

NuGO PPS1 mouse study 1: preliminary statistical analysis

M. Baccini · G. Tonini · A. Biggeri

Received: 8 October 2008 / Accepted: 12 November 2008 / Published online: 26 November 2008
© Springer-Verlag 2008

Abstract Time course experiments are aimed at characterizing the dynamic regulation of gene expression in biological systems. Data are collected at different time points to monitor the dynamic behaviour of gene expression. The NuGO PPS Mouse Study 1 investigates the development of high fat-induced insulin resistance (IR) over time in APOE*3Leiden (E3L) mice. The study consists in a series of analyses at time points, which are crucial in the development of central and peripheral IR. Affymetrix arrays have been made on critical organs. We present the results of the preliminary statistical analysis on these microarray data. We used a non-parametric approach to identify genes the expression of which changed over time, separately for three tissues: liver, muscle and white adipose tissue. We specified for each gene a basic ANOVA model, in order to check the null hypothesis that gene expression did not vary over time. We addressed the multiple tests problem calculating positive false discovery rate and q values for the F test statistics. The appropriateness of the hypothesis of homogeneous variances over time was investigated by mean of the Bartlett's test for homoscedasticity. This is a relevant point because heteroscedasticity could be indicative of outlying behaviour of some individuals at specific time points. The necessity to use a moderated F test was evaluated. We found that a considerable part of the genes varied expression over time. For part of the genes, the variance of the response was not homogeneous over time. Response differed by tissue.

Keywords Microarray experiments · Time course · ANOVA

Introduction

Time course experiments are aimed at characterizing dynamic regulation of gene expression in biological systems. In time course experiments array data are collected at different time points to monitor the dynamic behaviour of gene expression. They necessitate the development of approaches to identify genes that show changes in temporal patterns of expression among varying biological conditions. The statistical challenge is that there are few replicates for each time point, but a large number of genes. The ultimate goal is to identify time patterns in gene regulation, which implies the identification of dynamic genetic regulatory networks.

In this paper, we present a preliminary analysis of the time course experiment NuGO PPS Mouse Study 1. The general aim of the study is to establish the time course of fat mass increase and insulin resistance (IR) development in liver, adipose tissue and muscle using a humanized IR model: the APOE*3Leiden (E3L) mouse fed a high fat diet (HFD). The study resulted to be a series of analyses at time points which are crucial in the development of central and peripheral IR. Affy arrays were made on tissue from the three organs. Here, we want to establish if genes expression varies by time course and, in particular, which genes are changing.

Data

The experiment consisted in feeding 40 mice with a high fat diet. After 0, 1, 6, 9, and 12 weeks of HFD intervention,

M. Baccini · G. Tonini (✉) · A. Biggeri
Department of Statistics, University of Florence,
V.le Morgagni 59, 59100 Florence, Italy
e-mail: g.tonini@ispo.toscana.it

M. Baccini · G. Tonini · A. Biggeri
Biostatistics Unit, ISPO, Via Cosimo il Vecchio 2,
59100 Florence, Italy

E3L mice were sacrificed ($n = 8$ per time point) and liver, white adipose tissue and skeletal muscle collected. Due to technical problems some sample of muscle and white adipose tissue were not available for the analysis. In the end, the data for 37 mice for muscle tissue and 38 for white adipose tissue were available. RNA was isolated, labelled and hybridized to Affymetrix MOE430-2.0 microarrays. Microarray data consisted in gene expressions from the three different tissues (liver, muscle and white adipose tissue) at the five time points. Microarrays that passed the quality control were normalized using *gc-rma* and annotated into *Entrez gene-ids* using the custom MBNI CDF-file, version 9.0.1. The number of genes was 12,251 for liver, 10,776 for muscle and 12,492 for white adipose tissue.

