Supplementary Materials

Overview

The supplementary materials contain six sections. Section S1 presents the derivation of the marginal posterior distribution of the latent states $\boldsymbol{S}$ under the graphical ms-BB model. Section S2 presents the posterior marginal alternative probabilities (PMAPs) for the DNase-seq example when no within-group variation is allowed under the ms-BB model (i.e., $\nu(A) = \infty$ for all windows A). Section S3 presents simulation results for 3-group comparison. Section S4 reports a sensitivity analysis on the choice of the maximum resolution K. Section S5 investigates the accuracy numerical approximation adopted in computing the posterior. Finally, Section S6 provides an example of sampling from the joint posterior in the context of the DNase-seq example.

S1. The marginal posterior of $\boldsymbol{S}$ under the graphical ms-BB model

We shall derive the marginal posterior of $\boldsymbol{S}$ in two steps. First we establish the result for a finite nested dyadic partition (NDP) $\mathcal{T}$ with a maximum resolution level K. Then we extend the result to infinite NDPs. Suppose $\mathcal{T}$ is a finite NDP with maximum resolution K. For each $A \in \mathcal{T}$ and $s \in \{0, 1\}$, let

$$\phi_s(A) := \Pr(\boldsymbol{x}(A) \,|\, A, S(A_p) = s) \Big/ \prod_{A \in \mathcal{T}(A)} M_0(A)$$

where $\mathcal{T}(A)$ denotes the subtree of $\mathcal{T}$ consisting of A (as the root) and all its descendants. When $\sum_j n_{ij}(A) > 0$ for no more than one i, it is easy to check that $M_0(\tilde{A}) = M_1(\tilde{A})$ for all $\tilde{A} \in \mathcal{T}(A)$, and so

$$\mathrm{E}\left(\prod_{A \in \mathcal{T}^{(k)}(A)} M_{S(A)}(A) \,|\, S(A_p) = s \right) = \prod_{A \in \mathcal{T}^{(k)}(A)} M_0(A) \quad \text{for all } k.$$

Hence in that case $\phi_0(A) = \phi_1(A) = 1$.

When $\sum_j n_{ij}(A) > 0$ for at least two i's, and if A has children A_l and A_r, we have

$$\Pr(\boldsymbol{x}(A) \,|\, A, S(A_p) = s) = \sum_{s'} \rho_{s,s'}(A) M_{s'}(A) \Pr(\boldsymbol{x}(A_l) \,|\, A, S(A) = s') \cdot \Pr(\boldsymbol{x}(A_r) \,|\, A, S(A) = s').$$

Thus

$$\begin{aligned}\phi_s(A) &= \sum_{s'} \rho_{s,s'}(A) \cdot \frac{M_{s'}(A)}{M_0(A)} \cdot \frac{\Pr(\boldsymbol{x}(A_l) \,|\, A, S(A) = s')}{\prod_{A \in \mathcal{T}(A_l)} M_0(A_l)} \cdot \frac{\Pr(\boldsymbol{x}(A_r) \,|\, A, S(A) = s')}{\prod_{A \in \mathcal{T}(A_r)} M_0(A_r)} \\ &= \rho_{s,0}(A)\phi_0(A_l)\phi_0(A_r) + \rho_{s,1}(A) \cdot \mathrm{BF}(A) \cdot \phi_1(A_l)\phi_1(A_r).\end{aligned}$$

When A has no children, i.e., $A \in \mathcal{T}^K$, the above equality holds with $\phi_s(A_l) = \phi_s(A_r) = 1$.

By Bayes theorem,

$$\begin{aligned}P(S(A) = s' | S(A_p) = s, \boldsymbol{x}) &= P(S(A) = s', \boldsymbol{x}(A) \,|\, A, S(A_p) = s) / P(\boldsymbol{x}(A) \,|\, A, S(A_p) = s) \\ &= \rho_{s,s'}(A) \cdot \frac{M_{s'}(A)}{M_0(A)} \cdot \phi_{s'}(A_l)\phi_{s'}(A_r)/\phi_s(A).\end{aligned}$$

Letting $\boldsymbol{\phi}(A) = (\phi_0, \phi_1)'$ and putting the above equality in matrix form establishes the result for the case when $\mathcal{T}$ is finite with a maximum resolution K.

