

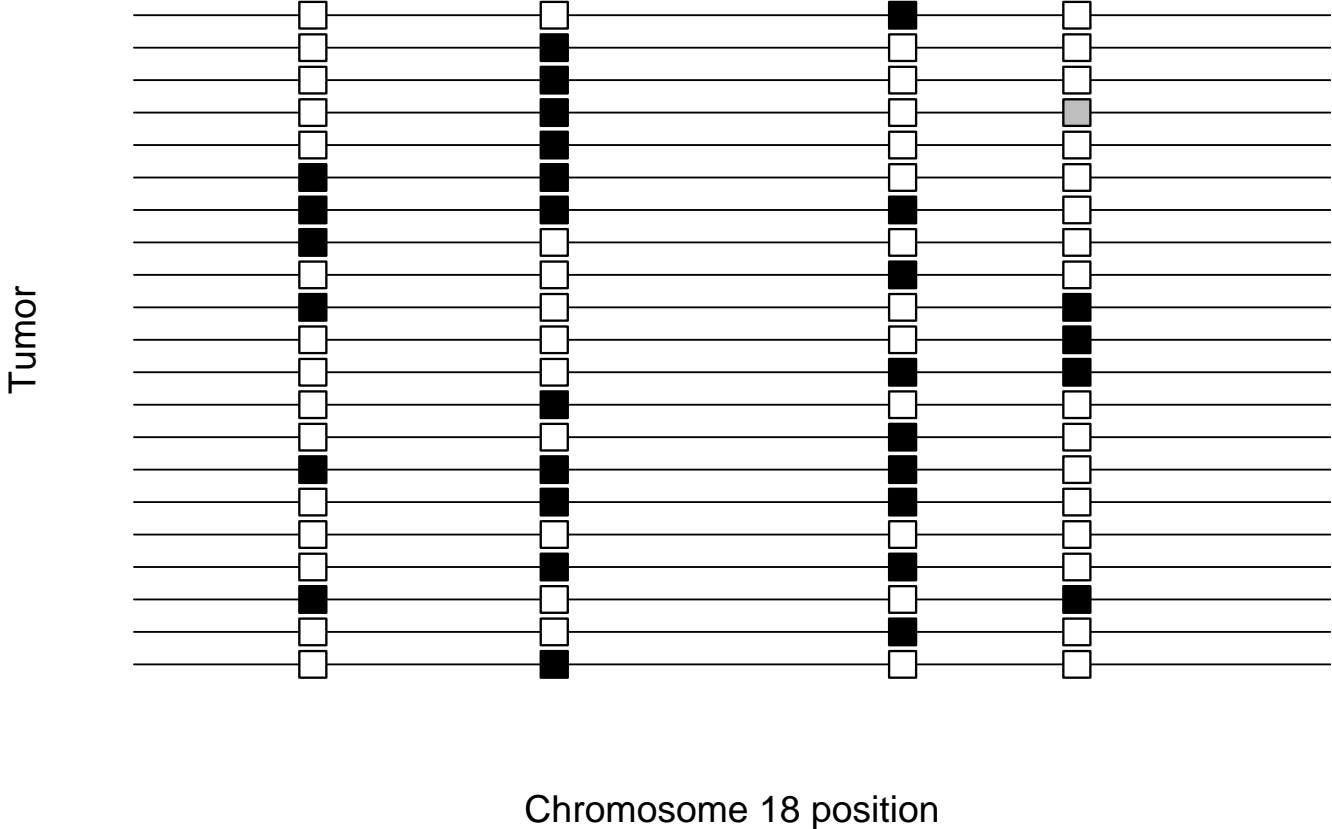
Modeling, analysis, and MCMC with LOH data

Michael Newton

IMA September 2003

Loss of heterozygosity

Data from Rat Mammary Cancer Study; Shull et al.



