1 Linear regression approach

Linear regression model for each observation $y_{ij} \in \mathcal{D}$ (probe i and tissue sample j) is of the form $y_{ij} = \mathbf{p}_{0j}^T \mathbf{x}_{ic(j)} + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathrm{N}(0, 1/\lambda_i)$. By choosing one experimental condition $1 \leq c \leq C$ and selecting only those tissue samples $j^* \in \{j_1^{(c)}, j_2^{(c)}, \ldots\}$ for which $c(j) = c$, we can express the linear regression model for each experimental condition c in vector format as $\mathbf{y}_{ic} = P_{0c}\mathbf{x}_{ic} + \boldsymbol{\epsilon}_{ic}$, where

$$\mathbf{y}_{ic} = \begin{pmatrix} y_{i,j_1^{(c)}} \\ y_{i,j_2^{(c)}} \\ \vdots \end{pmatrix}, P_{0c} = \begin{pmatrix} p_{0,1,j_1^{(c)}} & p_{0,2,j_1^{(c)}} & \cdots & p_{0,T,j_1^{(c)}} \\ p_{0,1,j_2^{(c)}} & p_{0,2,j_2^{(c)}} & \cdots & p_{0,T,j_2^{(c)}} \\ \vdots & \vdots & \ddots & \vdots \end{pmatrix}, \mathbf{x}_{ic} = \begin{pmatrix} x_{1,i,c} \\ x_{2,i,c} \\ \vdots \\ x_{T,i,c} \end{pmatrix}$$

and

$$\mathcal{D} = \bigcup_{i=1}^{I} \bigcup_{c=1}^{C} \{\mathbf{y}_{ic}\}.$$

The least-squares solution for each $\mathbf{x}_{ic}$ is therefore

$$\hat{\mathbf{x}}_{ic} = (P_{0c}^T P_{0c})^{-1} P_{0c}^T \mathbf{y}_{ic},$$

and residuals of the model can be written as

$$e_{ij} = y_{ij} - P_{0c(j)}\hat{\mathbf{x}}_{ic(j)},$$

from where we deduce precision estimates

$$\hat{\nu}_{ij} = e_{ij}^{-2}.$$

Next, we compute the likelihood function for precision estimates, assuming that each precision estimate comes from a common gamma density $\mathrm{Gamma}(\alpha, \beta)$:

$$L(\alpha, \beta | \hat{\boldsymbol{\nu}}) = \prod_{i=1}^{I} \prod_{j=1}^{J} f(\hat{\nu}_{ij} | \alpha, \beta), \quad \hat{\nu}_{ij} | \alpha, \beta \sim \mathrm{Gamma}(\alpha, \beta).$$

Maximum Likelihood (ML) estimates are sought for both α and β:

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \arg\max_{(\alpha,\beta)} \left\{ L(\alpha, \beta | \hat{\boldsymbol{\nu}}) \right\},$$

which can be performed with Newton-Raphson algorithm.

2 Prior specification (DSection)

The parameter estimates for α, β, and $\mathbf{x}_{ic}$, derived from the linear regression model, are plugged into DSection in the following manner:

$$\alpha := \hat{\alpha}, \quad \beta := \hat{\beta}, \quad \nu := \hat{\alpha}/\hat{\beta}, \quad \forall(t, i, c)(\mu_{tic} := \hat{x}_{tic}).$$

From now on, expression for cell type t, probe i under experimental condition c, x_{tic}, has a prior density

$$x_{tic}|\mu_{tic}, \nu \sim \text{Normal}(\mu_{tic}, 1/\nu),$$

precision of probe i, λ_i, has density

$$\lambda_i|\alpha, \beta \sim \text{Gamma}(\alpha, \beta),$$

and cell type proportion vector for tissue sample j, $\mathbf{p}_j$, has density

$$\mathbf{p}_j|w_0, \mathbf{p}_{0j} \sim \text{Dirichlet}(w_0\mathbf{p}_{0j}).$$

3 Posterior specification (DSection)

Next, we calculate the posterior densities. Let j^*, again, denote the running index for which $c(j^*) = c$. Furthermore, the following short-hand notation will be adopted:

