Time Course Experiments

Biostatistics 140.688
Rafael A. Irizarry

Today material courtesy of Terry Speed

Outline

Types and features
Design, including replication
(with examples)
Identifying the genes of interest
(with examples)

2

Types and features of microarray time course experiments

Types/features

- Typically short series: k = 4-10 time points for shorter,and 11-20 time points for longer series; often irregularly spaced; with no or few (< 5) replications.
- Can be periodic, as in the cell cycle: Cho *et al.* (1998), Spellman *et al.*,(1998), or circadian rhythms, Storch *et al.* (2002),

  OR
- May have no particular pattern, as in developmental time courses: Chu et al. (1998), Wen et al. (1998), Tamayo et al. (1999).

4

Types/features, cont.

- May be longitudinal, where mRNA samples at different times are extracted from the same unit (cell line, tissue or individual), but more commonly cross-sectional, where mRNA samples are from different units.
- Gene expression values at different time points may be correlated, especially in a longitudinal study, or when a common reference design is used for a cross-sectional study. At other times, the experimental design induces correlations in cross-sectional studies.

5

Types/features, completed.

- Two general types of hypotheses of interest: the one-sample (or one-class) problem: which genes are changing in time? and the 2 or >2 sample (or class) problem: which genes are changing differently in time across the samples (or classes)?
- Two broad types of mRNA samples: from cells or cell lines which give reasonably repeatable responses within classes, and whole organism (mice, humans), where there is a lot of response variability within classes.

6

Design of microarray time course experiments

Most important issues

The first issue is: longitudinal or cross-sectional? The question revolves on whether it is important to measure change within units.

For two-channel (cDNA or long oligo) arrays, a major question is whether or not to use a reference design. Most frequently, the answer is yes.

For very short two-channel time courses, the possibility arises of optimizing the design for contrasts of interest.

Important design issues include not just assignment of mRNA to arrays, but also the actual conduct of the experiment, including preparation of the sample mRNA, the times of hybridizations, and the equipment, reagents and personnel used.

8

First illustrative example:
A plant's response to a pathogen.

.

Healthy *Arabidopsis thaliana* (mustard weed) plant

9

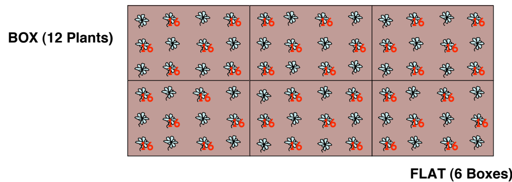
A. thaliana response to infection by *E. orontii*

- Two lines of plants: *Columbia*, *Col-0* = wildtype (wt), and an enhanced disease susceptibility line *eds16* (mutant).
- **Objective**: to identify genes whose temporal expression patterns following infection differ between wt (*Col-0)* and mutant (*eds16).*

10

Layout of the At experiments, I

BOX (12 Plants)

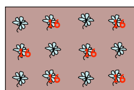


FLAT (6 Boxes)

- Plants are evenly positioned
- Wild type and mutant intermixed
- Same numbered leaves sampled each time point (leaves 5, 6, 7)
- "Random" sampling of plants from two flats for each time point
- New plants used for each time point (can't resample)
- Sampling occurs at same time of day (with exception of 6hr)
- Each RNA sample contains leaves from 2 plants

11

Layout of the At experiments, II

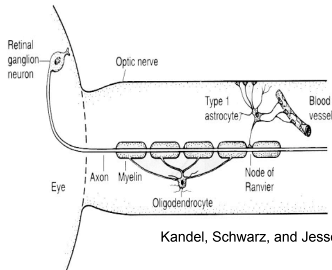


- Plants are grown in environmental chambers: Temp, Rel Hum, Light Intensity
- 4 week old plants were infected with heavy innoculum of powdery mildew (spores from 3 heavily infected leaves per box)
- Infection was performed using a 3 ft settling tower with mesh
- Uninfected plants were kept in similar environmental chambers
- Three separate experiments were performed
- Triplicate samples were harvested for each experiment at each time point
- Samples-0 (uninfected), 6 hr, 1 d, 3 d, 5d, 7dpi, 7d (uninfected)

12

Second illustrative example

Oligodendrocytes (OL) myelinate Central Nervous System axons……and develop from migrating oligodendrocyte precursor cells (OPC)