Now consider the case when $\mathcal{T}$ does not terminate at a finite maximum resolution K. Then we use a limiting truncation argument as follows. Now for each $k = 1, 2, \ldots$, and each $A \in \mathcal{T}^{(k)}$

$$\phi_s^{(k)}(A) := \Pr(\boldsymbol{x}(A) \,|\, A, S(A_p) = s, \mathcal{T} = \mathcal{T}^{(k)}) \Big/ \prod_{A \in \mathcal{T}^{(k)}(A)} M_0(A)$$

where $\mathcal{T}^{(k)}(A)$ is the subtree of $\mathcal{T}^{(k)}$ rooted at A. Now define

$$\phi_s^{(\infty)}(A) = \lim_{k \to \infty} \phi_s^{(k)} = \lim_{k \to \infty} \Pr(\boldsymbol{x}(A) \,|\, A, S(A_p) = s, \mathcal{T} = \mathcal{T}^{(k)}) \Big/ \prod_{A \in \mathcal{T}^{(k)}(A)} M_0(A).$$

Because when $\sum_j n_{ij}(A) > 0$ for no more than one i, $M_0(\tilde{A}) = M_1(\tilde{A})$ for all $\tilde{A} \in \mathcal{T}(A)$, in that case we have $\phi_s^{(\infty)} = 1$. Now for $\sum_j n_{ij}(A) > 0$ for at least two i's,

$$\begin{aligned}&\phi_s^{(\infty)}(A) \\ &= \lim_k \sum_{s'} \rho_{s,s'}(A) \cdot \frac{M_{s'}(A)}{M_0(A)} \cdot \frac{\Pr(\boldsymbol{x}(A_l) \,|\, A, S(A) = s', \mathcal{T} = \mathcal{T}^{(k)})}{\prod_{A \in \mathcal{T}^{(k)}(A_l)} M_0(A_l)} \cdot \frac{\Pr(\boldsymbol{x}(A_r) \,|\, A, S(A) = s', \mathcal{T} = \mathcal{T}^{(k)})}{\prod_{A \in \mathcal{T}^{(k)}(A_r)} M_0(A_r)} \\ &= \rho_{s,0}(A)\phi_0^{(\infty)}(A_l)\phi_0^{(\infty)}(A_r) + \rho_{s,1}(A) \cdot \mathrm{BF}(A) \cdot \phi_1^{(\infty)}(A_l)\phi_1^{(\infty)}(A_r)\end{aligned}$$

where the last equality follows from moving the limit into the sum. Letting $\boldsymbol{\phi}(A) = (\phi_0^{(\infty)}, \phi_1^{(\infty)})'$ establishes the result for the case with infinite $\mathcal{T}$.

S2. PMAPs for the ms-BB with $\nu(A) = \infty$ for the DNase-seq example

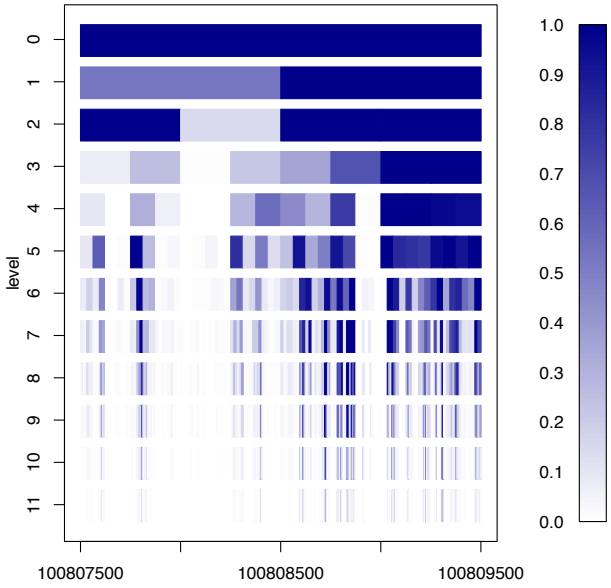

Figure S1: PMAPs for the DNase-seq example from the ms-BB model with $\nu(A) = \infty$. Ignoring in-group variation results in a large number of falsely identified windows.