$$A_{tij^*} = \sum_{t' \neq t} p_{t'j^*} x_{t'ic}.$$

The posterior for x_{tic} is

$$
\begin{aligned}
f(x_{tic}|\cdot) &\propto f(x_{tic}|\mu_{tic},\nu)\prod_{j^*} f\left(y_{ij^*}\Big|\sum_{t'=1}^{T} p_{t'j^*}x_{t'ic},\lambda_i\right)\\
&\propto \exp\left\{-\frac{1}{2}\left[\lambda_i\sum_{j^*}\left(y_{ij^*}-\sum_{t'=1}^{T}p_{t'j^*}x_{t'ic}\right)^2+\nu(x_{tic}-\mu_{tic})^2\right]\right\}\\
&\propto \exp\left\{-\frac{1}{2}\left[\lambda_i\sum_{j^*}\left(\left(\sum_{t'=1}^{T}p_{t'j^*}x_{t'ic}\right)^2-2y_{ij^*}\sum_{t'=1}^{T}p_{t'j^*}x_{t'ic}\right)+\nu(x_{tic}^2-2\mu_{tic}x_{tic})\right]\right\}\\
&\propto \exp\left\{-\frac{1}{2}\left[\lambda_i\sum_{j^*}\left((p_{tj^*}x_{tic}+A_{tij^*})^2-2y_{ij^*}p_{tj^*}x_{tic}\right)+\nu(x_{tic}^2-2\mu_{tic}x_{tic})\right]\right\}\\
&\propto \exp\left\{-\frac{1}{2}\left[\lambda_i\sum_{j^*}\left(x_{tic}^2p_{tj^*}^2+2p_{tj^*}x_{tic}A_{tij^*}-2y_{ij^*}p_{tj^*}x_{tic}\right)+\nu(x_{tic}^2-2\mu_{tic}x_{tic})\right]\right\}\\
&= \exp\left\{-\frac{1}{2}\left[x_{tic}^2\left(\lambda_i\sum_{j^*}p_{tj^*}^2+\nu\right)-2x_{tic}\left(\lambda_i\sum_{j^*}(p_{tj^*}y_{ij^*}-p_{tj^*}A_{tij^*})+\nu\mu_{tic}\right)\right]\right\}\\
&= \exp\left\{-\frac{1}{2}\left[x_{tic}^2Q_{tic}-2x_{tic}P_{tic}\right]\right\}\\
&= \exp\left\{-\frac{Q_{tic}}{2}\left[x_{tic}^2-2x_{tic}\frac{P_{tic}}{Q_{tic}}\right]\right\}\\
&\propto \exp\left\{-\frac{Q_{tic}}{2}\left[x_{tic}^2-2x_{tic}\frac{P_{tic}}{Q_{tic}}+\left(\frac{P_{tic}}{Q_{tic}}\right)^2\right]\right\}\\
&= \exp\left\{-\frac{Q_{tic}}{2}\left[x_{tic}-\frac{P_{tic}}{Q_{tic}}\right]^2\right\}\\
&\Leftrightarrow x_{tic}|\cdot \sim \text{Normal}\left(\frac{P_{tic}}{Q_{tic}},\frac{1}{Q_{tic}}\right)\\
&\Leftrightarrow x_{tic}|\cdot \sim \text{Normal}\left(\frac{\lambda_i\sum_{j^*}(p_{tj^*}y_{ij^*}-p_{tj^*}\sum_{t'\neq t}p_{t'j^*}x_{t'ic})+\nu\mu_{tic}}{\lambda_i\sum_{j^*}p_{tj^*}^2+\nu},\frac{1}{\lambda_i\sum_{j^*}p_{tj^*}^2+\nu}\right),
\end{aligned}
$$

