Supplemental Material

	hyperparameters
α_a	1
β_a	1
p_a^0	0.1
p_a^1	0.1
α_b	1
β_b	1
p_b^0	0.1
p_b^1	0.1
ν_r	P + 1
ν_ρ	P + 1
α_ξ	1
β_ξ	1
c^2_{max}	1

Table 1: Default values of the hyperparmeters in the Bayesian model for a dataset with P studies. For each of the simulated and real datasets in this paper, we used the default hyperparameters with $P = 3$.

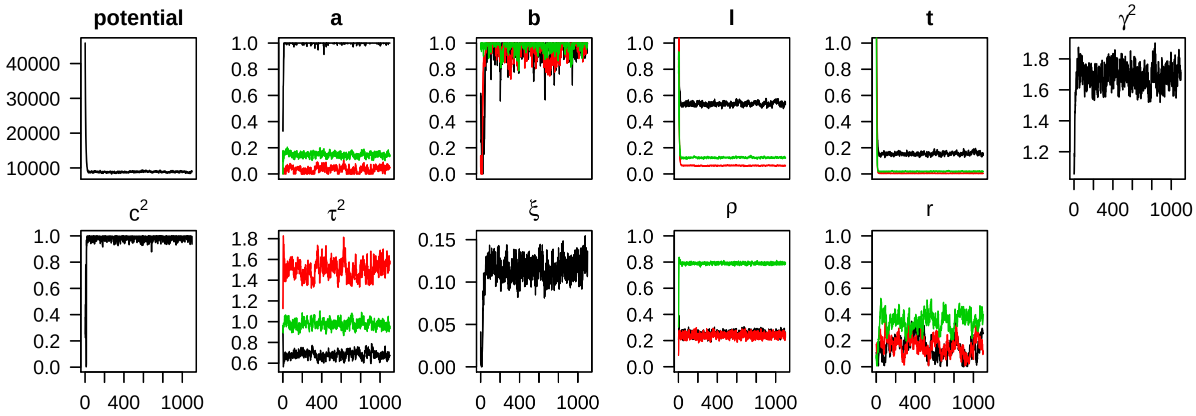


Figure 1: Trace plots of the negative log likelihood (panel 1, top row) and Metropolis-Hastings parameters obtained from fitting the Bayesian model to a simulated dataset of three studies. The parameters used to simulate the data are provided in row A of Table 1. A thinning interval of 20 was used in this plot, hence only 1100 out of 22,000 iterations are plotted. The first 4000 iterations were discarded before calculating posterior statistics of interest.

