Quantile-based permutation thresholds for QTL hotspot analysis: a tutorial

Elias Chaibub Neto* and Brian S Yandell†

September 17, 2012

1 Motivation

QTL hotspots, groups of traits co-mapping to the same genomic location, are a common feature of genetical genomics studies. Genomic locations associated with many traits are biologically interesting since they may harbor influential regulators. Nonetheless, non-genetic mechanisms, uncontrolled environmental factors and unmeasured variables are capable of inducing a strong correlation structure among clusters of transcripts, and as a consequence, whenever a transcript shows a spurious linkage, many correlated transcripts will likely map to the same locus, creating a spurious QTL hotspot. Permutation approaches that do not take into account the phenotypic correlation tend to underestimate the size of the hotspots that might appear by change in these situations (Breitling et al. 2008).

This issue motivated the development of permutation tests that preserve the correlation structure of the phenotypes in order to determine the significance of QTL hotspots (Breitling et al. 2008, Chaibub Neto et al. 2012). In this tutorial we present software tools implementing the NL-method (Chaibub Neto et al. 2012), the N-method (Breitling et al. 2008), and the Q-method (West et al. 2007, Wu et al. 2008) permutation approaches.

2 Overview

This tutorial illustrates the application of the NL-, N- and Q-methods, implemented in the `qtlhot` R package, to a few toy examples. The `qtlhot` package is built over the `R/qtl` package (Broman et al. 2003), and we assume the reader is familiar with it.

3 Basic functionality with No Real Hotspots

In this section we consider two toy simulated examples. In the first we simulate highly correlated phenotypes. In the second, we simulate uncorrelated phenotypes.

*Department of Computational Biology, Sage Bionetworks, Seattle WA
†Department of Statistics, University of Wisconsin-Madison, Madison WI

```
> library(qtlhot)
```

We start by simulating a "null backcross" data set composed of 1,000 phenotypes, 4 chromosomes, 51 equally spaced genetic markers per chromosome, and 100 individuals, with the `sim.null.cross` function. The `latent.eff` parameter control the amount of correlation among the phenotypes. Each phenotype k is generated according to the model $Y_k = \theta L + \epsilon_k$, where $L \sim N(0, \sigma^2)$ is a latent variable, θ represents the effect (`latent.eff`) of the latent variable on the phenotype, and $\epsilon_k \sim N(0, \sigma^2)$ represents a residual error term with σ^2 set to `res.var`. Note that we do not simulate any QTLs in a "null cross" and any linkages we might detect in such a data set are due entirely to chance.

```
> ncross1 <- sim.null.cross(chr.len = rep(100, 4),
+                            n.mar = 51,
+                            n.ind = 100,
+                            type = "bc",
+                            n.phe = 1000,
+                            latent.eff = 3,
+                            res.var = 1,
+                            init.seed = 123457)
```

The function `include.hotspots` takes the "null cross" as an input and includes 3 hotspots of size `hsize` at position `hpos` of chromosome `hchr` into it. Explicitly, it simulates each one of the hotspots according to the model $Y_k^* = \beta M + Y_k$, where Y_k is the phenotype generated by the `generate.null.cross` function; $M = \gamma \, Q + \epsilon_M$ is a master regulator that affects all phenotypes in the hotspot; Q is a QTL located at position `hpos` of chromosome `hchr`; γ represents the QTL effect (`Q.eff`); $\epsilon_M \sim N(0, \sigma^2)$; and β is computed such that the association between Y_k^* and Q, measured by the LOD score, is given (theoretically) by a valued sampled from the user specified LOD score range (`lod.range.1` and etc).

```
> cross1 <- include.hotspots(cross = ncross1,
+                             hchr = c(2, 3, 4),
+                             hpos = c(25, 75, 50),
+                             hsize = c(100, 50, 20),
+                             Q.eff = 2,
+                             latent.eff = 3,
+                             lod.range.1 = c(2.5, 2.5),
+                             lod.range.2 = c(5, 8),
+                             lod.range.3 = c(10, 15),
+                             res.var = 1,
+                             n.phe = 1000,
+                             init.seed = 12345)
```

Note that by choosing `latent.eff = 3` we generate highly correlated phenotype data. The distribution of the correlation values for each pair of phenotypes is given below.

```
> ncor1 <- cor(cross1$pheno)
> summary(ncor1[lower.tri(ncor1)])
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.4145  0.8517  0.8929  0.8649  0.9063  0.9691
```

```
> rm(ncor1)
```

Next we obtain standard permutation thresholds (Churchill and Doerge 1994) for single trait QTL mapping analysis for the sequence `alphas`, representing target genome wide error rates (GWER).

```
> set.seed(123)
> pt <- scanone(ncross1, method = "hk", n.perm = 1000)
```

```
> alphas <- seq(0.01, 0.10, by=0.01)
> lod.thrs <- summary(pt, alphas)
> lod.thrs
```

```
LOD thresholds (1000 permutations)
     lod
1%  3.11
2%  2.89
3%  2.68
4%  2.57
5%  2.44
6%  2.34
7%  2.26
8%  2.20
9%  2.15
10% 2.11
```

```
> lod.thr <- lod.thrs[5]
```

We perform QTL mapping analysis for 1,000 phenotypes using Haley-Knott regression, and keep only the `drop.lod` = 1.5 LOD support interval (Manichaikul et al. 2006) around significant peaks (above `lod.thr`) at each chromosome for each trait. LOD support intervals are the most commonly used interval estimate for the location of a QTL.

```
> scan1 <- scanone(cross1, pheno.col = 1:1000, method = "hk")
```

The routine `highlod` just saves the LOD support intervals for significant peaks, which dramatically reduces the object size from the result of `scanone`. We can examine the maximum hotspot size for each possible single trait threshold using the `max` method for `highlod` objects. The first column gives the counts associated with QTL mapping threshold of 3.11, whereas the last one shows the counts based on the more liberal threshold 2.11. Note how the counts increase as the QTL mapping thresholds decrease.

```
> high1 <- highlod(scan1, lod.thr = min(lod.thrs), drop.lod = 1.5)
> max(high1, lod.thr = lod.thrs)
```

```
       chr pos max.N lod.thr
D3M38    3  74    50    3.11
D2M12    2  22    69    2.89
D2M13    2  24    89    2.68
D2M131   2  24    93    2.57
D2M121   2  22    95    2.44
D2M132   2  24    98    2.34
D2M133   2  24    99    2.26
D2M122   2  22    99    2.20
D1M34    1  66   110    2.15
D1M341   1  66   126    2.11
```

