Supplementary Materials for

PairedFB: a full hierarchical Bayesian model for paired RNA-seq data with heterogeneous treatment effects

Yuanyuan Bian[1], Chong He[1], Jie Hou[2], Jianlin Cheng[2], and Jing Qiu[*3]

[1]Department of Statistics, University of Missouri, Columbia, Missouri, 65211.

[2]Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, Missouri, 65211.

[3]Department of Applied Economics and Statistics, University of Delaware, Newark, Delaware, 19716.

Contents

*qiujing@udel.edu

1 Supplementary Discussion

1.1 Transformation of the marginal distribution of $Y_i^g \mid Y_i^g + Y_i^{'g}$

To ensure the convergence when calculating the likelihood, we use the following transformation formulas for ${}_2F_1(a, b; c; z)$ (Abramowitz and Stegun, 1964, pp. 559):

$$
\begin{aligned}
{}_2F_1(a, b; c; z) &= (1-z)^{c-a-b}\,{}_2F_1(c-a, c-b; c; z) && (1)\\
&= (1-z)^{-b}\,{}_2F_1(b, c-a; c; \frac{z}{1-z}), && (2)
\end{aligned}
$$

and rewrite the likelihood function by:

$$
\begin{aligned}
& p(Y_i^g \mid Y_i^g + Y_i^{'g}, \pi^g, \phi^g) \\
= \; & \frac{(Y_i^g + Y_i^{'g})!}{Y_i^g! Y_i^{'g}!} \frac{B\left(Y_i^g + \pi^g \frac{1-\phi^g}{\phi^g}, Y_i^{'g} + (1-\pi^g)\frac{1-\phi^g}{\phi^g}\right)}{B\left(\pi^g \frac{1-\phi^g}{\phi^g}, (1-\pi^g)\frac{1-\phi^g}{\phi^g}\right)} \times \qquad (3)\\
& \begin{cases}
\left(\frac{L_i}{L_i'}\right)^{(1-\pi^g)\frac{1-\phi^g}{\phi^g}} \; {}_2\mathrm{F}_1\left(\frac{1-\phi^g}{\phi^g},\; Y_i^{'g} + (1-\pi^g)\frac{1-\phi^g}{\phi^g};\; Y_i^g + Y_i^{'g} + \frac{1-\phi^g}{\phi^g};\; 1 - \frac{L_i}{L_i'}\right) & \text{when } L_i < L_i' \\
\left(\frac{L_i'}{L_i}\right)^{\pi^g \frac{1-\phi^g}{\phi^g}} \; {}_2\mathrm{F}_1\left(\frac{1-\phi^g}{\phi^g},\; Y_i^g + \pi^g \frac{1-\phi^g}{\phi^g};\; Y_i^g + Y_i^{'g} + \frac{1-\phi^g}{\phi^g};\; 1 - \frac{L_i'}{L_i}\right) & \text{when } L_i \geq L_i'.
\end{cases}
\end{aligned}
$$

The two expressions guarantees the convergence of hypergeometric function under $L_i < L_i'$ and $L_i \geq L_i'$, respectively. Through simple algebra we can see that, if $L_i = L_i'$, the distribution would reduce to a beta-binomial$(Y_i^g + Y_i^{'g}; \pi^g, \phi^g)$.

1.2 Mean and variance of $Y_i^g \mid Y_i^g + Y_i^{'g}$

The mean and variance of Y_i^g given the total count can be derived through relationship between $p_i^g = \frac{\pi_i^g L_i}{\pi_i^g L_i + (1-\pi_i^g)L_i'}$ and π_i^g. Note that when π_i^g follows beta(α^g, β^g), then p_i^g would follow a generalized three-parameter beta distribution introduced by and denoted by G3B$(\alpha^g, \beta^g, \lambda_i)$, where $\lambda_i = \frac{L_i'}{L_i}$. The pdf of G3B$(\alpha, \beta, \lambda)$ is given by

$$
p(p|\alpha, \beta, \lambda) = \frac{\lambda^\alpha p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)[1-(1-\lambda)p]^{\alpha+\beta}}.
$$

The derivation is given as below. If $X \sim$ G3B(α, β, λ), the mth moment can be derived

from:

$$
\begin{aligned}
E(X^m) &= \int_0^1 \frac{\lambda^\alpha x^{m+\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)[1-(1-\lambda)x]^{\alpha+\beta}}\mathrm{d}x && (4)\\
&= \frac{\lambda^\alpha}{B(\alpha,\beta)}\int_0^1 \frac{x^{m+\alpha-1}(1-x)^{\beta-1}}{[1-(1-\lambda)x]^{\alpha+\beta}}\mathrm{d}x && (5)\\
&= \frac{B(m+\alpha,\beta)}{B(\alpha,\beta)}\lambda^\alpha\ {}_2\mathrm{F}_1(\alpha+\beta, m+\alpha; m+\alpha+\beta; 1-\lambda) && (6)\\
&= \frac{\Gamma(\alpha+m)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta+m)}\ {}_2\mathrm{F}_1(m,\beta; m+\alpha+\beta; 1-\lambda). && (7)
\end{aligned}
$$

Note that for large $\alpha,\ \beta$, we have $\frac{\Gamma(\beta+i)}{\Gamma(\beta)} \approx \beta^i$ and $\frac{\Gamma(\alpha+\beta+m+i)}{\Gamma(\alpha+\beta+m)} \approx (\alpha+\beta)^i$, thus:

$$
\begin{aligned}
{}_2\mathrm{F}_1(m,\beta; m+\alpha+\beta; 1-\lambda) &= \sum_{i=0}^{\infty}\frac{\Gamma(\beta+i)\Gamma(\alpha+\beta+m)\Gamma(m+i)}{\Gamma(\beta)\Gamma(\alpha+\beta+m+i)\Gamma(m)}\cdot\frac{(1-\lambda)^i}{i!} && (8)\\
&\approx \sum_{i=0}^{\infty}\frac{\Gamma(m+i)}{\Gamma(m)\Gamma(i+1)}\frac{\beta^i(1-\lambda)^i}{(\alpha+\beta)^i} && (9)\\
&= \sum_{i=0}^{\infty}\binom{m-1+i}{i}\left[\frac{(1-\lambda)\beta}{\alpha+\beta}\right]^i && (10)\\
&= \left[\frac{1}{1-\frac{(1-\lambda)\beta}{\alpha+\beta}}\right]^m = \left[\frac{\alpha+\beta}{\alpha+\lambda\beta}\right]^m. && (11)
\end{aligned}
$$

As a result,

$$
E(X^m) \approx \frac{\Gamma(\alpha+m)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta+m)}\left[\frac{\alpha+\beta}{\alpha+\lambda\beta}\right]^m.
$$

So that the first and second moment can be directly approximated as:

$$
\begin{aligned}
E(X) &\approx \frac{\alpha}{\alpha+\lambda\beta} && (12)\\
E(X^2) &\approx \frac{\alpha(\alpha+1)(\alpha+\beta)}{(\alpha+\beta+1)(\alpha+\lambda\beta)^2}, && (13)
\end{aligned}
$$

which gives

$$
Var(X) \approx \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\lambda\beta)^2}.
$$

Coming back to the likelihood model for our method, it becomes

$$
\begin{aligned}
E(Y_{ij}^g \mid Y_{ij}^g + Y_{ij}^{'g}, \pi_i^g, \phi_i^g) &= E[E(Y_{ij}^g|p_{ij}^g)] = (Y_{ij}^g + Y_{ij}^{'g})E(p_{ij}^g) && (14)\\
&\approx (Y_{ij}^g + Y_{ij}^{'g})\frac{\alpha_i^g}{\alpha_i^g + \lambda_{ij}\beta_i^g} && (15)\\
&= \frac{(Y_{ij}^g + Y_{ij}^{'g})\pi_i^g}{\pi_i^g + (1-\pi_i^g)\lambda_{ij}}; && (16)
\end{aligned}
$$

and

$$
\begin{aligned}
Var(Y_{ij}^g \mid Y_{ij}^g + Y_{ij}^{'g}, \pi_i^g, \phi_i^g) &= E[Var(Y_{ij}^g|p_{ij}^g)] + Var[E(Y_{ij}^g|p_{ij}^g)] && (17)\\
&= E[(Y_{ij}^g + Y_{ij}^{'g})p_{ij}^g(1-p_{ij}^g)] + Var[(Y_{ij}^g + Y_{ij}^{'g})p_{ij}^g] && (18)\\
&\approx \frac{(Y_{ij}^g + Y_{ij}^{'g})\alpha_i^g\beta_i^g[\lambda_{ij}(\alpha_i^g+\beta_i^g+1)-1] + (Y_{ij}^g + Y_{ij}^{'g})^2\alpha_i^g\beta_i^g}{(\alpha_i^g+\lambda_{ij}\beta_i^g)^2(\alpha_i^g+\beta_i^g+1)} && (19)\\
&= (Y_{ij}^g + Y_{ij}^{'g})\frac{\alpha_i^g\beta_i^g}{(\alpha_i^g+\lambda_{ij}\beta_i^g)^2}\left[\frac{Y_{ij}^g + Y_{ij}^{'g}-1}{(\alpha_i^g+\beta_i^g+1)}+1\right] && (20)\\
&= \frac{(Y_{ij}^g + Y_{ij}^{'g})\pi_i^g(1-\pi_i^g)L_{ij}^2}{[\pi_i^g L_{ij} + (1-\pi_i^g)L'_{ij}]^2}[(Y_{ij}^g + Y_{ij}^{'g}-1)\phi_i^g+1]; && (21)
\end{aligned}
$$

Then using the result of moments of G3B distribution above, for large α^g and β^g, we get approximation of the mean and variance of data as:

$$
\begin{aligned}
E(Y_i^g \mid Y_i^g + Y_i^{'g}, \pi^g, \phi^g) &\approx (Y_i^g + Y_i^{'g})\frac{\pi^g L_i}{\pi^g L_i + (1-\pi^g)L'_i}; && (22)\\
Var(Y_i^g \mid Y_i^g + Y_i^{'g}, \pi^g, \phi^g) &\approx (Y_i^g + Y_i^{'g})[(Y_i^g + Y_i^{'g}-1)\phi^g+1]\frac{\pi^g(1-\pi^g)L_i^2}{[\pi^g L_i + (1-\pi^g)L'_i]^2} && (23)
\end{aligned}
$$

1.3 MCMC algorithm

We would build a hybrid Gibbs sampler to get the posterior samples $(\pi^{*g}, \phi^{*g}, \gamma^g)$ for each gene and $(p, \mu_0, \mu, \sigma_0^2, \sigma^2)$, where we use Gibbs sampler for components with full conditional distributions available for direct sampling and Metropolis-Hasting algorithm for parameters with complicated full conditional distributions.