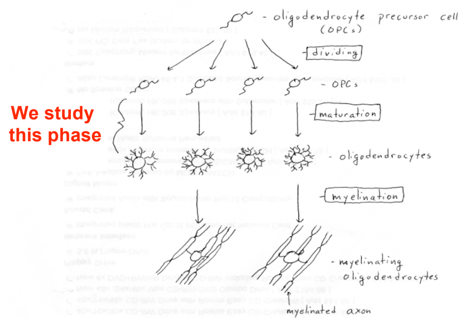


Kandel, Schwarz, and Jessel; 1991

13

The development from OPC to OL in vitro

We study this phase

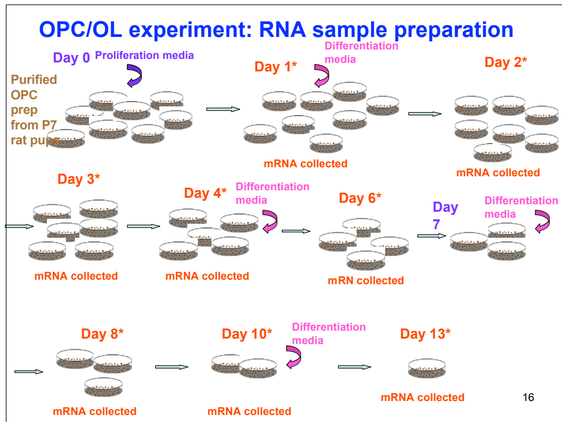


14

Purpose of OPC/OL experiment

- Broad purpose: to examine gene regulation in cultured oligodendrocyte precursor cells (OPC) as they develop into oligodendrocytes (OL).
- Narrower purpose: to identify a subset of genes with up-regulated timecourses. Candidate genes predicted to be secreted will be assayed for their ability to cluster sodium channels along cultured retinal ganglion axons.

15



Back to generalities: Replication

We can have biological, technical and probe set (spot) replicates

Replication is a good thing. With it we get estimates of variability relative to which temporal changes and/or condition differences can be assessed.

Biological replicates are best, as they permit conclusions to be extrapolated, something not possible with tech. reps.

With unreplicated experiments, inference to a wider population is not possible, and analysis is less straightforward, being dependent on unverifiable assumptions, as no estimate of pure error is available.

When we do have replicates, it is better to use the variation between them in the analysis, and not simple average them.

Today I will discuss only replicated time course experiments.

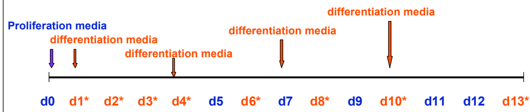
17

Replication in the At experiments

- Three experiments - effectively biological replicates - were conducted using the wt and mutant lines, and within each, 3 technical replicate series. Not all have been hybridized to chips. Later we use one series from experiments I and III, and two from experiment II.
- These experiments are longitudinal at the level of experiment, but cross-sectional at the level of mRNA sample (from separate leaves). The blurring of these distinctions is not unusual.

18

Replication in the OPC/OL experiment



mRNA samples at the 8 *-d time points were collected

4 independent preparations were performed, each of which generated mRNA for every time point. We view this as 4 biological replicates of a longitudinal study. Again it is not clearcut. For each biological replicate, a dye-swap pair of technical replicates was done.

19

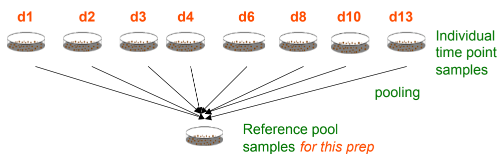
Hybridizations for the *At* experiments

- Initially we hybridized mRNA from just one of the technical replicate series from experiments I and III, and two from experiment II.
- The Affymetrix *Arabidopsis* 24K GeneChip® was used. In all 2(genotypes)x6(times)x4(experiments) = 48 chips were hybed.
  In addition, data from 2(genotypes)x4(experiments) = 8 chips for day 7 uninfected samples are plotted.
- Low level analysis (background, normalization, probe set summarization) done by RMA.

20

Pooling in the OPC/OL experiment

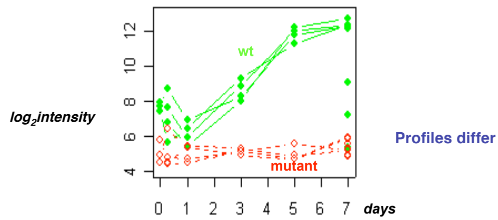
Each prep has its own reference pool, which is the pool of all the individual time point mRNA samples of *that* prep.



21

Statistical question for At experiments:

Find genes whose expression profiles differ between genotypes?