Next we infer the hotspot architecture at varying QTL mapping thresholds. In other words, for each genomic position, we count the number of traits that map to it with a LOD score equal or higher than the threshold in `lod.thr`. The `hots1` object is a `scanone` object with added attributes and specialized summary and plot methods. As an illustration, we show the counts for the 5 first markers on chromosome 2.

```
> hots1 <- hotsize(high1, lod.thr = lod.thr)
> summary(hots1)
```

```
hotsize elements:  chr pos max.N
LOD threshold: 2.438697

      chr pos max.N
D1M48   1  94    41
D2M12   2  22    95
D2M13   2  24    95
D3M35   3  68    50
D3M36   3  70    50
D3M37   3  72    50
D3M38   3  74    50
D3M39   3  76    50
D4M21   4  40    20
D4M22   4  42    20
D4M23   4  44    20
D4M24   4  46    20
D4M25   4  48    20
D4M26   4  50    20
D4M27   4  52    20
```

We plot the hotspot architecture inferred using the single trait permutation threshold 2.44 ($\alpha = 0.05$). Figure 1 shows the counts across the genome. Recall that in the call of function

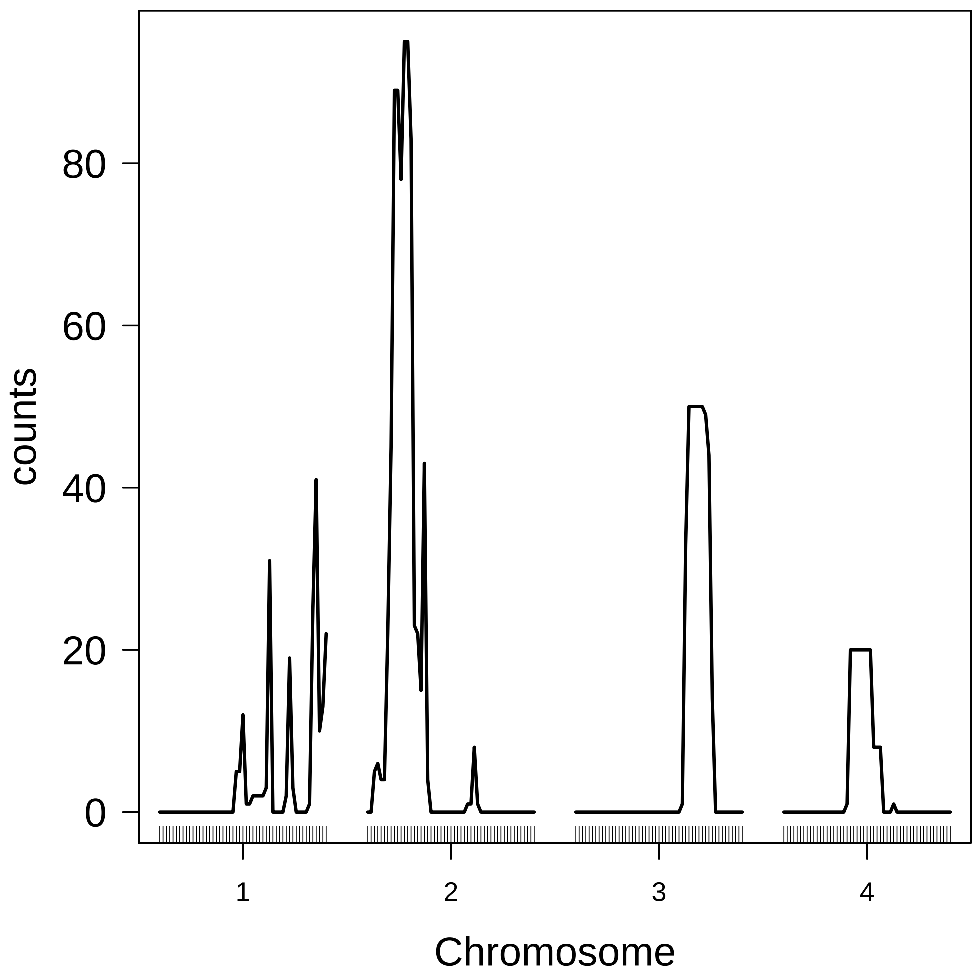


Figure 1: Hotspot architecture associated with QTL mapping threshold of 2.44 in example 1.

`include.hotspots` we set to simulate 3 hotspots: (1) a hotspot of size 100 at position 25cM of chromosome 2 with LOD scores around 2.5; (2) a hotspot of size 50 at position 75cM of chromosome 3 with LOD scores ranging from 5 to 8; and (3) a hotspot of size 20 at position 50cM of chromosome 4 with LOD scores ranging from 10 to 15. Nonetheless, Figure 1 shows several spurious peaks on chromosome 1, that arise because of the high correlation of the phenotypes.

```
> plot(hots1, cex.lab = 1.5, cex.axis = 1.5)
```