Since the model selection parameter π^{*g} has prior involving mixture of point mass and is difficult to be integrated out in presence of complex likelihood, to avoid reducibility of Markov chain, we re-parameterize it into the likelihood with the model indicator γ^g by:

$$\pi^{*g} = I(\gamma^g = 1)\tilde{\pi}^{*g},$$

where $\tilde{\pi}^{*g}$ is a free parameter following Gaussian prior $N(\mu_0, \sigma_0^2)$. The likelihood can therefore be rewritten as:

$$p(\mathbf{Y}^g|\pi^{*g}, \phi^g) = p(\mathbf{Y}^g|\tilde{\pi}^{*g}, \gamma^g, \phi^g), \quad (24)$$

with prior of $(\tilde{\pi}^{*g}, \gamma^g)$:

$$p(\tilde{\pi}^{*g}, \gamma^g) = N(\tilde{\pi}^{*g}|\mu_0, \sigma_0^2)p^{\gamma^g}(1-p)^{1-\gamma^g}. \quad (25)$$

We prove the equivalence between the original parameterization and the re-parameterization as below.

Proof. To validate the above re-parameterization, we essentially have to make sure $p(\gamma^g|\mathbf{Y})$ is the same. Since:

$$p(\gamma^g|\mathbf{Y}) = \int p(\gamma^g|\mathbf{p},\mathbf{Y})p(\mathbf{p}|\mathbf{Y})\mathrm{d}\mathbf{p} \propto \int p(\mathbf{Y}^g|\gamma^g)p(\gamma^g|\mathbf{p})p(\mathbf{p}|\mathbf{Y})\mathrm{d}\mathbf{p}, \tag{26}$$

while other terms remain the same, we only have to prove both parameterization give the same $p(\mathbf{Y}^g|\gamma^g)$.

Note that, for the original parameterization:

$$p(\mathbf{Y}^g|\gamma^g) = \iint p(\mathbf{Y}^g|\pi^{*g},\phi^{*g})p(\pi^{*g}|\gamma^g)\mathrm{d}\pi^{*g}p(\phi^{*g})\mathrm{d}\phi^{*g},$$

where

$$\int p(\mathbf{Y}^g|\pi^{*g},\phi^{*g})p(\pi^{*g}|\gamma^g)\mathrm{d}\pi^{*g} = \begin{cases} \int p(\mathbf{Y}^g|\pi^{*g},\phi^{*g})N(\pi^{*g}|\mu_0,\sigma_0^2)\mathrm{d}\pi^{*g} & \gamma^g = 1 \\ p(\mathbf{Y}^g|0,\phi^{*g}) & \gamma^g = 0 \end{cases}$$

On the other hand, for the re-parameterized form:

$$p(\mathbf{Y}^g|\gamma^g) = \iint p(\mathbf{Y}^g|\tilde{\pi}^{*g},\gamma^g,\phi^{*g})p(\tilde{\pi}^{*g})\mathrm{d}\tilde{\pi}^{*g}p(\phi^{*g})\mathrm{d}\phi^{*g},$$

where

$$\int p(\mathbf{Y}^g|\tilde{\pi}^{*g},\gamma^g,\phi^{*g})p(\tilde{\pi}^{*g})\mathrm{d}\tilde{\pi}^{*g} = \begin{cases} \int p(\mathbf{Y}^g|\tilde{\pi}^{*g},\phi^{*g})N(\tilde{\pi}^{*g}|\mu_0,\sigma_0^2)\mathrm{d}\tilde{\pi}^{*g} & \gamma^g = 1 \\ p(\mathbf{Y}^g|0,\phi^{*g})\int N(\tilde{\pi}^{*g}|\mu_0,\sigma_0^2)\mathrm{d}\tilde{\pi}^{*g} & \gamma^g = 0 \end{cases}$$

So the equivalence of the two parameterizations is established. □

After such re-parameterization, the model parameters now become $\tilde{\mathbf{\Theta}}^{*g} = (\tilde{\pi}^{*g},\gamma^g,\phi^{*g})$, where our main interest is the posterior samples of γ^g and $\tilde{\pi}^{*g}$ is an auxiliary parameter that has nothing to do with posterior inference but makes the sampling procedure feasible without complicated calculation of integral.

With the new parameterization $\tilde{\pi}^{*g}\gamma^g = \pi^{*g} = \log\left(\frac{\pi^g}{1-\pi^g}\right)$, and $\phi^{*g} = \log\left(\frac{\phi^g}{1-\phi^g}\right)$, based on the original likelihood shown in the main text Equation (5), we can now rewrite the

likelihood function for each gene as:

$$p(\mathbf{Y}^g|\tilde{\pi}^{*g},\gamma^g,\phi^{*g}) \propto \frac{B\left[Y_i^g+(1+e^{-\tilde{\pi}^{*g}\gamma^g})^{-1}e^{\phi^{*g}},Y_i'^g+(1+e^{-\tilde{\pi}^{*g}\gamma^g})^{-1}e^{\phi^{*g}-\tilde{\pi}^{*g}\gamma^g}\right]}{B\left[(1+e^{-\tilde{\pi}^{*g}\gamma^g})^{-1}e^{\phi^{*g}},(1+e^{-\tilde{\pi}^{*g}\gamma^g})^{-1}e^{\phi^{*g}-\tilde{\pi}^{*g}\gamma^g}\right]} \times {}_2F_1\left(Y_i^g+Y_i'^g,Y_i^g+(1+e^{-\tilde{\pi}^{*g}\gamma^g})^{-1}e^{\phi^{*g}};Y_i^g+Y_i'^g+e^{\phi^{*g}};1-\frac{L_i'}{L_i}\right), \quad (27)$$

Then the joint posterior of $(\tilde{\mathbf{\Theta}}^*,\boldsymbol{\eta})$ is given by:

$$p(\tilde{\mathbf{\Theta}}^*,p,\boldsymbol{\eta}|\mathbf{Y}) \propto \left\{\prod_{g=1}^{G}p(\mathbf{Y}^g|\tilde{\mathbf{\Theta}}^*,p)\cdot p(\tilde{\mathbf{\Theta}}^{*g}|\boldsymbol{\eta})\right\}\pi(p)\pi(\boldsymbol{\eta}) \quad (28)$$

$$\propto \prod_{g=1}^{G}\left[p(\mathbf{Y}^g|\tilde{\pi}^{*g},\gamma^g,\phi^{*g})N(\tilde{\pi}^{*g}|\mu_0,\sigma_0^2)N(\phi^{*g}|\mu,\sigma^2)p^{\gamma^g}(1-p)^{1-\gamma^g}\right] \times p^{\delta_1-1}(1-p)^{\delta_2-1}\times e^{-\frac{\mu^2}{2}}\times(\sigma_0^2)^{-a_0-1}e^{-\frac{b_0}{\sigma_0^2}}\times(\sigma^2)^{-a-1}e^{-\frac{b}{\sigma^2}}, \quad (29)$$

with $\delta_1=\delta_2=1$, $a_0=b_0=a=b=0.5$.

