Differential expression analysis of microarray experiments

Bioconductor 2007

Gordon Smyth

Walter and Eliza Hall Institute

Getting started

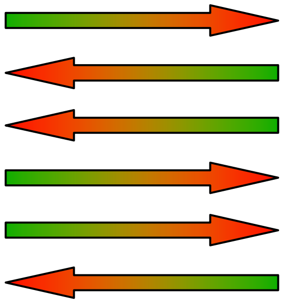
- Copy the directory ‘bioc2007limma’ from the flashdisk to a convenient place on your computer, e.g., c:/bioc2007limma
- Open c:/bioc2007limma/html/index.html in your browser
- Make c:/bioc2007limma/data the working directory of your R session

limma package documentation

- Function help pages
- Class help pages
- Group help pages
- User’s Guide

Example 1:
Integrin beta7+ vs beta7–

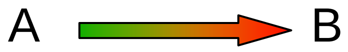
beta7-



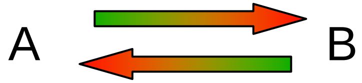
beta7+

- Reading two-color data
- Control spots
- Background correction
- Dye-swaps
- Empirical Bayes differential expression

Designs → Linear Models

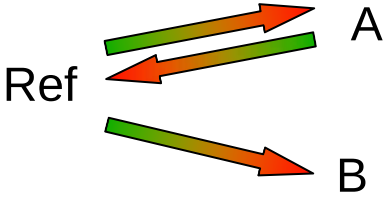


$$y = \log_2(R/G) \equiv B - A$$



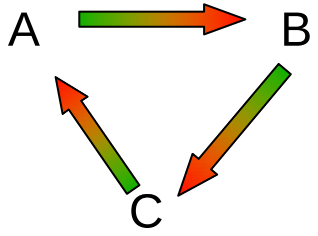
$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \beta$$

$$\beta \equiv B - A$$



$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

$$\beta_1 \equiv A - \mathrm{Ref}$$
$$\beta_2 \equiv B - A$$



$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

$$\beta_1 \equiv B - A$$
$$\beta_2 \equiv C - A$$

Linear Model Estimates

Obtain a linear model for each gene *g*

$$E(y_g) = X b_g$$

$$\mathrm{var}(y_g) = W_g^{-1} s_g^2$$

Estimate models to get

coefficients	$\hat{b}_{gj}$
standard deviations	s_g
standard errors	$\mathrm{se}(\hat{b}_{gj})^2 = c_{gj} s_g^2$

Hierarchical model for variances

Data $$s_g^2 : s_g^2 \frac{c^2_{d_g}}{d_g}$$

Prior $$\frac{1}{s_g^2} : s_0^2 \frac{c^2_{d_0}}{d_0}$$

Posterior $$E\left(\frac{1}{s_g^2} \mid s_g^2\right) = \frac{d_0 + d_g}{s_0^2 d_0 + s_g^2 d_g}$$

Posterior Statistics

Posterior variance estimators

$$\tilde{s}_g^2 = \frac{s_0^2 d_0 + s_g^2 d_g}{d_0 + d_g}$$

Moderated t-statistics

$$\tilde{t}_{gj} = \frac{\hat{b}_{gj}}{\tilde{s}_g \sqrt{c_{gj}}}$$

Baldi & Long 2001, Wright & Simon 2003, Smyth 2004

Exact distribution for moderated t

An unexpected piece of mathematics shows that, under the null hypothesis,

$$\tilde{t}_g \sim t_{d_0+d_g}$$

The degrees of freedom add!

The Bayes prior in effect adds d_0 extra arrays for estimating the variance.