Next, we perform permutation tests to assess the statistical significance of the hotspots detected on Figure 1. We consider the N- and NL-methods. [The Q method of West and Wu is documented in the `ww.perm` manual page.] The `hotperm` function implements the N- and NL-methods' permutation schemes (see Chaibub Neto et al. 2012, for details). The parameter `n.quant` sets the maximum hotspot size to be analyzed by the NL-method. The parameter drop controls the magnitude of the LOD support interval computation during the LOD profile processing step. The function's output is a list with two elements: `max.lod.quant` and `max.N`.

```
> set.seed(12345)
> hotperm1 <- hotperm(cross = cross1,
+                n.quant = 300,
+                n.perm = 100,
+                lod.thrs = lod.thrs,
+                alpha.levels = alphas,
+                drop.lod = 1.5,
+                verbose = FALSE)
```

```
> names(hotperm1)
```

```
[1] "max.lod.quant" "max.N"
```

```
> summary(hotperm1)
```

```
max.N: hotspot threshold by single-trait LOD threshold and significance level
       0.99   0.98   0.97   0.96   0.95   0.94   0.93   0.92   0.91   0.9
3.11  51.27  42.12  38.12  29.36  27.10   9.14   7.07   7.00   6.09   6.0
2.89  79.41  71.08  61.30  58.12  50.40  21.80  18.14  16.16  15.09  12.3
2.68 155.83 150.02 133.51  95.56  58.85  50.42  43.49  40.24  27.26  20.6
2.57 208.98 190.28 187.09 119.80  76.15  70.24  66.28  59.56  34.43  30.2
2.44 293.80 274.30 269.15 176.84 118.85 109.42  91.33  78.96  62.53  59.2
2.34 374.45 344.52 331.39 226.36 162.15 153.36 126.03  94.56  87.45  86.1
2.26 440.01 406.60 378.84 269.52 202.30 199.00 161.80 115.76 112.00 110.2
2.2  484.88 451.60 425.78 305.96 239.25 232.24 200.38 144.64 138.18 131.7
2.15 515.75 496.32 471.75 333.72 266.25 256.42 233.68 168.52 159.36 152.7
2.11 552.46 531.36 503.84 372.44 305.25 286.02 265.47 195.00 180.81 177.3

max.lod.quant: LOD threshold by hotspot size quantile and significance level
          1    2    3    4    5    6    7    8    9   10   11   12   13   14
max    5.95 5.22 5.06 4.96 4.89 4.68 4.58 4.53 4.52 4.49 4.43 4.37 4.35 4.34
75%    3.38 3.15 3.24 3.24 3.22 3.18 3.15 3.09 3.06 3.04 3.14 3.13 3.26 3.24
median 3.02 2.82 2.83 2.92 2.88 2.86 2.86 2.88 2.82 2.79 2.82 2.84 2.84 2.84
25%    2.72 2.59 2.64 2.62 2.61 2.58 2.55 2.54 2.50 2.59 2.64 2.67 2.68 2.66
min    2.46 2.46 2.44 2.44 2.45 2.44 2.46 2.45 2.44 2.45 2.45 2.45 2.53 2.51
         15   16   17   18   19   20   21   22   23   24   25   26   27   28
max    4.33 4.33 4.31 4.28 4.28 4.25 4.20 4.19 4.15 4.14 4.14 4.14 4.12 4.12
75%    3.23 3.21 3.21 3.20 3.19 3.19 3.17 3.17 3.16 3.16 3.13 3.13 3.12 3.12
median 2.83 2.82 2.80 2.79 2.72 2.70 2.67 2.65 2.65 2.64 2.63 2.66 2.68 2.67
25%    2.63 2.60 2.59 2.59 2.62 2.61 2.60 2.60 2.56 2.54 2.54 2.56 2.57 2.56
min    2.49 2.47 2.46 2.45 2.49 2.49 2.48 2.47 2.45 2.45 2.44 2.44 2.44 2.44
         29   30   31   32   33   34   35   36   37   38   39   40   41   42
```

max	4.09	4.06	4.05	4.05	4.05	4.05	4.05	4.05	4.04	4.02	4.01	4.01	4.00	4.00
75%	3.11	3.10	3.09	3.09	3.09	3.07	3.07	3.09	3.09	3.08	3.07	3.09	3.09	3.09
median	2.69	2.68	2.73	2.72	2.71	2.71	2.70	2.71	2.70	2.71	2.71	2.71	2.71	2.70
25%	2.55	2.54	2.54	2.53	2.53	2.53	2.52	2.51	2.50	2.51	2.51	2.50	2.50	2.49
min	2.45	2.45	2.49	2.49	2.47	2.45	2.45	2.45	2.44	2.44	2.44	2.46	2.46	2.45
	43	44	45	46	47	48	49	50	51	52	53	54	55	56
max	3.99	3.99	3.98	3.97	3.95	3.95	3.94	3.94	3.94	3.94	3.93	3.93	3.93	3.93
75%	3.08	3.07	3.07	3.07	3.06	3.05	3.05	3.04	3.00	3.00	2.99	2.98	2.98	2.97
median	2.70	2.71	2.70	2.70	2.69	2.69	2.68	2.82	2.78	2.76	2.83	2.77	2.75	2.71
25%	2.49	2.51	2.50	2.49	2.49	2.49	2.48	2.55	2.54	2.54	2.62	2.62	2.61	2.60
min	2.44	2.45	2.45	2.44	2.44	2.44	2.44	2.45	2.44	2.44	2.46	2.46	2.45	2.45
	57	58	59	60	61	62	63	64	65	66	67	68	69	70
max	3.93	3.92	3.92	3.90	3.89	3.89	3.89	3.89	3.88	3.88	3.88	3.87	3.86	3.86
75%	2.95	2.95	2.93	2.94	2.94	2.92	2.92	2.92	2.91	2.91	2.90	2.90	2.90	2.90
median	2.68	2.66	2.66	2.77	2.77	2.88	2.88	2.86	2.86	2.86	2.86	2.85	2.84	2.84
25%	2.60	2.59	2.58	2.59	2.59	2.63	2.62	2.62	2.60	2.59	2.59	2.59	2.59	2.58
min	2.45	2.44	2.44	2.44	2.44	2.56	2.56	2.55	2.55	2.54	2.53	2.53	2.52	2.52
	71	72	73	74	75	76	77	78	79	80	81	82	83	84
max	3.86	3.86	3.86	3.85	3.85	3.85	3.85	3.85	3.85	3.84	3.83	3.83	3.83	3.82
75%	2.89	2.89	2.89	2.88	2.87	2.87	2.87	2.87	2.87	2.87	2.87	2.86	2.86	2.85
median	2.83	2.82	2.82	2.81	2.81	2.81	2.80	2.79	2.81	2.80	2.79	2.79	2.78	2.77
25%	2.56	2.56	2.55	2.55	2.54	2.54	2.54	2.53	2.54	2.54	2.54	2.54	2.53	2.53
min	2.51	2.51	2.48	2.48	2.48	2.45	2.44	2.44	2.49	2.49	2.49	2.47	2.47	2.47
	85	86	87	88	89	90	91	92	93	94	95	96	97	98
max	3.82	3.82	3.82	3.81	3.81	3.80	3.80	3.80	3.80	3.80	3.80	3.79	3.79	3.79
75%	2.85	2.84	2.84	2.83	2.83	2.82	2.82	2.81	2.81	2.81	2.81	2.80	2.80	2.80
median	2.76	2.75	2.75	2.75	2.74	2.74	2.78	2.78	2.78	2.78	2.77	2.77	2.77	2.77
25%	2.52	2.52	2.52	2.52	2.51	2.51	2.60	2.60	2.60	2.59	2.58	2.58	2.57	2.57
min	2.47	2.46	2.46	2.44	2.44	2.44	2.51	2.51	2.51	2.50	2.50	2.50	2.49	2.49
	99	100	101	102	103	104	105	106	107	108	109	110	111	112
max	3.78	3.78	3.78	3.78	3.77	3.77	3.77	3.76	3.76	3.76	3.75	3.75	3.74	3.74
75%	2.79	2.79	2.79	2.78	2.78	2.78	2.78	2.78	2.77	2.77	2.77	2.77	2.76	2.76
median	2.76	2.76	2.76	2.76	2.75	2.75	2.75	2.74	2.74	2.74	2.74	2.75	2.75	2.74
25%	2.56	2.56	2.55	2.55	2.55	2.54	2.54	2.54	2.54	2.53	2.53	2.60	2.59	2.59
min	2.48	2.48	2.48	2.48	2.47	2.46	2.46	2.46	2.45	2.44	2.44	2.46	2.45	2.45
	113	114	115	116	117	118	119	120	121	122	123	124	125	126
max	3.73	3.72	3.71	3.70	3.70	3.70	3.70	3.68	3.68	3.68	3.67	3.67	3.67	3.67
75%	2.76	2.76	2.76	2.75	2.75	2.75	2.75	2.75	2.74	2.74	2.74	2.74	2.73	2.73
median	2.74	2.74	2.74	2.74	2.75	2.75	2.74	2.74	2.74	2.73	2.73	2.73	2.72	2.72
25%	2.59	2.59	2.58	2.58	2.72	2.72	2.72	2.72	2.72	2.72	2.71	2.71	2.71	2.70
min	2.45	2.44	2.44	2.44	2.58	2.57	2.56	2.55	2.54	2.54	2.54	2.54	2.54	2.53
	127	128	129	130	131	132	133	134	135	136	137	138	139	140
max	3.66	3.66	3.65	3.65	3.65	3.65	3.64	3.64	3.63	3.63	3.63	3.63	3.63	3.63
75%	2.73	2.73	2.72	2.72	2.72	2.72	2.72	2.71	2.71	2.71	2.71	2.71	2.71	2.71