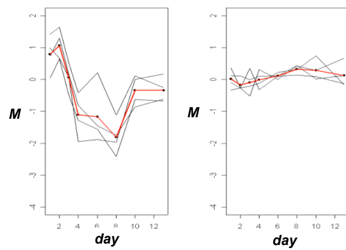


Extra dots at day 7 from uninfected samples.

22

Statistical question for the OPC/OL experiment

Find genes whose expression levels ***change*** over time?



Change over time **No change over time**

23

Identifying the genes of interest

Clustering
Pairwise comparisons
ANOVA with time as a factor
Empirical Bayes methods

Clustering with time-course data: brief literature review

Clustering methods have been widely used in this context to find groups of genes with interesting and similar patterns.

Hierarchical clustering: Eisen *et al* (1998)

Self-organizing maps: Tamayo *et al* (1999), Saban *et al* (2001), Burton *et al* (2002).

k-means clustering: Tavazoie *et al* (1999)

Bayesian model-based clustering: Bar-Joseph *et al* (2002, 2003), Ramoni *et al* (2002)

HMM clustering: Schliep *et al* (2003).

25

Some drawbacks of clustering methods

They make no explicit use of the replicate information. They either use all the slides or means of the replicates.

Clustering does not provide a ranking for the individual genes based on the magnitude of change in expression levels over time.

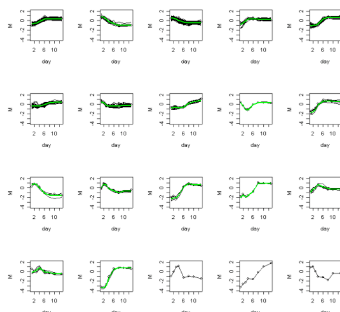
When the number of genes becomes large, clustering methods may not provide clear group patterns.

Cluster analysis may fail to detect changing genes that belong to clusters for which most genes do not change (Bar-Joseph et al. 2003).

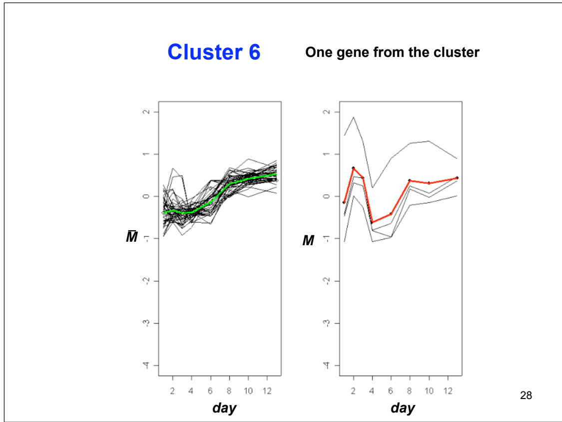
There is the perennial question: How many clusters?

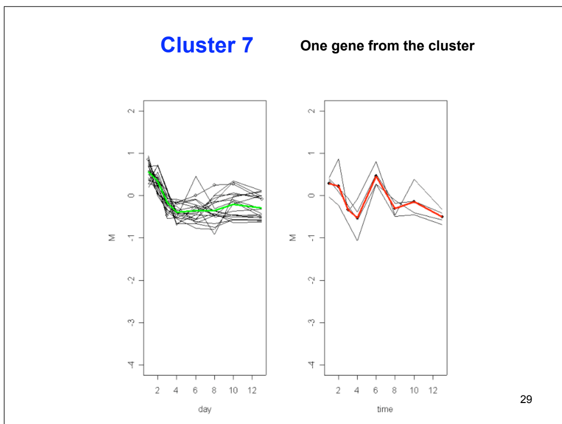
26

Hierarchical clustering of the top 500 genes OPC/OL expt, filtered by their variance



27





Pairwise comparisons

One strategy is to make many or all univariate pairwise comparisons, e.g. of consecutive times: days

1 vs 2, 2 vs 3, 3 vs 4, 4 vs 6, 6 vs 8, 8 vs 10, 10 vs 13

Illustration on the OPC/OL data: *t*-tests, univariate posterior odds : e.g. the *LOD* statistic, Lönnstedt and S (2002), Smyth (2004), the moderated *t* statistic, Smyth (2004),

