Web-based Supplementary Materials for "Tail posterior probability for inference in pairwise and multiclass gene expression data" by N.Bochkina and S.Richardson

Web Appendix A

Microarray Experiment

The data considered in this paper derives from the sets of experiments on the H2Kb muscle cell line of mouse (Tatoud et al., 2006), as a part of BAIR project (http://www.bair.org.uk), studying reaction of cells to insulin to advance further in understanding the development of diabetes. The cells were treated with insulin and metformin, an insulin-replacement drug, and observed after the treatment at two time points, 2 and 12 hours. For comparison, non-treated cells were grown and collected at the time of treatment (0 hours) and at 2 and 12 hours. This experiment has 3 biological replicates for each of the seven conditions. RNA from each of the 21 samples was hybridised on an Affymetrix MOE430A chip (a single channel array), making a total of 21 microarrays, with 22690 genes per chip. The aim of the experiment was to find genes changing their expression in the cells due to treatment by insulin or metformin. The data we use below was background-corrected and the gene expression index was calculated using RMA software (Irizarry et al., 2003) implemented in R (http://www.r-project.org) with intensity-dependent loess normalisation which results in data being transformed on the $\log_2$ scale. For biological interpretation of the results in this and the previous sections see Tatoud et al.(2006).

Web Appendix B

Sensitivity to the prior distribution of b.

In this section we discuss sensitivity to the choice of prior for the hyperparameter b, the scale parameter of the inverse gamma prior distribution of the variances $\sigma_g^{-2} \sim \Gamma(a, b)$ in model (1).

To choose the prior family for b, we follow suggestions in Gelman (2006). There is no direct equivalence between the random effect model considered by Gelman (2006), and the hierarchical model for the variance σ_g^2 considered here. The commonality is that in both cases, a non-informative prior distribution for inverse variance of a random effect model in the former case and precision parameter b in our case is being sought. Therefore, since Gelman (2006) recommends a uniform prior on the standard deviation, the equivalent of this for b here is $1/\sqrt{b}$. Another issue in support of this choice is that if we consider a non-informative prior for b with singularity at $b = 0$ (e,g, $f(b) = 1/b$), it is easy to show that this will result in improper posterior, also with singularity at $b = 0$. Therefore, we want to exclude a small neighbourhood of $b = 0$, i.e. we want to bound $1/b$ (and thus $1/\sqrt{b}$) from above. These arguments lead us to the following prior: $1/\sqrt{b} \sim U[0, u]$. Note that this is equivalent to the Pareto prior on b: $f(b) = k\epsilon^k b^{-k-1} I(b > \epsilon)$ with parameter $k = 1/2$ and $\epsilon = 1/u^2$.

One needs to take care in choosing u. If u is too small, we may exclude values of b supported by the likelihood, and if u is too large it may lead to negative "bias" in the posterior estimate of b (see discussion in Gelman, 2006). Therefore, we consider several values of u and compare the resulting posterior distributions.

To choose a data-driven value of ϵ (and thus of u), we used the following heuristics. Since $b = (a-1)E(s_g^2 \mid a, b)$, we can choose the lower bound on b proportional to $(a-1)\min_g s_g^2$, with a plug-in value of a. One way is to plug-in a reasonably small value of a, e.g. $a = 2$, and another way is to plug-in an estimate of a based on the distribution $f(s_g^2 \mid a, b)$, e.g. a moment estimator. After obtaining the posterior distribution of b, it is always useful to check if its density is high around the value $1/u^2$ which implies that the chosen value of u was too small.

For the H2Kb data, $\min_g s_g^2 = 0.00169$, and for the simulated data $\min_g s_g^2 = 0.00024$, and we take $u = (\min_g s_g^2)^{-1/2}$, which is $u_{H2Kb} = 24.4$ for the H2Kb data and $u_{sim} = 64.5$ for the simulated data. To check sensitivity to the choice of u, we also consider other values of u, roughly twice and half of the selected values.

We applied this hyperprior $1/\sqrt{b} \sim U[0, u]$ to H2Kb muscle cell data with $u = 14.4, 24.4, 48$ and to the simulated data with $u = 30, 64.5, 129$, together with the priors for the other parameters for the multiclass model as specified in Section 5. We compare posterior distributions of b, posterior estimates of σ_g^2 and the tail posterior probabilities.