median	2.72	2.72	2.72	2.72	2.72	2.71	2.71	2.71	2.71	2.71	2.71	2.71	2.71	2.71
25%	2.70	2.70	2.70	2.70	2.70	2.69	2.69	2.66	2.66	2.66	2.66	2.66	2.65	2.65
min	2.53	2.53	2.52	2.52	2.52	2.52	2.52	2.51	2.51	2.51	2.51	2.51	2.51	2.51
	141	142	143	144	145	146	147	148	149	150	151	152	153	154
max	3.63	3.63	3.62	3.62	3.61	3.61	3.61	3.61	3.59	3.59	3.59	3.59	3.58	3.58
75%	2.70	2.70	2.70	2.70	2.69	2.69	2.69	2.69	2.68	2.68	2.68	2.68	2.67	2.67
median	2.70	2.70	2.70	2.69	2.69	2.69	2.68	2.68	2.68	2.68	2.67	2.67	2.67	2.67
25%	2.65	2.65	2.64	2.64	2.64	2.64	2.64	2.64	2.64	2.63	2.63	2.63	2.62	2.62
min	2.51	2.50	2.50	2.50	2.50	2.50	2.49	2.49	2.49	2.49	2.49	2.49	2.48	2.48
	155	156	157	158	159	160	161	162	163	164	165	166	167	168
max	3.58	3.58	3.58	3.58	3.57	3.57	3.57	3.56	3.56	3.56	3.56	3.55	3.55	3.54
75%	2.67	2.67	2.66	2.66	2.66	2.66	2.66	2.66	2.65	2.65	2.65	2.65	2.65	2.64
median	2.67	2.67	2.66	2.66	2.65	2.65	2.65	2.64	2.64	2.64	2.63	2.63	2.62	2.62
25%	2.62	2.62	2.62	2.61	2.61	2.61	2.61	2.61	2.61	2.61	2.60	2.60	2.60	2.60
min	2.48	2.48	2.47	2.47	2.47	2.47	2.47	2.46	2.46	2.46	2.46	2.46	2.45	2.45
	169	170	171	172	173	174	175	176	177	178	179	180	181	182
max	3.54	3.53	3.53	3.53	3.53	3.52	3.52	3.52	3.52	3.51	3.51	3.51	3.51	3.49
75%	2.64	2.64	2.64	2.64	2.64	3.08	3.08	3.08	3.07	3.07	3.07	3.06	3.06	3.06
median	2.62	2.62	2.62	2.62	2.61	2.62	2.62	2.62	2.62	2.62	2.61	2.61	2.61	2.61
25%	2.60	2.60	2.60	2.60	2.60	2.60	2.60	2.60	2.60	2.60	2.60	2.59	2.59	2.59
min	2.45	2.44	2.44	2.44	2.44	2.60	2.60	2.59	2.59	2.59	2.59	2.59	2.59	2.58
	183	184	185	186	187	188	189	190	191	192	193	194	195	196
max	3.49	3.49	3.48	3.48	3.47	3.47	3.46	3.44	3.44	3.43	3.43	3.42	3.42	3.42
75%	3.05	3.05	3.05	3.04	3.04	3.03	3.03	3.02	3.02	3.01	3.01	3.01	3.01	3.00
median	2.60	2.60	2.60	2.60	2.59	2.59	2.59	2.59	2.59	2.58	2.58	2.58	2.58	2.58
25%	2.58	2.58	2.58	2.58	2.58	2.58	2.57	2.57	2.57	2.57	2.56	2.56	2.56	2.56
min	2.58	2.58	2.58	2.58	2.57	2.57	2.57	2.57	2.57	2.57	2.56	2.56	2.56	2.56
	197	198	199	200	201	202	203	204	205	206	207	208	209	210
max	3.42	3.42	3.42	3.41	3.41	3.40	3.40	3.40	3.40	3.40	3.39	3.39	3.39	3.39
75%	3.00	3.00	3.00	3.00	3.00	2.99	2.99	2.99	2.99	2.98	2.98	2.98	2.98	2.98
median	2.58	2.57	2.57	2.57	2.57	2.57	2.57	2.56	2.56	2.56	2.56	2.56	2.56	2.55
25%	2.56	2.56	2.56	2.56	2.55	2.55	2.55	2.55	2.54	2.54	2.54	2.54	2.54	2.54
min	2.56	2.56	2.56	2.55	2.55	2.55	2.55	2.54	2.54	2.54	2.54	2.54	2.53	2.53
	211	212	213	214	215	216	217	218	219	220	221	222	223	224
max	3.39	3.39	3.39	3.39	3.38	3.38	3.38	3.37	3.37	3.37	3.37	3.37	3.36	3.36
75%	2.98	2.97	2.97	2.97	2.97	2.96	2.96	2.96	2.96	2.95	2.95	2.95	2.95	2.95
median	2.55	2.55	2.55	2.55	2.54	2.54	2.54	2.54	2.53	2.53	2.53	2.53	2.53	2.52
25%	2.54	2.54	2.53	2.53	2.53	2.53	2.52	2.52	2.52	2.52	2.52	2.52	2.52	2.51
min	2.53	2.53	2.53	2.53	2.52	2.52	2.52	2.52	2.52	2.52	2.52	2.52	2.52	2.51
	225	226	227	228	229	230	231	232	233	234	235	236	237	238
max	3.36	3.36	3.36	3.35	3.35	3.34	3.34	3.34	3.34	3.34	3.34	3.33	3.33	3.33
75%	2.94	2.94	2.94	2.93	2.93	2.93	2.93	2.93	2.93	2.93	2.92	2.92	2.92	2.92
median	2.52	2.52	2.52	2.51	2.51	2.51	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50
25%	2.51	2.51	2.51	2.50	2.50	2.50	2.49	2.49	2.49	2.49	2.49	2.49	2.49	2.49

```
min    2.50 2.50 2.50 2.50 2.50 2.50 2.49 2.49 2.49 2.49 2.49 2.49 2.49 2.49
        239  240  241  242  243  244  245  246  247  248  249  250  251  252
max    3.33 3.33 3.33 3.33 3.33 3.32 3.32 3.31 3.31 3.31 3.31 3.31 3.31 3.30
75%    2.92 2.92 2.92 2.92 2.91 2.91 2.91 2.90 2.90 2.90 2.90 2.90 2.90 2.89
median 2.50 2.49 2.49 2.49 2.49 2.49 2.49 2.48 2.48 2.48 2.48 2.48 2.48 2.48
25%    2.48 2.48 2.48 2.48 2.48 2.48 2.48 2.47 2.47 2.47 2.47 2.46 2.46 2.46
min    2.48 2.48 2.48 2.48 2.48 2.48 2.48 2.47 2.47 2.47 2.47 2.46 2.46 2.46
        253  254  255  256  257  258  259  260  261  262  263  264  265  266
max    3.30 3.30 3.30 3.30 3.29 3.29 3.29 3.29 3.28 3.28 3.28 3.28 3.28 3.27
75%    2.89 2.89 2.89 2.89 2.89 2.89 2.89 2.89 2.88 2.88 2.88 2.88 2.88 2.87
median 2.48 2.48 2.48 2.47 2.47 2.47 2.47 2.47 2.47 2.47 2.47 2.47 2.47 2.46
25%    2.46 2.46 2.46 2.46 2.46 2.46 2.46 2.46 2.45 2.45 2.45 2.45 2.45 2.45
min    2.46 2.46 2.46 2.46 2.46 2.45 2.45 2.45 2.45 2.45 2.45 2.45 2.44 2.44
        267  268  269  270  271  272  273  274  275  276  277  278  279  280
max    3.27 3.27 3.26 3.26 3.26 3.26 3.26 3.26 3.26 3.26 3.26 3.26 3.26 3.26
75%    2.87 2.87 2.87 2.87 2.87 2.86 2.86 2.86 3.26 3.26 3.26 3.26 3.26 3.26
median 2.46 2.46 2.46 2.47 2.47 2.47 2.46 2.46 2.86 2.86 2.86 2.86 2.86 2.85
25%    2.45 2.44 2.44 2.46 2.46 2.45 2.45 2.45 2.46 2.46 2.46 2.46 2.46 2.45
min    2.44 2.44 2.44 2.45 2.44 2.44 2.44 2.44 2.46 2.46 2.46 2.46 2.46 2.45
        281  282  283  284  285  286  287  288  289  290  291  292  293  294
max    3.26 3.25 3.25 3.25 3.25 3.25 3.25 3.25 3.25 3.25 3.25 3.24 3.24 3.24
75%    3.26 3.25 3.25 3.25 3.25 3.25 3.25 3.25 3.25 3.25 3.25 3.24 3.24 3.24
median 2.85 2.85 2.85 2.85 2.85 2.85 2.85 2.84 2.84 3.25 3.25 3.24 3.24 3.24
25%    2.45 2.45 2.45 2.45 2.44 2.44 2.44 2.44 2.44 3.25 3.25 3.24 3.24 3.24
min    2.45 2.45 2.45 2.45 2.44 2.44 2.44 2.44 2.44 3.25 3.25 3.24 3.24 3.24
        295  296  297  298  299  300
max    3.24 3.24 3.24 3.23 3.23 3.23
75%    3.24 3.24 3.24 3.23 3.23 3.23
median 3.24 3.24 3.24 3.23 3.23 3.23
25%    3.24 3.24 3.24 3.23 3.23 3.23
min    3.24 3.24 3.24 3.23 3.23 3.23
```