30

Use of moderated *t* and posterior odds (LOD)

$$s_*^2 = \frac{(n-1)s^2 + \nu\lambda^2}{n-1+\nu}$$

Moderated s^2 of *M* values

$$t_*^2 = \frac{\overline{M}^2}{s_*^2/n}$$

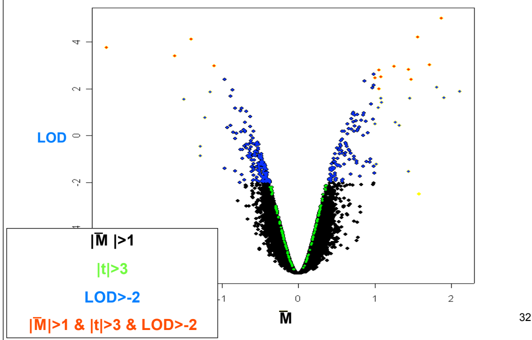
Moderated *t*

$$LOD = c + \left(\frac{n+\nu}{2}\right)\log_{10}\left\{\frac{t_*^2 + n - 1 + \nu}{t_*^2 \dfrac{1/n}{1/n + 1/\kappa} + n - 1 + \nu}\right\}$$

Log_{10} of posterior odds against differential expression

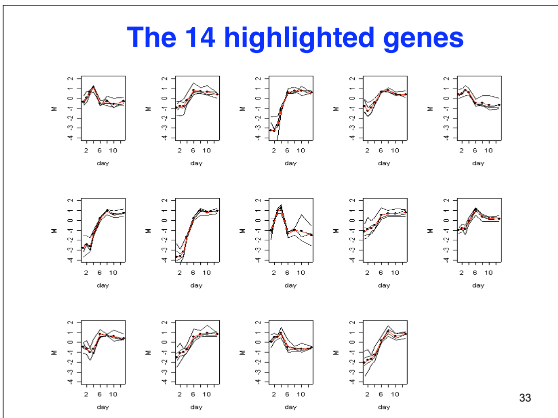
31

OPC/OL experiment: day 6 vs. day 4



32

The 14 highlighted genes



33

Pairwise comparisons: some drawbacks

As the previous slide shows, the strategy works, but….

- It involves a large number of tests for each gene, and there are over 10,000 genes in a typical microarray experiment: a two-way multiple testing problem.
- Merging all the lists of genes can be a tricky problem.
- We still cannot rank the genes according to the overall amount of change, which is often felt to be desirable.

34

ANOVA with time as a factor

At experiment: treat time and genotypes as factors with 6 and 2 levels, resp., and form the ratio of the *times* x *genotypes* MS to *residual* MS, giving an $F_{5,d}$ under the null, where $d = 2 \times 6 \times (4-1) = 36$ are the residual d.f. Since there is pairing of wt and mutant, we should include that too, giving a kind of split-plot anova with 3 d.f. for *reps*, and 33 residual d.f., with *times, genotypes* and *times* x *genotypes* as before.

OPC/OL experiment: here we simply regard the 8 *times* as defining 8 "groups", and use anova to test the hypothesis of all times means being equal, 4 replicate measurements for each time ("group").

35

Drawbacks with ANOVA

First, this approach does not deal adequately with correlations across time, if the experiment have a longitudinal component.

Second, just as with the t-statistic we illustrated in the pariwise comparisons, an element of moderation is desirable.

Despite these reservations, anova can and does provide an adequate analysis, although we feel it can be improved by attempting to deal with the above two issues.

One further point is this: with cross-sectional data, we can include regression modelling under the heading of anova, see later.

36

In general, what do we want?

We prefer a formula to rank genes, in order to

- find those changing or not similarly expressed
- provide a cut off for clustering

We feel that this formula should be

- *t*-like or F-like,

i.e. involve standardized measures of effects,

- multivariate, where appropriate, and
- moderated.

37

Why moderation?

- We seek genes with large absolute or relative amounts of change over time, in relation to their replicate variances, and covariances where relevant.
- Variances and covariances are poorly estimated in this context.
- Some sort of smoothing, borrowing strength, or empirical Bayes approach is called for. Simulations show that this helps, i.e. doing so improves the identification of genes of interest.
- We use multivariate normals with conjugate priors, as we want usable formulae, and not to have to use MCMC.

38

Multivariate approaches for longitudinal time course experiments

Here we treat one entire series as a random *k*-vector

Notation and models

We denote by $\boldsymbol{X}_{g,1},\ldots,\boldsymbol{X}_{g,n}$ the replicate random k-vectors representing the observed time series for a single gene.

For the *At* data, $n = 4$ and $k = 6$, and the $X_{g,i,t}$ are differences of log intensities, i.e. *log ratios* of *mutant* to *wt*.

For the OPC/OL data, $n = 4$ and $k = 8$, and the $X_{g,i,t}$ are *log ratios* of *experimental* to *reference pool* intensities.