Thus, for each gene g, $g=1\ldots G$, the full conditionals of model parameters $\tilde{\mathbf{\Theta}}^{*g}=(\tilde{\boldsymbol{\pi}}^{*g},\gamma^g,\phi^{*g})$ are given by:

$$P(\gamma^g=1|\cdots) = \frac{p(\mathbf{Y}^g|\tilde{\pi}^{*g},\gamma^g=1,\phi^{*g})p}{p(\mathbf{Y}^g|\tilde{\pi}^{*g},\gamma^g=1,\phi^{*g})p+p(\mathbf{Y}^g|\tilde{\pi}^{*g},\gamma^g=0,\phi^{*g})(1-p)} \quad (30)$$

$$p(\tilde{\pi}^{*g}|\cdots) \propto I(\gamma^g=1)p(\mathbf{Y}^g|\tilde{\pi}^{*g},\phi^{*g})N(\tilde{\pi}^{*g}|\mu_0,\sigma_0^2) \quad (31)$$
$$+I(\gamma^g=0)N(\tilde{\pi}^{*g}|\mu_0,\sigma_0^2)$$

$$p(\phi^{*g}|\cdots) \propto p(\mathbf{Y}^g|\gamma^g,\tilde{\pi}^{*g},\phi^{*g})N(\phi^{*g}|\mu,\sigma^2). \quad (32)$$

The parameter of DE proportion is updated by:

$$p(p|\cdots)\sim\text{Beta}\left(\delta_1+\sum_{g=1}^{G}I(\gamma^{g(m)}=1),\delta_2+\sum_{g=1}^{G}I(\gamma^{g(m)}=0)\right). \quad (33)$$

And the full conditional for the hyper-parameters are:

$$p(\mu_0|\cdots) \sim N\left(\bar{\tilde{\pi}}^*,\frac{\sigma_0^2}{G}\right) \quad (34)$$

$$p(\mu|\cdots) \sim N\left(\bar{\phi}^*,\frac{\sigma^2}{G}\right) \quad (35)$$

$$p(\sigma_0^2|\cdots) \sim IG\left(\frac{G}{2}+a_0,\frac{\sum_{g=1}^{G}(\tilde{\pi}^{*g}-\mu_0)^2}{2}+b_0\right) \quad (36)$$

$$p(\sigma^2|\cdots) \sim IG\left(\frac{G}{2}+a,\frac{\sum_{g=1}^{G}(\phi^{*g}-\mu)^2}{2}+b\right), \quad (37)$$

where $\bar{\tilde{\pi}}^*=\frac{1}{G}\sum_{g=1}^{G}\tilde{\pi}^{*g}$, $\bar{\phi}^*=\frac{1}{G}\sum_{g=1}^{G}\phi^{*g}$.

So we conclude the MCMC procedure by following steps after giving initial values. For the m-th iteration:

Step 1. For the g-th gene, $g = 1 \dots G$:

(a) Sample $\gamma^{g(m)}$ directly from Bernoulli(pp^g), where

$$pp^g = \frac{p(\mathbf{Y}^g|\tilde{\pi}^{*g(m-1)}, \phi^{*g(m-1)})p^{(m-1)}}{p(\mathbf{Y}^g|\tilde{\pi}^{*g(m-1)}, \phi^{*g(m-1)})p^{(m-1)} + p(\mathbf{Y}^g|0, \phi^{*g(m-1)})(1-p^{(m-1)})};$$

(b) Sample $\tilde{\pi}^{*g(m)}$ from the full conditional distribution in (31) by Metropolis-Hasting algorithm:

- sample the proposal $\tilde{\pi}^{*g(prop)}$ from $N(\mu_0^{(m-1)}, \sigma_0^{2(m-1)})$,
- calculate ratio $r_1 = \frac{p(\mathbf{Y}^g|\tilde{\pi}^{*g(prop)}, \phi^{*g(m-1)}, \gamma^{g(m)})}{p(\mathbf{Y}^g|\tilde{\pi}^{*g(m-1)}, \phi^{*g(m-1)}, \gamma^{g(m)})}$
- update $\tilde{\pi}^{*g(m)} = \begin{cases} \tilde{\pi}^{*g(prop)} & \text{if } r_1 > r^* \\ \tilde{\pi}^{*g(m-1)} & \text{if } r_1 \leq r^* \end{cases}$, where r^* is sampled from Unif(0,1).

(c) Sample $\phi^{*g(m)}$ from the full conditional distribution in equation (32) by Metropolis-Hasting algorithm:

- sample the proposal $\phi^{*g(prop)}$ from $N(\mu^{(m-1)}, \sigma^{2(m-1)})$,
- calculate ratio $r_2 = \frac{p(\mathbf{Y}^g|\tilde{\pi}^{*g(m)}, \phi^{*g(prop)}, \gamma^{g(m)})}{p(\mathbf{Y}^g|\tilde{\pi}^{*g(m)}, \phi^{*g(m-1)}, \gamma^{g(m)})}$
- update $\phi^{*g(m)} = \begin{cases} \phi^{*g(prop)} & \text{if } r_2 > r^* \\ \phi^{*g(m-1)} & \text{if } r_2 \leq r^* \end{cases}$, where r^* is sampled from Unif(0,1).

Step 2. Sample $p^{(m)}$ from Beta $\left(\delta_1 + \sum_{g=1}^G I(\gamma^{g(m)} = 1), \delta_2 + \sum_{g=1}^G I(\gamma^{g(m)} = 0)\right)$.

Step 3. Sample $\mu_0^{(m)}$ from $N\left(\bar{\tilde{\pi}}^{*(m)}, \frac{\sigma_0^{2(m-1)}}{G}\right)$.

Step 4. Sample $\mu^{(m)}$ from $N\left(\bar{\phi}^{*(m)}, \frac{\sigma^{2(m-1)}}{G}\right)$.

Step 5. Sample $\sigma_0^{2(m)}$ from $IG\left(\frac{G}{2} + a_0, \frac{\sum_{g=1}^G (\tilde{\pi}^{*g(m)} - \mu_0^{(m)})^2}{2} + b_0\right)$.

Step 6. Sample $\sigma^{2(m)}$ from $IG\left(\frac{G}{2} + a, \frac{\sum_{g=1}^G (\phi^{*g(m)} - \mu^{(m)})^2}{2} + b\right)$.

1.4 Proof of the non-decreasing monotonicity of the estimated peFDR

First, denote the ranked (from largest to smallest) l-th estimated posterior probability of $\gamma^g = 1$ as $\hat{p}^{(l)}$, that is $\hat{p}^{(l)} \geq \hat{p}^{(l+1)}$. We want to prove

$$\hat{\text{peFDR}}^{(l)} \leq \hat{\text{peFDR}}^{(l+1)}, \tag{38}$$

where $\hat{\text{peFDR}}^{(l)} = \frac{\sum_{i=1}^{l}(1-\hat{p}^{(i)})}{l}$.

Proof. Since $\hat{\text{peFDR}}^{(l+1)} = \frac{\sum_{i=1}^{l}(1-\hat{p}^{(i)})+(1-\hat{p}^{(l+1)})}{l+1}$, in order to prove

$$\frac{\sum_{i=1}^{l}(1-\hat{p}^{(i)})}{l} \leq \frac{\sum_{i=1}^{l}(1-\hat{p}^{(i)})+(1-\hat{p}^{(l+1)})}{l+1}, \tag{39}$$

we need to show

$$\sum_{i=1}^{l}(1-\hat{p}^{(i)}) \leq l(1-\hat{p}^{(l+1)}), \tag{40}$$

i.e. $\sum_{i=1}^{l}\hat{p}^{(i)} \geq l \times \hat{p}^{(l+1)}$.

This is true because $\hat{p}^{(1)} \geq \ldots \geq \hat{p}^{(l)} \geq \hat{p}^{(l+1)}$ indicates $\sum_{i=1}^{l}\hat{p}^{(i)} \geq l \times \hat{p}^{(l+1)}$. □

1.5 The simulation setting

For the gth gene, the paired read counts of the ith pair Y_i^g and $Y_i^{\prime g}$ are sampled from two independent Poisson distributions as follows:

$$Y_i^g \sim \text{Poi}(\lambda_i^g L_i \pi_i^g) \tag{41}$$

$$Y_i^{\prime g} \sim \text{Poi}(\lambda_i^g L_i^{'}(1-\pi_i^g)). \tag{42}$$

By setting the normalization factor at 1 for all libraries, we simulate the library sizes L_i and L_i' from discrete uniform distribution from 8 to 23 million (These numbers were obtained from the range of the estimated library sizes of the human tumor dataset.) When analyzing the simulated data, we use the true simulated library sizes for all the competitive methods in order to eliminate the effect of normalization. The parameters μ_i^g and $\mu_i^{\prime g}$ are generated by the equations $\mu_i^g = \lambda_i^g \pi_i^g$ and $\mu_i^{\prime g} = \lambda_i^g(1-\pi_i^g)$, respectively. Note that $\mu_i^g + \mu_i^{\prime g} = \lambda_i^g$ and $\pi_i^g = \frac{\mu_i^g}{\mu_i^g + \mu_i^{\prime g}}$. Here λ_i^g is the total expression levels of gene g in pair i after adjusting for

the library size, which is sampled from the empirical distribution of the methods of moment estimators based on human tumor dataset. Specifically, it is easy to obtain $\lambda_i^g = \frac{EY_i^g}{L_i} + \frac{EY_i'^g}{L_i'}$. Thus the moment estimator of λ_i^g would be

$$\hat{\lambda}_i^g = \frac{y_i^g}{\hat{L}_i} + \frac{y_i'^g}{\hat{L}_i'}, \tag{43}$$

where y_i^g and $y_i'^g$ are the observed read counts of treatment and control group for gene g and pair i, $\hat{L}_i$ and $\hat{L}_i'$ are the estimated library sizes using the weighted trimmed mean of M-values (TMM) method by Robinson and Oshlack (2010).

The parameter π_i^g is the proportion of the treatment mean out of the overall pair mean for the i-th pair and is generated by $\pi_i^g \sim \text{beta}\left(\pi^g\left[\frac{1-\phi^g}{\phi^g}\right], (1-\pi^g)\left[\frac{1-\phi^g}{\phi^g}\right]\right)$, allowing treatment effects to vary among different pairs. The average treatment effect π^g is set to be $\frac{1}{2}$ for non-DE genes so that $\pi^{*g} = 0$, and π^{*g} of DE genes (i.e. π_{DE}^{*g}) is simulated from various settings including parametric and empirical distributions. The parameter ϕ^g controls the heterogeneity level of treatment effects and is also simulated from parametric or empirical settings. Note that this Poisson-beta simulation setting can also arise from the negative binomial or Poisson-gamma mixture setting, see the Supplementary Discussion Section 1.6 for detailed discussion.