The `max.N` element of the `hotperm1` object stores the output of the N-method's permutations and is given by a matrix with 100 rows representing the permutations, and 10 columns representing the QTL mapping thresholds. Entry ij stores the maximum genome wide hotspot size detected at permutation i using the QTL mapping threshold j. The `max.N` element of the `summary` is a 10 by 10 matrix with rows indexing the QTL mapping thresholds and columns indexing the target genome wide error rates. Each entry ij shows the hotspot size above which a hotspot is considered significant at a GWER j using the QTL mapping threshold i. As before, our interest focus on the diagonal, and the N-method's threshold that controls the hotspot GWER at a 5% level when the QTL mapping was controlled at a GWER of 5% is 200.55. Note that according to the N-method, none of the hotspots on Figure 1 is significant.

The `max.lod.quant` element of the `hotperm1` object stores the output of the NL-method's permutations and is given by a matrix with 100 rows representing the permutations, and 300 columns representing the hotspot sizes analyzed. Entry ij stores the maximum genome wide $qLOD(n)$ value–the nth LOD score in a sample ordered from highest to lowest–computed at permutation i using the QTL mapping threshold j. The `max.lod.quant` element of the `summary` shows just the extremes and quartiles.

The `max.lod.quant` element of the `quantile` contains the sliding LOD quantiles from 1 up to 300 for each single trait LOD threshold. This can be used to plot how many traits pass the sliding LOD threshold, and are significant, at each locus in the genome. Figure 3 shows the hotspot significance profile for the thresholds targeting GWER at a 5% level. Note that we only consider LODs above `lod.thr`, and chromosome 1 has none.