Our underlying model is that these $\boldsymbol{X}_{g,i}$ are i.i.d. $N(\mu_g,\Sigma_g)$, and we make different assumptions about μ_g and Σ_g.

40

Hypotheses

- With the *At* data, we are interested in testing the null hypothesis H_g: $\mu_g = 0, \Sigma_g > 0$, against the alternative K_g: $\mu_g \neq 0, \Sigma_g > 0$.
- With the *OPC/OL* data, we are interested in testing the null hypothesis H_g: $\mu_g = constant, \Sigma_g > 0$, against the alternative K_g: μ_g *not constant*, $\Sigma_g > 0$.

41

Notation and models, cont.

For our empirical Bayes (EB) approach, we have priors for μ_g and Σ_g reflecting the indicator status $I = I_g$ of the gene, where $I_g = 1$ if H_g is true, and $I_g = 0$ otherwise, i.e. if K_g is true.

We suppose that $Pr(I_g = 1) = p$, independently for every gene, for a hyperparameter p, $0 < p < 1$.

From now on, we drop the subscripts g wherever possible.

42

Notation and models, completed

With this background, our prior for Σ is inverse Wishart with degrees of freedom ν and matrix parameter $(\nu\Lambda)^{-1}$, where $\Lambda > 0$ is positive definite. When we are dealing with a variance σ^2, we use an *inverse Gamma* prior with analogous parameters λ^2 and ν.

Our priors for μ will be different depending on whether $I=0$ or $I=1$, but in all cases are multivariate normal, and will involve Λ (or λ^2). We omit the details.

Finally, the data $X_1,\ldots, X_n$ are supposed i.i.d. given I, Σ and μ, with $X_i \mid I, \Sigma, \mu \sim N(\mu, \Sigma)$.

The multivariate normality assumption is reasonable, but not precise. However, we judge our results by their utility, not on goodness-of-fit of the models. 43

Summary of results for the At experiment; formulae for the OPC/OL experiment are similar.

Our moderated S is

$$\tilde{S} = [E(\Sigma^{-1} \mid S)]^{-1} = \frac{(n-1)S + \nu\Lambda}{n-1+\nu},$$

our moderated t-statistic is

$$\tilde{t} = n^{1/2}\tilde{S}^{-1/2}\overline{X}.$$

Finally, $O = \frac{P(I=1 \mid data)}{P(I=0 \mid data)} = \left(\frac{p}{1-p}\right)\frac{P(\tilde{t} \mid I=1)}{P(\tilde{t} \mid I=0)}$

is an increasing function of $\tilde{T}^2 = \tilde{t}'\tilde{t}$.

We write $MB = log_{10}O$ for our multivariate B-statistic. 44

Likelihood Ratio statistic

For the likelihood ratio (LR) test, we simply test the null H against the alternative K in the usual way. We calculate:

$$LR = 2(l_K^{\max} - l_H^{\max}) = n\log(1 + \frac{n}{n-1}\overline{X}^T S^{-1}\overline{X})$$
$$= n\log(1 + T^2/(n-1))$$

where S is assumed non-singular. Here T^2 is Hotelling's statistic. In our case, n < k and S is singular. If we plug in $\tilde{S}$, our moderated S, we get the moderated Hotelling statistic, $\tilde{T}^2$, just seen. 45

Hyperparameter estimation

There are $k(k+1)/2 + 3$ parameters in the prior: Λ, p, ν, and η.

We simply choose $p = 0.02$, although clearly more could be done here. Neither p nor η enter into $\tilde{T}^2$.

Estimates of the hyperparameters ν *and* η are developed using the univariate approach of Smyth (2004): η using the $p/2$ genes with the highest $\tilde{T}^2$ values, and ν using all the genes. We omit the details.