Wright and Simon 2003, Smyth 2004

Hierarchical model for means

Data

$$\hat{b}_{gj} \;:\; N(b_{gj}, c_{gj} s_g^2)$$

Prior

$$P(b_{gj} \;{}^{1}\; 0) = p$$

$$b_{gj} \mid b_{gj} \;{}^{1}\; 0 : \; N(0, c_{0j} s_g^2)$$

Lönnstedt and Speed 2002, Smyth 2004

Posterior Odds

Posterior odds of differential expression

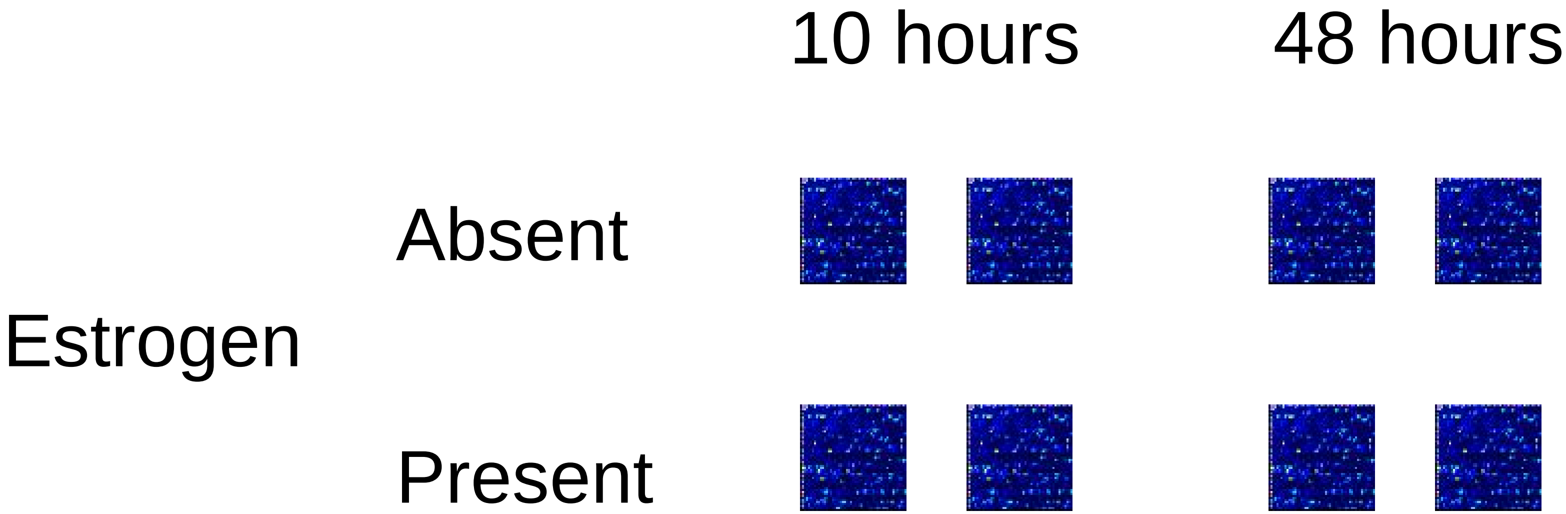
$$\frac{p(b \neq 0 \mid \hat{b}, s^2)}{p(b = 0 \mid \hat{b}, s^2)} = \frac{p}{1-p}\left(\frac{c}{c+c_0}\right)^{1/2}\left\{\frac{\tilde{t}^2 + d + d_0}{\tilde{t}^2 \frac{c}{c+c_0} + d + d_0}\right\}^{\frac{1+d+d_0}{2}}$$