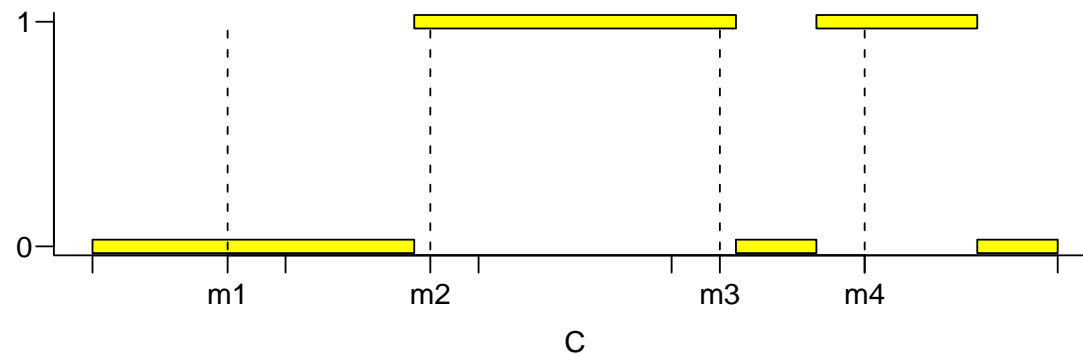
Plus single marker data from 5 other chromosomes.

Statistical Problems

- characterize deletion hot spots: significance, location, effect

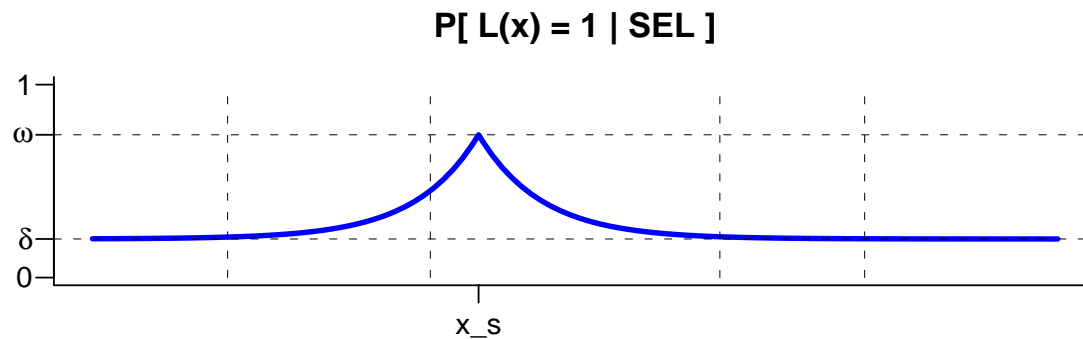
Instability

- Continuous 'time' binary Markov process $\{L(x) : x \in C\}$
- $P[L(x + dx) = 1 | L(x) = 0] = \lambda \delta dx + o(dx)$
- $P[L(x + dx) = 0 | L(x) = 1] = \lambda(1 - \delta) dx + o(dx)$



Selection

- special location $x_s \in C$ (putative suppressor gene locus)
- $P[\text{SEL} | L(x_s) = 1] \geq P[\text{SEL} | L(x_s) = 0]$
- plus conditional independence [one gene assumption]
- all observations are conditional on SEL

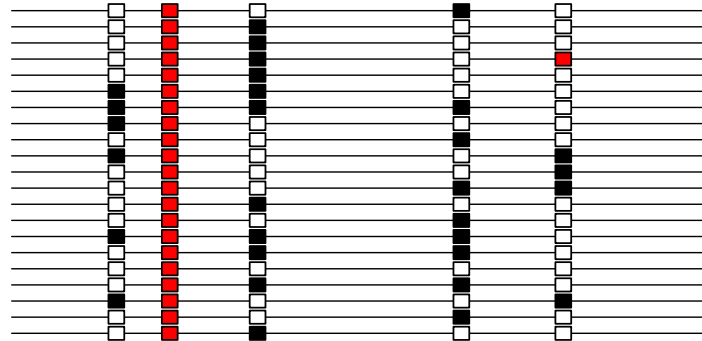


Instability - Selection

- snapshot of a dynamic biological system [tumor growth]
- aberrations emerge randomly [genetic instability]
- beneficial aberrations survive [cell-level selection]

Inference

Data: $D = \text{LOH at all markers, all tumors}$
 Unknown Parameters: $\theta = (\delta, \omega, \lambda, x_s)$
 Missing Data: $Z = L(x_s) \forall \text{ tumors}$
 and L at any noninformative markers