Λ is estimated by the method of moments using the formula

$$E(\mathbf{S}) = (\nu-k-1)^{-1}\nu\Lambda.$$

46

Illustrative results for our At experiment

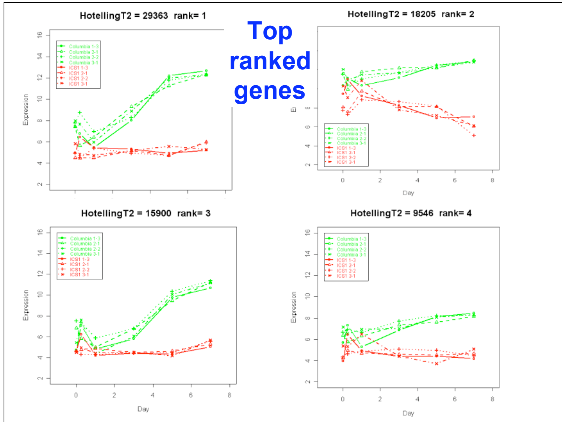
Estimate of Λ for the *At* experiment

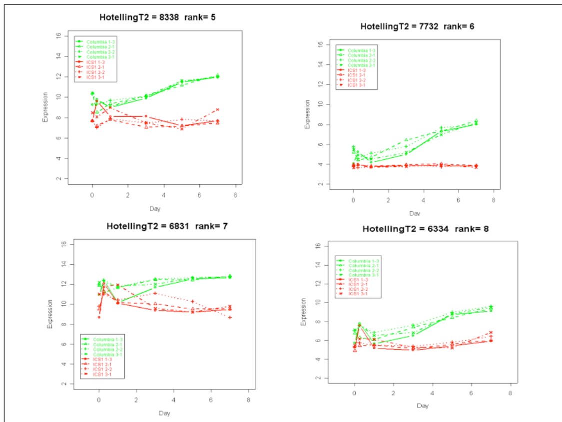
100×SD: 14, 17, 15, 13, 16, 16.

Correlation matrix

1.00					
.15	1.00				
-.01	.15	1.00			
.12	.07	.13	1.00		
-.09	-.01	.02	-.02	1.00	
.05	.06	.02	.15	-.16	1.00

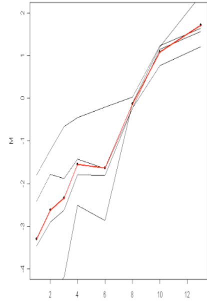
48





Illustrative results for our OPC/OL experiment

Evidence of autocorrelation

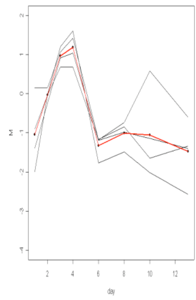


One gene's sample covariance matrix:

$$
\begin{matrix}
2.67 & 2.40 & 2.36 & 1.31 & 1.63 & 0.04 & -0.11 & -0.69 \\
2.40 & 2.17 & 2.14 & 1.20 & 1.48 & 0.05 & -0.10 & -0.61 \\
2.36 & 2.14 & 2.18 & 1.25 & 1.57 & 0.07 & -0.04 & -0.53 \\
1.31 & 1.20 & 1.25 & 0.74 & 0.93 & 0.06 & 0.00 & -0.26 \\
1.63 & 1.48 & 1.57 & 0.93 & 1.19 & 0.07 & 0.02 & -0.31 \\
0.04 & 0.05 & 0.07 & 0.06 & 0.07 & 0.01 & 0.02 & 0.02 \\
-0.11 & -0.10 & -0.04 & 0.00 & 0.02 & 0.02 & 0.05 & 0.08 \\
-0.69 & -0.61 & -0.53 & -0.26 & -0.31 & 0.02 & 0.08 & 0.25
\end{matrix}
$$

If $\boldsymbol{S}=\boldsymbol{UDV}'$, $\boldsymbol{D}= diag(8.88, 0.37, 0.02, 0, 0, 0, 0, 0)$

Further evidence of autocorrelation



Another gene's sample covariance matrix:

$$
\begin{matrix}
.83 & .10 & -.04 & -.05 & .19 & .24 & 1.02 & .69 \\
.10 & .03 & .01 & .03 & .00 & .01 & .14 & .04 \\
-.04 & .01 & .05 & .09 & -.04 & -.03 & .00 & -.07 \\
-.05 & .03 & .09 & .17 & -.08 & -.06 & .04 & -.11 \\
.19 & .00 & -.04 & -.08 & .09 & .09 & .19 & .21 \\
.24 & .01 & -.03 & -.06 & .09 & .11 & .27 & .26 \\
1.02 & .14 & .00 & .04 & .19 & .27 & 1.31 & .80 \\
.69 & .04 & -.07 & -.11 & .21 & .26 & .80 & .66
\end{matrix}
$$

If $\boldsymbol{S}=\boldsymbol{UDV}'$, $\boldsymbol{D}= diag(2.80, 0.37, 0.07, 0, 0, 0, 0, 0)$

