alleles were set to come from a single cross parent.

After a segregating pool was simulated, individuals were sampled from the top 5% tail of the phenotypic distribution and marker allele frequencies across the genome were assessed in these pools of selected individuals. The genotypes of the selected individuals were checked at markers placed every cM. Peaks were defined as significant if the allele counts at a simulated QTL surpassed a Bonferroni-corrected threshold of $\alpha = 0.05$ using the χ^2 statistic, where the number of tests applied was calculated as the length of the genome in cM. Perfect counting of allele frequencies in groups of selected individuals was assumed, so these results indicate a theoretical maximum of how well X-QTL could work. Each combination of parameters was simulated at least 50 times.

Our simulation results suggest that X-QTL could be useful under a variety of scenarios in both haploid and diploid organisms when very large populations can be grown and selected, and precise and quantitative genome-wide genotyping is available. X-QTL can be conducted in yeast with segregating pools containing more than 10^5 haploid individuals, suggesting that this approach should be extremely powerful in this organism.

The numerical simulations reported here were conducted using custom Python scripts.

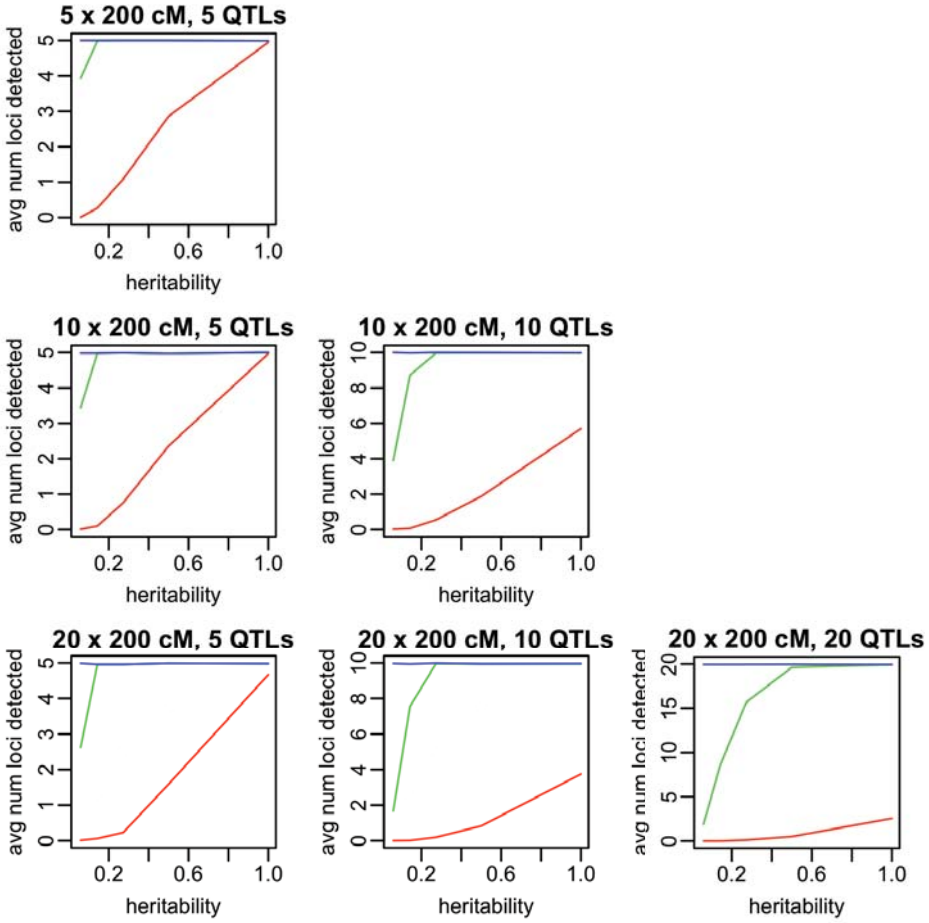

Figure S1B. Average number of loci detected by X-QTL for haploids with 200 cM chromosomes under a variety of scenarios. Populations of 10^3 (red), 10^4 (green), and 10^5 (blue) cross progeny are shown.

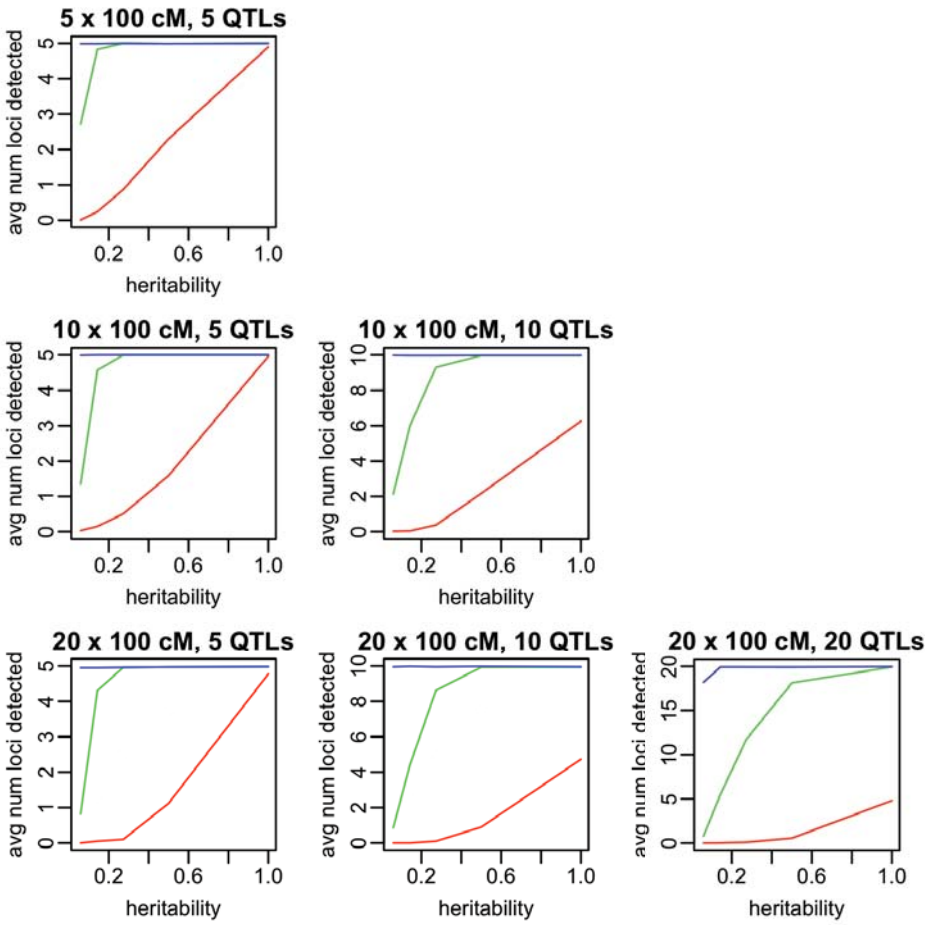

Figure S1C. Average number of loci detected by X-QTL for diploids with 100 cM chromosomes under a variety of scenarios. Populations of 10^3 (red), 10^4 (green), and 10^5 (blue) cross progeny are shown.

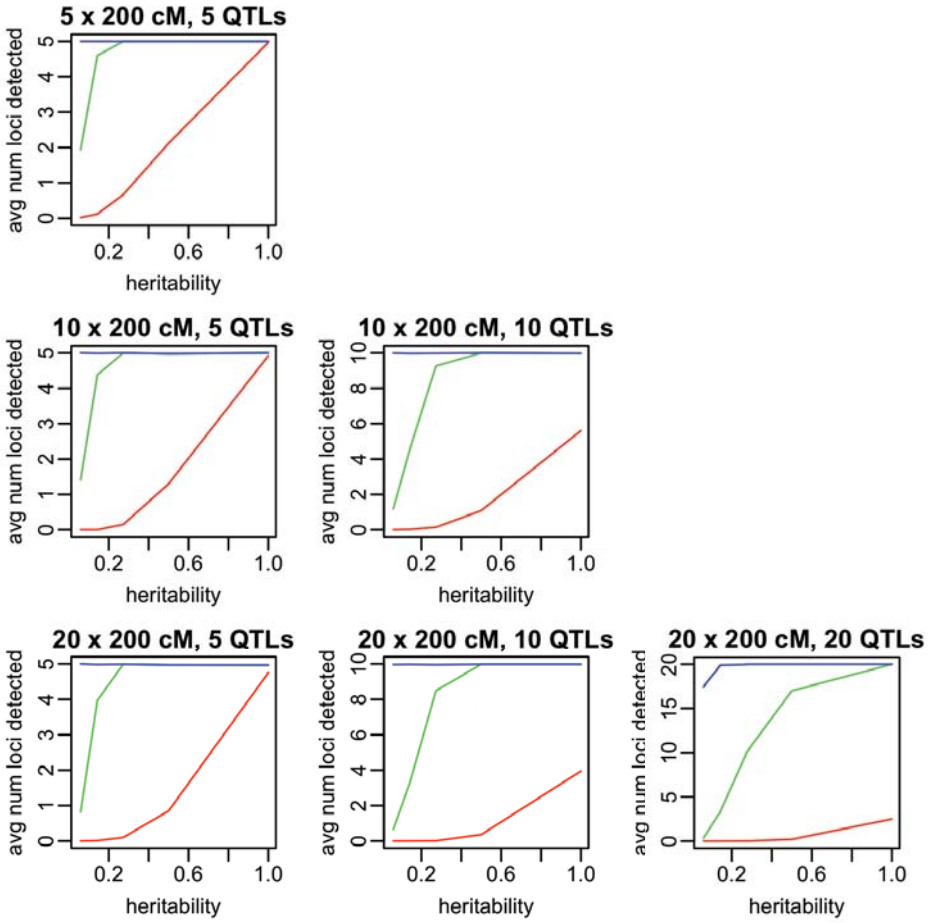

Figure S1D. Average number of loci detected by X-QTL for diploids with 200 cM chromosomes under a variety of scenarios. Populations of 10^3 (red), 10^4 (green), and 10^5 (blue) cross progeny are shown.