With state $S = (\theta, Z)$, working conditionally on $I = \cap \text{SEL}$,
 the target is:

$$\begin{aligned}
 P(S|D, I) &\propto P(S, D|I) \\
 &= P(D, Z, \theta|I) \\
 &= \underbrace{P(D|Z, \theta)}_{\text{M-chains}} \underbrace{P(Z|\theta, I)}_{\text{Bernoulli}(\omega)} \underbrace{P(\theta|I)}_{\text{prior}}
 \end{aligned}$$

Prior

$$P(\theta|I) = \underbrace{P(\delta)}_{\text{Unif}(0,1)} \underbrace{P(\omega|\delta)}_{\text{Unif}(\delta,1)} \underbrace{P(\lambda)}_{\text{Exponential}} \underbrace{P(x_s)}_{\text{Unif}(C)}$$

MCMC

Generate states S_1, S_2, \dots aiming at target $P(S|D, I)$

Propose $S^* \sim q(S, \cdot)$ according to some move type.

Accept proposal w.p. $\min(1, r)$ where

$$r = \frac{P(S^*|D, I) q(S^*, S)}{P(S|D, I) q(S, S^*)}.$$

Use simple move types for computational feasibility.

Use multiple move types for irreducibility.

Use well-chosen move types for statistical efficiency.

Subsample to simplify output analysis.

Run output analysis on saved states.

Move Types

- ω : Propose $\omega^* \sim \text{Beta}(a, b)$ restricted to $(\delta, 1)$
- δ : Propose $\delta^* \sim \text{Unif}(\delta - \epsilon, \delta + \epsilon)$, reflected into $(0, \omega)$.
- λ : Propose $\lambda^* \sim \text{Exponential}(\lambda)$
- x_s : Sample a neighboring (or current) marker gap; sample x_s^* uniformly within the gap.
- Z : Select an entry of missing data; propose opposite.

```
> names(res)
[1] "mcmc"      "acrate"    "prior"     "thsave"    "logpost"
```

```
> res$mcmc
```

```
$nskip
```

```
[1] 100
```

```
$nsave
```

```
[1] 5000
```

```
$eps
```