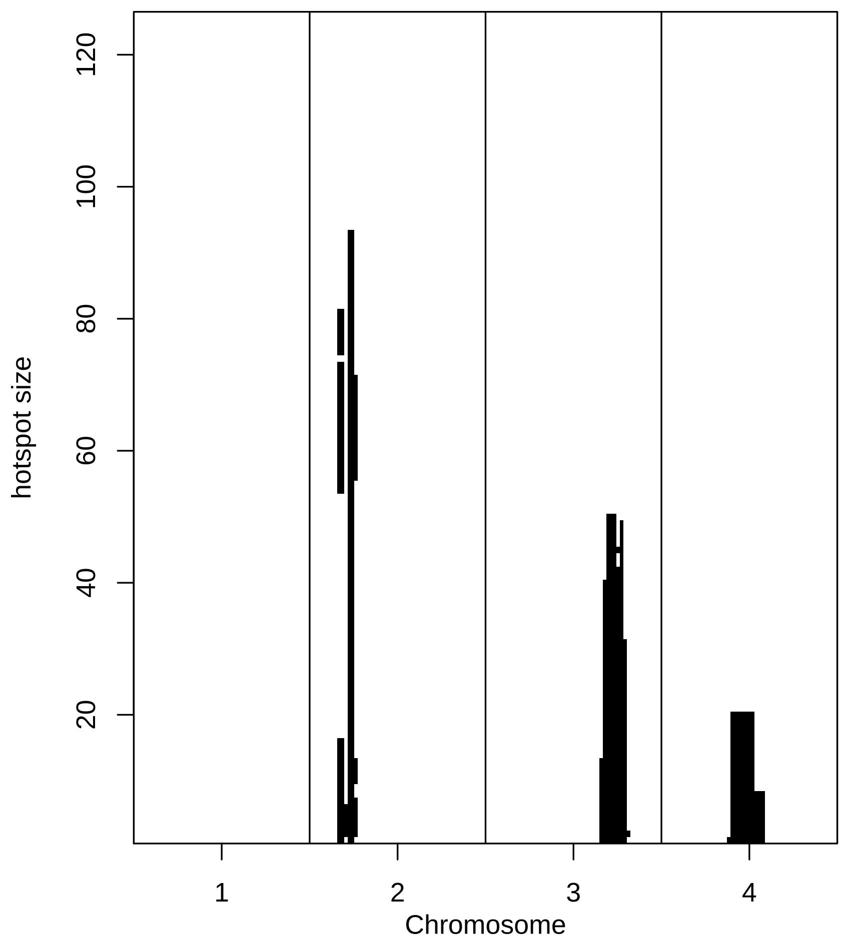


Figure 2: Hotspot size significance profile targeting GWER at a 5% level for example 1.

```
> quant1 <- quantile(hotperm1, 0.05, lod.thr = lod.thr)
> plot(high1, quant.level = quant1, sliding = TRUE)
```

Figure 3 depicts a sliding window of hotspot size thresholds ranging from $n = 1, \ldots, N$, where $N = 100$ corresponds to the (approximate) hotspot size threshold derived from the N-method. For each genomic location this figure shows the hotspot sizes at which the hotspot was significant, that is, at which the hotspot locus had more traits than the hotspot size threshold on the left mapping to it with a LOD score higher than the threshold on the right than expected by chance. For example, the hotspot on chromosome 3 was significant up to size 20, meaning that more than 1 trait mapped to the hotspot locus with LOD higher than 3.99, more than 11 traits mapped to the hotspot locus with LOD higher than 3.36, and so on up to hotspot size 49 where more than 49 traits mapped to the hotspot locus with LOD higher than 2.84.

The N-method that did not detect any hotspots, but the NL-method's sliding window correctly detected the simulated hotspots and showed that the apparent hotspots on chromosome 1 were noisy artifacts.

Here is another way to visualize this, using the sliding LOD threshold. We will also show a 5-cM smoothign window applied to the counts. Note that with the blue curve, no hotspots are detected on chr 1, and hotspots of size 20 are found on chr 2 and 4, with a large hotspot of size 50 on chr 3.

4 Example with Uncorrelated Phenotypes

Next we consider a second toy example with uncorrelated phenotype data. We repeat the simulation and analysis steps presented previously changing `latent.eff` to zero.

```
> ncross2 <- sim.null.cross(chr.len = rep(100,4),
+                           n.mar = 51,
+                           n.ind = 100,
+                           type = "bc",
+                           n.phe = 1000,
+                           latent.eff = 0,
+                           res.var = 1,
+                           init.seed = 123457)
> cross2 <- include.hotspots(cross = ncross2,
+                            hchr = c(2, 3, 4),
+                            hpos = c(25, 75, 50),
+                            hsize = c(100, 50, 20),
+                            Q.eff = 2,
+                            latent.eff = 0,
+                            lod.range.1 = c(2.5, 2.5),
+                            lod.range.2 = c(5, 8),
+                            lod.range.3 = c(10, 15),
+                            res.var = 1,
+                            n.phe = 1000,
+                            init.seed = 12345)
> ncor2 <- cor(cross2$pheno)
> summary(ncor2[lower.tri(ncor2)])
```

```
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-0.471200 -0.067160  0.001263  0.002224  0.070240  0.666900
```

```
> rm(ncor2)
> scan2 <- scanone(cross2, pheno.col = 1:1000, method = "hk")
> high2 <- highlod(scan2, lod.thr = lod.thr, drop.lod = 1.5)
> hots2 <- hotsize(high2)
```

```
> set.seed(12345)
> hotperm2 <- hotperm(cross = cross2,
+                 n.quant = 300,
+                 n.perm = 100,
+                 lod.thrs = lod.thrs,
+                 alpha.levels = alphas,
+                 drop.lod = 1.5,
+                 verbose = FALSE)
```

```
> quant2 <- quantile(hotperm2, 0.05, lod.thr = lod.thr)
```

The N-method, as expected, gave rise to much smaller thresholds in this second example with uncorrelated phenotypes. Additionally, inspection of Figure 4 shows no spurious hotspots on chromosome 1.

```
> plot(high2, lod.thr = lod.thr, quant.level = quant2, sliding = TRUE)
```

Figure 5 presents the hotspot significance profile targeting 5% GWER. For this second example, all methods correctly detected the simulated hotspots.

```
> plot(high2, quant.level = quant2)
```

Figure 6 shows another way to represent significant hotspots. We overlay the largest significant hotspot counts using the sliding quantiles in red on top of the curve on Figure 4. Notice that the large sizes are all significant, but only small sizes corresponding to larger LOD scores are significant. We add a right axis with the sliding LOD thresholds.