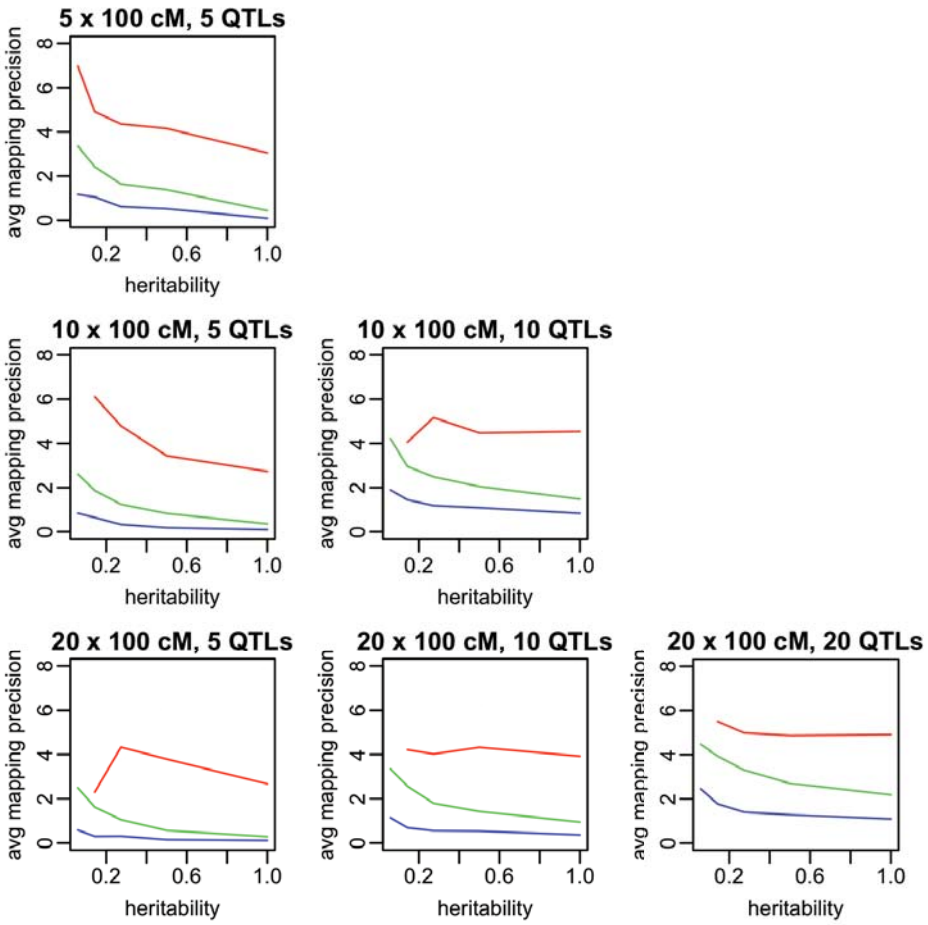

Figure S2A. Average mapping precision of X-QTL (in cM) for haploids with 100 cM chromosomes under a variety of scenarios. Starting populations of 10^3 (red), 10^4 (green), and 10^5 (blue) cross progeny are shown. The mapping precision of an individual simulated QTL was measured as the distance from the actual peak to the summit of the closest observed peak within 10 cM in either direction from the actual peak. The average mapping precision (y axis) was computed by taking the mean of the mapping precisions across all detected QTLs under a given parameter combination. Only loci detected in Figure S1 were included in this analysis. Erratic behavior is observed when 10^3 individuals are used because in many cases zero or only a very small number of peaks were detected.

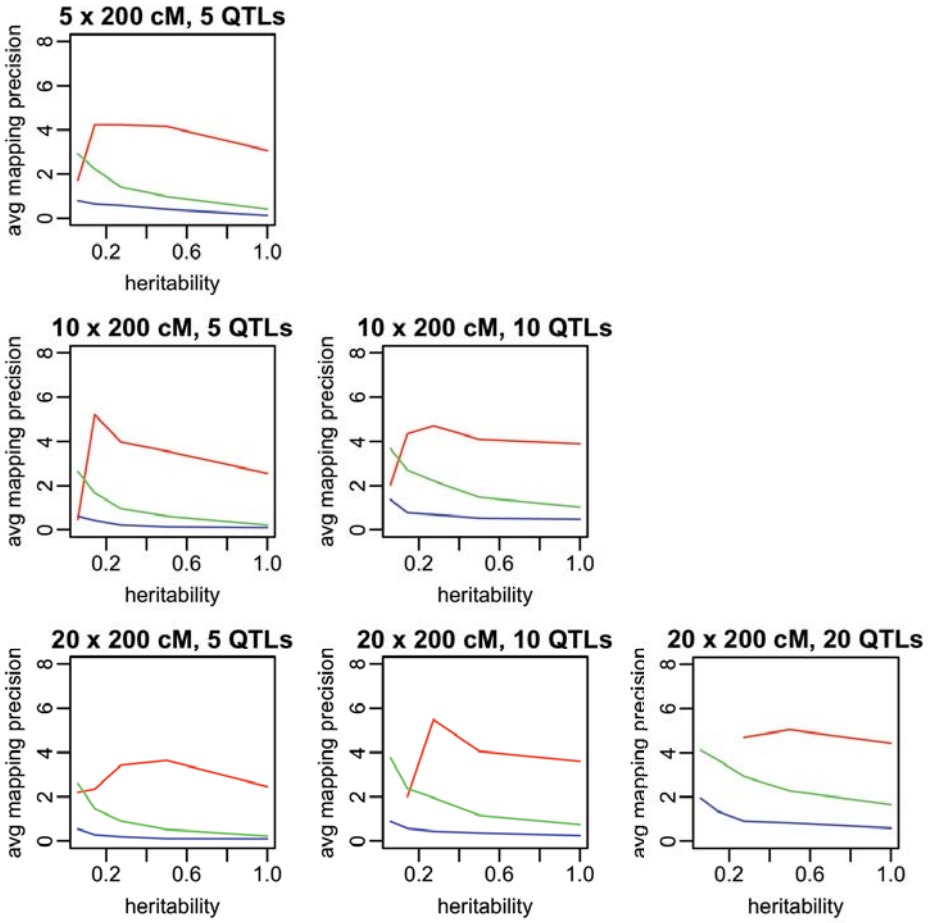

Figure S2B. Average mapping precision of X-QTL (in cM) for haploids with 200 cM chromosomes under a variety of scenarios. Populations of 10^3 (red), 10^4 (green), and 10^5 (blue) cross progeny are shown.

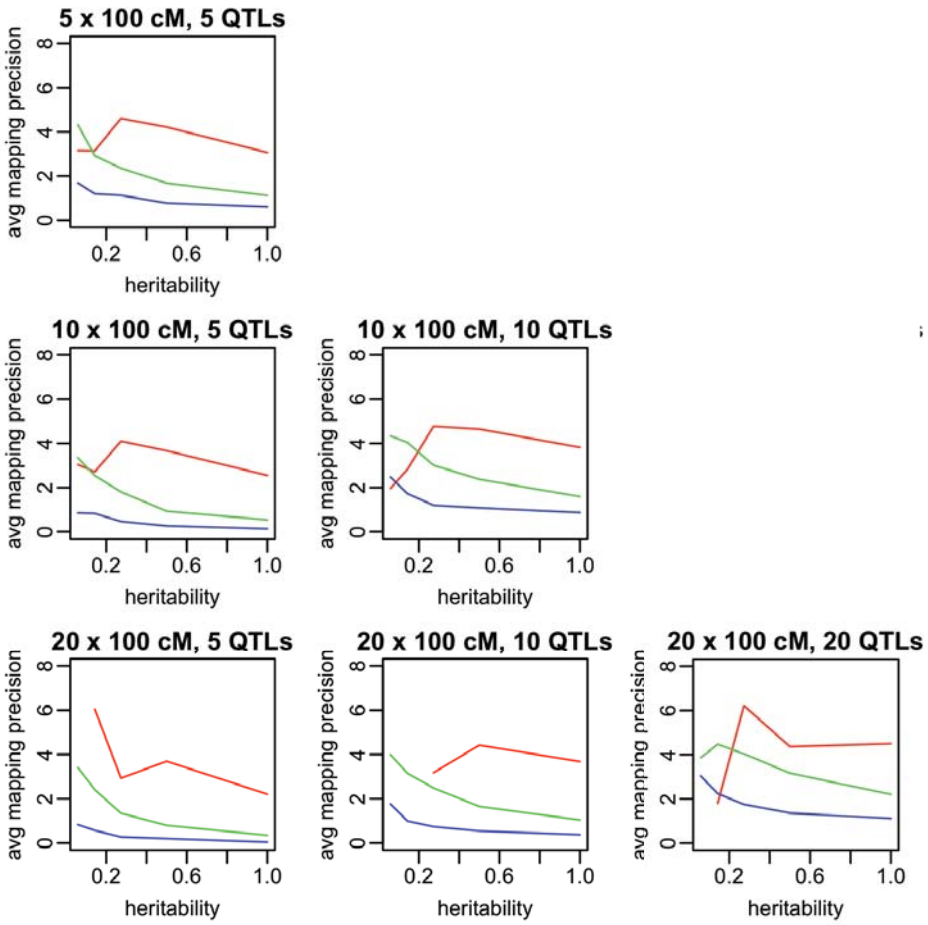

Figure S2C. Average mapping precision of X-QTL (in cM) for diploids with 100 cM chromosomes under a variety of scenarios. Populations of 10^3 (red), 10^4 (green), and 10^5 (blue) cross progeny are shown.