We generate π_{DE}^{*g} and ϕ^g from five different distribution combinations.

Setting E: Both π_{DE}^{*g} and ϕ^g are drawn from the empirical distribution (see Supplementary Figures S1, S2 for density curves of the empirical estimates of π^{*g} and ϕ^{*g}).

Settings SE1 and SE2: ϕ^g is drawn from empirical distribution but π_{DE}^{*g} is drawn from a scaled empirical distribution with a scalar of $\sqrt{5}$ and $\sqrt{10}$, respectively.

Settings P1 and P2: $\pi_{DE}^{*g} \sim N(0, 5 \times 0.68^2)$, $\phi^g \sim \text{beta}(1, b)$ where $b = 12$ and 5, respectively.

The first three setting are the empirical settings where **E** represents empirical and **SE** represent scaled empirical. The last two settings are the parametric settings, which are represented by **P**. For settings SE1 and SE2, we scale the empirical distribution for π_{DE}^{*g} without changing the mean. In other words, we generate π_{DE}^{*g} from empirical values of $\hat{\pi}_{new}^{*g} = \sqrt{a}(\hat{\pi}_o^{*g} - \hat{\mu}_o) + \hat{\mu}_o$,

where $\hat{\pi}_o^{*g}$ is the original estimates from real data, $\hat{\mu}_o$ is the mean of $\hat{\pi}_0^{*g}$s' and $\sqrt{a}$ is a scalar to amplify the percentage of large signals. The larger a is, the larger percentage of large signal it produces. See supplementary Figure S1 for the density curves of $\hat{\pi}_{new}^{*g}$ under various values of a and note $\hat{\pi}_{new}^{*g} = \hat{\pi}_o^{*g}$ when $a = 1$. For the parametric settings P1 and P2, the normal variance $\sigma^2 = 0.68$ is estimated from the human tumor dataset by Tuch *et al.* (2010)and we amplify the signal by scaling by $\sqrt{5}$. The overdispersion parameter phi^g is sampled from beta$(1, b)$ where beta$(1, 12)$ tries to match with the empirical distribution of ϕ^g and beta$(1, 5)$ represents the case when there is higher level of heterogeneity in the treatment effects (see Supplementary Figure S2). For each of the five settings, we consider six different combinations of sample size and proportion of DE genes. In particular, we consider three sample sizes ($n = 3, 5, 10$) and two proportions of DE genes ($p = 10\%, 20\%$).

To estimate empirical values of π^g from human tumor dataset, i.e. the mean treatment effect of gene g across all the libraries, we use the fact that $\pi^g = E(\pi_i^g) = E\left(\frac{\mu_i^g}{\mu_i^g + \mu_i'^g}\right)$. Since $\hat{\pi}_i^g = \frac{\hat{\mu}_i^g}{\hat{\mu}_i^g + \hat{\mu}_i'^g} = \frac{y_i^g/\hat{L}_i}{y_i^g/\hat{L}_i + y_i'^g/\hat{L}_i'}$, the moment estimator of π^g is given by:

$$\hat{\pi}^g = \bar{\hat{\pi}}_i^g = \frac{1}{n}\sum_{i=1}^{n} \hat{\pi}_i^g \tag{44}$$

The overdispersion parameter ϕ^g is defined based on the reparameterization of Beta distribution as $\phi^g = \frac{Var(\pi_i^g)}{E(\pi_i^g)[1-E(\pi_i^g)]}$. and therefore its moment estimator would be:

$$\hat{\phi}^g = \frac{\widehat{Var}(\hat{\pi}_i^g)}{\bar{\hat{\pi}}_i^g(1 - \bar{\hat{\pi}}_i^g)}. \tag{45}$$

We also check the robustness of the proposed full Bayesian model by generating data using a different likelihood from our model assumption. Please see Supplementary Discussion Section 1.7 for details.

1.6 Poisson-gamma mixture can result in our Poisson-beta simulation setting

To see this, let

$$\begin{aligned} Y_i^g|\ \mu_i^g &\sim \text{Poi}(\mu_i^g L_i) \\ Y_i'^g|\ \mu_i'^g &\sim \text{Poi}(\mu_i'^g L_i'). \end{aligned}$$

If $\mu_i^g = \lambda_i^g \theta_i^g$ and $\mu_i'^g = \lambda_i^g \theta_i'^g$ with $\theta_i^g \sim Ga(a^g, b^g)$ and $\theta_i'^g \sim Ga(a'^g, b^g)$, then $\pi_i^g = \frac{\mu_i^g}{\mu_i^g + \mu_i'^g} = \frac{\theta_i^g}{\theta_i^g + \theta_i'^g} \sim \text{beta}(a^g, a'^g)$.

1.7 Robustness studies for the PairedFB procedure when the likelihood function is misspecified

In the main text, we have considered the simulation studies under the beta binomial framework with different prior distributions for the parameters ϕ^g and π_{DE}^{*g}. To check the robustness of the proposed full Bayesian model, we follow the strategy of Hardcastle and Kelly (2013) to replace the beta distribution of π_i^g by the minimax distribution. Specifically, instead of generating π_i^g from beta $\left(\pi^g \left[\frac{1-\phi^g}{\phi^g}\right], (1-\pi^g)\left[\frac{1-\phi^g}{\phi^g}\right]\right)$, we sample π_i^g from the minimax distribution (Jones, 2004) with density as below:

$$f(x|a,b) = abx^{a-1}(1-x^a)^{b-1}, \quad 0 < x < 1 \tag{46}$$

for $a > 0,\ b > 0$. The minimax distribution has the mean of $bB(1+1/a, b)$ and the variance of $bB(1+2/a, b) - [bB(1+1/a, b)]^2$. We match the mean and variance used in the beta distribution for the minimax distribution by numerically solving a and b according to the value of π^g and ϕ^g sampled from either empirical or parametric setting.

For robustness test, we considered the two scaled empirical settings SE1 and SE2 for π^g and ϕ^g. The results are shown in Figures S25 to S32. Based on the FD plots and ROC plots of these figures, the PairedFB is the best procedure in terms of ranking the true DE genes on the top and Chung's method being the least favorable. The advantage of PairedFB over other procedures becomes more pronounced for larger sample sizes and larger proportion of DE genes. As expected, the PairedFB fails to control FDR at nominal levels due to model misspecification in many cases. However, the actual FDR of our method remains the lowest among all the methods for almost all cases. Note for small sample size ($n = 3$) and low DE proportion (top left panels of Figure S27-S28, S31-S32), all the methods fail to deliver reasonable results in terms of FDR and power. As the sample size or DE proportion increases, the actual FDR of our method converges to the nominal level (see bottom panels of Figure S27-S28, S31-S32). Under such simulation settings, the limma voom method cannot control FDR at small sample sizes anymore and still has the undesirable property of increased FDR

level with the sample size. The result shows that our procedure is very robust in terms of its relative performance with existing methods. We would suggest to be cautious on the analysis result for all the methods when sample size is small, as no statistical magic would happen with only few information.

1.8 The computational problems of `ShrinkBayes` package

We also try to compare with the `ShrinkBayes` package proposed by Van De Wiel *et al.* (2013) by building generalized linear mixed model (GLMM) with random pair effect and assuming data follow Poisson distribution. However, we have encountered numerous computational problems with this software. First all, we failed to install this software (version 2.13.4) in at least three computers including a university linux server, one Macbook air and one Mac pro. We were finally able to install it successfully in one Windows PC (with 4 CPUs) and conducted the simulation study using `ShrinkBayes`. However, we kept getting error messages such as failure in convergence of the Newton-Ralphson optimizer. This was explained by the installation guide of the software, "when running ShrinkBayes you may see *** WARNINGS *** from INLA (e.g. on eigenvalues, or on convergence, or even something like 18500 Aborted...). They can currently not be surpressed, because they are produced by C-code. Please ignore them." Based on 50 runs of simulation, we decided to exclude this method in our comparison since the results consistently showed overly large false positive rate (e.g. the method would declare 80% DE genes at 5% nominal level when the true DE proportion is only 10%). We are not sure such a high false positive rates is due to the method itself or due to the internal error of the INLA package used by the software. After contacting the author of ShrinkBayes, we also tried to fit the GLMM with Negative Binomial distribution for the data, but the program was not responsive after 4 iterations in the ShrinkSeq step mainly because of the memory limit of the computer. Due to all the problems we encountered with using ShrinkSeq, we decided not to include the results of ShrinkBayes (Van De Wiel *et al.*, 2013) in our comparison for the simulation study and real data analysis.