Comparison of the posterior distributions of b on the simulated data shows that for all prior distributions, the true value of $b = 0.05$ is well recovered, since it is close to the point of maximum of the posterior densities (see Web Figure 1a). Note that for $u = 30$ and $u = 64.5$ the estimated posterior densities of b coincide. For the H2Kb data, we observe that the posterior distributions of b for all priors are similar (Web Figure 1b), and for $u = 24.4$ and $u = 48$ the estimated posterior densities of b coincide. It is not surprising that the choice of prior distribution for b does not affect much its posterior distribution since there are thousands of values of s_g^2 which contribute to estimating 2 parameters, a and b.

Comparison of the posterior estimates of σ_g^2 (Web Figure 2) and of the tail posterior probabilities (Web Figures 3 and 4) shows very good agreement, as they all lie on the diagonal and the difference between the values is comparable to MCMC errors of the estimates (the figures of the posterior probabilities are given only for two comparisons, insulin vs control at 2 and at 12 hours).

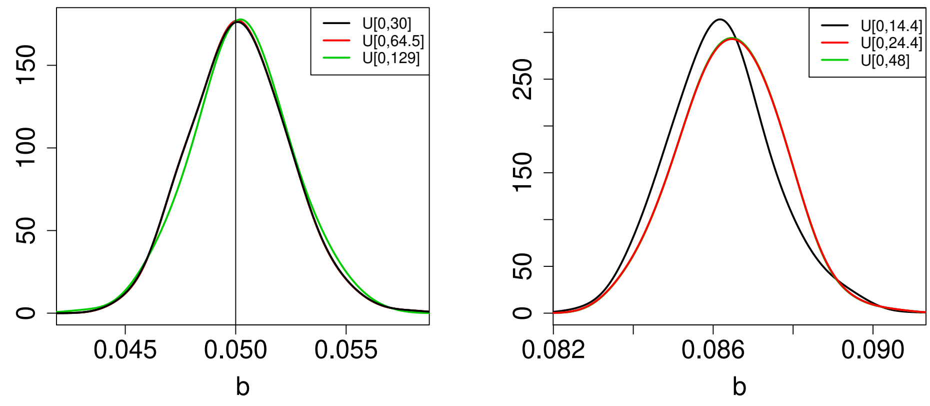
We also checked the sensitivity to other families of prior distributions for b, such as gamma distributions and folded t distributions considered by Gelman (2006), and we did not find much difference between the estimates of σ_g^2 and the posterior probabilities for both data sets.

Therefore, we conclude that in the considered cases, specifying different prior distributions for b does not affect much either of the posterior distribution of b, posterior estimates of σ_g^2 or the tail posterior probabilities.

Web Appendix C

Calibration of $p(\delta_g, \delta_g^{(\alpha)})$ under H_0.

In this section we treat $p(\delta_g, \delta_g^{(\alpha)})$ as a test statistic in the frequentist sense.



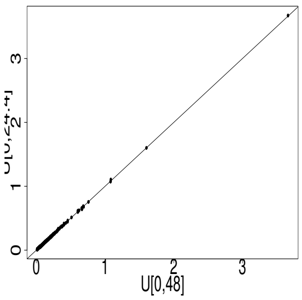
(a) Simulated data (b) H2Kb data

Figure 1. Posterior distributions of b.

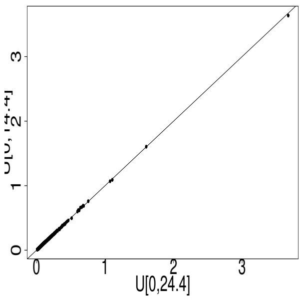
Taking into account that for any x, $G(0, -x + y) + G(0, -x - y)$ is an increasing function of $y > 0$, i.e. that $G(0, -x + y_2) + G(0, -x - y_2) > G(0, -x + y_1) + G(0, -x - y_1)$ is equivalent to $y_2 > y_1 > 0$, taking $x = \delta_g^{(\alpha)}$ and $y_2 = |\bar{y}_g|$ and $y_1 = \delta_g^{(\alpha)}$, we obtain the following:

$$
\begin{aligned}
& P\{p(\delta_g, \delta_g^{(\alpha)}) > 0.5 + G(0, -2\delta_g^{(\alpha)}) \mid H_0\} = \\
= \; & P\{G(0, -\delta_g^{(\alpha)} + \bar{y}_g) + G(0, -\delta_g^{(\alpha)} - \bar{y}_g) > \\
& G(0, -\delta_g^{(\alpha)} + \delta_g^{(\alpha)}) + G(0, -\delta_g^{(\alpha)} - \delta_g^{(\alpha)}) \mid H_0\} \\
= \; & P\{|\bar{y}_g| > \delta_g^{(\alpha)} \mid H_0\} = 2(1 - G(0, \delta_g^{(\alpha)})) = \alpha
\end{aligned}
$$

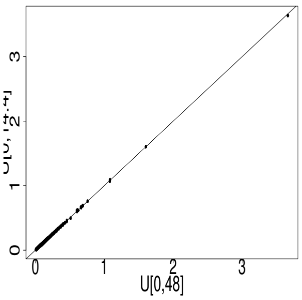
by definition of $\delta_g^{(\alpha)}$. Here we also used the following property of $G(x, y)$: $0.5 = G(0, 0) = G(0, \delta_g^{(\alpha)} - \delta_g^{(\alpha)})$. Since $G(0, -2\delta_g^{(\alpha)})$ is small for small values of α (for example, it is approximately $\alpha/2^{2A+1}$ for $w_g^{-2} \mid s_{gs}^2 \sim \Gamma(A, B_g)$), the



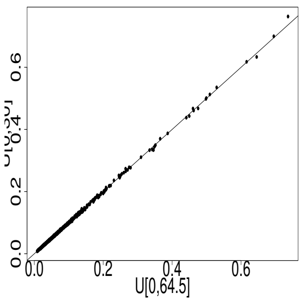
(a) H2Kb data, U(0,48) vs U(0,24.4)



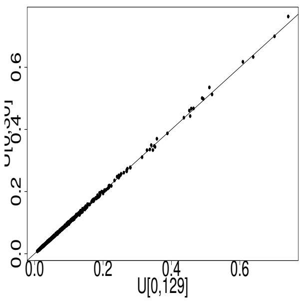
(b) H2Kb data, U(0,24.4) vs U(0,14.4)



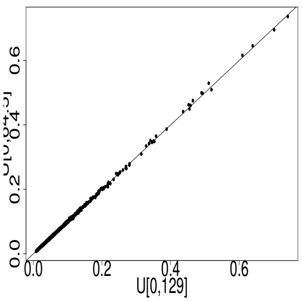
(c) H2Kb data, U(0,48) vs U(0,14.4)



(d) Sim data, U(0,30) vs U(0,64.5)

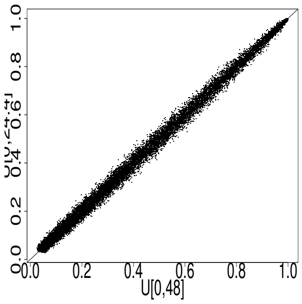


(e) Sim data, U(0,30) vs U(0,129)

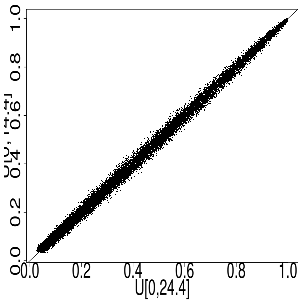


(f) Sim data, U(0,64.5) vs U(0,129)

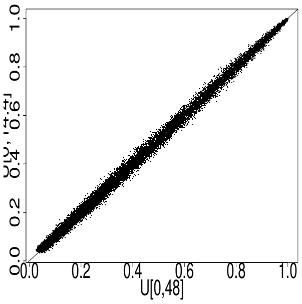
Figure 2. Posterior estimates of σ_g^2.



