EDA Quiz

James Bullard, Kasper Hansen, Margaret Taub

July 8, 2008

1. Given the following time-series microarray data, construct a barplot for the gene expression through the 12 time points. Add a line to the barplot which is the trimmed average with a window size of $K = 3$. Note, one thing to remember is that barplot returns a vector of x positions. Howver, it does so with the function invisible, which we haven't seen, so check out the help: ?invisible. (See Figure (1) for the plot you want to emulate.)

```
> states <- numeric(12)
> states[1] <- rpois(1, lambda = 4) + 1
> for (i in 2:12) states[i] <- rpois(1, lambda = states[i - 1]) +
+     1
> dta <- rexp(12, 1/states)
```

2. Add other smooth lines to the plot above. Namely, try a lowess smoothed line. Are there transformations of the data which help make any patterns stand out?

3. As in most microarray experiments, the measurements above were actually produced from a summary statistic of a number of lower-level measurements, i.e., we have J probes for each gene and we summarize these to produce a number which is our "gene-level" measure. We have taken averages above to produce the data, but we are concerned that this is too dramatically affected by outliers. Using the data below explore the distributions of the probe level measures for the timepoints, and produce a plot like Figure (2). In the plot, the red line connects the boxplot means and the yellow line connects the boxplot running means ($K = 3$ again). It should be easy to produce both if you have solved Question 1.

```
> rawData <- sapply(states, function(s) {
+     rexp(22, 1/s)
+ })
```

4. The file pvals.dta contains three sets of p-values from a genome-wide association study, where three methods were used to calculate significance. Please reproduce the p-value plot shown in Figure 3 to the best of your ability.
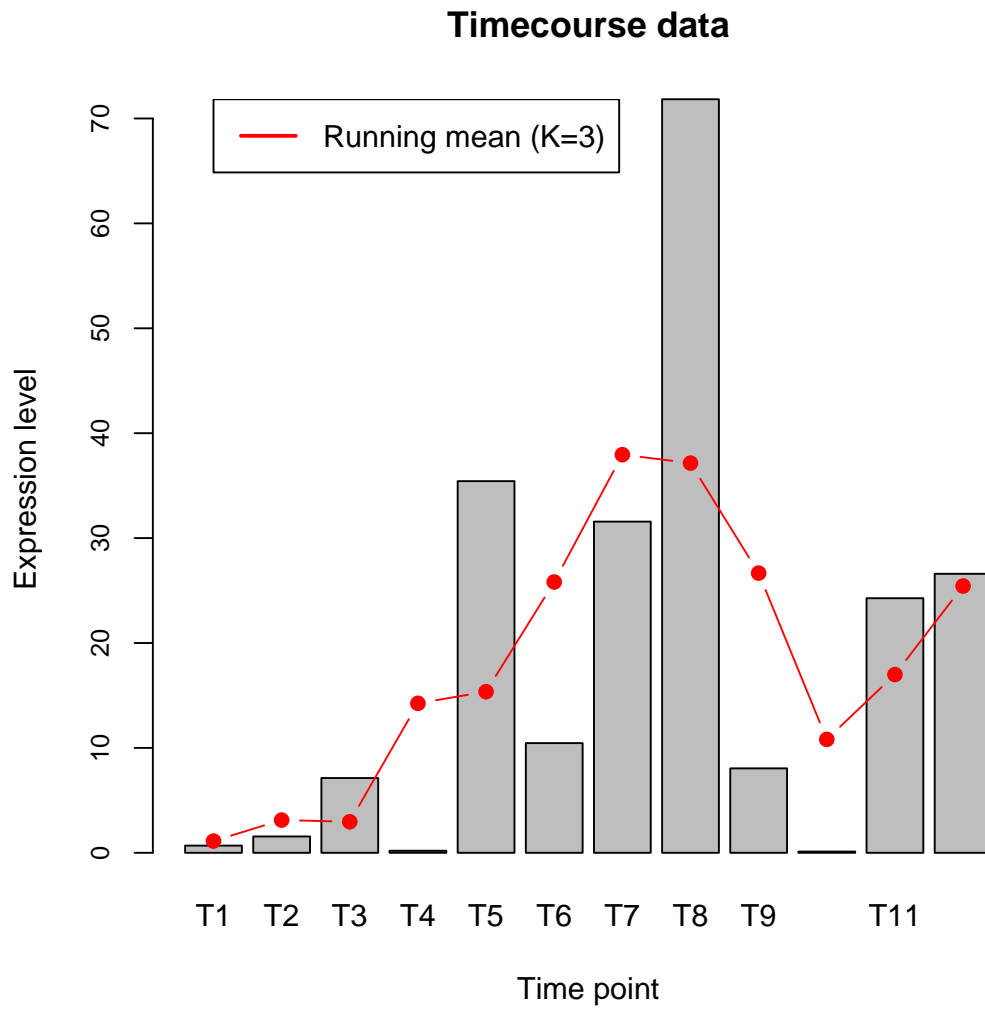
Timecourse data