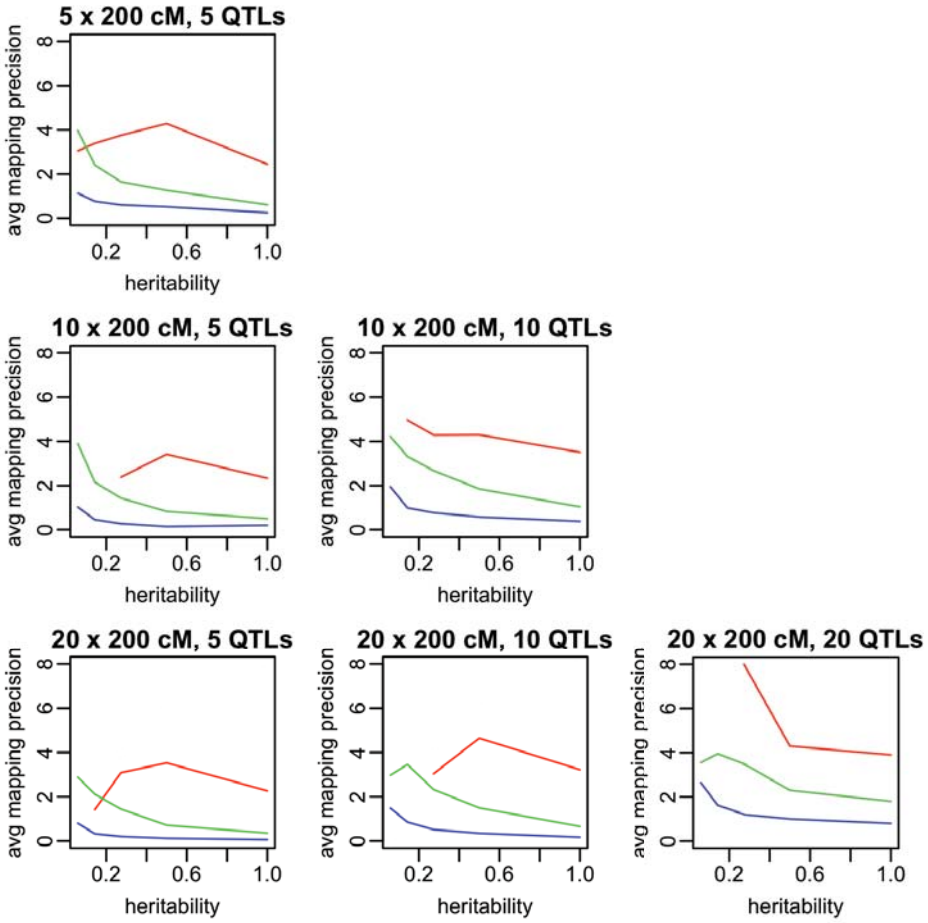

Figure S2D. Average mapping precision of X-QTL (in cM) for diploids with 200 cM chromosomes under a variety of scenarios. Populations of 10^3 (red), 10^4 (green), and 10^5 (blue) cross progeny are shown.

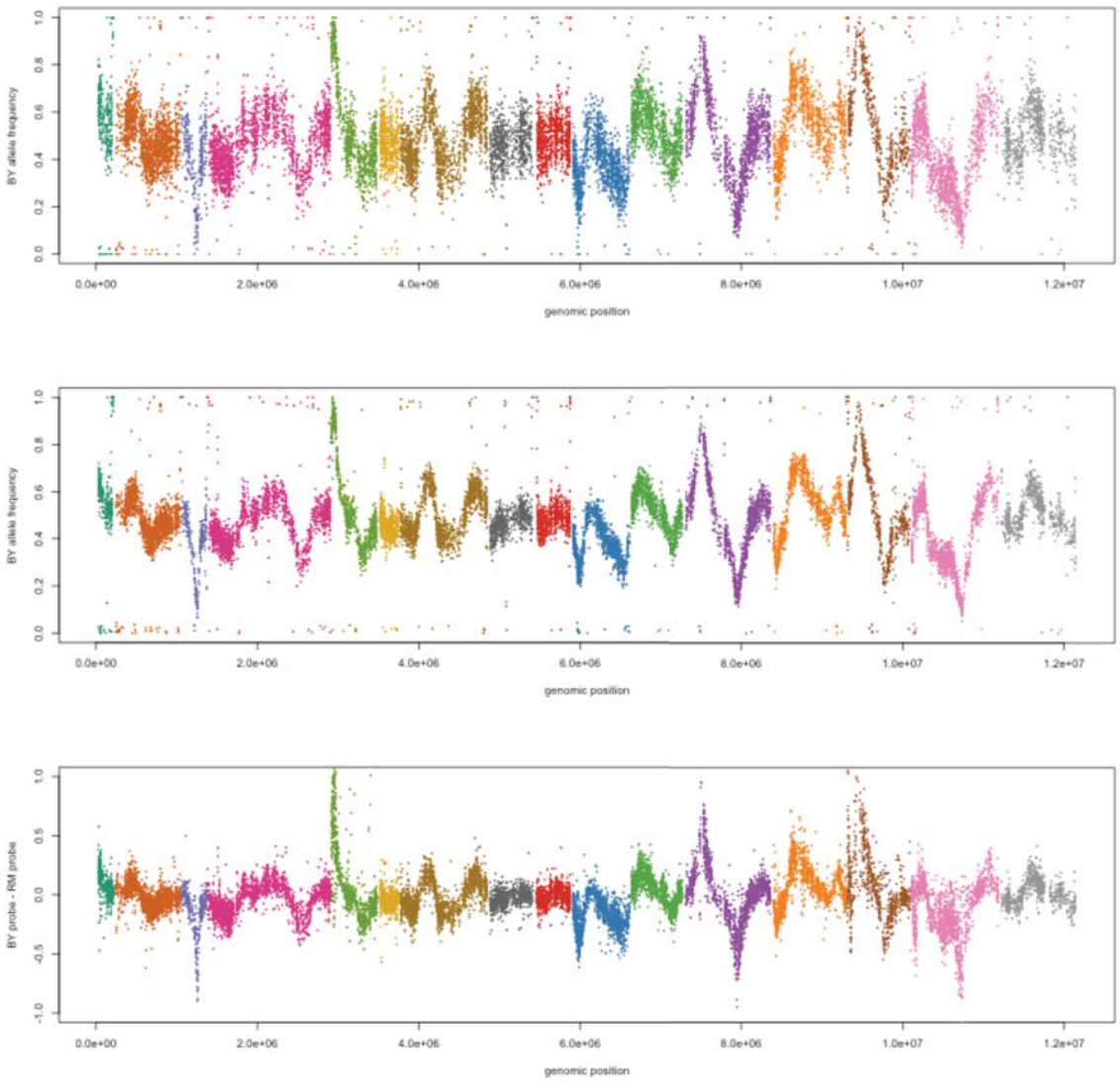

Figure S3. Comparison of data from one lane of the Illumina Genome Analyzer (above), four lanes of the Genome Analyzer (middle), and our microarray (below). One lane of 75 bp reads, which produced ~60X coverage of the yeast genome, was used to generate the top plot. Four sequencing lanes of 75 bp reads, which produced ~240X coverage, were used to generate the middle plot. The bottom plot shows results from a single microarray hybridization of the same 4-NQO-selected pool that was sequenced. The same SNPs are used in both plots. SNPs showing BY allele frequencies of 0 or 1 in the above plot may either not be polymorphic in our strains or may not have been mapped correctly by ELAND. Such SNPs will have a value of 0 on an array when the BY and RM allele-specific probes are compared.