5 References

1. Breitling R., Y. Li, B. M. Tesson, J. Fu, C. Wu, et al., 2008 Genetical genomics: spotlight on QTL hotspots. PLoS Genetics **4:** e1000232.

2. Broman K. W., W. Wu, S. Sen, G. A. Churchill, 2003 R/qtl: QTL mapping in experimental crosses. Bioinformatics **19**: 889-890.

3. Chaibub Neto et al., 2012 Quantile-based permutation thresholds for QTL hotspots. Genetics (under review).

4. Churchill G. A., and R. W. Doerge, 1994 Empirical threshold values for quantitative trait mapping. Genetics **138**: 963-971.

5. Manichaikul A., J. Dupuis, S. Sen, and K. W. Broman, 2006 Poor performance of bootstrap confidence intervals for the location of a quantitative trait locus. Genetics **174:** 481-489.

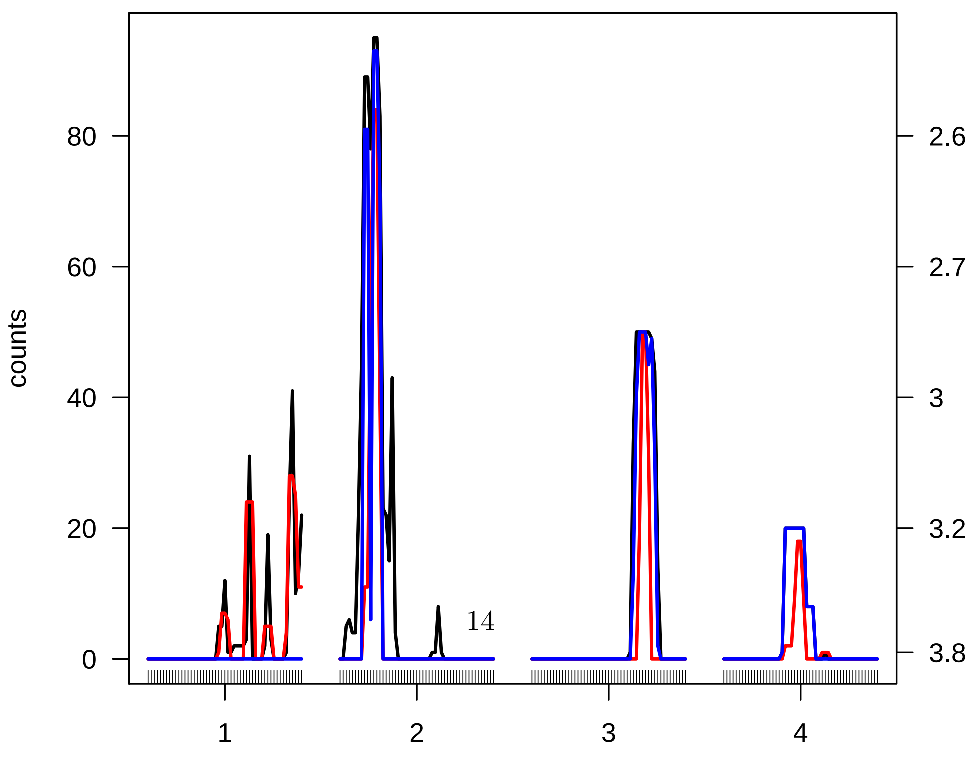
6. West M. A. L., K. Kim, D. J. Kliebenstein, H. van Leeuwen, R. W. Michelmore, R. W. Doerge, D. A. St. Clair 2007 Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis. Genetics **175:** 1441-1450.

7. Wu C., D. L. Delano, N. Mitro, S. V. Su, J. Janes, et al. 2008 Gene set enrichment in eQTL data identifies novel annotations and pathway regulators. PLoS Genetics **4:** e1000070.

```
> hotsq1 <- hotsize(high1, lod = lod.thr, window = 5, quant.level = quant1)
> plot(hotsq1)
> summary(hotsq1)

hotsize elements:  chr pos max.N max.N.window quant
LOD threshold: 2.438697
smooth window: 5
quantile level summary:

      chr pos max.N max.N.window quant
D1M48   1  94    41           28     0
D2M12   2  22    95           84    93
D2M13   2  24    95           84    93
D3M35   3  68    50            0    40
D3M36   3  70    50           19    50
D3M37   3  72    50           50    50
D3M38   3  74    50           50    50
D3M39   3  76    50           31    45
D4M21   4  40    20            2    20
D4M22   4  42    20            2    20
D4M23   4  44    20            2    20
D4M24   4  46    20            9    20
D4M25   4  48    20           18    20
D4M26   4  50    20           18    20
D4M27   4  52    20            9    20
```



```
> par(mar=c(4.1,4.1,0.1,0.1))
> plot(hots2, cex.lab = 1.5, cex.axis = 1.5)
```