Methods

The first point of our analysis was to understand if gene expression varied by time course, and in particular which genes were changing. To answer this question, we performed a one-way ANOVA for each gene, separately for each tissue. It should be stressed that in this analysis, gene expression measurements at different time points were independent. We tested the null hypothesis that the average level of expression was constant over time, comparing the between time variance with the error variance. The resulting F test statistics was:

$$F = \frac{D_{\text{between}}/(T-1)}{D_{\text{error}}/(N-T)} = \frac{\sum_{t=1}^T n_t (\bar{g}_t - \bar{g})^2 / (T-1)}{\sum_{t=1}^T \sum_{i=1}^{n_t} (g_{it} - \bar{g}_t)^2 / (N-T)},$$

where T is the number of time points ($T = 5$), n_t is the number of animals at the t th time point and N is the total number of animals; g_{it} indicates the gene expression level for the i th animal at the t th time point, \bar{g}_t is the mean at the t th time point and \bar{g} is the overall mean. The F statistics under the null hypothesis followed a Fisher's F distribution with $T - 1$ and $N - T$ degrees of freedom.

We addressed the multiple tests problem, arising from simultaneously testing a large number of independent hypotheses, following the approach proposed by Storey et al. [3]. This approach is based on positive false discovery rate (pFDR) and q values calculation. Let π_0 be the a priori probability of the null hypothesis to be true. The pFDR is defined as:

$$\text{pFDR}(c) = \Pr(H_0 | F \geq c) = \frac{\pi_0 \Pr(F \geq c | H_0)}{\Pr(F \geq c)}$$

It represents the probability for the null hypothesis to be true given that the observed value of the F statistics is

larger than a specific critical value c , which defines the rejection region. The probability π_0 can be estimated from the empirical distribution of the observed p values (for detail see 3).

The q value is a quantity which can be calculated for each gene. It is defined as

$$q(f) = \inf_{c \geq f} \{\Pr(H_0 | F \geq c)\},$$

where f is the observed F statistics for a specific gene. It represents the probability that the null hypothesis is true (proportion of false-positives) when that particular test is called significant.

As a comparison, we also calculated the Benjamini–Hochberg FDR [2] which corresponds to the Storey q values when π_0 is set equal to one (3).

The ANOVA model relies on the homoscedasticity assumption. In our case, we assumed that the gene expressions variance was constant over time. Heteroscedasticity could be indicative that gene expression patterns over time are heterogeneous among mice, with possible outlying “response” profiles. To evaluate the homoscedasticity assumption, we performed for each gene the Bartlett's test for homogeneity of variances. pFDR and q values associated to the gene-specific test statistics were calculated.

In the presence of small sample size, error variances in the denominator of the F test statistics can be poorly estimated. This can lead to false-positives if error variances are severely underestimated and to false-negatives if error variances are severely overestimated.

In order to avoid these problems, variance-stabilized versions of the F test can be used (see for example [1, 4]). In our analysis, we evaluated the need of using a stabilized approach inspecting the relationship between the error variance estimates and the corresponding F test statistics.

Results

Liver tissue

A strong deviation from the expected uniform distribution of p values under the null hypothesis was found (Fig. 1). This indicates that a large proportion of genes varied over time. According to the Storey's approach, our estimate of the a priori probability π_0 was 0.536.

As shown in the QQ plot of the gene-specific F statistics, many genes deviated from the null hypothesis (Fig. 2). In particular, 197 genes resulted significant at a pFDR of 0.001, 673 at a pFDR of 0.01 and 3548 at a pFDR of 0.1 (Table 1). Using the Benjamini–Hochberg (BH) FDR approach, which implicitly assumes a prior belief of zero true discoveries, we found 150 genes declared significant at

Fig. 1 Histogram of p values of F test for liver, muscle and white adipose tissue