1.9 Negative binomial model can accommodate heterogeneous treatment effect implicitly

The negative binomial model can actually accommodate the heterogeneous treatment effects implicitly to some extent. One can see this by viewing the negative binomial distributions arising from the Poisson-gamma mixture distribution. Let

$$\begin{aligned} Y_i^g|\ \mu_i^g &\sim \text{Poi}(\mu_i^g L_i) \\ Y_i'^g|\ \mu_i'^g &\sim \text{Poi}(\mu_i'^g L_i'). \end{aligned}$$

Here $\mu_i^g = \lambda^g Q_i^g \theta_i^g$ and $\mu_i'^g = \lambda'^g Q_i^g \theta_i'^g$, with λ^g and λ'^g the baseline expression level of gene g in the first and second treatment after adjusting for the library size, respectively, and Q_i^g the pair effect. If $\theta_i^g \sim Ga(a^g, b^g)$ and $\theta_i'^g \sim Ga(a'^g, b'^g)$, then the treatment effect $\pi_i^g = \frac{\mu_i^g}{\mu_i^g + \mu_i'^g} = \frac{\lambda^g \theta_i^g}{\lambda^g \theta_i^g + \lambda'^g \theta_i'^g}$ is different for each pair i.

1.10 Difference between our real data analysis using edgeR and baySeq and the original papers

The number of DEGs detected by the paired baySeq according to our data analysis (1528) is quite different from what was reported in Hardcastle and Kelly (2013). The difference is due to different ways of formulating the problem. Here in our data analysis, we aim to test whether the fold-change between tumor and normal tissue is one by treating patients as replicates under the same model structure, while Hardcastle and Kelly (2013) tested the consistent fold change across patients by treating the patients as one factor of the experiment. The number of DEGs detected by the edgeR based on our data analysis (1269) is close to what was obtained for the same data analysis in edgeR user's guide of V3.12.1 (1272) and the minor inconsistency is due to different data preprocessing results (note in edgeR user's guide, they keep observations with the largest number of total read counts if there's any duplicated gene symbol).

1.11 Comparing PairedFB with competing methods at comparable FDR levels for real data analysis

Now, we compare the results of PairedFB with edgeR or baySeq at comparable FDR level. Since our simulation studies show that the actual FDRs of the paired baySeq and edgeR are highly inflated for sample size of 3, it is not fair to compare them with our method at the same nominal level. Instead, we want to compare our method at 5% nominal level with edgeR or paired baySeq at a comparable FDR level. Comparable FDR level means that the resulting actual FDR levels are comparable. Although for real data, we do not know the truth and cannot evaluate the actual FDR, we can use the simulation results to guide our choice of nominal FDR levels. Recall our simulation studies show the FD plots and the ROC curves of the three methods are similar for sample size 3, which implies that their actual FDR would be comparable when their powers are comparable. In other words, when the three methods detect similar amount of DEGs, their actual FDRs should be comparable. Therefore, we fix the nominal level of the PairedFB at 5% and pick nominal level of 3% for the paired baySeq and 4% for edgeR such that they detect similar amount of DEGs (as highlighted in bold in main Table 1). Such a choice is non-favorable to our method based on the simulation studies since it has been shown that the smallest actual FDR of the other two methods at nominal level 3% or 4% would be inflated to at least 7% when the sample size is 3, while the actual FDR of PairedFB is always close to or slightly less than 5% at nominal level 5%. One purpose of doing such a comparison is to remind readers that one can trust the nominal FDR level of PairedFB, while they need to be cautious about interpreting the nominal FDR level of the edgeR or paired baySeq method.

When the nominal levels of the PairedFB, the edgeR and the paired baySeq are set to be 5%, 4% and 3% respectively, they detect similar amounts of DEGs (around 1200) with a large percentage of overlapping genes (1043 overlapping DEGs are detected by all three methods, see the Venn diagram in Figure S40). Although they detect a lot of common genes at comparable FDR levels, there are still genes uniquely identified by each method. The top 20 enriched gene sets based on the overlapping genes detected by both PairedFB and edgeR (or baySeq) methods are mostly cancer or carcinoma related (bolded in Supplementary

Table S5), where the "RICKMAN HEAD AND NECK CANCER F" gene set is ranked among top 10. Therefore, the list of overlapping genes detected by both PairedFB and edgeR (or baySeq) is a reliable list with important biological significance, to which we could compare the top 20 enriched gene sets based on the DEG list of uniquely identified genes by either method. Supplementary Figure S41 draws the Venn diagram of the top 20 enriched gene sets based on DEGs detected by both PairedFB and edgeR (or baySeq), by PairedFB only, and by edgeR (or baySeq), respectively. The Venn diagrams in Supplementary Figure S41 show that the "PairedFB only" has 35% and 15% enriched gene sets in its top 20 list that are overlapped with those based on the commonly detected genes with edgeR and baySeq, most of which are top ranked enriched gene sets; while "edgeR only" and "baySeq only" have 10% and 5% overlapping with those based on the commonly detected genes, respectively. Furthermore, one important gene set "RICKMAN HEAD AND NECK CANCER C", which is directly related to the biological target of the dataset, is only enriched in list of genes uniquely identified by PairedFB at top 10th (see third column in Supplementary Table S5) when comparing with baySeq. This indicates the DEGs identified by PairedFB only at nominal level 5% are biologically more consistent with the DEGs identified by both PairedFB and edgeR (or baySeq) and tends to uniquely identify more biological related genes when compare with other competing methods at lower nominal levels.

2 Figures

2.1 Empirical distribution of the method of moment estimators for π^{*g} and ϕ^{*g}.

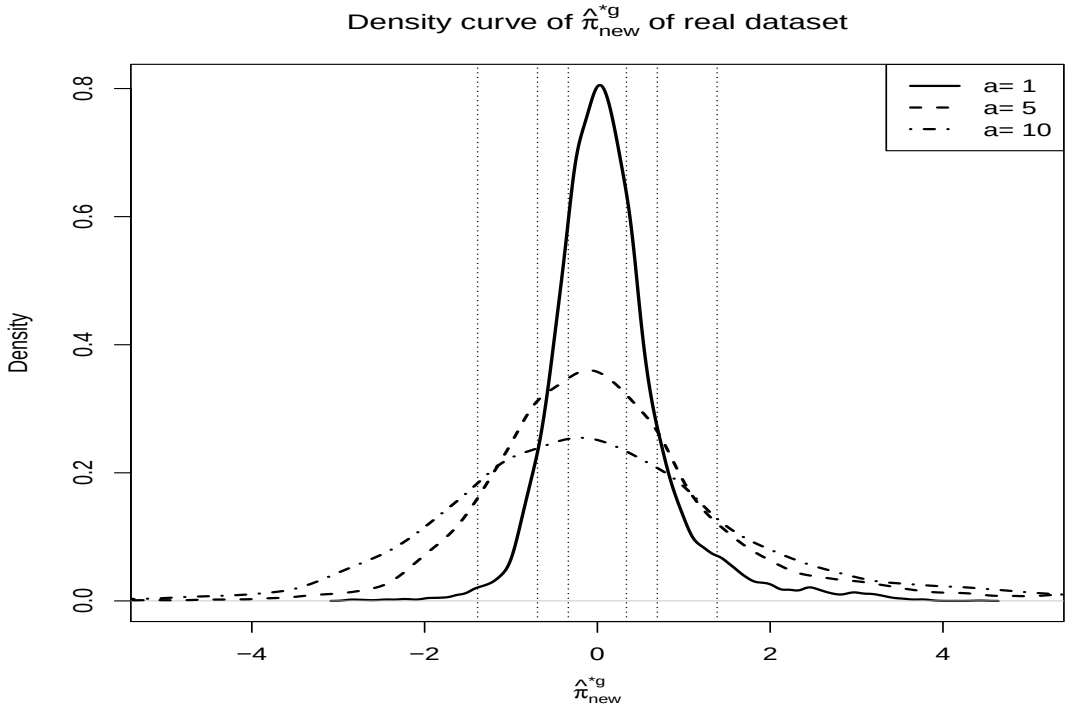


Figure S1: The scaled density curves of the method of moment estimates of π^{*g} based on the human tumor data by Tuch *et al.* (2010) with various scaling values ($sqrta = 1, \sqrt{5}, \sqrt{10}$). The vertical dotted lines indicate Fold Change $= 1.4, 2$, and 4, respectively.

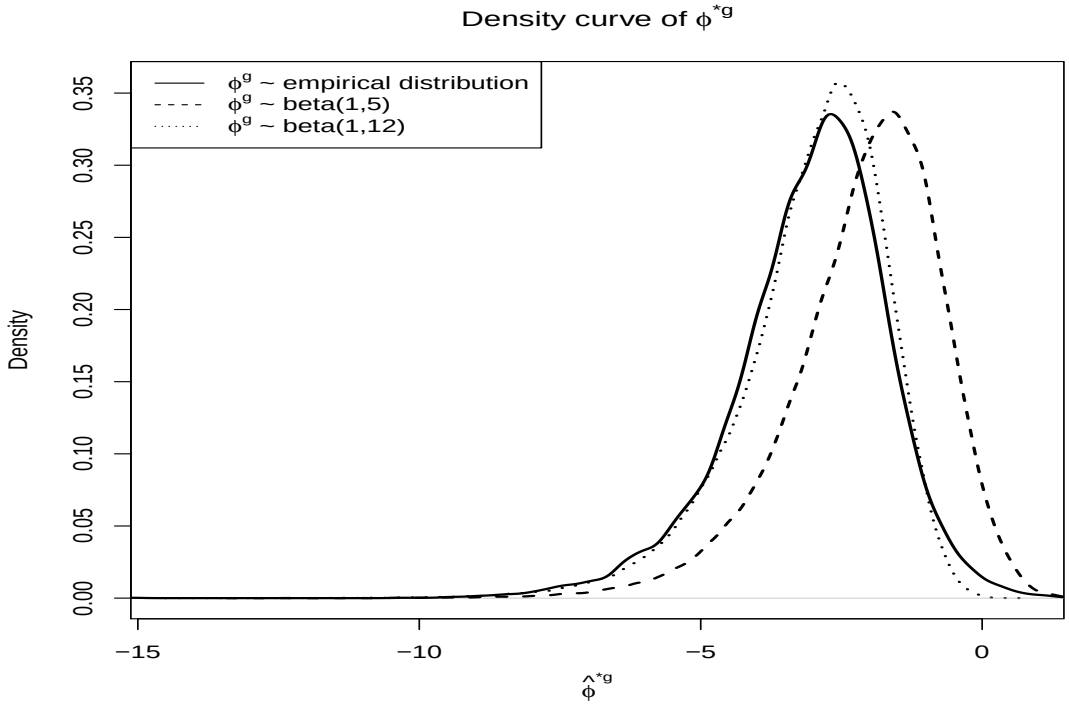


Figure S2: Density curves of ϕ^{*g} (logit of ϕ^g) when ϕ^g is sampled from the empirical values of the human tumor data by Tuch *et al.* (2010) and the parametric distributions of Beta(1, 5) and Beta(1, 12), respectively