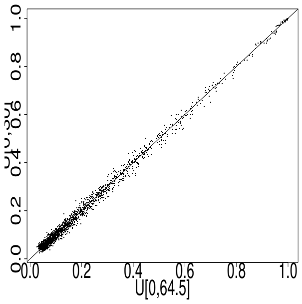
(a) H2Kb data, U(0,48) vs U(0,24.4)



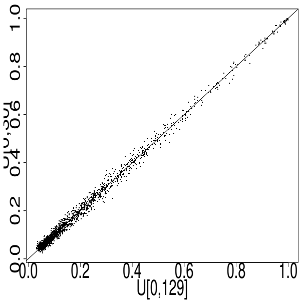
(b) H2Kb data, U(0,24.4) vs U(0,14.4)



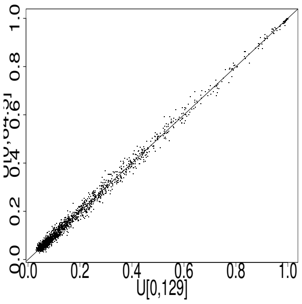
(c) H2Kb data, U(0,48) vs U(0,14.4)



(d) Sim data, U(0,30) vs U(0,64.5)

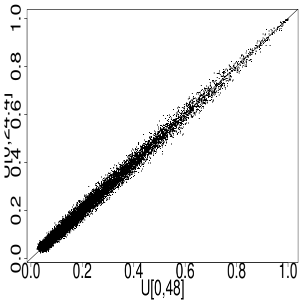


(e) Sim data, U(0,30) vs U(0,129)

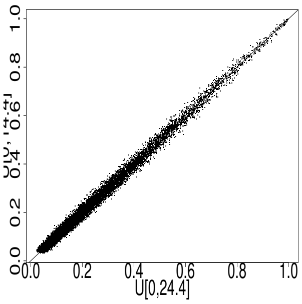


(f) Sim data, U(0,64.5) vs U(0,129)

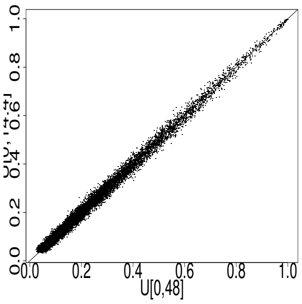
Figure 3. Posterior estimates of tail posterior probability, insulin vs control at 2 hrs.



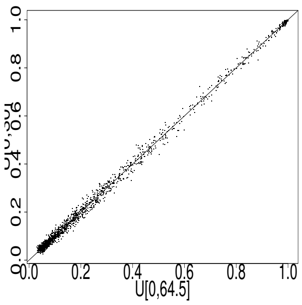
(a) H2Kb data, U(0,48) vs U(0,24.4)



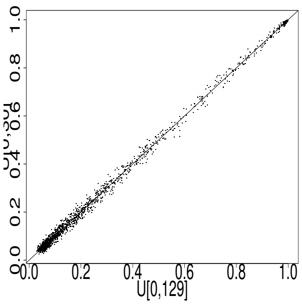
(b) H2Kb data, U(0,24.4) vs U(0,14.4)



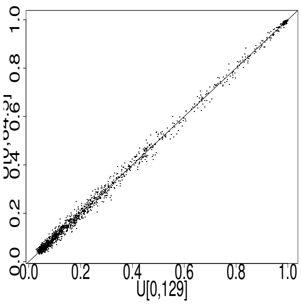
(c) H2Kb data, U(0,48) vs U(0,14.4)



(d) Sim data, U(0,30) vs U(0,64.5)



(e) Sim data, U(0,30) vs U(0,129)



(f) Sim data, U(0,64.5) vs U(0,129)

Figure 4. Posterior estimates of tail posterior probability, insulin vs control at 12 hrs.

tail posterior probability under H_0 satisfies

$$P\{p(\delta_g, \delta_g^{(\alpha)}) > 0.5 \mid H_0\} \approx \alpha.$$

Web Appendix D

Distribution of $p(t_g, t^{(\alpha)})$ under H_0.