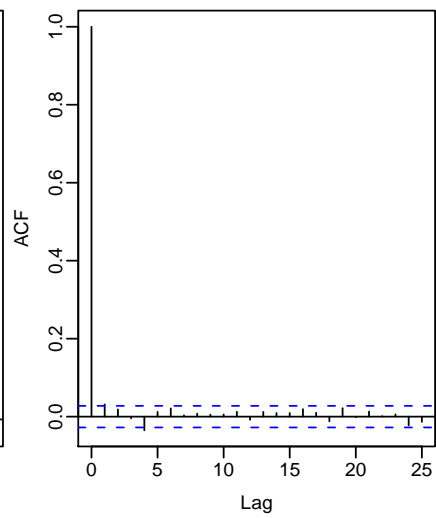
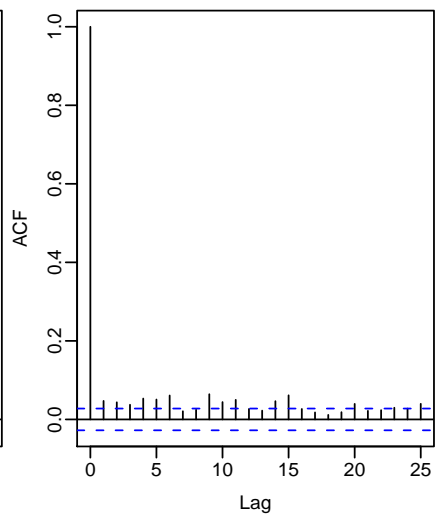
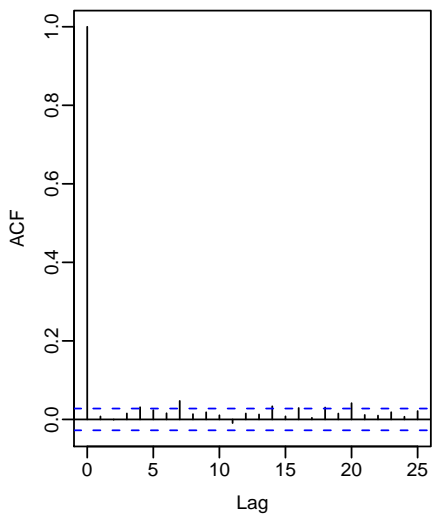
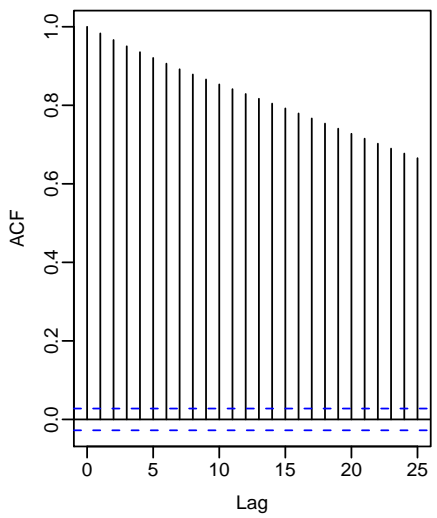
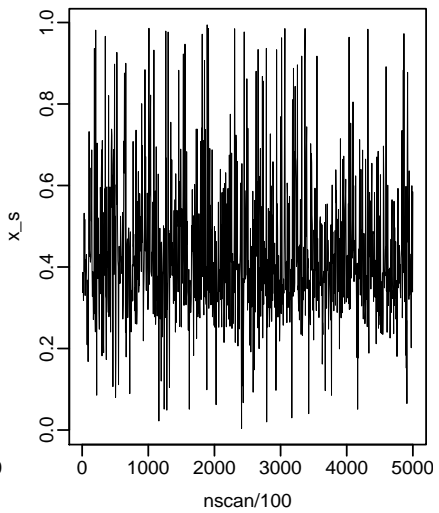
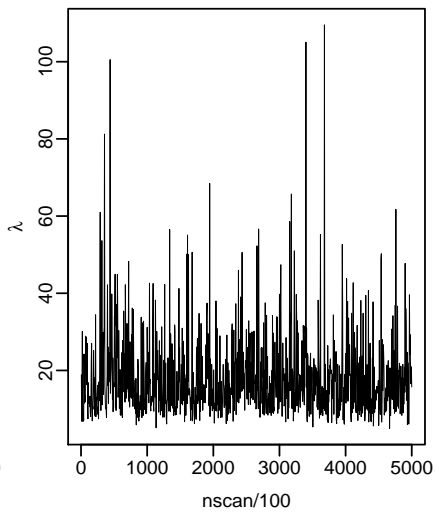
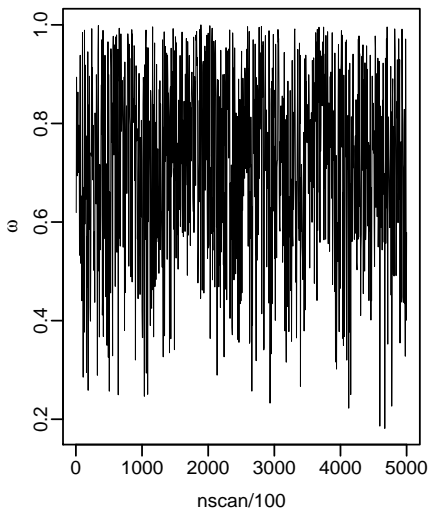
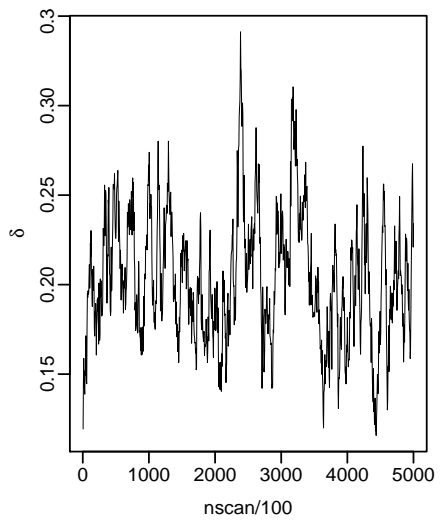
```
[1] 0.002
```

```
> res$acrate
```

```
      omega lambda&delta      xs      impute
[1,] 0.990132      0.316144 0.325568 0.1729838
```

```
> res$thsave[1:2,]
```

```
      delta      omega      lambda      xs
[1,] 0.1220636 0.6499343 7.079887 0.3926389
[2,] 0.1185316 0.5888894 8.516544 0.2771046
```