S3. Simulation studies for 3-group comparison

To evaluate the performance of our method for multi-group comparison, we repeated the same simulation studies done in Secton 3 for three-group comparisons. In these simulations, we use the same setup as the five scenarios studied in the paper, but add an additional null group. Figures S1 and S2 present the histograms of the test statistics and the ROC curves respectively for four different methods. There is no substantive difference for our method in 2-group or 3-group cases.

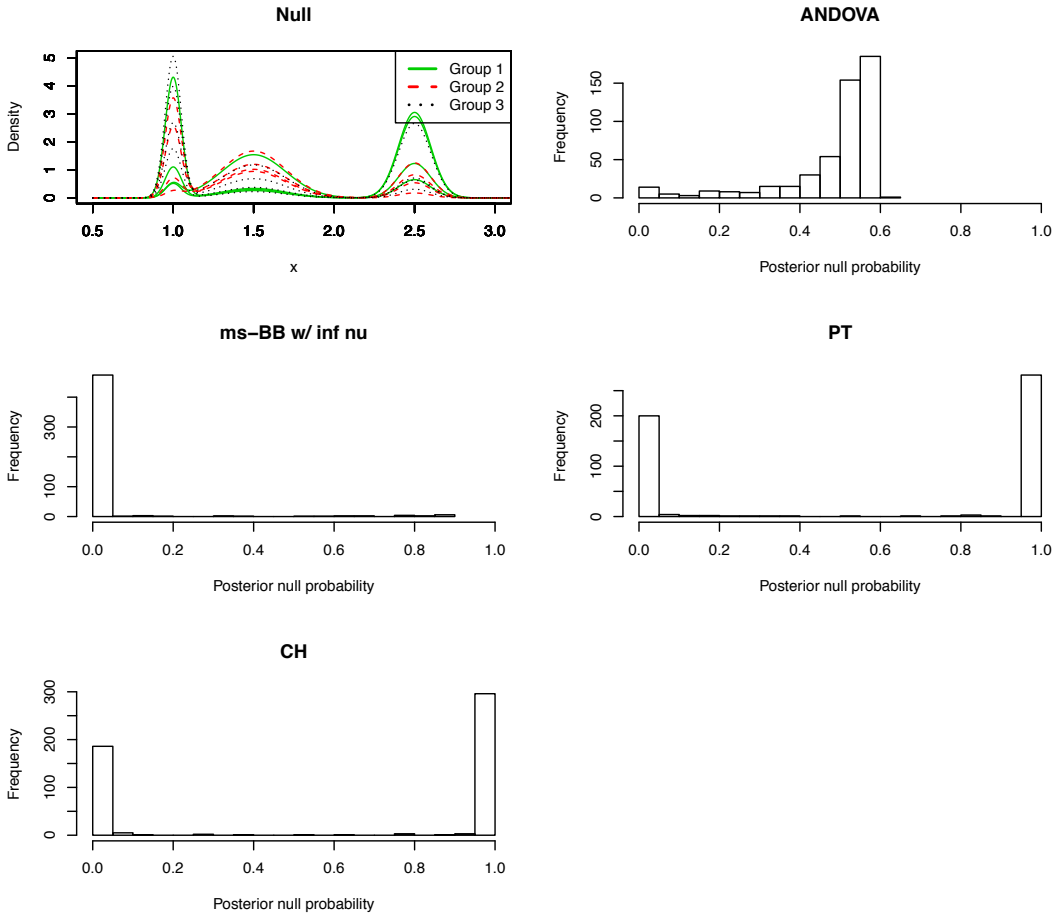

Figure S1: The histogram of test statistics under the null hypothesis of no three-group difference. The seven methods are ANDOVA based on graphical ms-BB, ms-BB with $\nu(A) = \infty$, PT, and CH.