for λ_i the posterior is

$$\begin{aligned}
f(\lambda_i|\cdot) &\propto f(\lambda_i|\alpha,\beta)\prod_{j=1}^{J} f\left(y_{ij}\Big|\sum_{t=1}^{T} p_{tj}x_{tic},\lambda_i\right)\\
&\propto \lambda_i^{\alpha-1}\exp\{-\lambda_i\beta\}\lambda_i^{J/2}\exp\left\{-\frac{\lambda_i}{2}\sum_{j=1}^{J}\left(y_{ij}-\sum_{t=1}^{T}p_{tj}x_{tic}\right)^2\right\}\\
&= \lambda_i^{\alpha+J/2-1}\exp\left\{-\lambda_i\left[\beta+\frac{1}{2}\sum_{j=1}^{J}\left(y_{ij}-\sum_{t=1}^{T}p_{tj}x_{tic}\right)^2\right]\right\}\\
&\Leftrightarrow \lambda_i|\cdot \sim \text{Gamma}\left(\alpha+\frac{J}{2},\beta+\frac{1}{2}\sum_{j=1}^{J}\left(y_{ij}-\sum_{t=1}^{T}p_{tj}x_{tic}\right)^2\right),
\end{aligned}$$

and, finally, for $\mathbf{p}_j$ the (un-normalized) posterior is

$$\begin{aligned}
f(\mathbf{p}_j|\cdot) &\propto f(\mathbf{p}_j|w_0\mathbf{p}_{0j})\prod_{i=1}^{I} f\left(y_{ij}\Big|\sum_{t=1}^{T} p_{tj}x_{tic},\lambda_i\right)\\
&\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{I}\lambda_i\left(y_{ij}-\sum_{t=1}^{T}p_{tj}x_{tic}\right)^2\right\}\prod_{t=1}^{T}p_{tj}^{w_0p_{0tj}-1}\\
&= \exp\left\{-\frac{1}{2}\sum_{i=1}^{I}\lambda_i\left(y_{ij}-\sum_{t=1}^{T}p_{tj}x_{tic}\right)^2+\sum_{t=1}^{T}(w_0p_{0tj}-1)\ln(p_{tj})\right\}.
\end{aligned}$$

4 Sampling (DSection)

Parameters x_{tic} and λ_i are sampled using Gibbs sampling, whereas for $\mathbf{p}_j$ Metropolis-Hastings is employed. An algorithmic representation for the sampling process is on Table 1.

for $s = 1 : (B + S)$ **do**
  for $j = 1 : J$ **do**
    $\mathbf{p}_j \sim K(\mathbf{p}_j^* \to \mathbf{p}_j)$
    $u \sim \mathrm{U}(0, 1)$
    if $u < \min\left\{1, \rho_j(\mathbf{p}_j^* \to \mathbf{p}_j)\right\}$ **then**
      $\mathrm{MCData}(s).\mathbf{p}_j \leftarrow \mathbf{p}_j$
    else
      $\mathrm{MCData}(s).\mathbf{p}_j \leftarrow \mathbf{p}_j^*$
    end if
  end for
  for $i = 1 : I$ **do**
    $\mathrm{MCData}(s).\lambda_i \sim f(\lambda_i|\cdot)$
    for $t = 1 : T$ **do**
      for $c = 1 : C$ **do**
        $\mathrm{MCData}(s).x_{tic} \sim f(x_{tic}|\cdot)$
      end for
    end for
  end for
end for
$\mathrm{MCData} \leftarrow \mathrm{MCData}((B + 1) : (B + S))$

Table 1: algorithmic representation of the sampling process for DSection. Variable s is the sampling index that runs from 1 to $B + S$, where B is the number of burn-in iterations to be discarded after sampling, and S is the length of the resulting Markov chain.

4.1 Initializing the DSection sampler

Before starting the sampling process in order to generate Markov chains for the model parameters, one needs to initialize the sampler with proper parameter values. Theoretically it does not matter how the sampler is initialized as the sampler will eventually converge to the posterior distribution. However, a poor initialization of the sampler may take a longer time to attain the posterior in contrast to careful initialization. We choose to take advantage of the already-computed least-squares solution for the respective linear regression model, where one basically obtains estimates for λ_i and x_{tic}. Moreover, the proportion vectors, $\mathbf{p}_j$, are initialized with the prior information, $\mathbf{p}_{0j}$, directly.

5 Deriving ground-truth for Affymetrix data

Sample (j)	$1-3$	$4-6$	$7-9$	$10-12$	$13-21$	$22-24$	$25-27$	$28-30$	$31-33$
Brain (p_{1j})	0.00	0.05	0.10	0.25	0.50	0.75	0.90	0.95	1.00
Heart (p_{2j})	1.00	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.00

Table 2: Known cell type proportions for each sample in Affymetrix data. For each mixing experiment (one column of the table), a triplet of measurements have been conducted except for samples $13-21$, which all have 50%/50% mixing ratio. Samples $10-12$ and $22-24$ were used for estimating cell type specific gene expression profiles, and the expression estimates were then compared to the pure cell type specific gene expressions (samples $1-3$ and $31-33$). Furthermore, we included samples $7-9$ and $25-27$ when testing how increasing the number of heterogeneous samples for analysis with DSection affects the model performance.