A second run

```
> res$mcmc
```

```
$nskip
```

```
[1] 100
```

```
$nsave
```

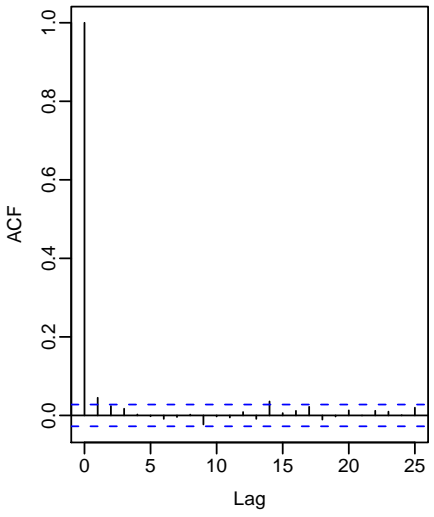
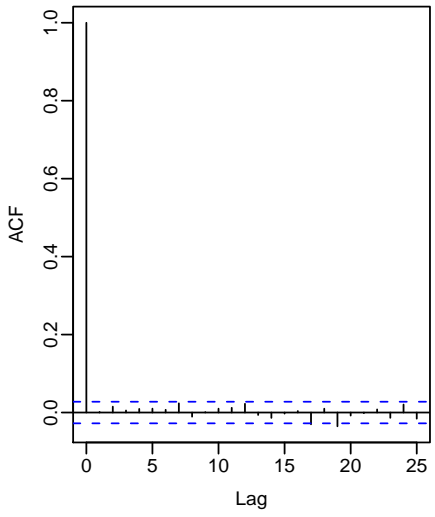
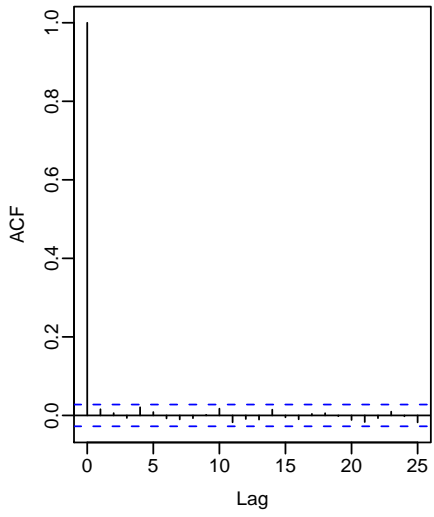
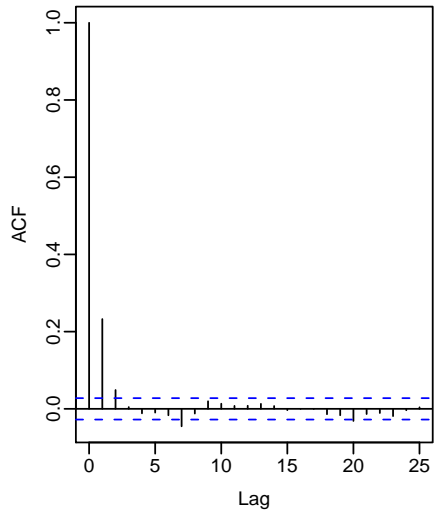
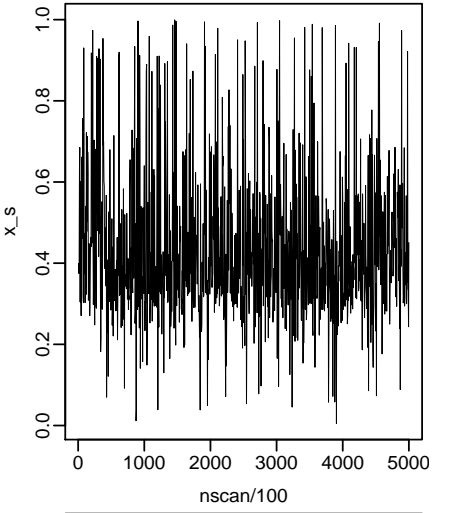
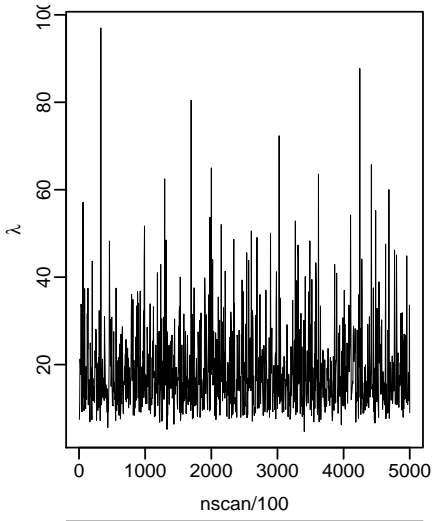
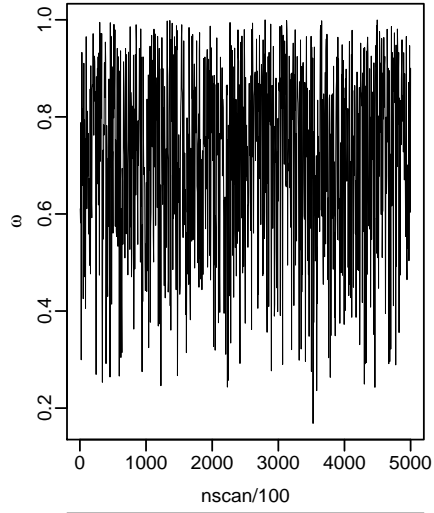
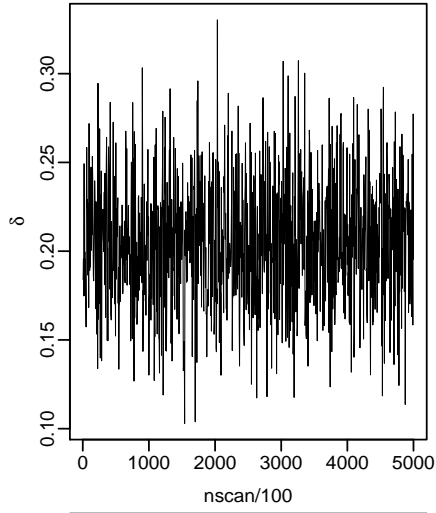
```
[1] 5000
```

```
$eps
```

```
[1] 0.02
```

```
> res$acrate
```

```
          omega lambda&delta          xs      impute  
[1,] 0.988586      0.303814 0.324792 0.1752827
```