Figure 1: Barplot of gene level expression measures from a microarray timecourse experiment.
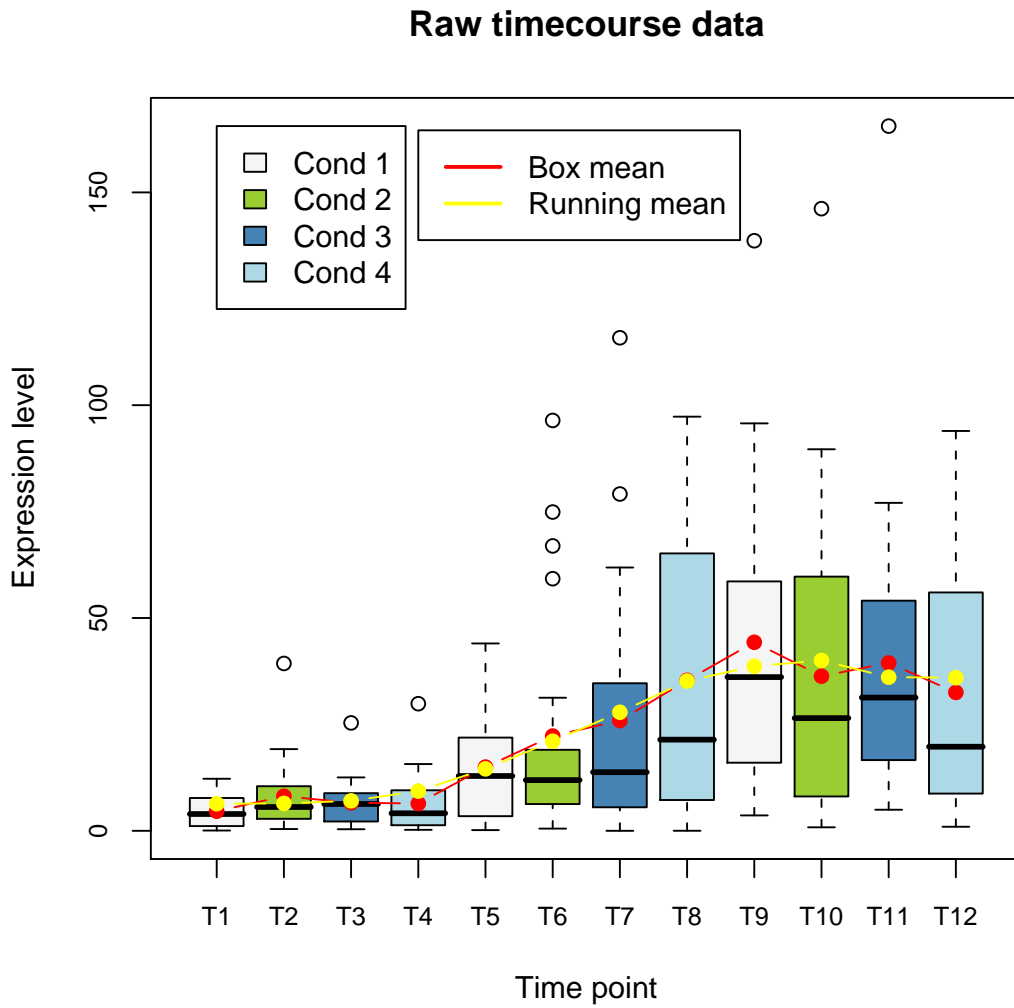