2.2 Simulation results for identifying statistical significant DE genes

2.2.1 Setting SE1: π_{DE}^{*g} is drawn from empirical distribution scaled by $\sqrt{5}$ and ϕ^g is drawn from empirical distribution

Actual FDR and Power comparison at nominal level 10% under setting I

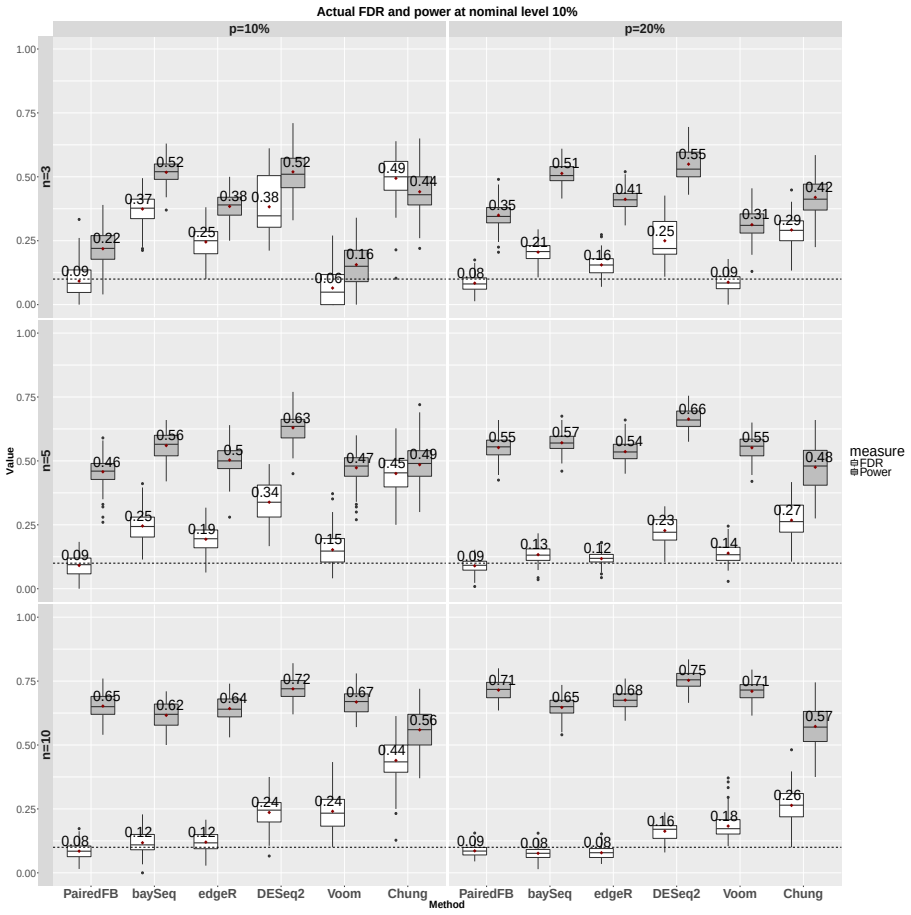


Figure S3: The boxplots of actual false discovery proportion and true detection rate over 100 simulations at nominal level 10% comparing all methods. Subplots in each row represent results for different sample sizes ($n = 3, 5, 10$) and column for different DE proportions ($p = 10\%, 20\%$). ϕ^g and π^{*g} of DE genes (π_{DE}^{*g}) are sampled from Setting SE1. The numbers besides each box indicate the mean value over 100 simulations. The horizontal black dashed line in each plot outlines the 10% nominal level.

2.2.2 Setting SE2: π_{DE}^{*g} is drawn from empirical distribution scaled by $\sqrt{10}$ and ϕ^g is drawn from empirical distribution

Small DE proportion $p = 10\%$ under Setting SE2

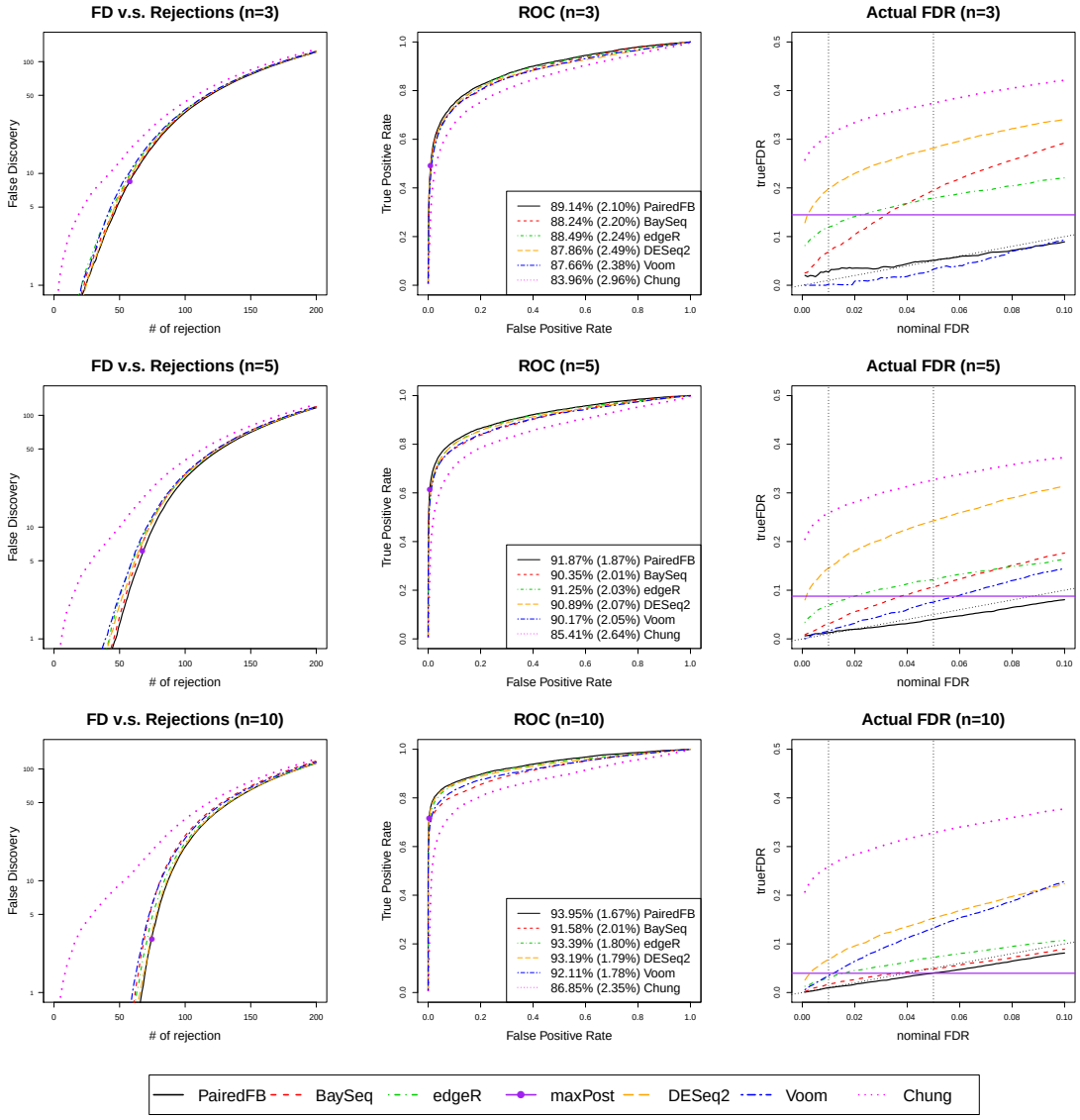


Figure S4: False discovery plots, ROC curves and the actual FDR curves comparing all methods under Setting SE2 when DE proportion $p = 10\%$. In the third column, the black dotted line indicates the true nominal FDR level and two vertical dotted lines indicates 1% and 5% nominal level.