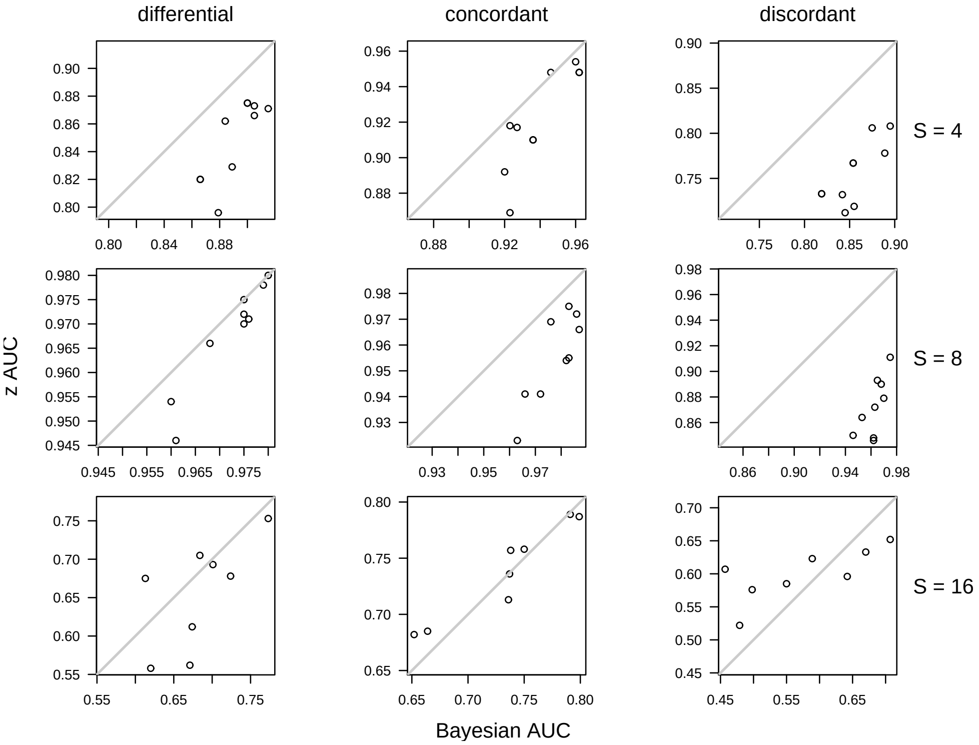
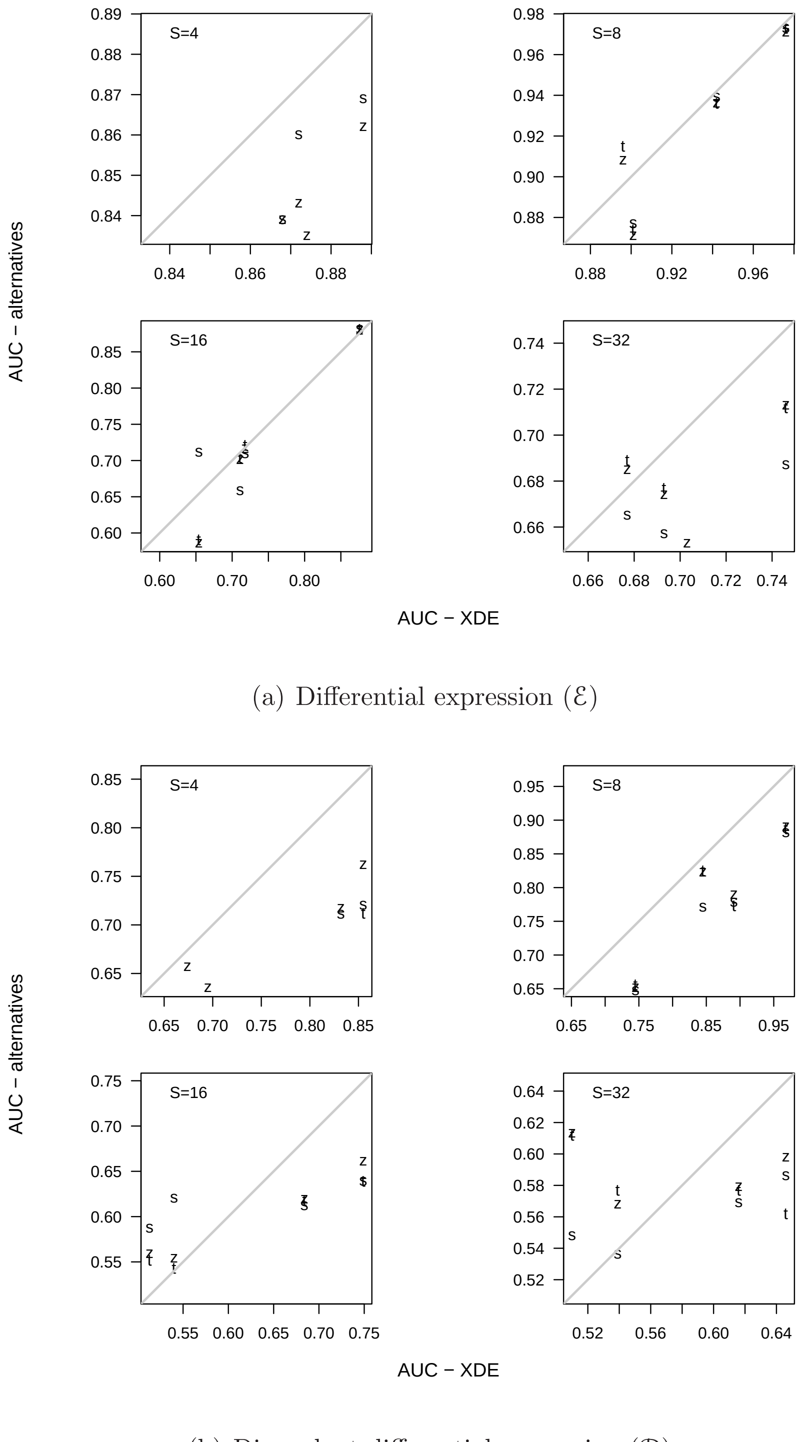