Monotonic function of $|\tilde{t}|$

Hence $\tilde{t}$ gives the best possible ranking of genes

Lönnstedt and Speed 2002, Smyth 2004

Example 2: Estrogen



10 hours

48 hours

Estrogen

Absent

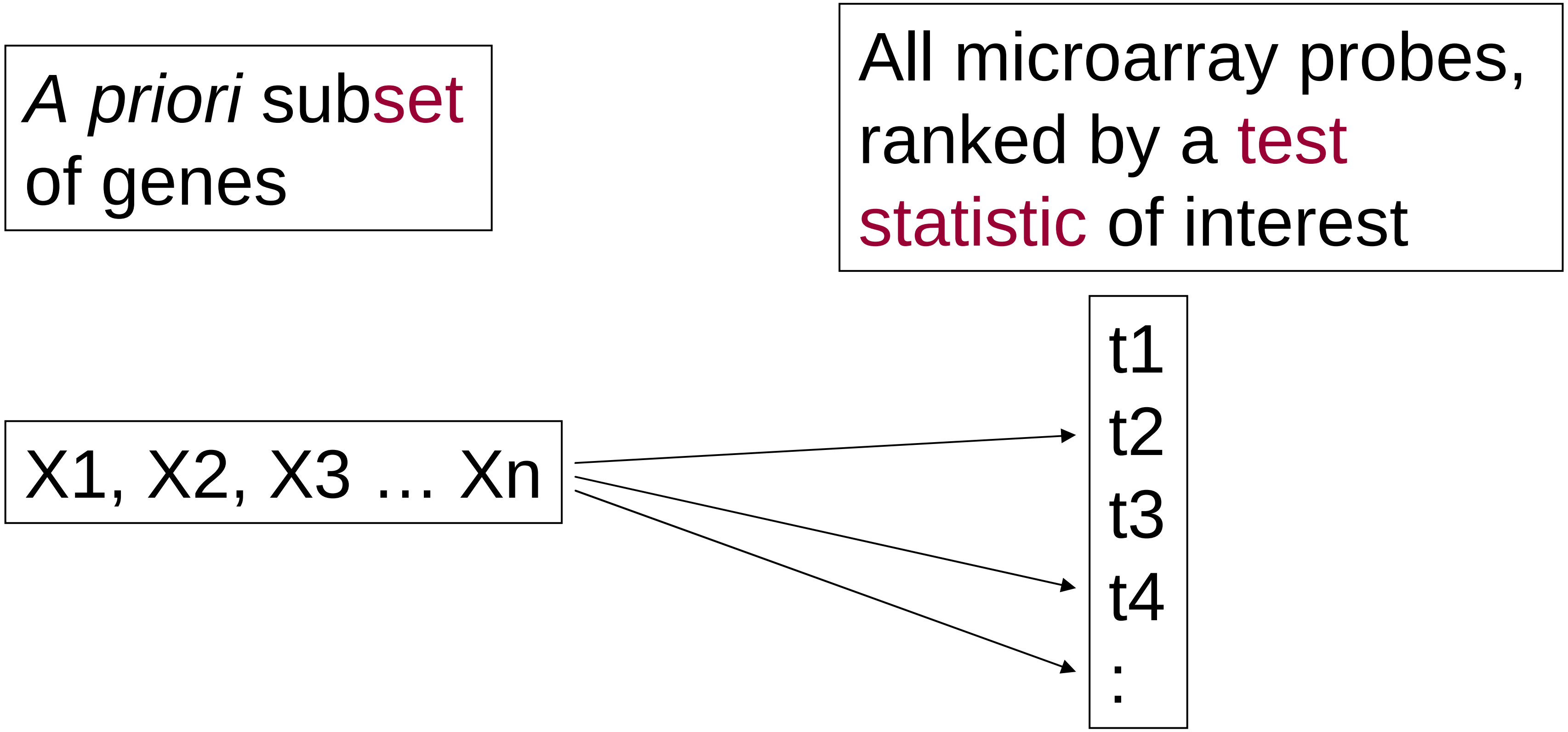
Present

- Reading Affymetrix data
- Factorial designs
- Gene set tests

Gene sets

- Test significance of a (prior specified) group of genes
- The genes might belong to a known pathway or might be the top genes from a related experiment
- The set might be significant even if individual genes are not
- Gene set enrichment analysis (GSEA) originated by Mootha et al PNAS 2003 and Subramanian et al PNAS 2005