53

Average single gene covariance matrix

$$
\bar{S} = \begin{matrix}
.10 & .06 & .05 & .04 & .03 & .03 & .03 & .02 \\
.06 & .11 & .06 & .05 & .04 & .04 & .04 & .03 \\
.05 & .06 & .11 & .05 & .04 & .04 & .04 & .03 \\
.04 & .05 & .05 & .09 & .04 & .04 & .04 & .03 \\
.03 & .04 & .04 & .04 & .09 & .04 & .04 & .03 \\
.03 & .04 & .04 & .04 & .04 & .10 & .05 & .04 \\
.03 & .04 & .04 & .04 & .04 & .05 & .09 & .04 \\
.02 & .03 & .03 & .03 & .03 & .04 & .04 & .07
\end{matrix}
$$

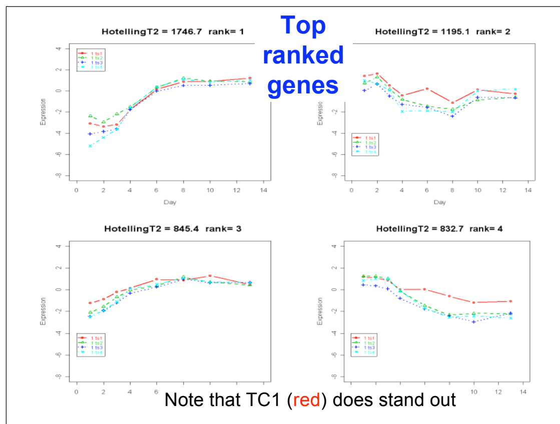
$$
\bar{S}^{-1} = \begin{matrix}
14.6 & -4.6 & -3.1 & -1.5 & -0.7 & -0.3 & -0.1 & -0.3 \\
-4.6 & 15.8 & -3.3 & -3.0 & -1.4 & -1.1 & -1.3 & -0.4 \\
-3.1 & -3.3 & 16.4 & -3.1 & -1.4 & -1.5 & -1.4 & -1.7 \\
-1.5 & -3.0 & -3.1 & 18.3 & -3.0 & -2.1 & -2.1 & -1.5 \\
-0.7 & -1.4 & -1.4 & -3.0 & 16.9 & -2.5 & -3.0 & -1.0 \\
-0.3 & -1.1 & -1.5 & -2.11 & -2.5 & 17.6 & -4.8 & -4.2 \\
-0.1 & -1.3 & -1.4 & -2.1 & -3.0 & -4.8 & 21.0 & -5.9 \\
-0.3 & -0.4 & -1.7 & -1.5 & -1.0 & -4.3 & -5.9 & 22.1
\end{matrix}
$$

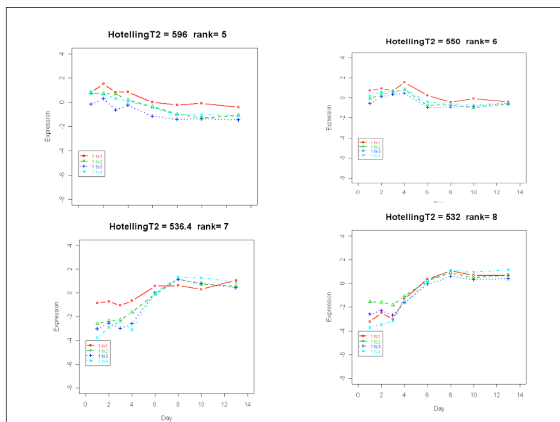
If $\bar{\boldsymbol{S}}=\boldsymbol{U\Lambda V}^T$, $\Lambda = diag(.38\ .09\ .06\ .05\ .05\ .05\ .04\ .04)$,
$\boldsymbol{u_1} = (.34\ .41\ .40\ .37\ .32\ .37\ .34\ .27)^T$

54

Observations on autocorrelation

The correlation structure exhibited in the average covariance matrix resembles that of a slow moving average process, which is perhaps not surprising given the way in which the samples of cells were taken and the use of a common reference mRNA source.





Conclusions

- Methods which rank genes (e.g. the *MB* statistic or the moderated Hotelling T^2) perhaps provide easier access to genes whose absolute or relative expression varies over time, than do multi-gene methods (e.g. cluster analysis).
- Among the single-gene methods, *MB* performs no worse than other methods in both real data and simulated data comparisons, and better than the *F*.
- The Hotelling T^2 statistic is a viable alternative to *MB*, but we still need the moderated ***S***.
- The *MB* statistic may be able to select interesting genes which are missed by other methods.

58

An important new paper

A Significance method for time Course Microarray Experiments Applied to Two Human Studies,

JD Storey, JT Leek, W Xiao, JY Dai and RW Davis.

University of Washington Biostatistics Working Paper Series Paper 232, 2004.