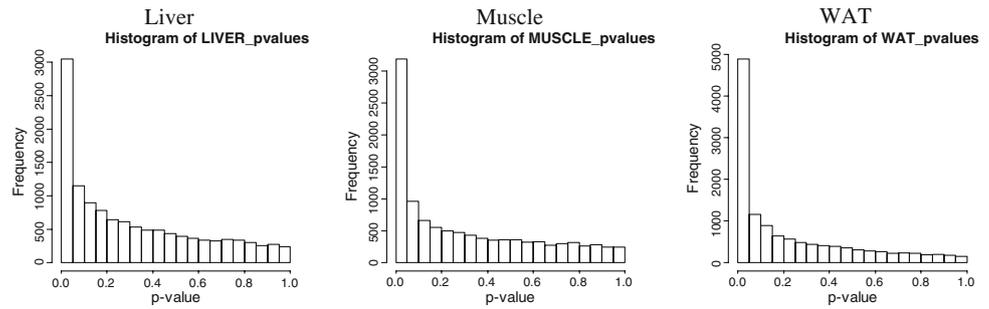


Table 1 Number and percentage of significant genes at different pFDR levels

	Liver			Muscle			WAT		
	q value			q value			q value		
	<0.1	<0.01	<0.001	<0.1	<0.01	<0.001	<0.1	<0.01	<0.001
# Genes	3,548	673	197	3,742	1,132	400	8,182	2,869	971
Percent (%)	29	5.5	1.6	34.7	10.5	3.7	65.5	22.9	7.8

BH FDR = 0.001, 400 at BH FDR = 0.01 and 1741 at BH FDR = 0.1.

For each gene we tested for homoschedasticity using the Bartlett’s chi square test statistic ($T - 1 = 4$ degrees of freedom). Overall, the test resulted significant for 59 genes at pFDR = 0.001, 95 at pFDR = 0.01 and 260 at pFDR = 0.1. This indicates that the response pattern over time was different among animals only for few genes (Fig. 3).

Figure 4 reports F statistics versus error variance on a double log scale. The green line corresponds to a pFDR of 0.01 for the F test. There is no evidence of an excess of false-negatives for large values of the estimated error variances, while there is a small evidence of an excess of false-positives for small values of the estimated error variance.

Muscle tissue

Results for muscle tissue showed a stronger evidence of deviation from the expected uniform distribution of p values under the null hypothesis (Fig. 1), which means a

larger proportion of genes varying over time. Our estimate of π_0 was 0.455.

We found that 400 genes resulted to be significant at a pFDR of 0.001; 1,132 at a pFDR of 0.01 and 3,742 at a pFDR of 0.1. Using the Benjamini–Hochberg FDR approach we found 284 genes declared significant at BH FDR = 0.001, 750 at BH FDR = 0.01 and 2,433 at BH FDR = 0.1 (Table 1; Fig. 2).

As for liver, we evaluated the homoschedasticity assumption using the Bartlett’s χ^2 statistic. The number of genes for which the homoschedasticity assumption was not valid was very low (21 genes at pFDR = 0.001, 85 at pFDR = 0.01 and 246 at pFDR = 0.1) (Fig. 3). There was no evidence of an excess of false-negative or -positive results for extreme values of the estimated error variance (Fig. 4).

White adipose tissue

White adipose tissue showed the largest proportion of genes varying over time (Fig. 1). Our estimate of π_0 was 0.260.

Fig. 2 QQ plot of empirical versus theoretical distribution of F statistics for liver, muscle and white adipose tissue. Genes resulting significant at different pFDR levels are shown in different colours. Bisector line (solid) and line which passes through the first and third quartiles (dashed) are added

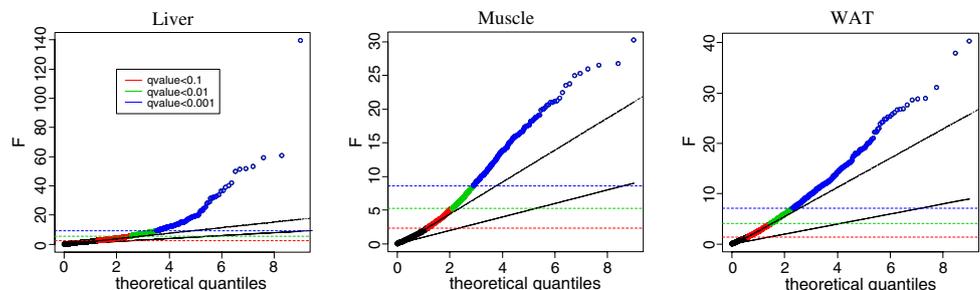


Fig. 3 QQ plot of empirical versus theoretical Bartlett's χ^2 statistics distribution for liver, muscle and white adipose tissue. Solid line passes through the first and third quartiles