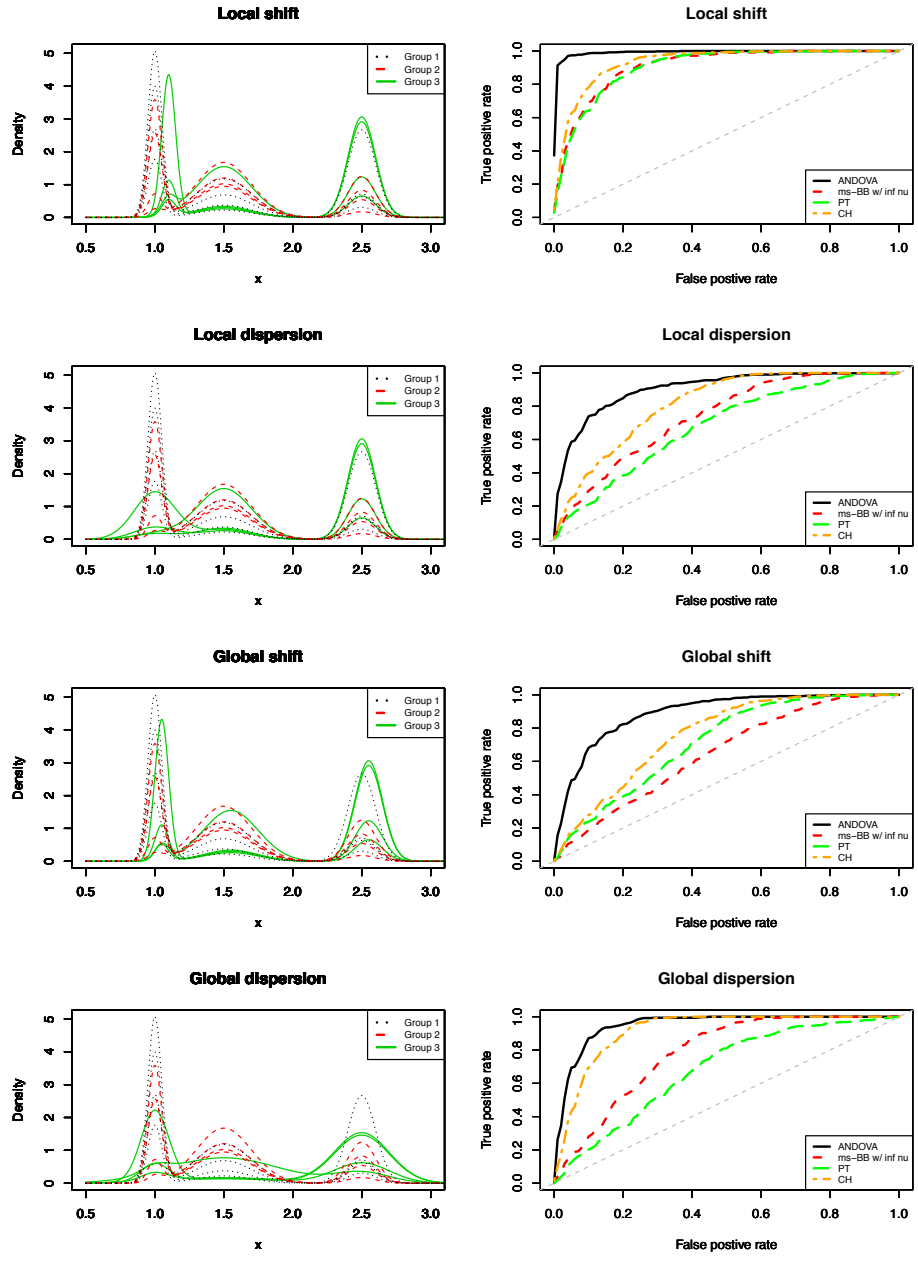

Figure S2: Performance comparison of different methods under the four scenarios with cross-group differences. Left panel: A realization of the true sampling densities under the simulation setting for each alternative scenario. Group 1 densities are indicated with dotted black lines, Group 2 dashed red lines, and Group 3 solid green lines. Each group consists of four replicate samples. Right: ROC curves of methods—ANDOVA based on graphical ms-BB, ms-BB with $\nu(A) = \infty$, PT and CH—for detecting the existence of cross-group differences.

S4. Sensitivity analysis on the choice of maximum resolution K

We evaluate the robustness of our method to the choice of K through a sensitivity analysis. In particular, we compare the testing performance of our method under the four alternative simulation scenarios considered in Section 3. The results (see Figures S3 and S4) show that our method is robust to the choice of K in identifying cross-group differences. As such, in practice one can simply take the largest level that could potentially contain interesting structure and is allowed by the computational resources.