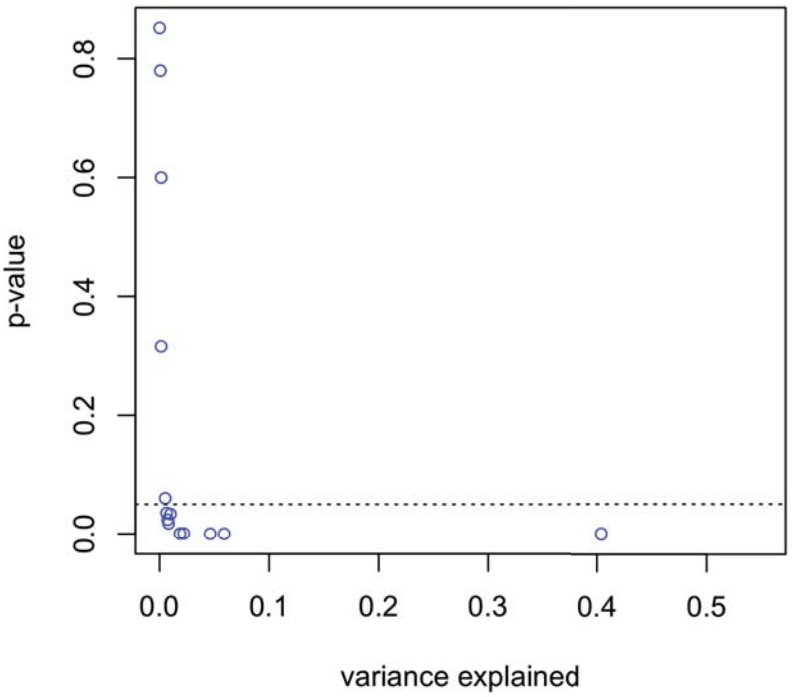

Figure S4. Effect sizes and p-values for the 14 4-NQO loci in a population of 96 individually genotyped and phenotyped BYxRM segregants. These segregants were generated from the same cross used to create the segregating pools and were genotyped at the loci identified for 4-NQO resistance by X-QTL using PCR and digestion of PCR products with restriction enzymes. Phenotyping was done by pinning a dilution series of each segregant on 0.1 ug/ml 4-NQO. Each segregant was grown up to saturation in an overnight culture of YPD, and then serially diluted in 11 increments of 50%. The lowest dilution at which the cells had grown after 20 hours was scored and used as the phenotype in our analysis. Each segregant was phenotyped in triplicate. The aov function in R was used to analyze these data. The heritability for 4-NQO-resistance reported in the main text was calculated from the following model

$$\text{trait} = \text{batch} + \text{strain} + \text{error}$$

by dividing the sum of squares (SS) for the strain effect by the total SS. Allele effects, which are plotted in the above figure, were estimated using the model

$$\text{trait} = \text{batch} + \text{locus1} + \text{locus2} + \ldots + \text{locus13} + \text{locus14} + \text{error}$$

The amount of variance explained by each locus was computed by dividing the sum of squares (SS) for that locus by the total SS. The dotted line denotes $p = 0.05$. Nine loci exhibited significant effects ($p < 0.05$). All loci had effects below 10% explained variance except for *RAD5*, which explained ~40% of the variance.

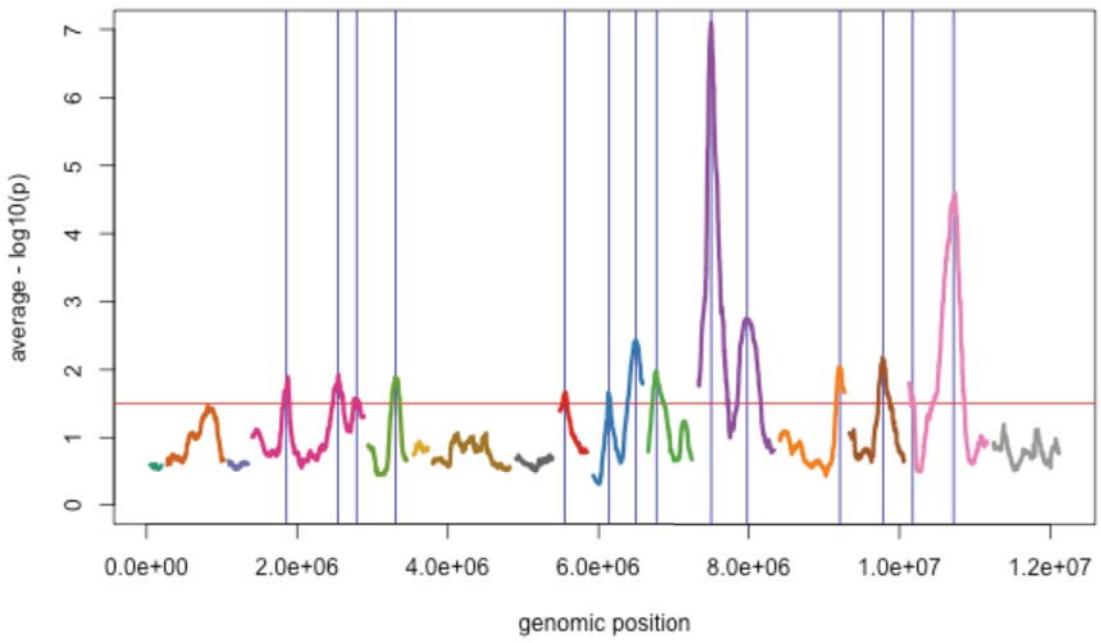

Figure S5A. Loci detected for 4-NQO at a global FDR of 0.05. The blue lines indicate detected peaks, while the red line indicates the threshold for peak identification. Results are represented in the same manner in Figures S5-21.

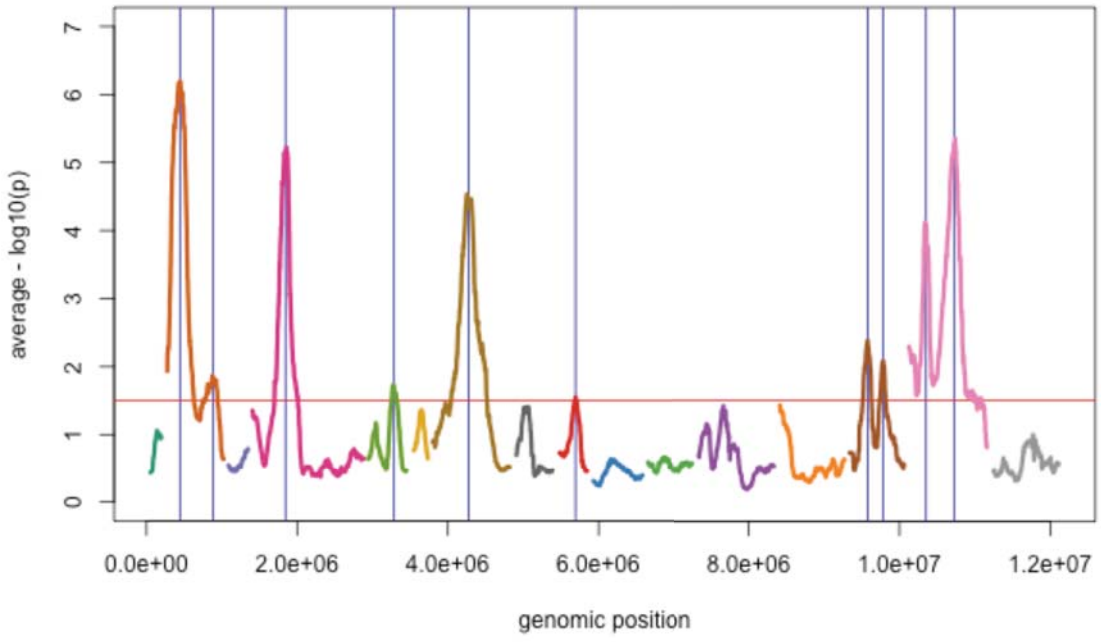

Figure S5B. Loci detected for benomyl at a global FDR of 0.05.

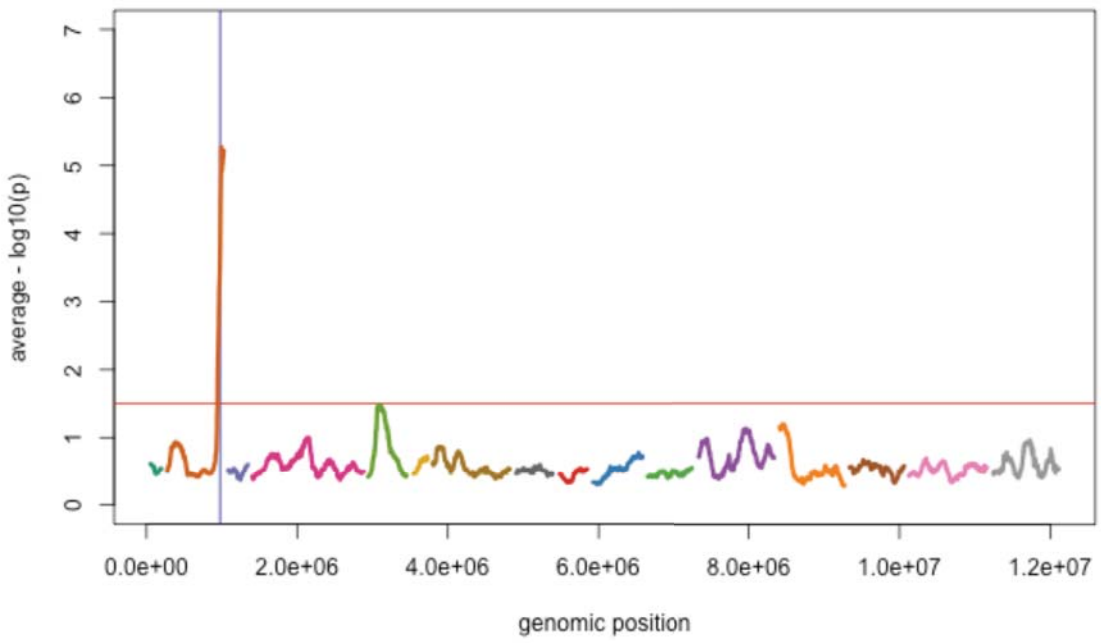

Figure S5C. Loci detected for cadmium chloride at a global FDR of 0.05.