Figure 2: Box plots of gene expression measures from a microarray timecourse experiment.
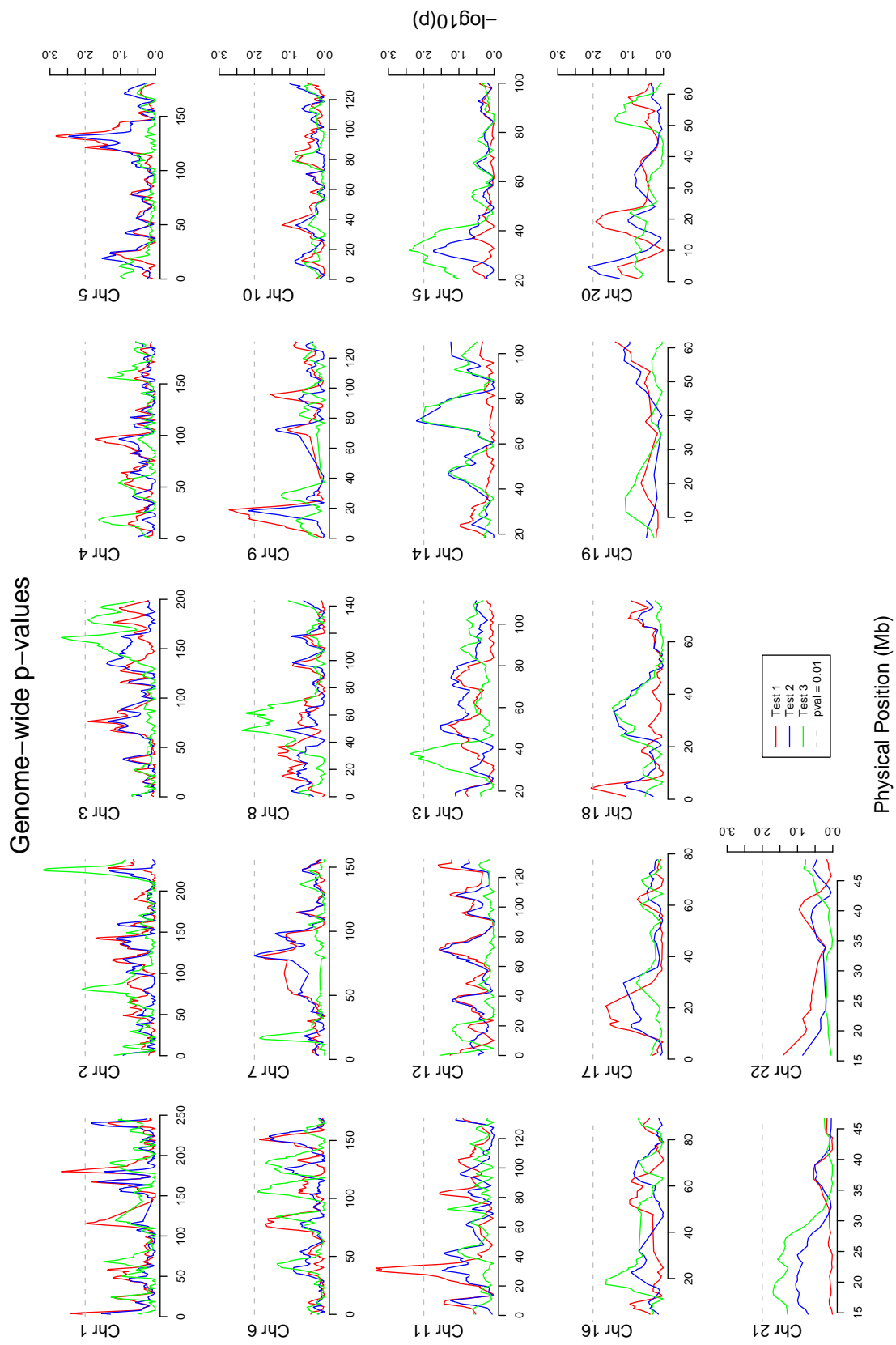
Figure 3: P-value plots by chromosome.