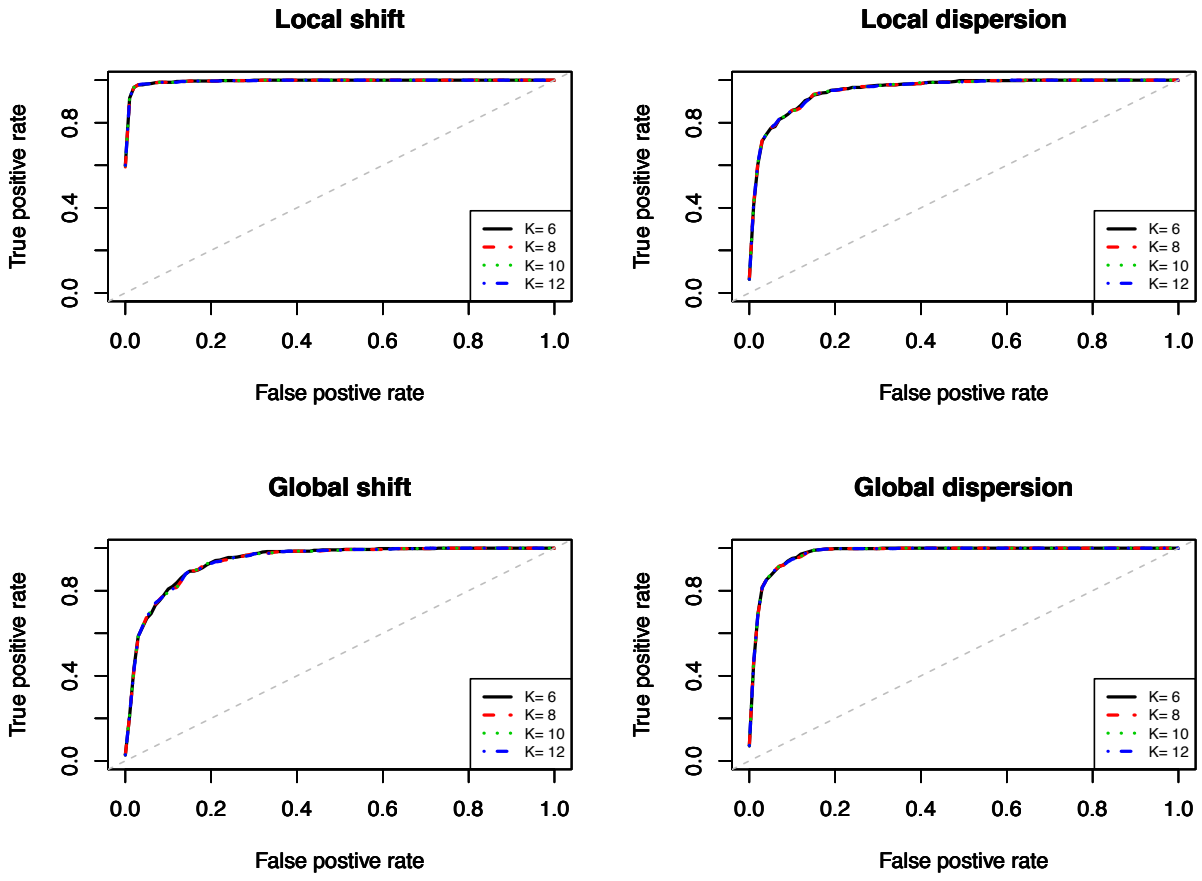

Figure S3: ROC curves of the ms-BB method for the four alternative scenarios at four different maximum resolution levels. The ROC curves at the four K levels are almost overlapping with each other.