Mean rank gene set tests



A priori subset of genes

All microarray probes, ranked by a test statistic of interest

X1, X2, X3 … Xn

t1
t2
t3
t4
:

Look for ranks for set genes amongst test statistics

Example 3:
Targets of SAHA and depsipeptide

Case Study

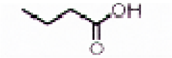
Peart, Smyth, van Laar, Richon, Holloway, Johnstone

Identification and functional significance of genes regulated by structurally diverse histone deacetylase inhibitors

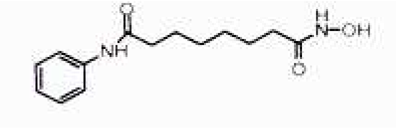
PNAS Feb 2005

Tumour cell growth inhibitors

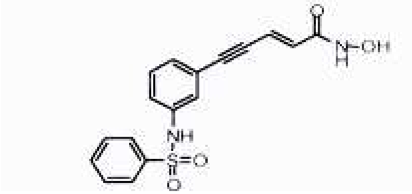
- Histone deacetylase inhibitors (HDACis) are anti-cancer agents that inhibit tumour cell growth and survival
- Not toxic to normal cells
- Genes active in biological effects are unknown



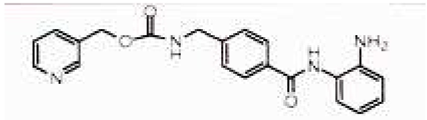
butyrate



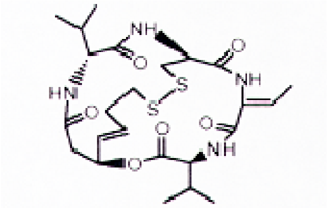
SAHA



oxamflatin



MS-27-275

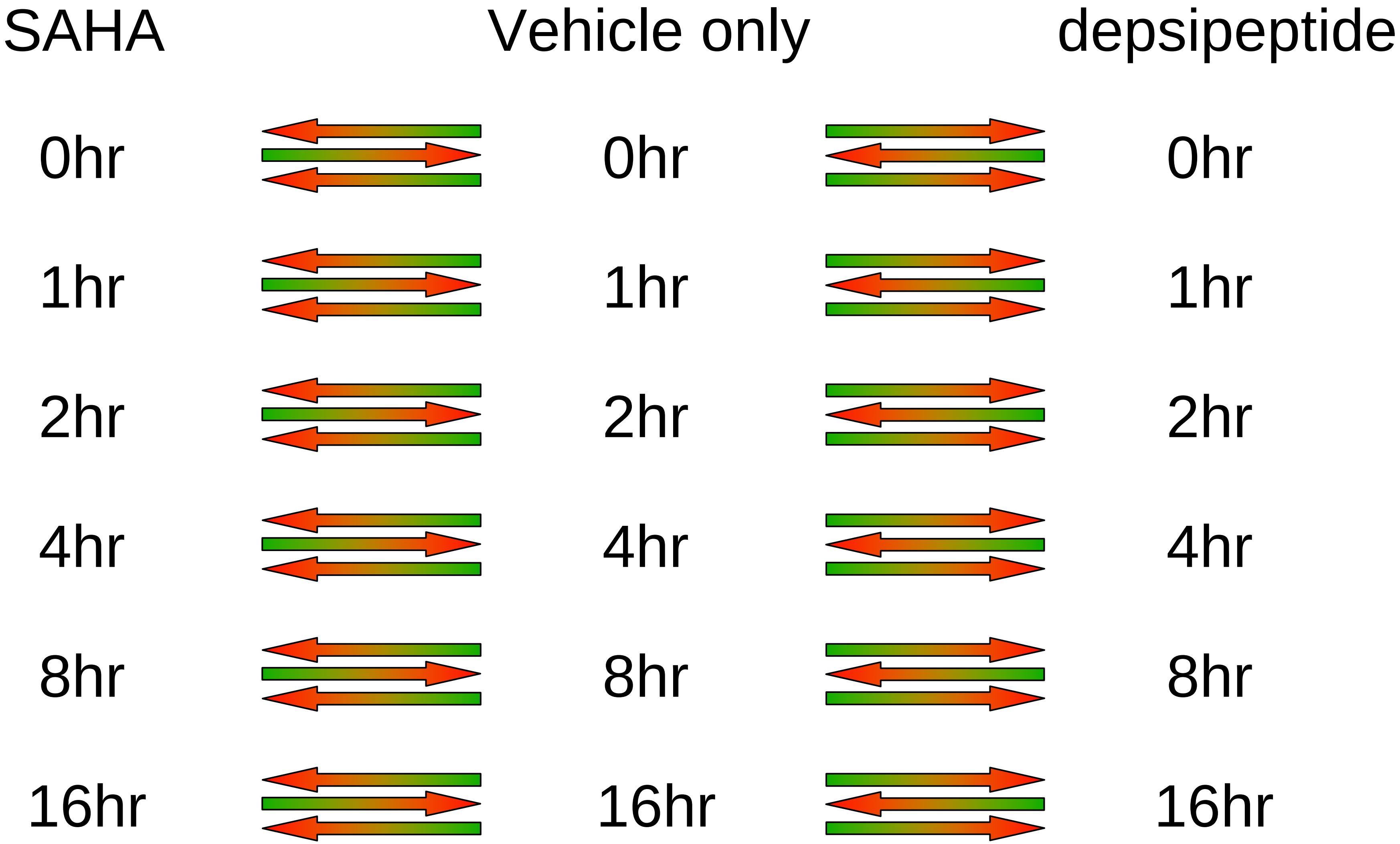


depsipeptide