Large DE proportion $p = 20\%$ under Setting SE2

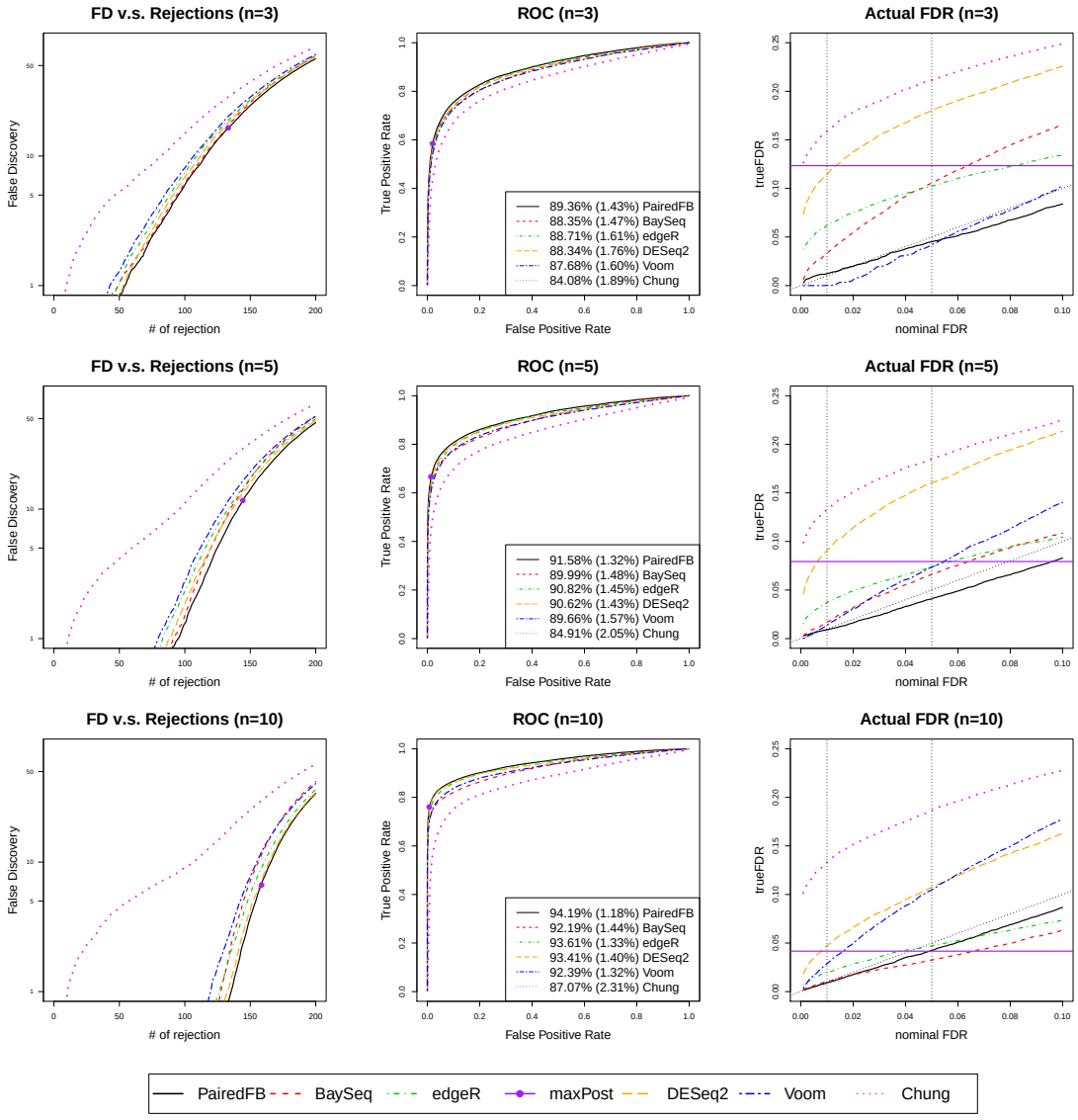


Figure S5: False discovery plots, ROC curves and the actual FDR curves comparing all methods under Setting SE2 when DE proportion $p = 20\%$. In the third column, the black dotted line indicates the true nominal FDR level and two vertical dotted lines indicates 1% and 5% nominal level.

Actual FDR and Power comparison at nominal level 5% under Setting SE2

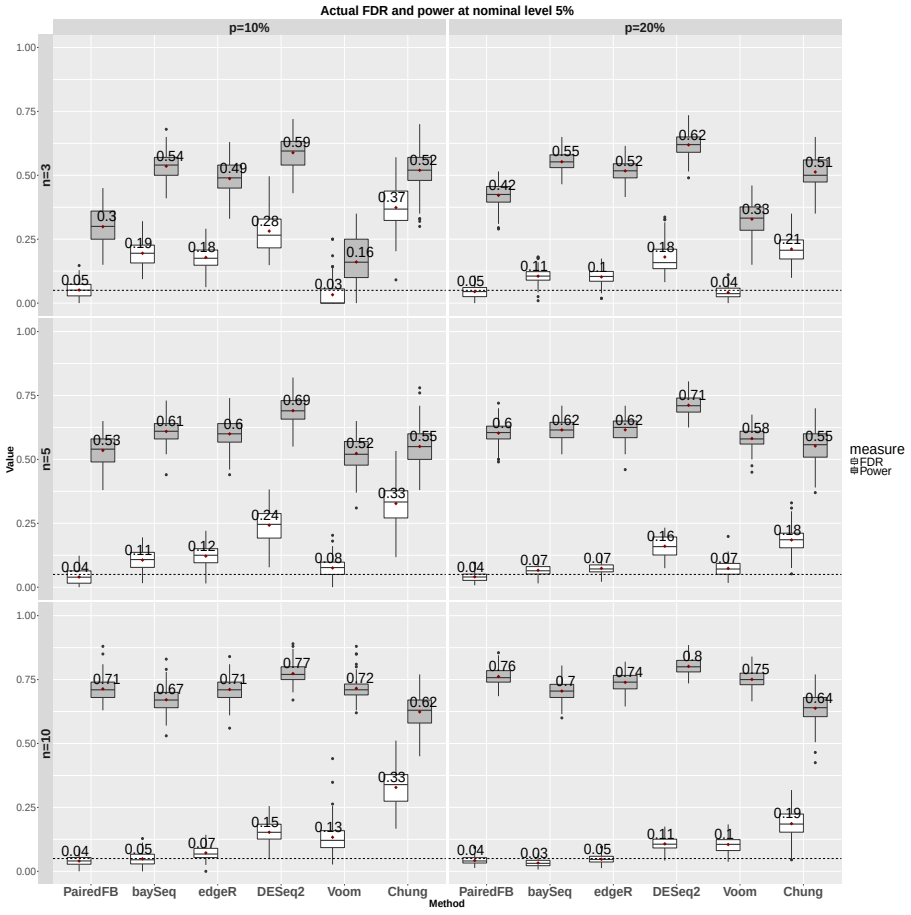


Figure S6: The box plots of the actual false discovery proportion and true detection rate over 100 simulations at nominal level 5% comparing all methods. Subplots in each row represent results for different sample sizes ($n = 3, 5, 10$) and column for different DE proportions ($p = 10\%, 20\%$). ϕ^g and π^{*g} of DE genes (π^{*g}_{DE}) are sampled under Setting SE2. The numbers besides each box indicate the mean value over 100 simulations. The horizontal black dashed line in each plot outlines the 5% nominal level.

Actual FDR and Power comparison at nominal level 10% under Setting SE2

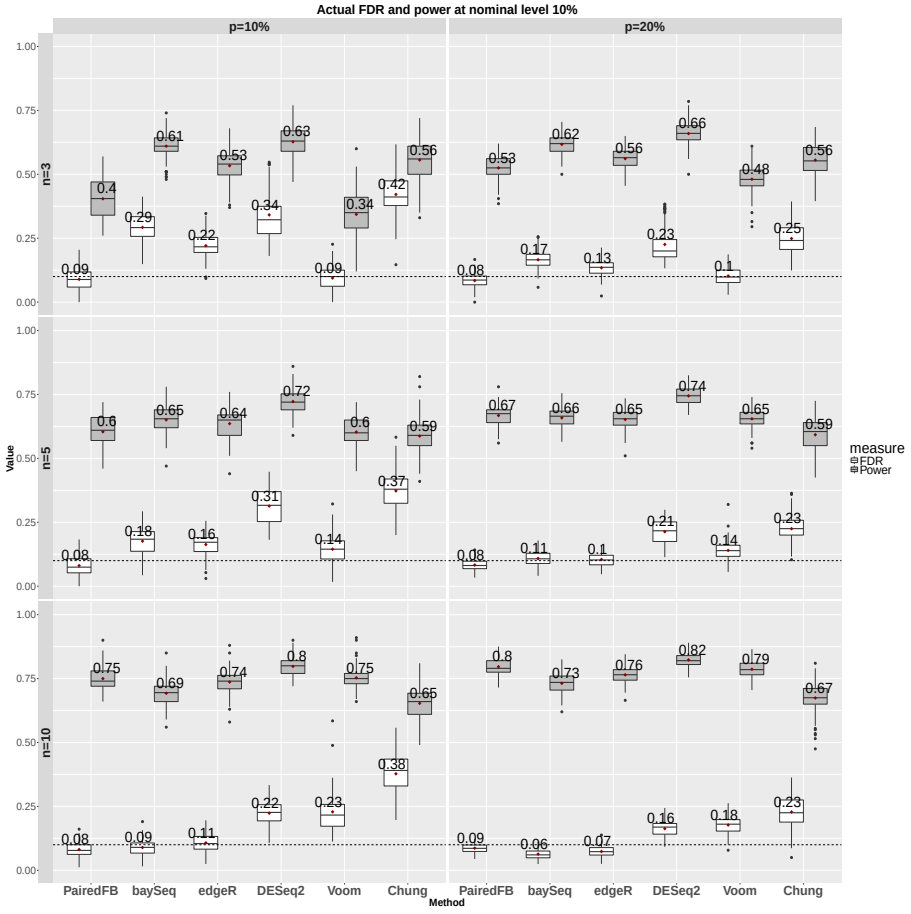


Figure S7: The box plots of the actual false discovery proportion and true detection rate over 100 simulations at nominal level 10% comparing all methods. Subplots in each row represent results for different sample sizes ($n = 3, 5, 10$) and column for different DE proportions ($p = 10\%, 20\%$). ϕ^g and π^{*g} of DE genes (π^{*g}_{DE}) are sampled under Setting SE2. The numbers besides each box indicate the mean value over 100 simulations. The horizontal black dashed line in each plot outlines the 10% nominal level.