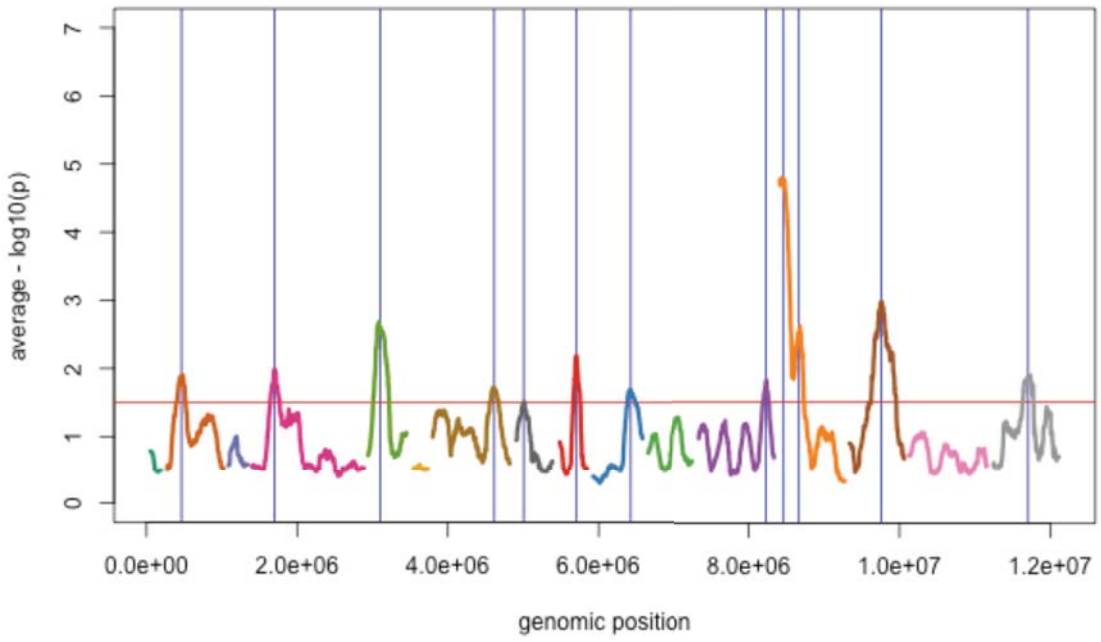

Figure S5D. Loci detected for chlorpromazine at a global FDR of 0.05.

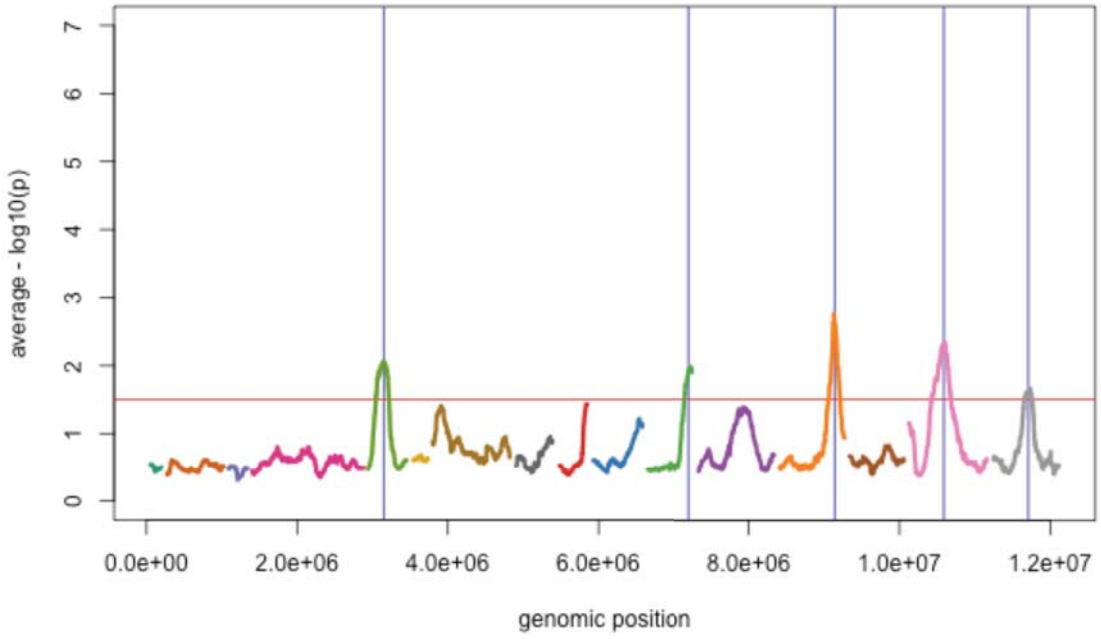

Figure S5E. Loci detected for cobalt chloride at a global FDR of 0.05.

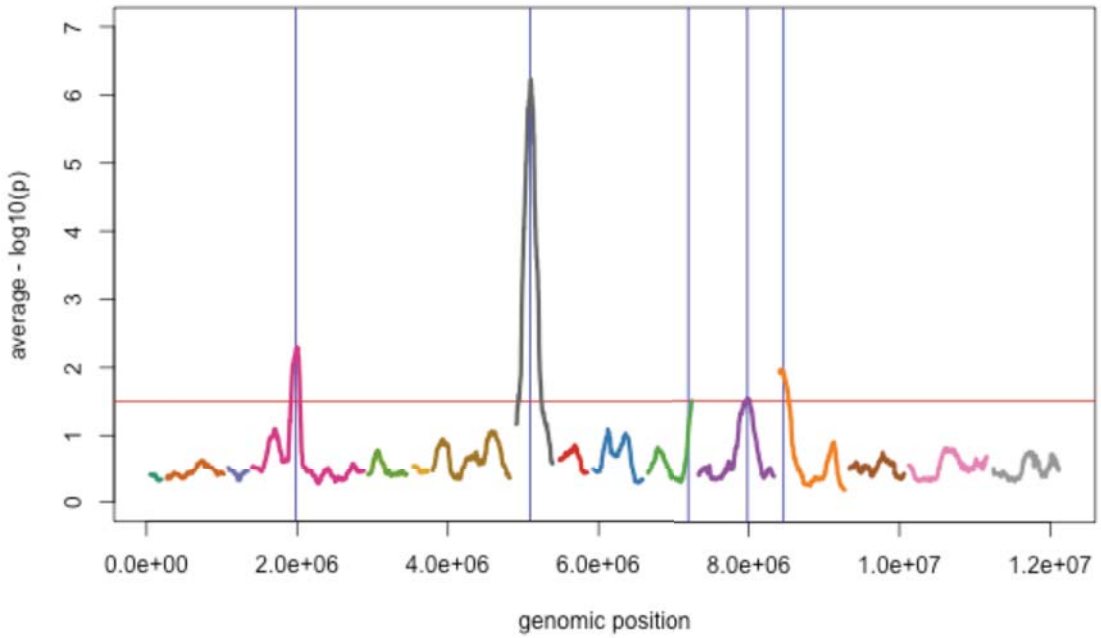

Figure S5F. Loci detected for copper sulfate at a global FDR of 0.05.

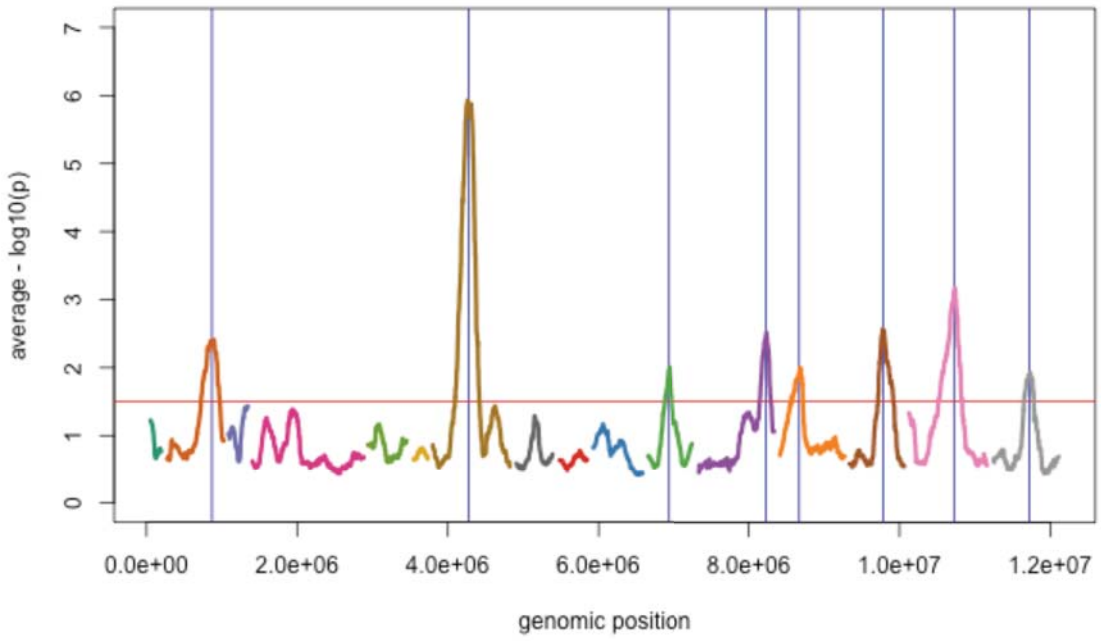

Figure S5G. Loci detected for cycloheximide at a global FDR of 0.05.