Brief Summary, 1

Method developed for two human studies, both using the Affymetrix human U133A and 133B chips.

Endotoxin study, monitoring gene expression responses to bacterial endotoxin in blood leukocytes. Four subjects were administered endotoxin, another four a placebo, and blood samples were taken at 2, 4, 6, 9, and 24 hours after infusion.

Kidney aging study, to investigate changes in gene expression in the human kidney across different agess. Samples from normal kidney tissue removed at nephrectomy or renal transplant biopsy from 72 patients with ages ranging from 27 to 92 years.

60

Brief Summary, 2

- The model used on each case has the following form for gene i on individual j at time t:

$$y_{ij}(t) = \mu_i(t) + \gamma_{ij}(t) + \varepsilon_{ij}(t).$$

where the population average curve is $\mu_i(t)$, individuals deviate from the population average curve by $\gamma_{ij}(t)$, *and* measurement error and the remaining sources of variation are modelled by the $\varepsilon_{ij}(t)$. It is the $\gamma_{ij}(t)$ which distinguishes this model from the ones we previously considered for model organisms with more repeatable expression profiles.

The observations are at times t_{ij}, and the $\mu_i(t)$ term is modeled by cubic splines.

61

Software availability

Programs implementing our multivariate methods will go into the open source **R**-based **Bioconductor** package before the end of this summer.

Available programs for some other approaches are listed in the handout for this afternoon's workshop.

Acknowledgements

Yu Chuan Tai, UC Berkeley

Mary Wildermuth, UC Berkeley

Jason Dugas, Stanford

Moriah Szpara, UC Berkeley & members of the Ngai lab

Gordon Smyth, WEHI

Monica Nicolau, Stanford University

63

Hybridizations for the OPC/OL experiments

The cDNA slides were made in the Ngai lab, UC Berkeley, using the RIKEN clone set, and the hybs done in '02/'03.

19,200 spots/slide, in 8x4 print-tip groups of size 25x24.

Some genes were replicated: their $M = log_2 R/G$ were averaged.

Two dye-swap technical replicate slides run on mRNA from each biological replicate: their M and $-M$ were averaged.

Time course (TC) 1 was done using slides from one batch, while TC 2-4 used slides from another batch.

The raw intensities were from an Axon scanner; the image analysis was done by Spot using a morph background.

Normalization was by print-tip lowess , followed by between array MAD scale normalization for TC1, as there was a lot of variation across time in this replicate.

64

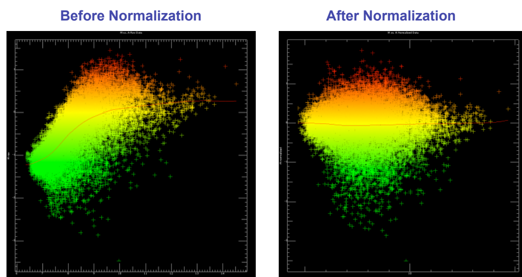
OPC/OL experiment: hybridization dates

	Rep1	Rep2
TC1	4/4/02 Cy5=pool, Cy3=time	4/11/02 Cy5=time, Cy3=pool
TC2	12/3/02 Cy5=pool, Cy3=time	1/14/03 Cy5=time, Cy3=pool
TC3	12/10/02 Cy5=time, Cy3=pool	2/10/03 Cy5=pool, Cy3=time
TC4	2/6/03 Cy5=pool, Cy3=time	2/8/03 Cy5=time, Cy3=pool

65

MA-plot for one hybridization in the OPC/OL experiment

Before Normalization   After Normalization



vertical axis: $M = log_2 R/G$, horizontal axis: $A = log_2 \sqrt{RG}$

66

QA/QC in the At experiments

In the At experiment, we checked the quality of all chips using `fitPLM()` in `AffyExtensions`. We found that 5 chips were of low quality, and these were repeated.

In addition the log_2 intensities of replicate 1-3 wt day 3 sample were inconsistent with those from the other wt experiments for that day, despite having no obvious QC problems with the chips. These were "adjusted" using median polish on all the wt data.

67

QA/QC in the OPC/OL experiments

Here quality was a greater concern, no doubt as a result of the wide spread of times over which the hybridizations were conducted. Also, the analysis was done some time later, and there was no possibility of repeating any of the hybridizations.

It turned out TC1 (data from a different chip batch) did stand out from the rest, but omitting this replicate was not an option, as there were concerns about aspects of the other hybridizations as well: attenuated response range.

In the end, we relied on visual examination of consistency of responses, and qrt-pcr follow-up to give us confidence in our conclusions.

68