Figure 2: To evaluate the extent to which random draws of δ^* and $\mathbf{\Delta}^*$ influence the performance of the different methods, we simulated 10 artificial datasets for Simulation $A^\dagger$ (top row), $\mathrm{E}^\dagger$ (row 2), and $\mathrm{J}^\dagger$ (row 3) using different seeds for the random number generator. In each panel, we plot the AUC from the Bayesian model (horizontal axis) against the AUC from the z-score (vertical axis). Not shown are the AUC from the SAM- and t-scores. The columns depict the three different ways to evaluate differential expression. Posterior averages for the Bayesian statistic (Section 4) were calculated from 1000 iterations (saving every 20$^{\mathrm{th}}$ iteration of 20,000 iterations) following a burnin of 2000 iterations.



(a) Differential expression ($\mathcal{E}$)

(b) Discordant differential expression ($\mathcal{D}$)

Figure 3: In each panel, we plot the AUC obtained from alternative methods on the vertical axis and the AUC from the Bayesian model (XDE) on the horizontal axis. Differential expression ($\mathcal{E}$) and discordant differential expression ($\mathcal{D}$) are considered separately. In several instances, the AUC corresponding to the t- and SAM-scores were lower than the limit used for the scatterplots and were not plotted.

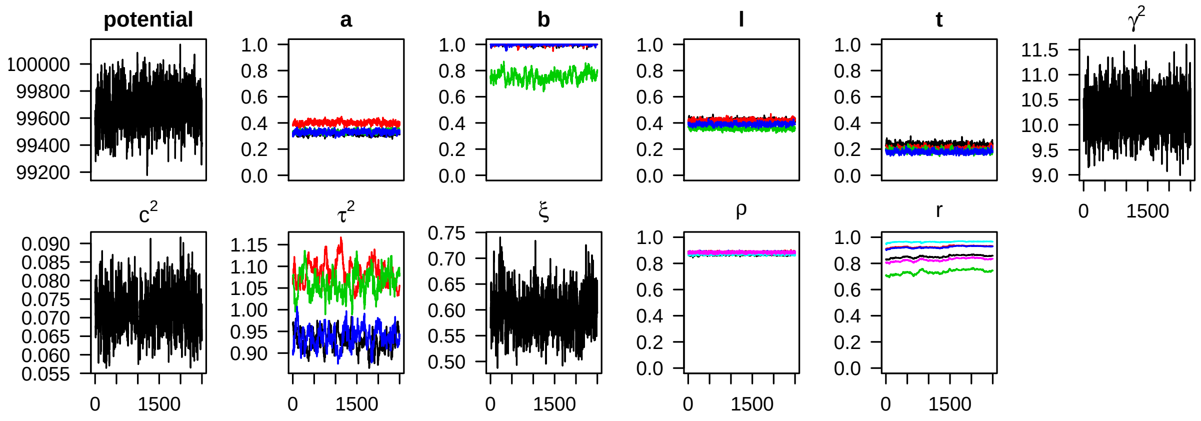
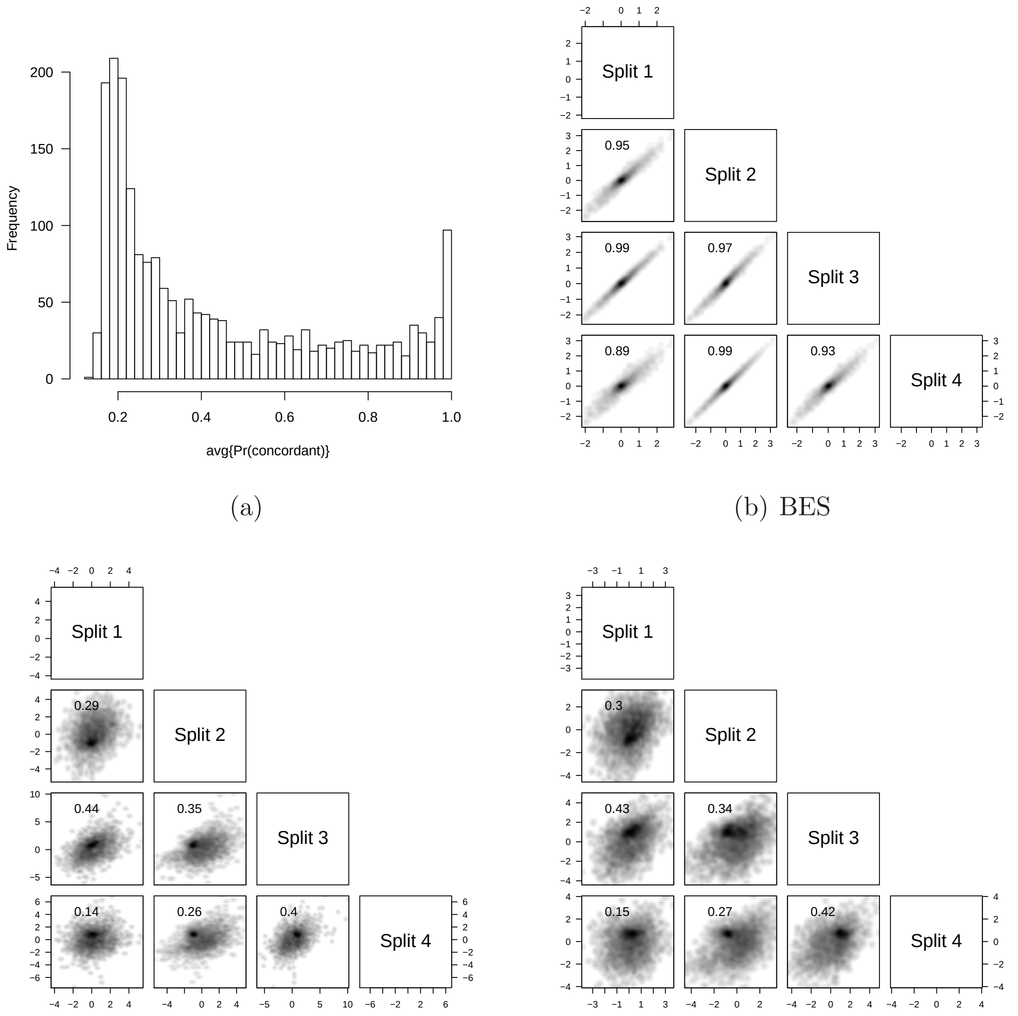