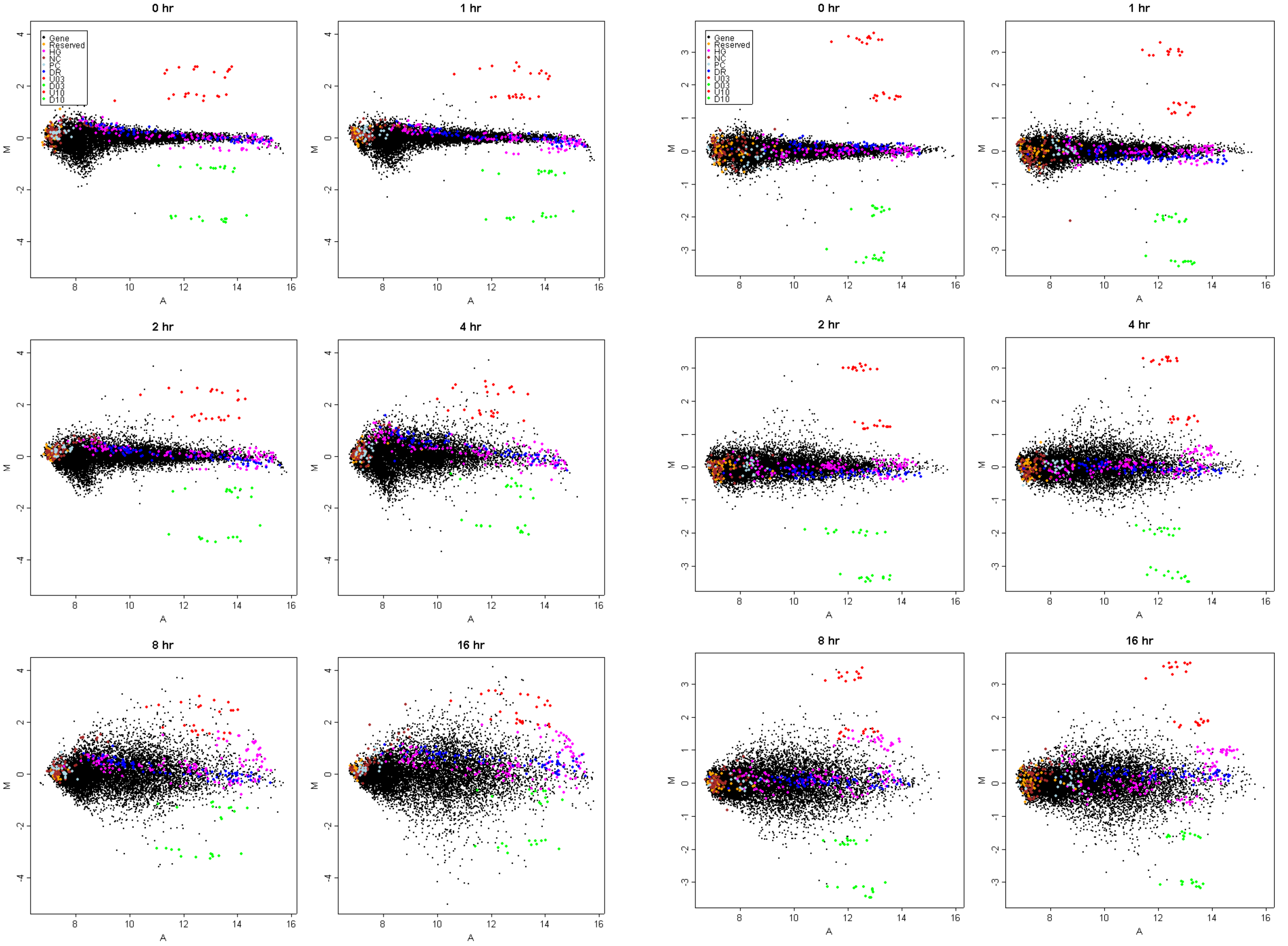
Target cell cultures

- Study effects of SAHA and depsipeptide on the acute T-cell leukemia cell line **CEM**
- **SAHA** and **depsipeptide** are structurally different but have similar biological effects (induce death through intrinsic apoptotic pathway)
- Prising out subtle differences is of great interest

Experimental design



SAHA	Vehicle only	depsipeptide
0hr	0hr	0hr
1hr	1hr	1hr
2hr	2hr	2hr
4hr	4hr	4hr
8hr	8hr	8hr
16hr	16hr	16hr



SAHA

depsipeptide

Aims of experiment

- Identify common responders: genes which respond similarly to SAHA and depsipeptide
- Identify specific responders: genes which respond to one of SAHA or depsipeptide, but not to the other
- Different responders, genes which respond to both SAHA and depsipeptide but differently, are of lesser interest

Classic ANOVA methods are applicable

- An F-test for time on 5 df will find genes which change at any time (simpler than a series of t-tests at each time)
- An F-test for drug x time interaction will find genes which react differently to the two drugs

Moderated F-Statistic

The idea of shrinking the variance extends immediately to multiple contrasts

Moderated F-statistic

$$\tilde{F}_g = \frac{\mathrm{MST}_g}{\tilde{s}_g} \sim F_{k,d_g+d_0}$$

MST=Mean Sum of squares between Treatments

Wright & Simon 2003, Smyth 2004

Linear model analysis

- Fit linear model to the M-values (log-ratios) for each gene
- Include effects for drug x time
- Allow for probe/drug specific dye-effects
- Treat each time series of 6 arrays as a randomized block, i.e., allow arrays hybridized together to be correlated

Classifying common and specific responders

Tests	Common	SAHA specific	depsi specific
Time (SAHA)	☻	☻	x
Time (depsi)	☻	x	☻
Drug x time interaction	x	☻	☻

☻ = significant, x = not significant

Acknowledgements

Peter MacCallum Cancer Centre

- Melissa Peart
- Ricky Johnstone

- Ryan van Laar
- Andy Holloway