Consider model (1) and prior settings (3) with $\sigma_{g1}^{-2} = \sigma_{g2}^{-2} = \sigma_g^{-2} \sim \Gamma(a, b)$. Then the posterior distribution of $w_g^2 = \frac{m_1+m_2}{m_1 m_2}\sigma_g^2$ conditioned on hyperparameters a and b is inverse gamma: $w_g^{-2} \sim \Gamma(2A, \tilde{B}_g)$, $A = a+(m_1+m_2)/2-1$, $\tilde{B}_g = \frac{m_1 m_2}{m_1+m_2}B_g$ and $B_g = b + [(m_1-1)s_{g1}^2 + (m_2-1)s_{g2}^2]/2$. Thus, the tail posterior probability $p(t_g, t^{(\alpha)})$ can be rewritten using (7) in the following way:

$$\begin{aligned}
p(t_g, t^{(\alpha)}) &= P\{|t_g| > t^{(\alpha)} \mid y_{gsr}\} \\
&= \int_0^\infty [\Phi(t_g - \bar{y}_g w_g^{-1}) + \Phi(-t_g - \bar{y}_g w_g^{-1})] f_\gamma(w_g^{-2} \mid 2A, \tilde{B}_g) d(w_g^{-2}) \\
&= \int_0^\infty [\Phi(t_g - v_g \bar{y}_g/\tilde{B}_g^{1/2}) + \Phi(-t_g - v_g \bar{y}_g/\tilde{B}_g^{1/2})] f_\gamma(v_g^2 \mid 2A, 1) d(v_g^2).
\end{aligned}$$

Since the distribution of $z_g = \bar{y}_g\sqrt{A/\tilde{B}_g}$ under the null hypothesis in the considered settings is t distribution with $2A$ degrees of freedom (see, e.g. Smyth, 2004), the distribution of the tail posterior probability $p(t_g, t^{(\alpha)})$ is the same as the distribution of

$$\int_0^\infty \left[\Phi\left(-t^{(\alpha)} + \bar{z}_g\sqrt{v_g^2/A}\right) + \Phi\left(-t^{(\alpha)} - \bar{z}_g\sqrt{v_g^2/A}\right)\right] f_\gamma(v_g^2, A, 1) dv_g^2.$$

Consequently, the integral expression above defines a gene-independent cumulative distribution function $F_0(x)$ for $p(t_g, t^{(\alpha)})$. To evaluate this distribution, we can either plug-in a sensible estimate of parameter A, or integrate it out using its posterior distribution.

Web Appendix E

Comparison of gradients

For larger values of $p(t_g, t^{(\alpha)})$, its gradient becomes greater than the gradient of $p(\delta_g, 0)$ at some change point $y_* = y_*(s_g^2)$. For example, for the considered microarray data, the gradient of the internal part of the volcano (i.e. of the curve $\{\min(|\bar{y}_g| : p_g = p), p\}$ for $p \in [0.2, 0.9]$) is approximately twice as small for $p(t_g, t^{(\alpha)})$ compared to the corresponding gradient for the two-sided $p(\delta_g, 0)$ (Figures 1a and 1b). In general, we show in the theorem below that in the considered settings with equal variances having inverse gamma prior distribution, for any fixed value of s_g^2, the gradient of $p(t_g, t^{(\alpha)})$ is smaller for values of $|\bar{y}_g| < y_*$ than the gradient of the two-sided p-value like selection rule, thus confirming the behaviour illustrated by the plots. Furthermore, for the considered data, the values of the tail posterior probability, as well as of $p(\delta_g, 0)$, corresponding to $y_*(s_g^2)$, are approximately constant for different values of s_g^2.

The next theorem compares the gradients of the tail posterior probability $p(t_g, t^{(\alpha)})$ and the two-sided p-value type selection rule $2\max\{p(\delta_g, 0), 1 - p(\delta_g, 0)\} - 1$ as functions of the observed difference $\bar{y}_g$ for the same values of s_{gs}^2. In the theorem, we drop the indices for convenience.