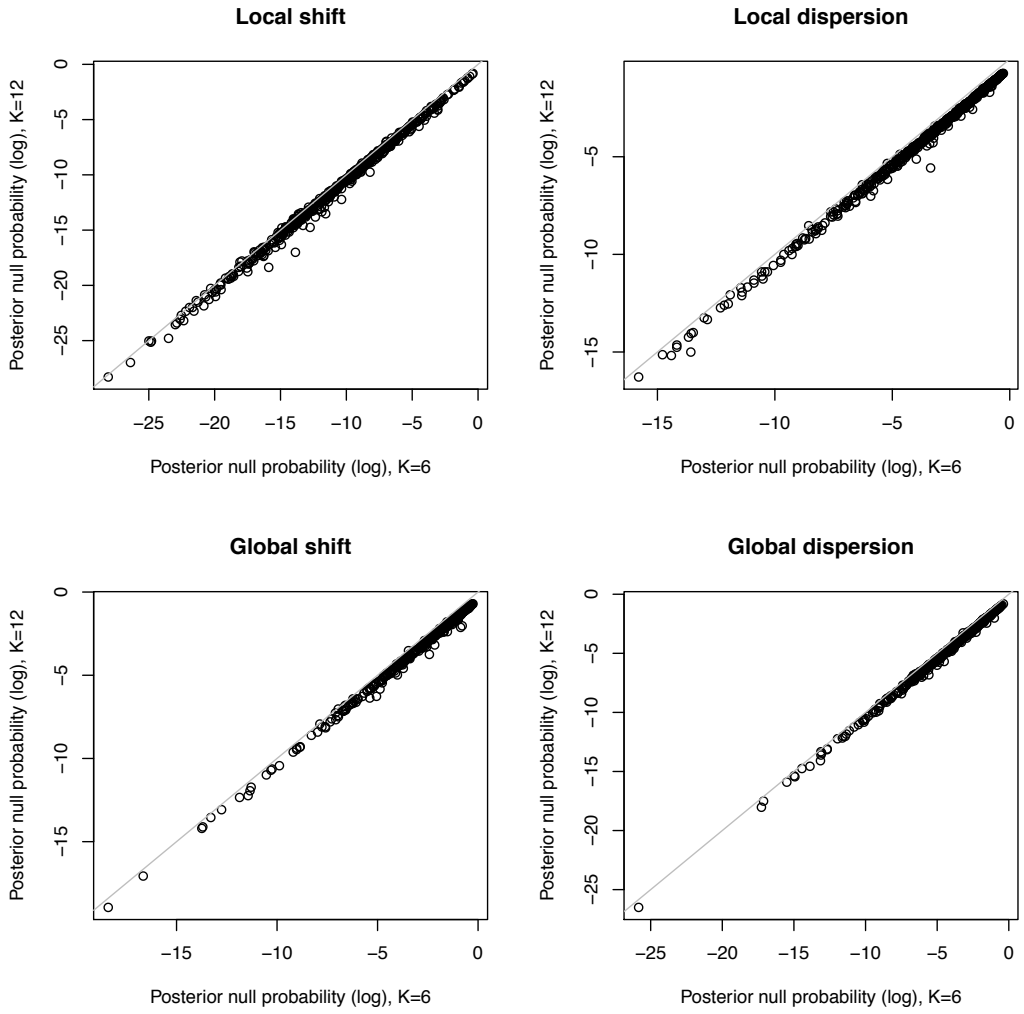

Figure S4: Posterior joint null probability under the four alternative scenarios at two different maximum resolution levels: $K = 6$ versus at $K = 12$. The values at the two maximum resolution levels are very similar to each other.

S5. Assessing the quality of numerical approximation

We formally evaluate the approximation accuracy of our model by comparing the resulting PMAPs to those from a very accurate calculation of the integrals. The latter are achieved through using highly accurate Riemann quadratures, which uses 100 grid points on both $\theta(A)$ and $\nu(A)$.

Because the PMAP on each scanning window is a direct transform of the marginal likelihood terms, we compare the PMAP on each window using our INLA approach to that calculated using the accurate quadrature as a way to gauge the accuracy of our numerical approximation. Moreover, in this comparison we aim to separate out two sources of approximation errors—those from the Laplace approximation and those from the Riemann approximation. To achieve this, we make two comparison, the first is to compare the "accurate PMAPs" to those calculated using Laplace approximation on $\theta(A)$ along with a very fine Riemann approximation on $\nu(A)$, so that the error is "entirely" due to the Laplace approximation on $\theta(A)$. The second is to compare the "accurate PMAPs" to those calculated without a Laplace approximation, but a very fine quadrature on $\theta(A)$, along with a coarse Riemann approximation—using 10 grid points—on $\nu(A)$. Hence the approximation error in the second comparison is "entirely" due to the Riemann step on $\nu(A)$ in the INLA.

We plot in Figure S5 the accurate PMAPs versus the two different types of approximate PMAPs for all windows under each of the five simulation scenarios studied in Section 3. As we can see, the approximation is generally satisfactory in these simulations. The Laplace approximation contributes more error to the approximation than the Riemann part. In fact, with as few as 10 grid points on $\nu(A)$, the Riemann approximation is already very accurate. Because of this, we have set the default number of grid points on $\nu(A)$ to be 10 in our software.