Figure 4: A single dataset, the Huang study, was split into four disjoint parts with 5 ER negative and 16 ER positive samples in each. Plotted are traces for the negative log likelihood (panel 1, top row) and a subset of the Metropolis-Hastings parameters.



(a)

(b) BES

(c) t-statistics

(d) z-scores

Figure 5: Top left: Distribution of the posterior probability for concordant differential expression, $\text{PM}_{\text{c}}(g)$. Panels 2-4 are scatterplots of study-specific measures of differential expression in the split-study validation. t and z statistics, estimated independently for each study, show considerable variation across studies with discordance that is probably within the noise of the experiment. The modest correlation of the study-specific statistics motivates an approach that more effectively models the inter-study and inter-gene relationships. The BES (top right) shows how noisy genes are shrunk towards zero, whereas genes in quadrants (+, +) and (-, -) that show some evidence of differential expression in each of the studies are shrunk less. As the $\text{PM}_{\text{c}}(g)$ is useful for ranking concordant differential expression in multiple studies or platforms, the highest ranked genes are typically genes whose differential expression was not platform- or study-dependent. As the goal of many microarray experiments is to select genes for subsequent validation by other platforms for measuring transcript abundance (such as qRT-PCR), a ranking that is not platform- and study-dependent may facilitate this effort.