Theorem 1. *For any $\theta > 0$ there exists $y_* = y_*(\theta, s^2) \geqslant 0$:*

$$\forall y : \quad y < y_*, \quad \left|\frac{\partial(G(-\theta, y) + G(-\theta, -y))}{\partial y}\right| < \left|\frac{\partial(2G(0, y) - 1)}{\partial y}\right|;$$
$$\forall y : \quad y > -y_*, \quad \left|\frac{\partial(G(-\theta, y) + G(-\theta, -y))}{\partial y}\right| < \left|\frac{\partial(2(1 - G(0, y)) - 1)}{\partial y}\right|.$$

Below we give a sketch of the proof.

Proof. (Theorem 1) Due to the asymmetry of the derivatives of $G(-\theta, y) + G(-\theta, -y)$ and of $2\max(G(0, y), 1 - G(0, y)) - 1$ with respect to y, it is sufficient to consider only the case $y > 0$. To prove the threorem, it is sufficient to show that $I(\theta, y) < 0$ for $0 < y < y_*$ where

$$\begin{aligned} I(\theta, y) &= \frac{\partial(G(-\theta, y) + G(-\theta, -y))}{\partial y} - 2\frac{\partial G(0, y)}{\partial y} \\ &= \int_0^\infty v[\varphi(-\theta + yv) - \varphi(\theta + yv) - 2\varphi(yv)]f(v^2|s^2)d(v^2). \end{aligned}$$

Note that for $\theta, z = yv > 0$, expression $i(\theta, z) = \varphi(-\theta+z) - \varphi(\theta+z) - 2\varphi(z)$ as a function of z changes its sign from negative to positive at the point $z^* = \theta/2 + \log(\sqrt{1 + e^{-\theta^2}} + 1)$, which corresponds to the solution with $z > 0$ of the quadratic equation $x^2 - 2xe^{\theta^2/2} - 1 = 0$, $x = e^{z\theta}$.

Therefore, we can represent $I(\theta, y)$ as a difference of two positive integrals:

$$I(\theta, y) = \int_{z^*/y}^\infty i(\theta, yv)vf(v^2|s^2)d(v^2) - \int_0^{z^*/y} (-i(\theta, yv))vf(v^2|s^2)d(v^2).$$

Since $\lim_{y\to 0} I(\theta, y) = -2\varphi(0) < 0$ and $\lim_{y\to\infty} I(\theta, y) > 0$ (due to the length of the integration interval tending to zero for the second integral and bounded $|i(\theta, yv)|$), there exists $y_* \in (0, \infty)$: for any $0 < y < y_*$, $I(\theta, y) < 0$. Thus, the statement of the theorem is proved.

Web Appendix F

WinBUGS code for fitting Bayesian model for multiclass data

To fit model (12) in WinBUGS, we split data into two parts: `y0[g,r]`$= y_{g00r}$ for the log intensity at the baseline ($t = 0$), and the rest of the data `y[g,c,t,r]`$= y_{gctr}$ for gene $g = 1, \ldots, n$, treatment $c = 1, \ldots, n.cond$, time $t = 1, \ldots, n.time - 1$ and replicate $r = 1, \ldots, n.repl$, which forms a balanced array due to equal number of replicates in each condition. Data sets `y0[g,r]` and `y[g,c,t,r]` and parameters n, $n.repl$, $n.cond$ and $n.time$ need to be specified in the data file.

With respect to the considered data set, we have $n.repl = n.cond = n.time = 3$ and the values of $c = 1, 2, 3$ correspond to treatments insulin, metformin and control respectively, and time points $t = 1, 2$ correspond to 2 and 12 hours. To simplify this code, we use explicitly that $n.time - 1 = 2$ splitting parameter δ_{gct} into two, one for each time point: `delta2h[g,c]`$= \delta_{g1c}$ and `delta12h[g,c]`$= \delta_{g2c}$. The reason for this is that the program for importing a summary from WinBUGS into R available to us is not very good in importing three-dimensional arrays. The posterior probabilities `ppTailPP[g,c,t]` (tail posterior probability based on the standardised difference), `ppFC2[g,c,t]` (posterior probability that $|\delta_{gct}| > \log_2 2$) and `ppFC0[g,c,t]` (posterior probability that $\delta_{gct} > 0$) calculated for each value of parameter δ_{gct}, are split in the same way, which we mark by adding the name of the time point after each name of the posterior probability, e.g. `ppTailPP.2h[g,c]`.