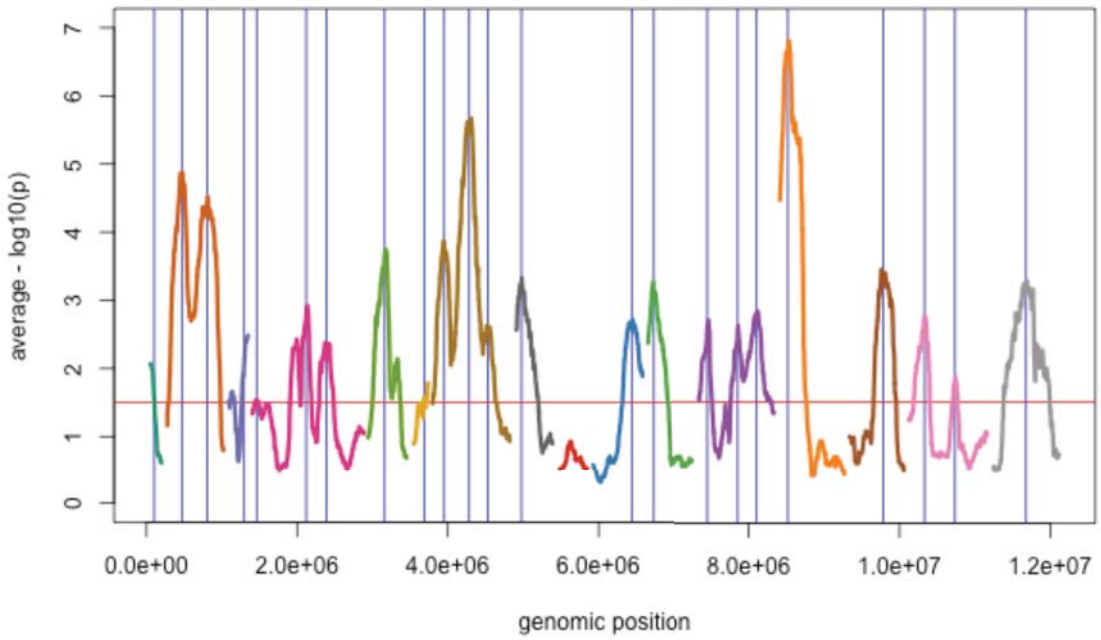

Figure S5H. Loci detected for diamide at a global FDR of 0.05.

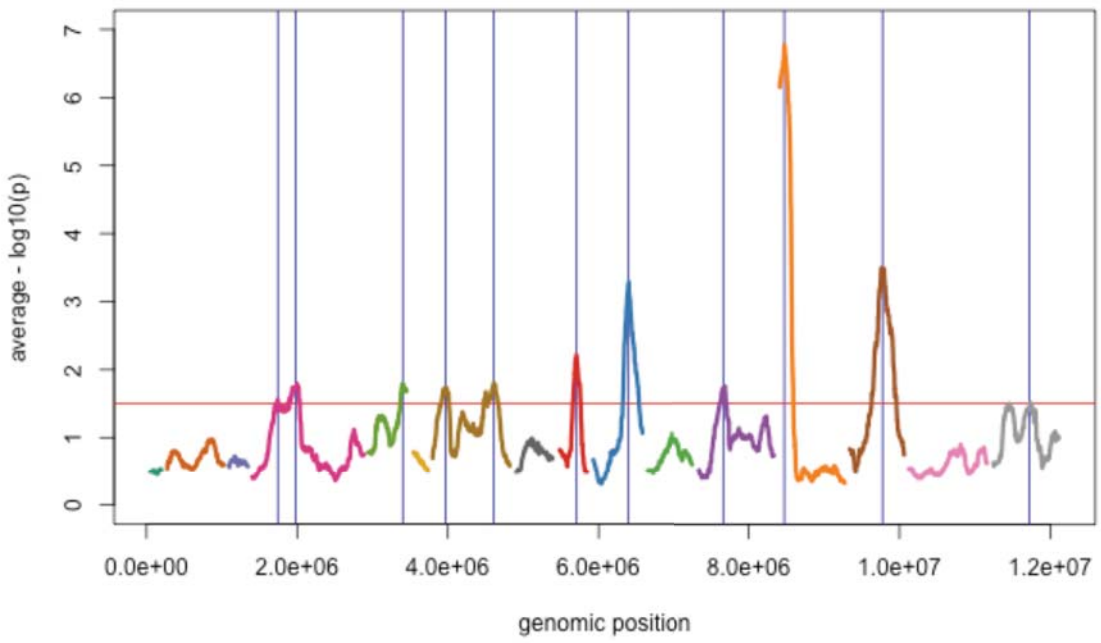

Figure S5I. Loci detected for E6-berbamine at a global FDR of 0.05.

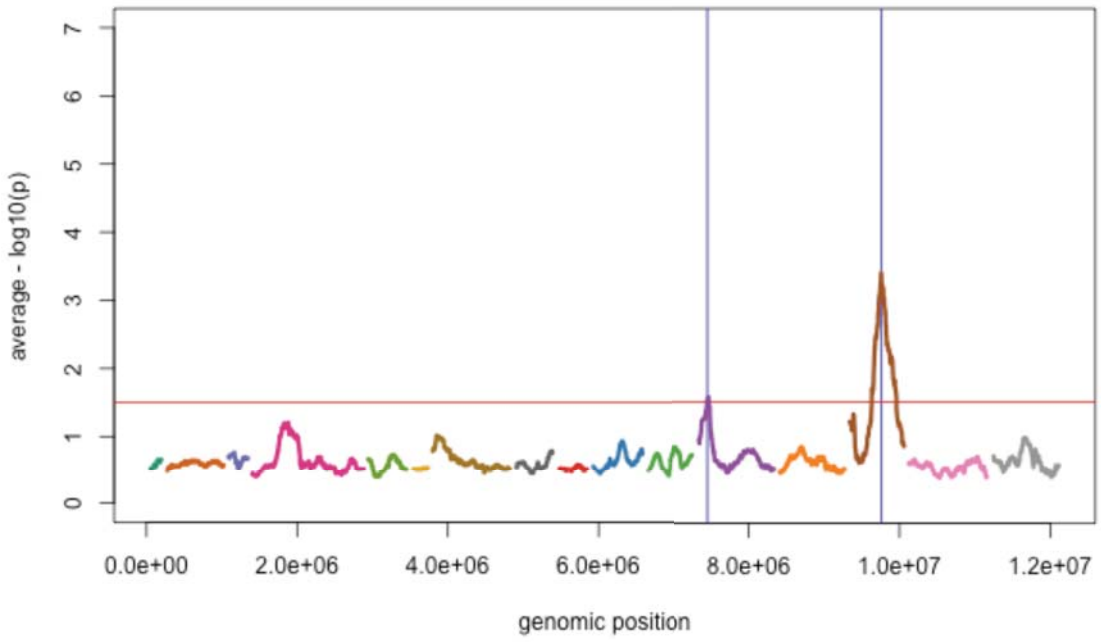

Figure S5J. Loci detected for ethanol at a global FDR of 0.05.

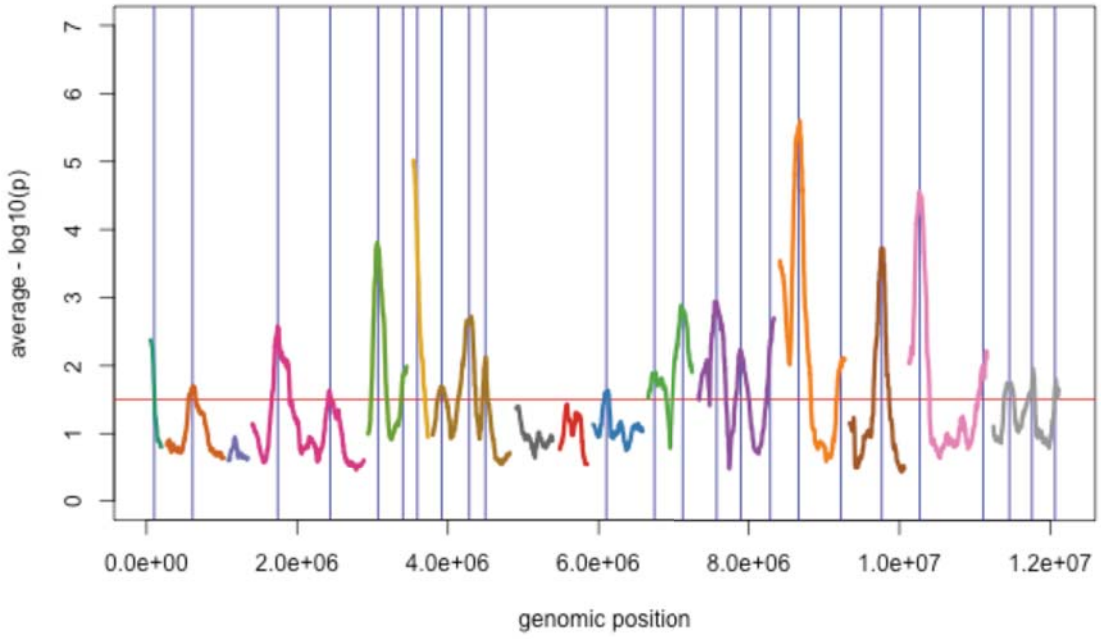

Figure S5K. Loci detected for hydrogen peroxide at a global FDR of 0.05.