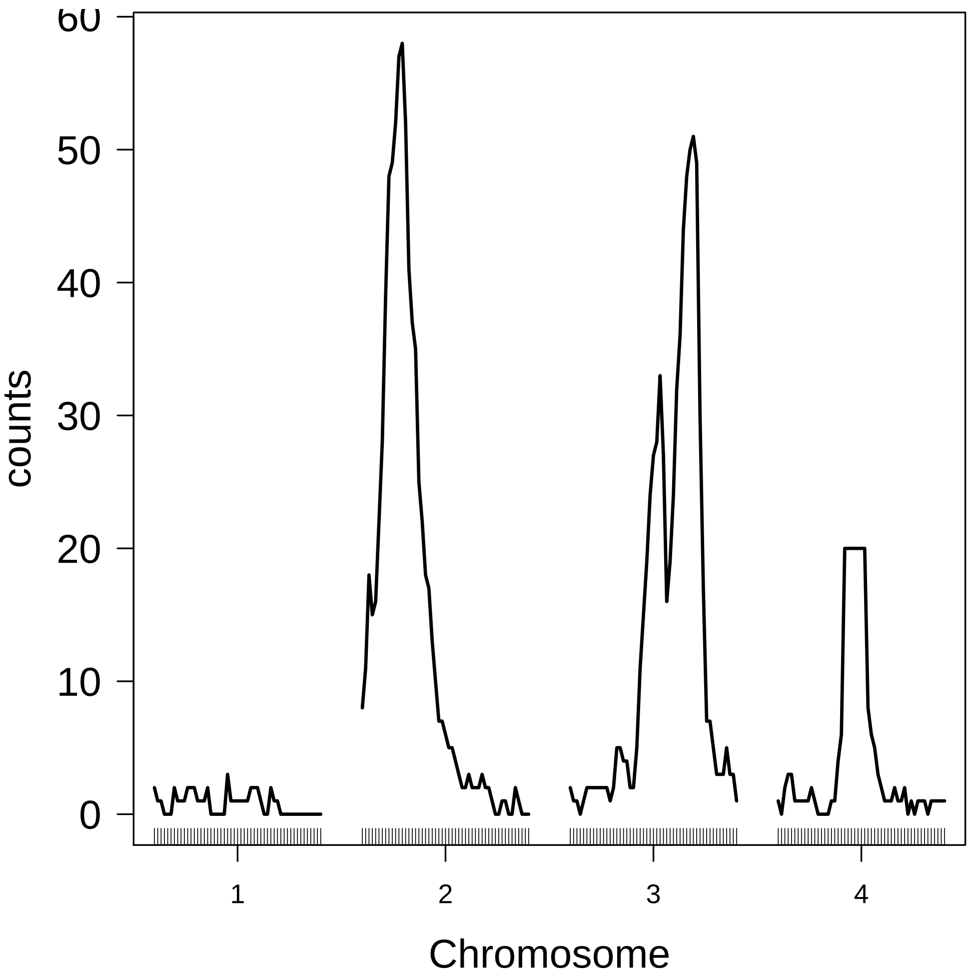


Figure 4: Hotspot architecture targeting 5% GWER for example 2.

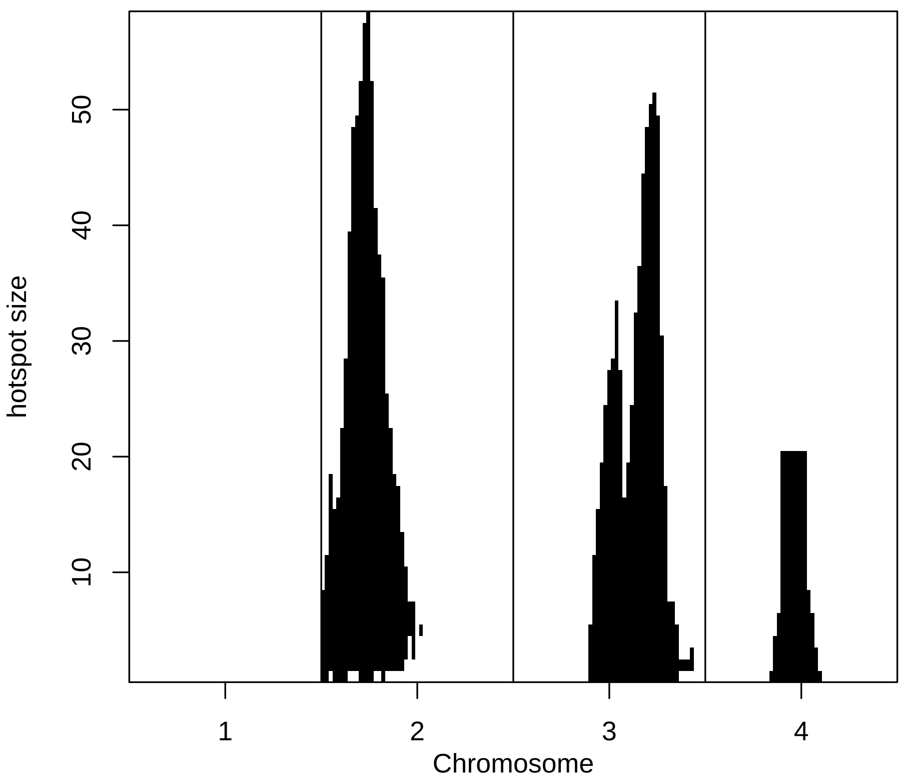


Figure 5: Hotspot significance profile targeting 5% GWER for example 2.

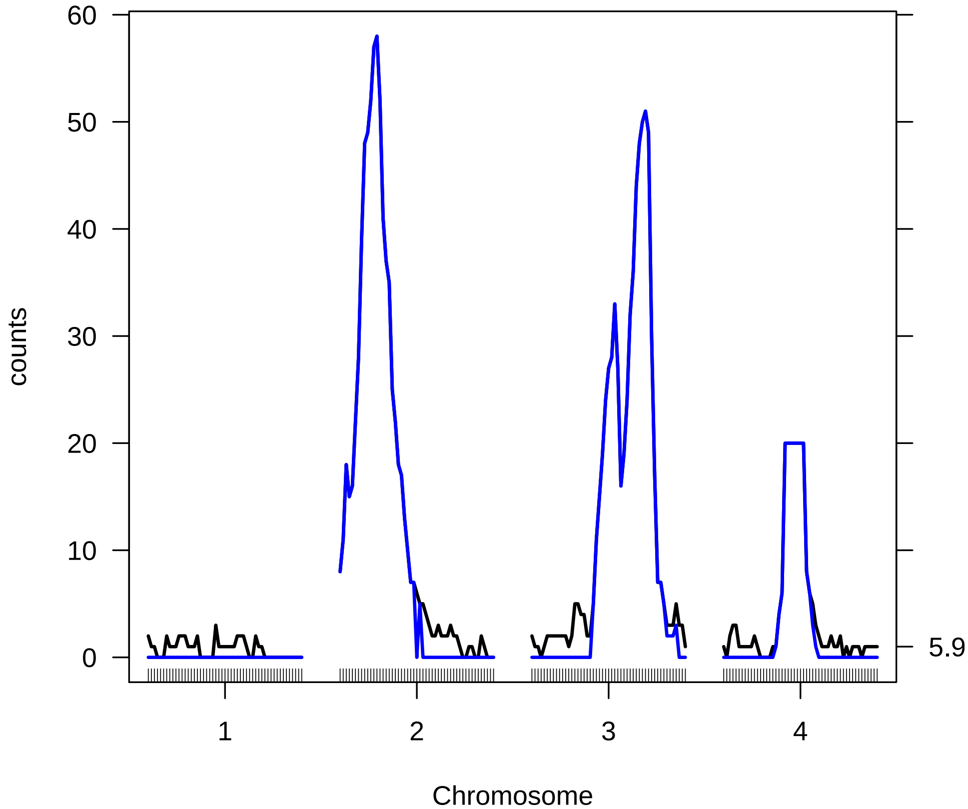


Figure 6: Hotspot significance scan targeting 5% GWER for example 2.