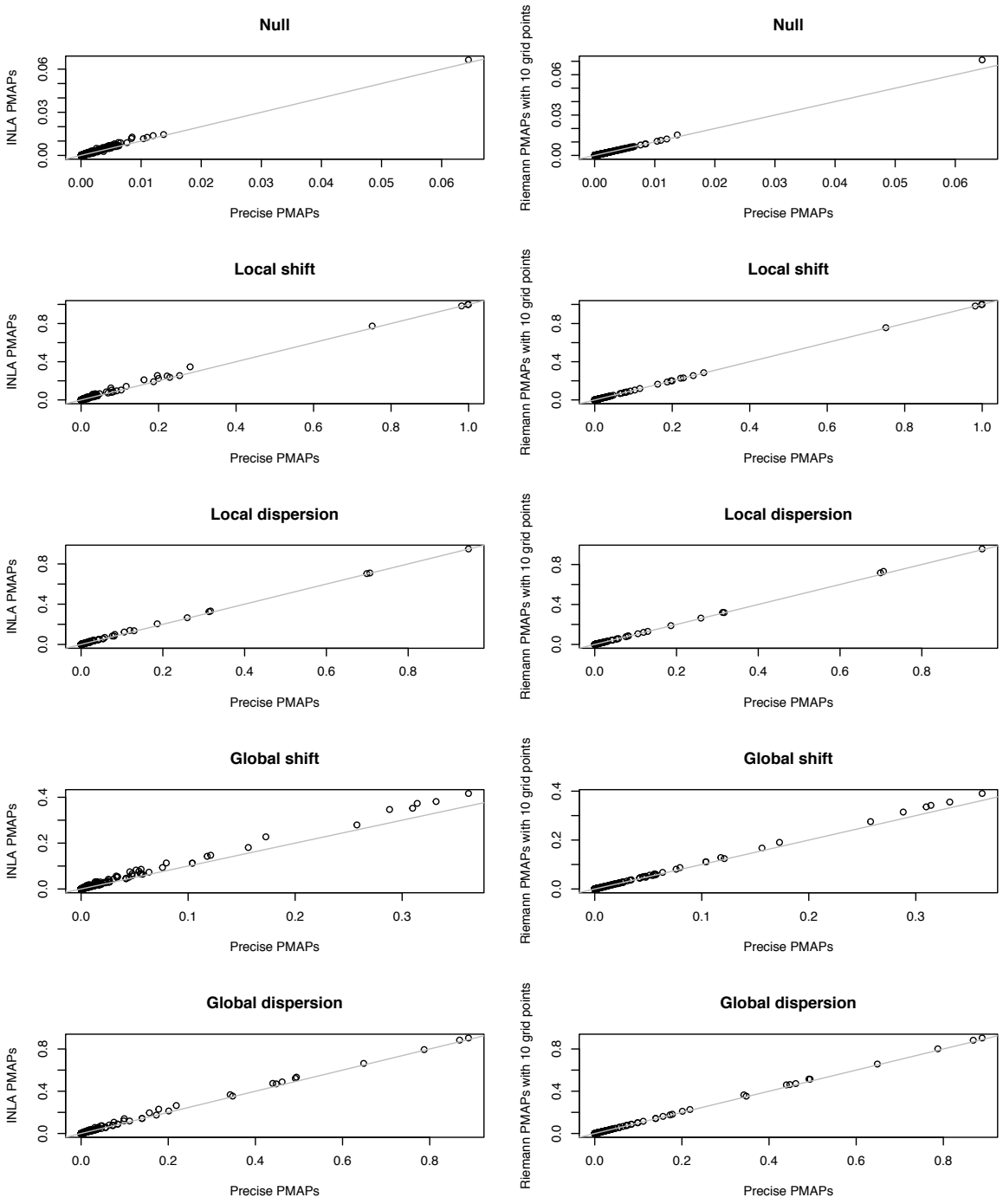

Figure S5: Accurate PMAPs versus approximate PMAPs under five simulation scenarios. The gray line is the 45 degree line. Points along this line indicates perfect approximation. Left: accurate PMAPs versus INLA PMAPs. Right: accurate PMAPs versus Rieman PMAPs with 10 grid points on $\nu(A)$.