Note that we approximate the non-informative prior distribution $f(x) = 1$ by "just proper" distribution $N(0, 10^4)$ which allows to use WinBUGS to fit the model.

```
model; {

   for(g in 1 : n)
   {
    for(r in 1 : n.repl)
    {
     for( c in 1 : n.cond)
     {
       for( t in 1 : (n.time-1))
       {
```

```
        y[g, c, t, r] ~ dnorm(x[g, c, t], tau[g])
      }
    }
    y0[g, r] ~ dnorm(alpha[g], tau[g])
   }
  }

# Intermediate level of hierarchy

 for( g in 1 : n)
 {
   for( c in 1 : n.cond )
   {
      x[g,c,1] <- alpha[g] + gamma[g,1] + delta2h[g,c]
      x[g,c,2] <- alpha[g] + gamma[g,2] + delta12h[g,c]
   }
 }
#------------------------------

# prior distributions for the parameters

#note that control is condition 3

   for(g in 1:n)
   {
    alpha[g] ~ dnorm(0.0, 0.0001)
    gamma[g,1] ~ dnorm(0.0, 0.0001)
    gamma[g,2] ~ dnorm(0.0, 0.0001)

   delta2h[g,3] <- 0
   delta12h[g,3] <- 0
   for(c in 1 : (n.cond-1))
   {
     delta2h[g, c] ~ dnorm(0.0, 0.0001)
     delta12h[g, c] ~ dnorm(0.0, 0.0001)
   }

# variance

    tau[g] ~ dgamma(a, b)
    sig2[g] <- 1.0/tau[g]
  }
```

```
# prior distributions for the hyperparameters

   a ~ dgamma(0.01,0.01)
   b ~ dgamma(1,0.01)

# posterior probabilities

   for(g in 1:n)
   {
    for(c in 1 : (n.cond-1))
    {
     ppTailPP.2h[g,c] <- step(abs(delta2h[g,c]) - 2*sqrt(2*sig2[g]/n.repl))
     ppFC2.2h[g,c] <- step(abs(delta2h[g,c]) - 1)
     ppFC0.2h[g,c] <- step(delta2h[g,c])

     ppTailPP.12h[g,c] <- step(abs(delta12h[g,c]) - 2*sqrt(2*sig2[g]/n.repl))
     ppFC2.12h[g,c] <- step(abs(delta12h[g,c]) - 1)
     ppFC0.12h[g,c] <- step(delta12h[g,c])
    }
   }

}
```

(a) Correlation between differences

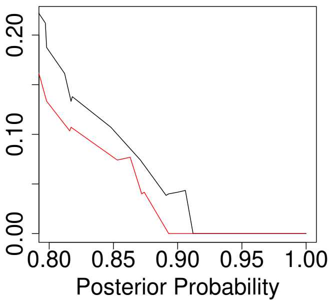
(b) Correlation between standardised differences

Figure 5. Histograms of correlation between differences δ_{g11} and δ_{g12} and correlation between corresponding standardised differences t_{g11} and t_{g12} (using their posterior distributions), H2Kb data.

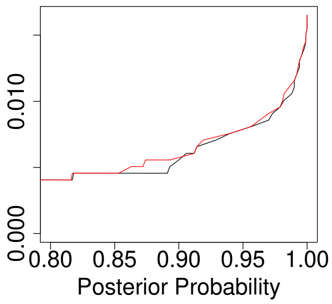
References

Irizarry, R., Bolstad, B., Collin, F., Cope, L., Hobbs, B., and Speed, T. (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Research* **31**(4), e15.

Tatoud, R., Bochkina, N., Jamain, A., Mande, K., Andersson, U., Blangiardo, M., Craig, A., Aitman, T., Richardson, S., Carling, D., and Scott, J. (2006) Statistical and Network-based analysis of the action of the antidiabetic drug Metformin in Muscle. (Working paper, Imperial College, London)



(a) FDR



(b) FNR

Figure 6. False discovery (FDR) and non-discovery (FNR) rates for different thresholds, simulated data. Joint probability - red, minimum of pairwise probabilities - black.