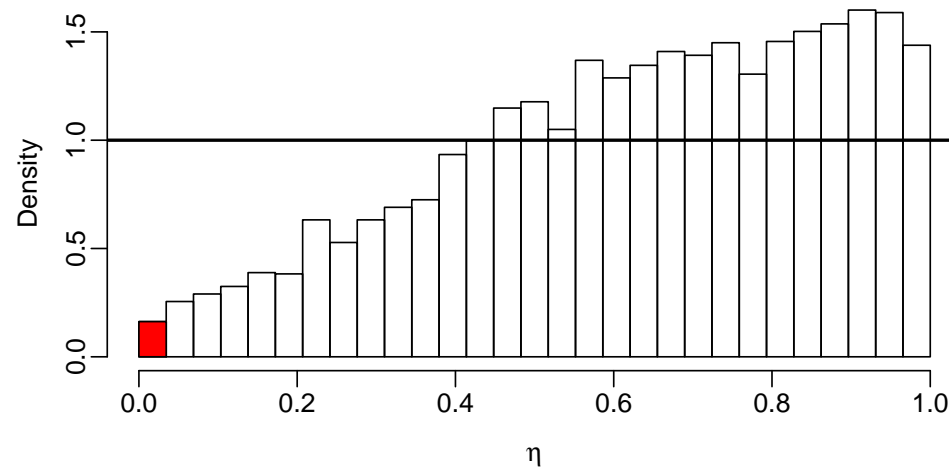


Bayes Factor for $H : \omega = \delta$

Introduce $\eta = (\omega - \delta)/(1 - \delta)$, so $H : \eta = 0$.