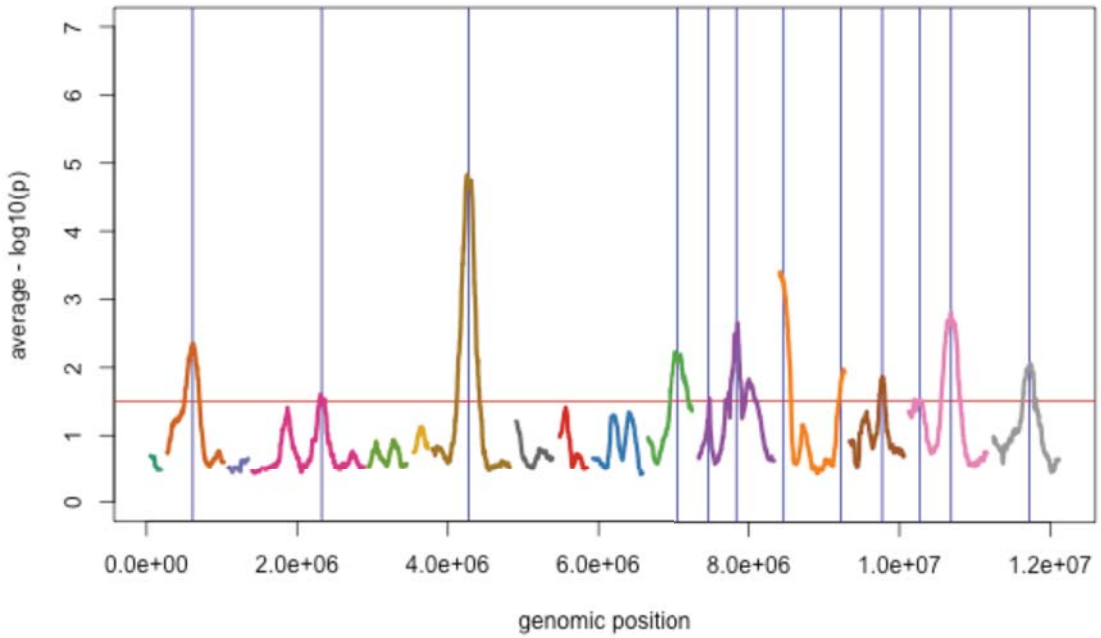

Figure S5L. Loci detected for manumycin at a global FDR of 0.05.

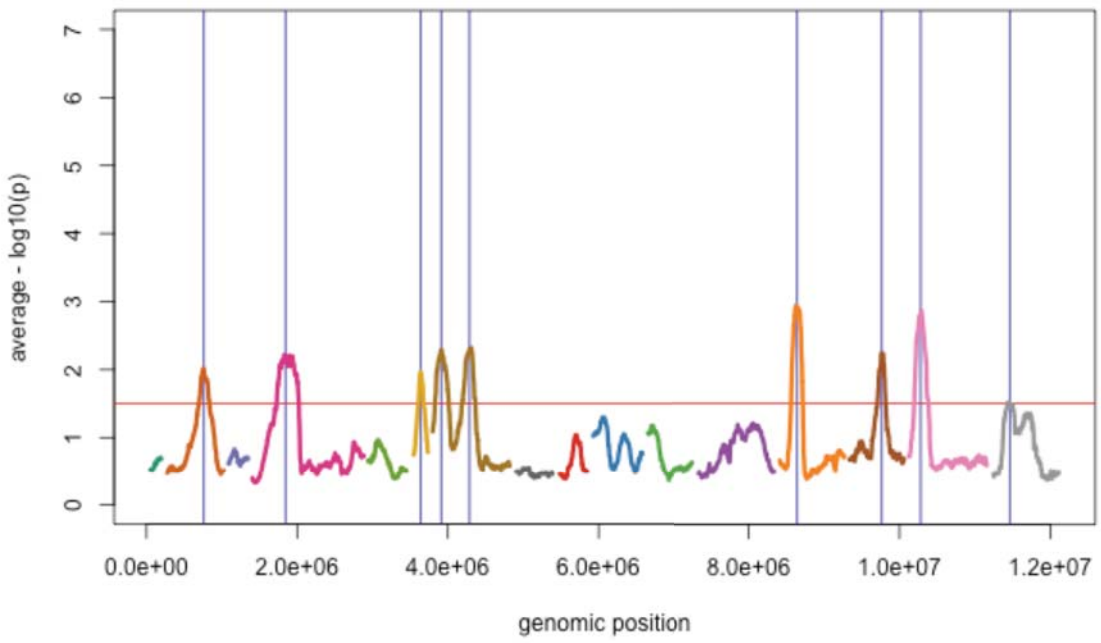

Figure S5M. Loci detected for neomycin at a global FDR of 0.05.

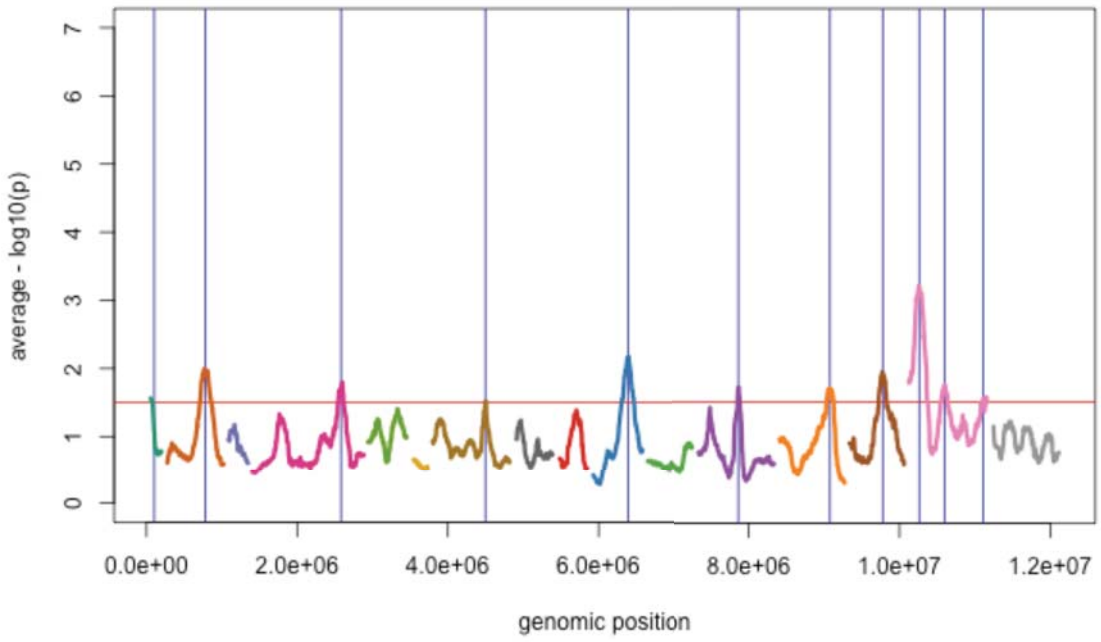

Figure S5N. Loci detected for paraquat at a global FDR of 0.05.