Although no ground-truth for replication variances of Affymetrix data truly exists, we estimate them on the basis of sample-groups having identical cell type proportions. That is, samples $1-3$ have identical cell type proportions (0%/100%), so do samples $4-6$ (5%/95%), etc., which – when grouped – yield sample variance estimates over which we consequently compute the sample mean. This "ground-truth" serves as our reference for making comparisons against model estimates similarly to that of our simulation case.

Moreover, no ground-truth for truly differentially and non-differentially expressed genes exist for Affymetrix data. However, based on the derived estimates for replication variances and non-heterogeneous expression measurements (samples $1-3$ for heart and $31-33$ for brain), both hidden from the actual analysis, we produced a binary list representing ground-truth differential expression. The list was obtained by considering, gene-wise, the average expressions in both cell types, brain and heart, and if the absolute difference of cell type specific expressions was bigger than $\sqrt{2/(3\lambda_i)}$, i.e., the denominator of the two-sample t-test with equal sample sizes of three, we declared that gene to be differentially expressed. Thus, decreasing gene-specific replication variance (increasing precision) and increasing sample size both decrease the threshold. We tried other thresholding criteria as well, such as $\sqrt{1/\lambda_i}$ (one-STD confidence) and $\sqrt{4/\lambda_i}$ (two-STD confidence), yielding very similar results to what are shown in the actual manuscript.

6 Adding noise to cell type proportions

In the simulation, ground-truth cell type proportions were first drawn from Dirichlet$(3,3,3)$, a density that is lightly concentrated around its mode, $(1/3,1/3,1/3)$. These proportion vectors were further transformed to noisy

priors which reflect the inaccuracy of, say, manual or automated image analysis performed on basis of tissue images. The transformation first maps each ground-truth proportion, p_{tj}, uniformly to $[0.5 * p_{tj}, 1.5 * p_{tj}]$, after which the vectors are normalized back to the T-simplex. This convention does not bias the proportions, i.e., the expected value of the randomization equals to the ground-truth.

7 Testing for differential expression

The test statistics of the two-sample t-test – assuming equal variance and unequal sample sizes, and that we are testing whether expression of probe i is differentially expressed between tissues t_1, t_2 and experimental conditions c_1, c_2 – is computed by

$$t^{\text{test}}(x_{t_1 i c_1} \quad \text{vs} \quad x_{t_2 i c_2}) = \frac{\hat{x}_{t_1 i c_1} - \hat{x}_{t_2 i c_2}}{\sqrt{\frac{1}{\sqrt{\hat{\lambda}_i}} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}, \tag{1}$$

where the hats denote the respective MCMC estimates, and n_1 and n_2 are the number of tissue samples ($j \in \{1, \ldots, j, \ldots, J\}$) for which $c(j) = c_1$ and $c(j) = c_2$, respectively.

7.1 p-values, multiple correction, false discovery rate

The t-test statistics can then be transformed into p-values, to which further multiple correction methods such as the well-known and widely used Bonferroni's, Benjamini-Hochberg's, and Storey's (q-value) methods can be applied, for controlling the number of false positives. As we demonstrate only the performance of the method itself over the whole range of significance thresholds, i.e., with the receiver operating characteristics (ROC) curves, the use of multiple correction methods would not have made any difference to the outcome of our analysis; thus, only "uncorrected" p-values were computed and used throughout the manuscript (as it is easier to threshold p-values than the original test statistics). However, in any actual analysis where a significance threshold needs to be carefully chosen so as to not invoke "too many" false positives in the data, such multiple correction methods are advised to be applied.

Based on the p-value histograms we computed from the extracted test statistics, Storey's q-values would have been well applicable for estimating FDR as the prior assumptions of the shapes of the histograms (peak near 0, flat over the support $[0, 1]$) seemed to hold in most cases.