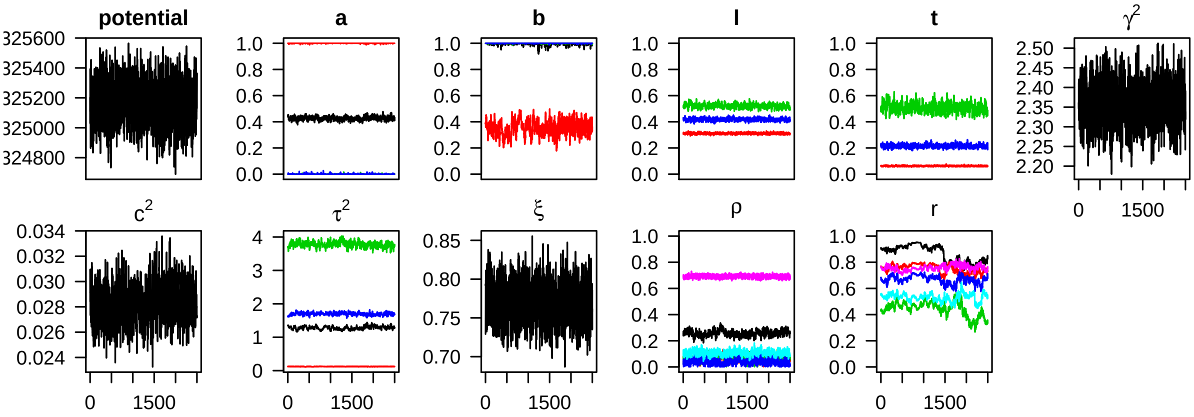


Figure 6: Traceplots for the negative log likelihood (panel 1, top row) and Metropolis-Hastings parameters for the experimental data example.

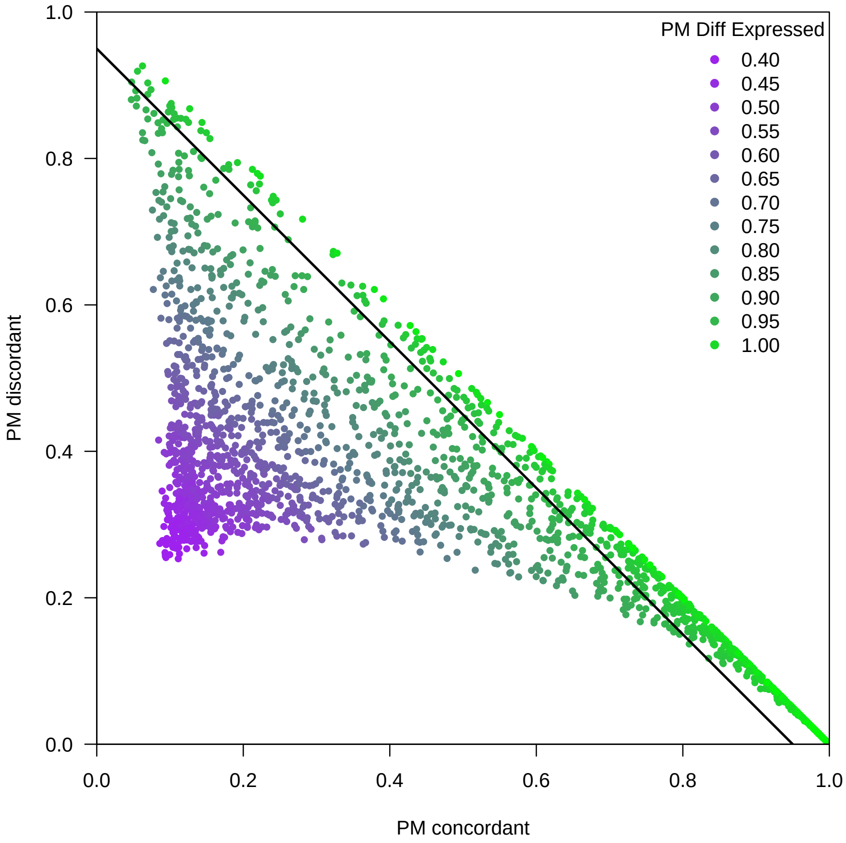


Figure 7: A scatterplot of the posterior means (PM) for the indicators of concordant (x-axis) and discordant (y-axis) differential expression. Plotting symbols are color coded by the gradient of the PM for the differential expression indicators. In purple, are genes for which the model is uncertain regarding differential expression. Here, the model has difficulty distinguishing between low levels of differential expression and no differential expression. Genes with strong evidence of differential expression above the diagonal $\mathrm{PM}_{\varepsilon}(g) = 0.95$ line are predominantly concordant across studies.