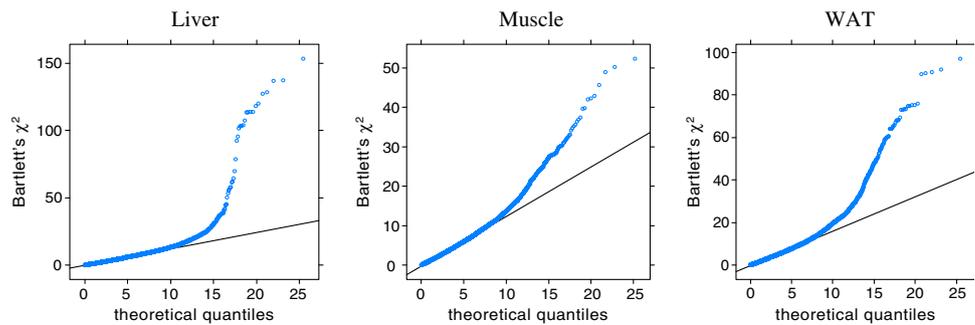
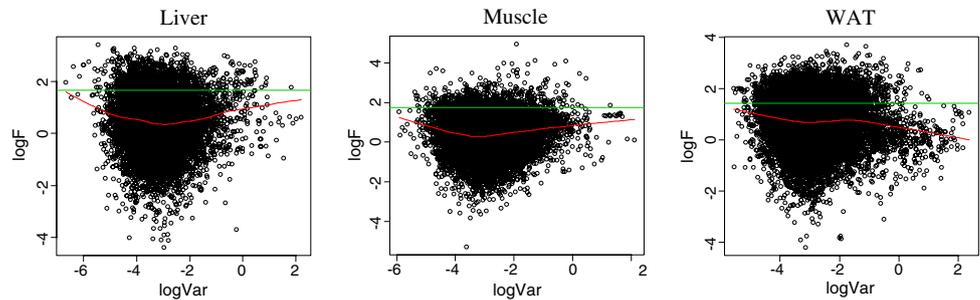


Fig. 4 Plot of F statistics versus error variance estimates on a double log scale for liver, muscle and white adipose tissue. The red line is a regression spline showing the relationship between F statistics and error variance. The green line corresponds to a q value of 0.001 for the F test



Nine hundred and seventy-one genes resulted significant at a pFDR of 0.001; 2,869 at a pFDR of 0.01 and 8,182 at a pFDR of 0.1. With the Benjamini–Hochberg FDR approach we found 483 genes declared significant at BH FDR = 0.001; 1,560 at BH FDR = 0.01 and 4,383 at BH FDR = 0.1 (Table 1, Fig. 2).

For white adipose tissue the homoscedasticity assumption was violated for a larger number of genes (224 genes at pFDR = 0.001, 569 at pFDR = 0.01 and 2,001 at pFDR = 0.1) (Fig. 3). This could be indicative that for some genes the response pattern over time was different among mice.

We found certain evidence that the F statistics was lower than the expected for large values of the estimated error variance (Fig. 4). In this case a moderated approach based on the addition of a positive term to the variance could be appropriate.

Discussion

Our results indicate that in all three tissues, there was a large percentage of genes the expression of which varied by time course. For some of them we observed a different variability over time. This can be indicative of heterogeneity of the underlying (and unobservable) mouse-specific patterns of expression over time, with mice that “respond”

to the diet at different times. Next step of the analysis will be to verify if adjustment for relevant covariates (for example insulin resistance) can partly explain the observed heteroscedasticity.

In order to evaluate robustness of our results, sensitivity analysis could be performed using variance-stabilized F statistics to avoid problems related to poor estimation of the error variance.

In this preliminary analysis, we detected the genes whose expression changed over time, but we did not classify genes according their expression pattern. This is a relevant point and it will be addressed in future analyses.

References

1. Baldi P, Long AD (2001) A Bayesian frame work for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes”. *Bioinformatics* 17(6):509–519
2. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: a practical and powerful approach to multiple testing. *J R Stat Soc* 57(1):289–300
3. Storey JD, Xiao W, Leek JT, Tompkins RG, Davis RW (2005) Significance analysis of time course microarray experiments. *PNAS* 102(36):12837–12842
4. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarray applied to the ionizing radiation response. *PNAS* 98:5116–5121