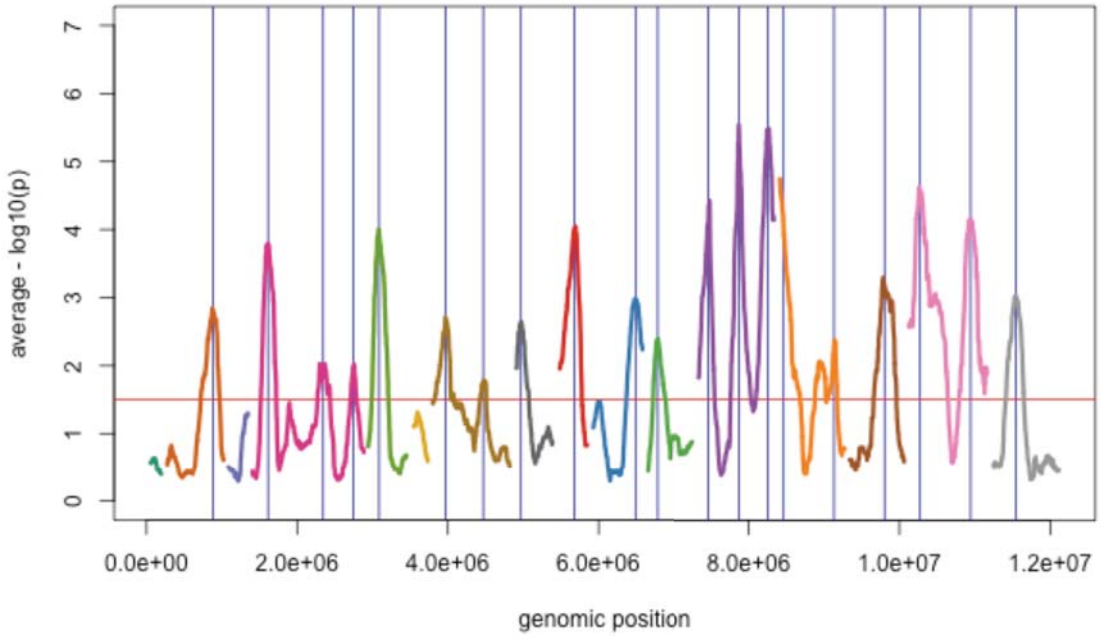

Figure S5O. Loci detected for sodium dodecyl sulfate at a global FDR of 0.05.

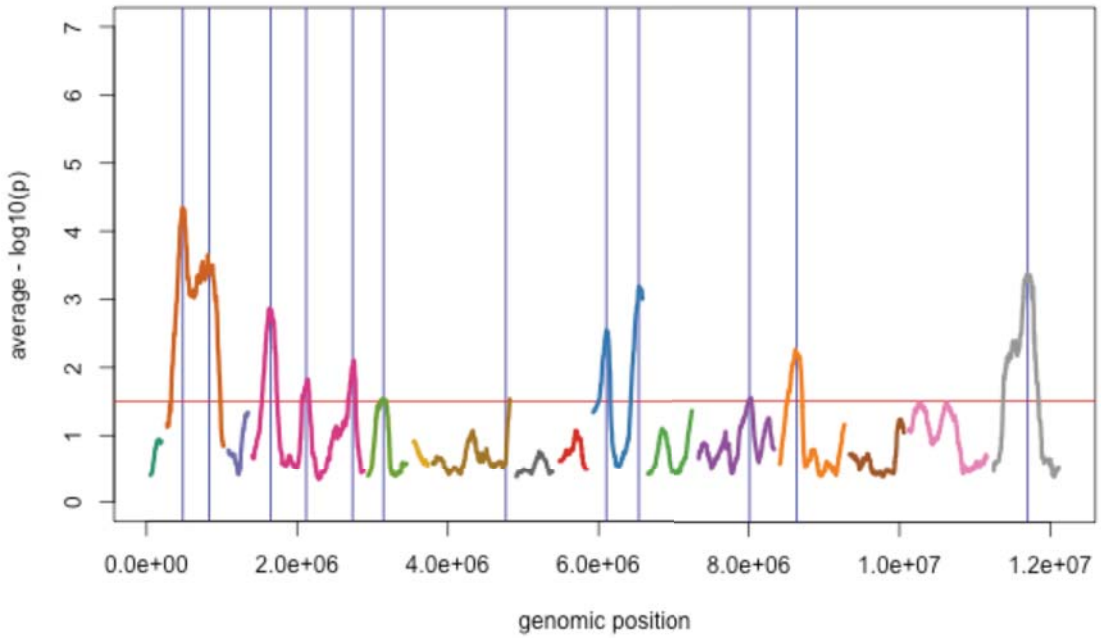

Figure S5P. Loci detected for tunicamycin at a global FDR of 0.05.

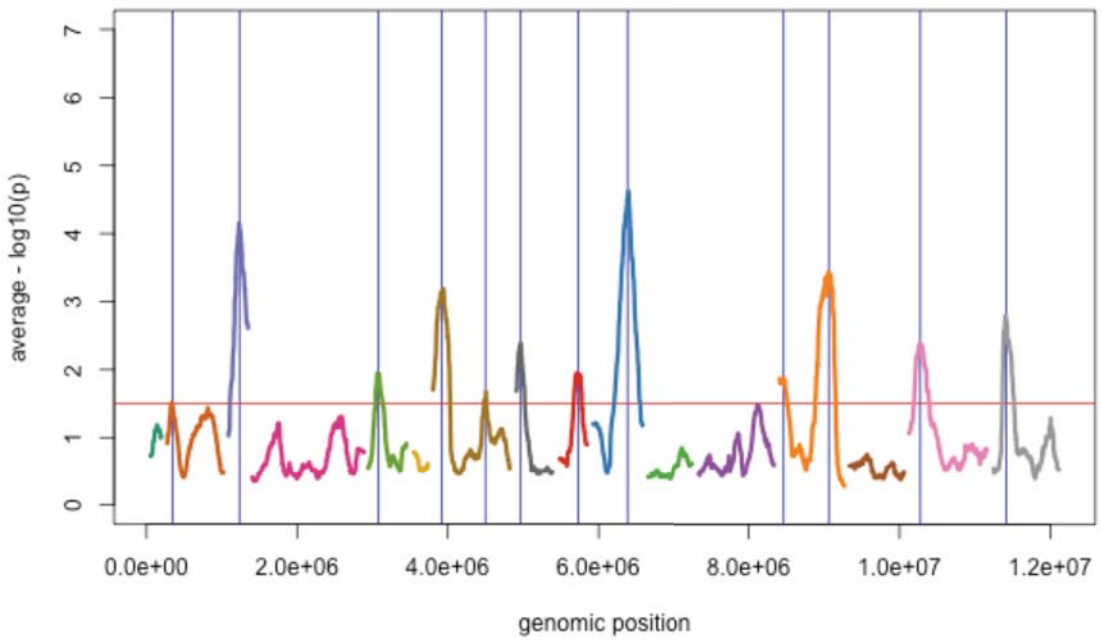

Figure S5Q. Loci detected for zeocin at a global FDR of 0.05.

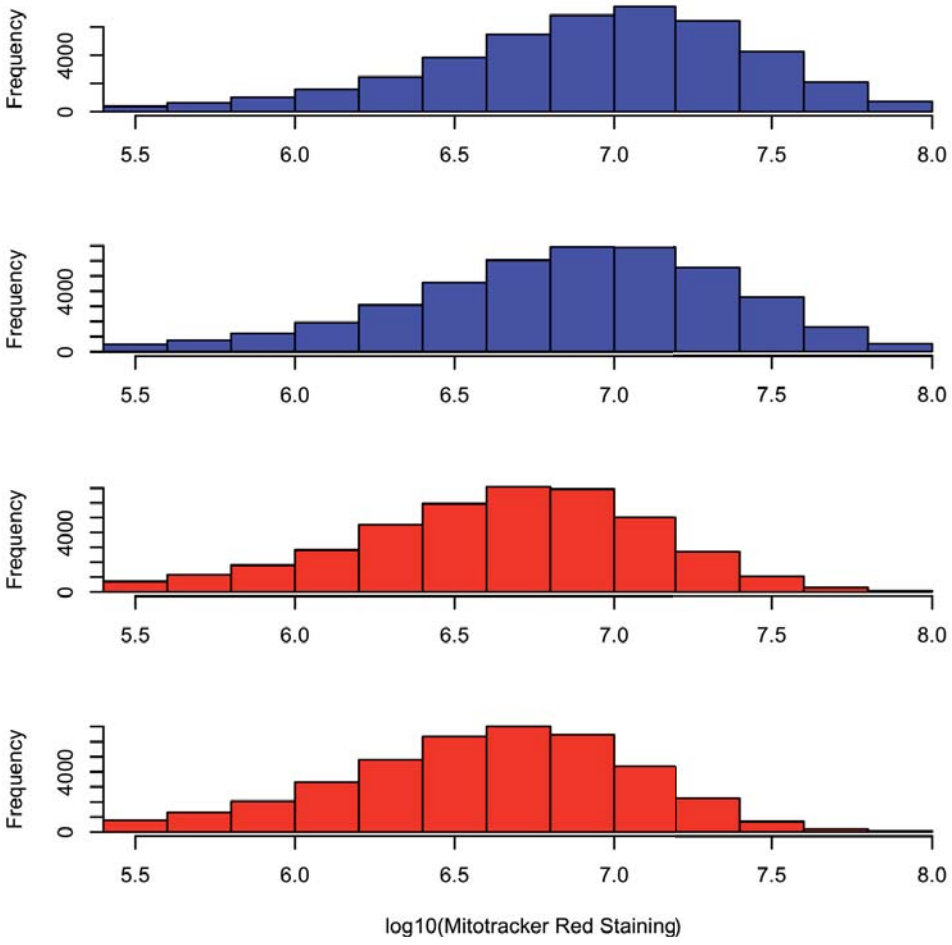

Fig S6. Mitotracker Red staining of BY4716 (red) versus BY4716 with a functional *HAP1* (blue). Two biological replicates are plotted for each strain. BY with a functional *HAP1* exhibits increased Mitotracker Red staining relative to BY with a nonfunctional *HAP1*. Seven biological replicates were tested for each strain. The strains exhibited a highly significant difference in Mitotracker Red staining ($p = 6 \times 10^{-7}$).