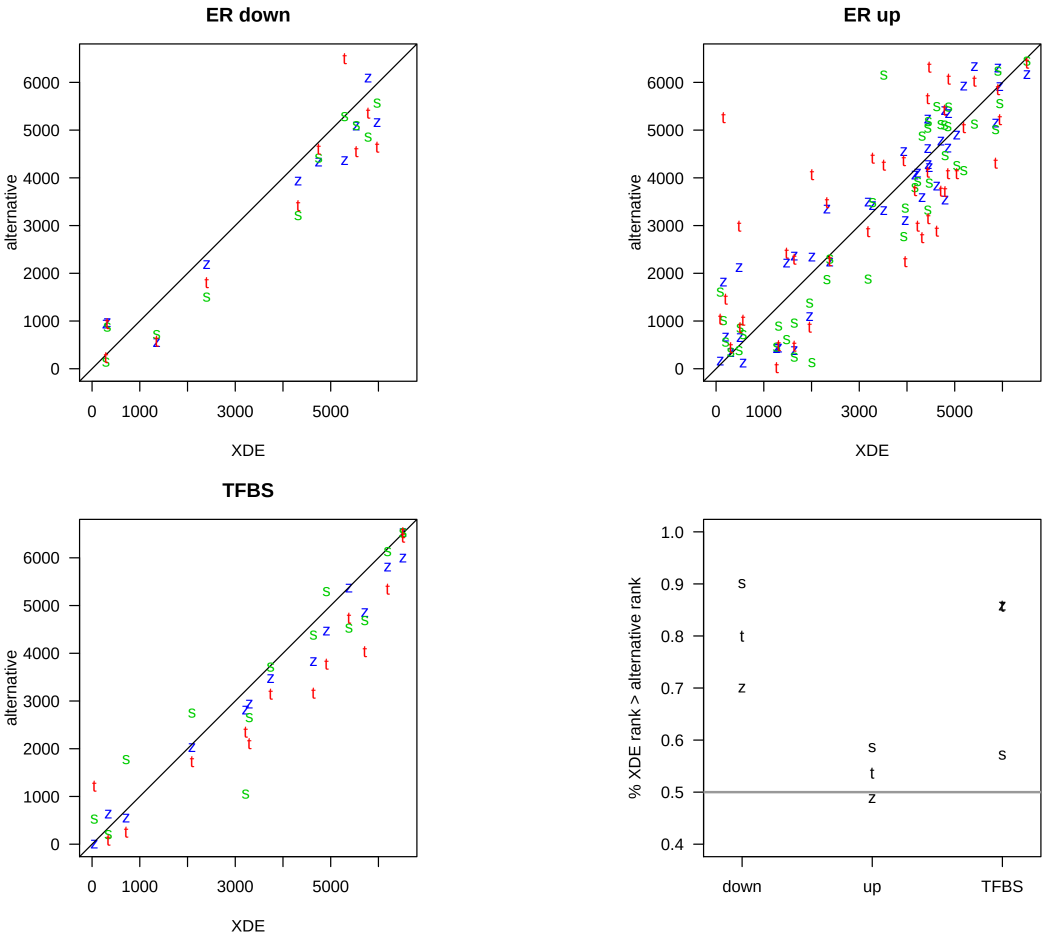


Figure 8: We performed split study validation on a reduced set of approximately 6500 genes in the Farmer study. We ranked three sets of estrogen receptor genes (up, down, and transcription factor binding site) using scores for concordant differential expression from our multiple indicator model (x-axis) and alternative methods (y-axis). At the bottom right, is the proportion of times the *XDE* model ranks an ER gene higher than an alternative method for each of the gene sets.