S6. Illustration of posterior sampling

We draw posterior samples from the full posterior, and plot the corresponding effect size of the Medullo D721 group versus the baseline Leukemia K562 group for 25 posteriors draws.

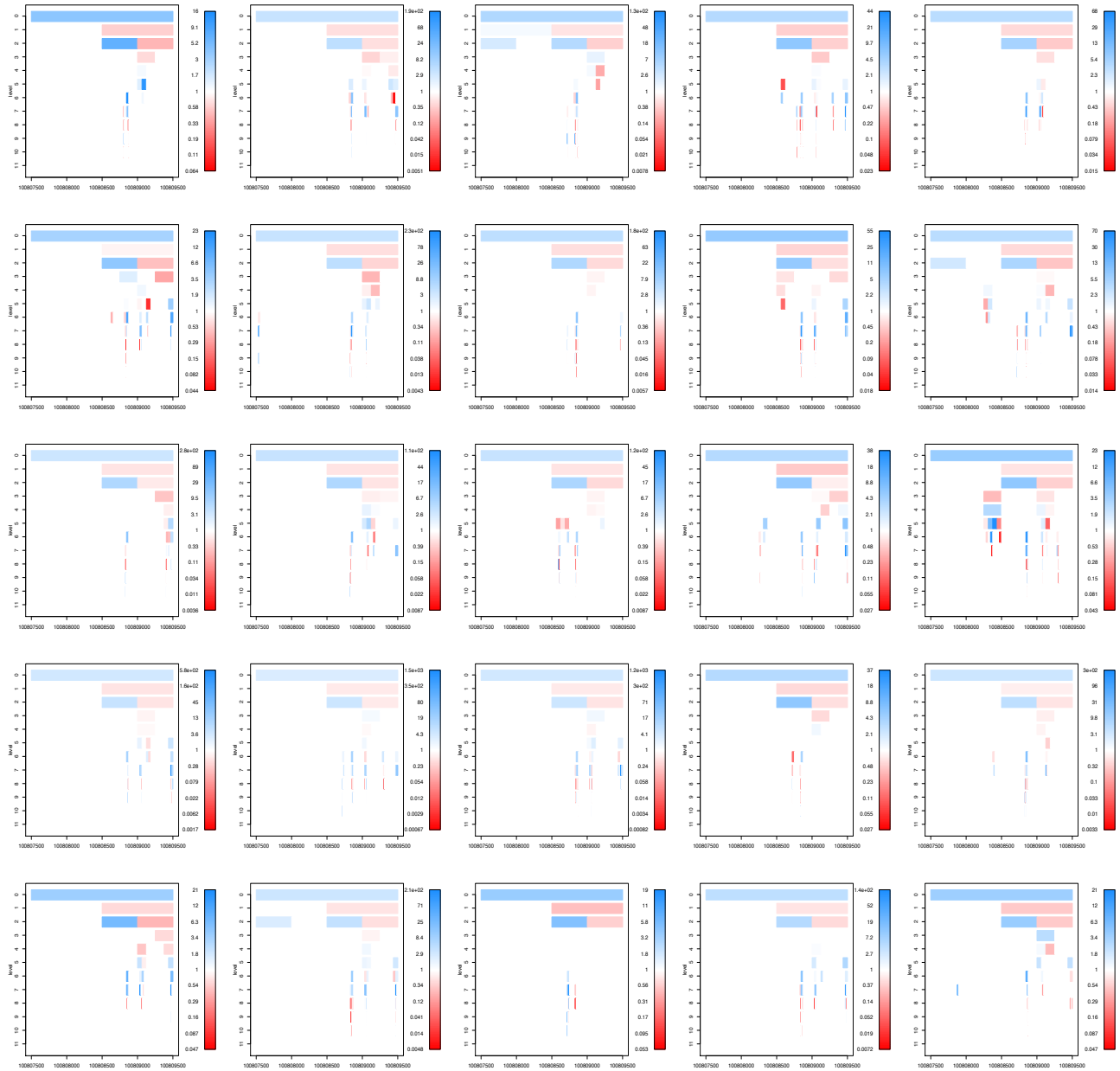

Figure S6: Posterior draws of the effect size of Medullo D721 versus Leukemia K562.