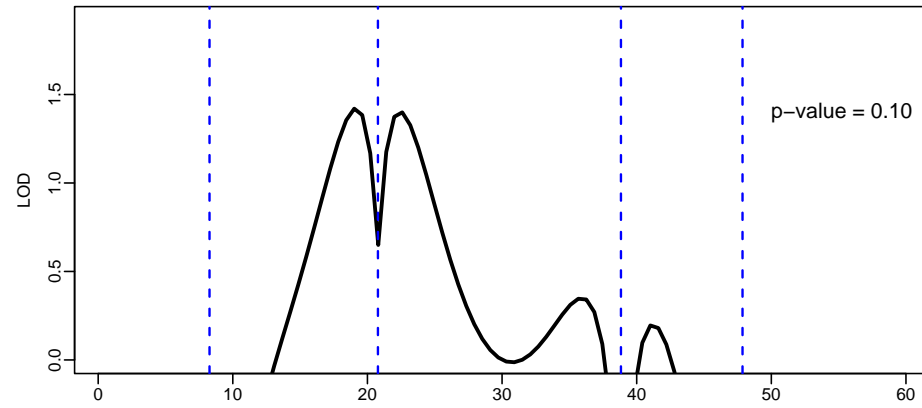
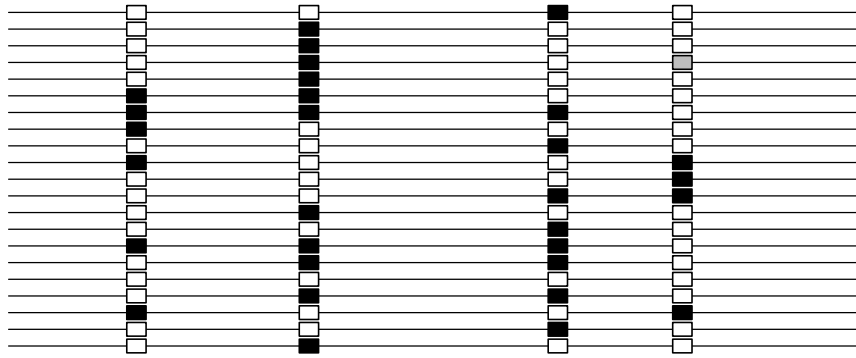
By Savage's density ratio,

$$\text{BF} = \frac{\pi(D|H)}{\pi(D|H^c)} = \frac{\pi_a(\eta = 0|D)}{\pi_a(\eta = 0)}$$

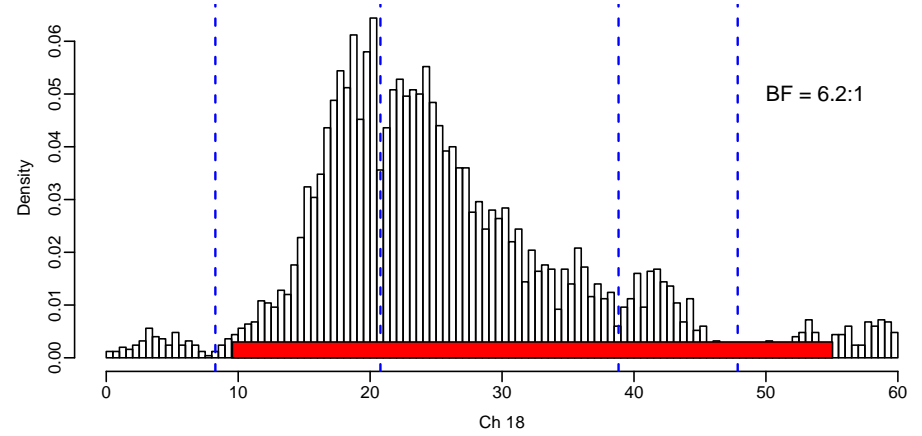


Here 6:1 favoring a gene.

Chromosome 18 Analysis



Ch 18: delta=0.20, omega=0.72



Citations

J.M. Dickey (1971). The weighted likelihood ratio, linear hypotheses on normal location parameters. *The Annals of Mathematical Statistics* **42**, 204-223.

M.A. Newton, M.N. Gould, C.A. Reznikoff, and J. D. Haag (1998). On the statistical analysis of allelic-loss data. *Statistics in Medicine*, **17**, 1425-45.

M.A. Newton and Y.J. Lee, 2000. Inferring the location and effect of tumor suppressor genes by instability-selection modeling of allelic-loss data. *Biometrics*, **56**, 1088-1097.