2.2.3 Setting E: π_{DE}^{*g} is drawn from the original empirical distribution and ϕ^g is drawn from empirical distribution

Small DE proportion $p = 10\%$ under Setting E

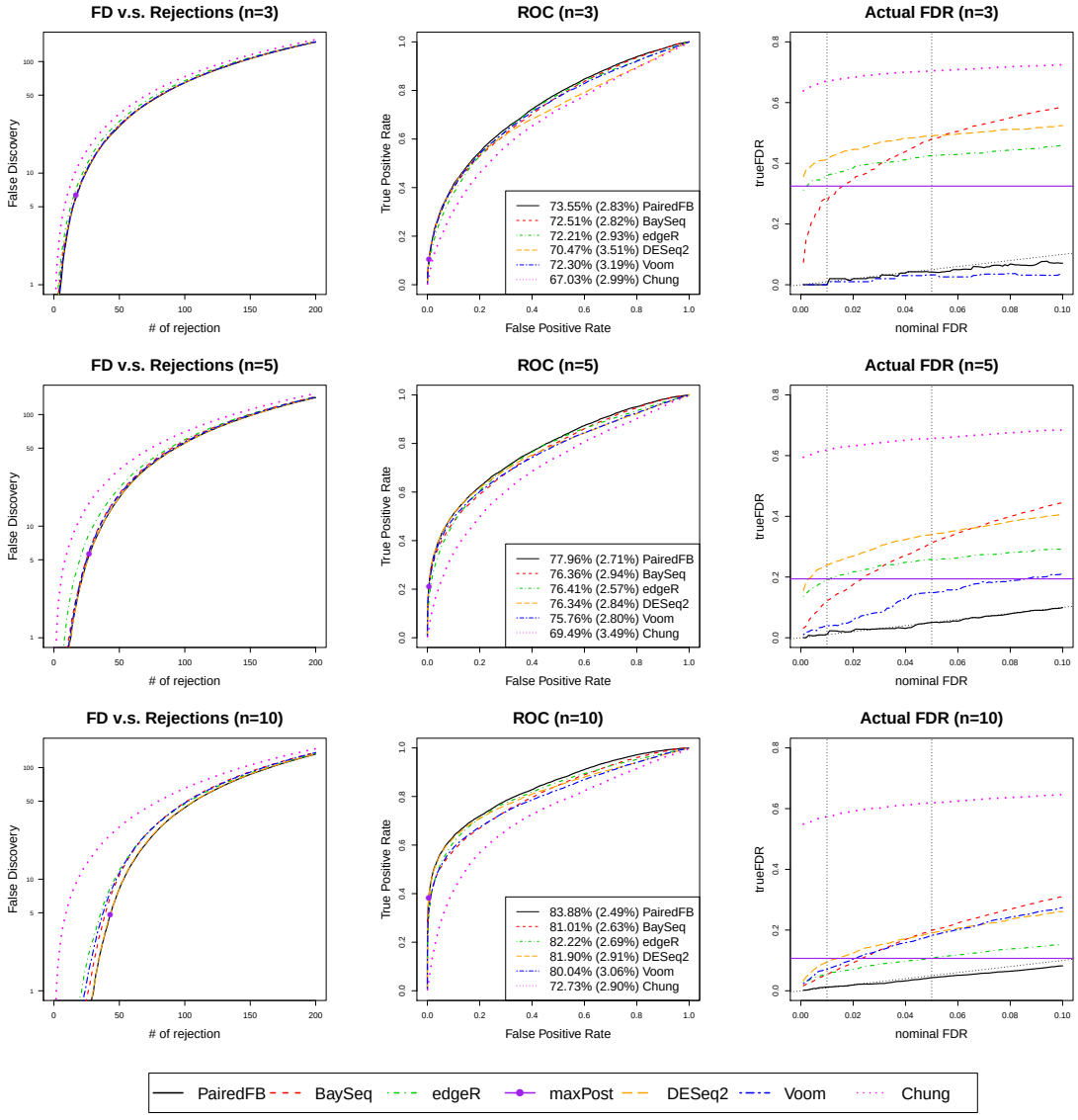


Figure S8: False discovery plots, ROC curves and the actual FDR curves comparing all methods under Setting E when DE proportion $p = 10\%$. In the third column, the black dotted line indicates the true nominal FDR level and two vertical dotted lines indicates 1% and 5% nominal level.

Large DE proportion $p = 20\%$ under Setting E

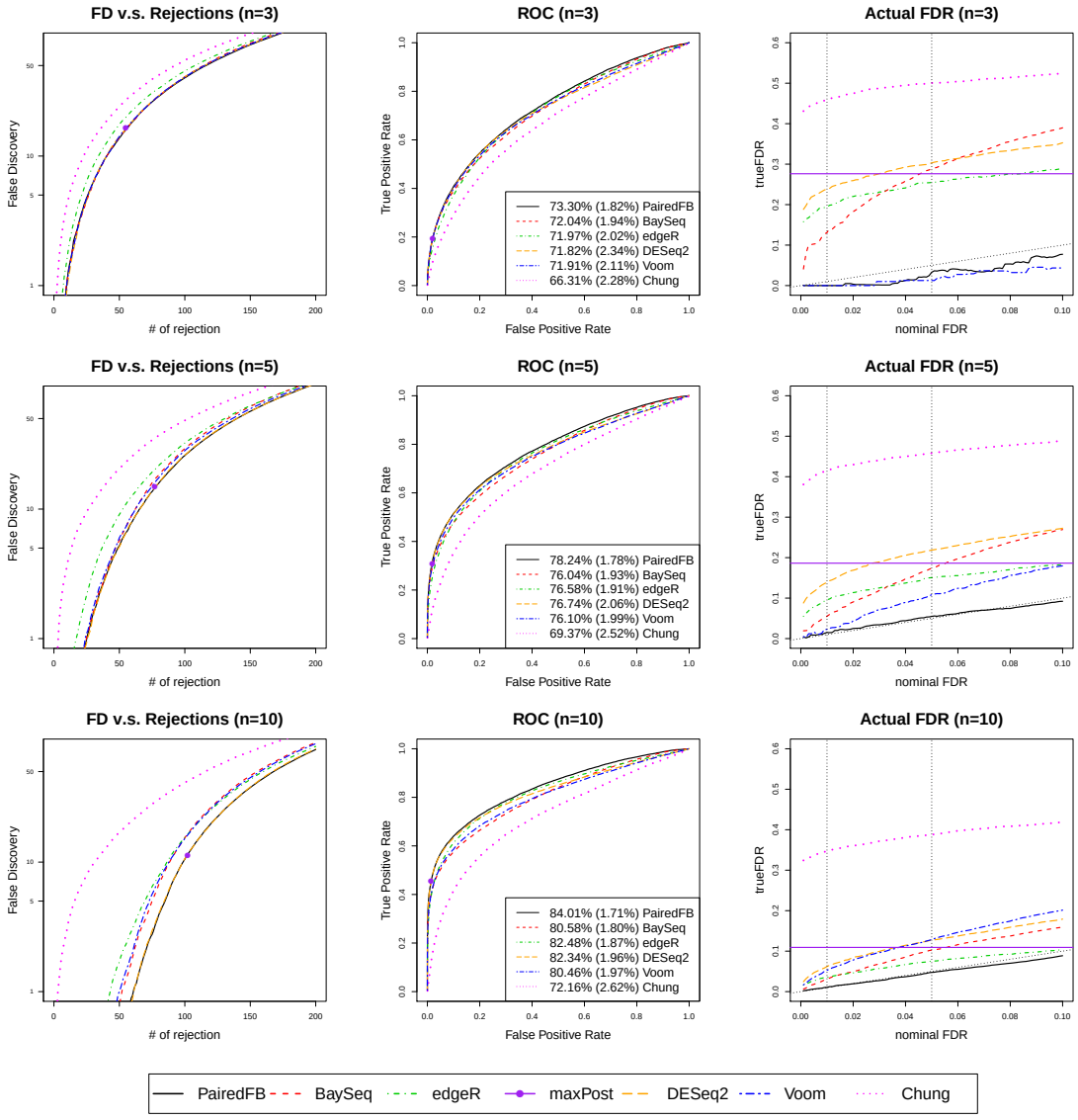


Figure S9: False discovery plots, ROC curves and the actual FDR curves comparing all methods under Setting E when DE proportion $p = 20\%$. In the third column, the black dotted line indicates the true nominal FDR level and two vertical dotted lines indicates 1% and 5% nominal level.

Actual FDR and Power comparison at nominal level 5% under Setting E

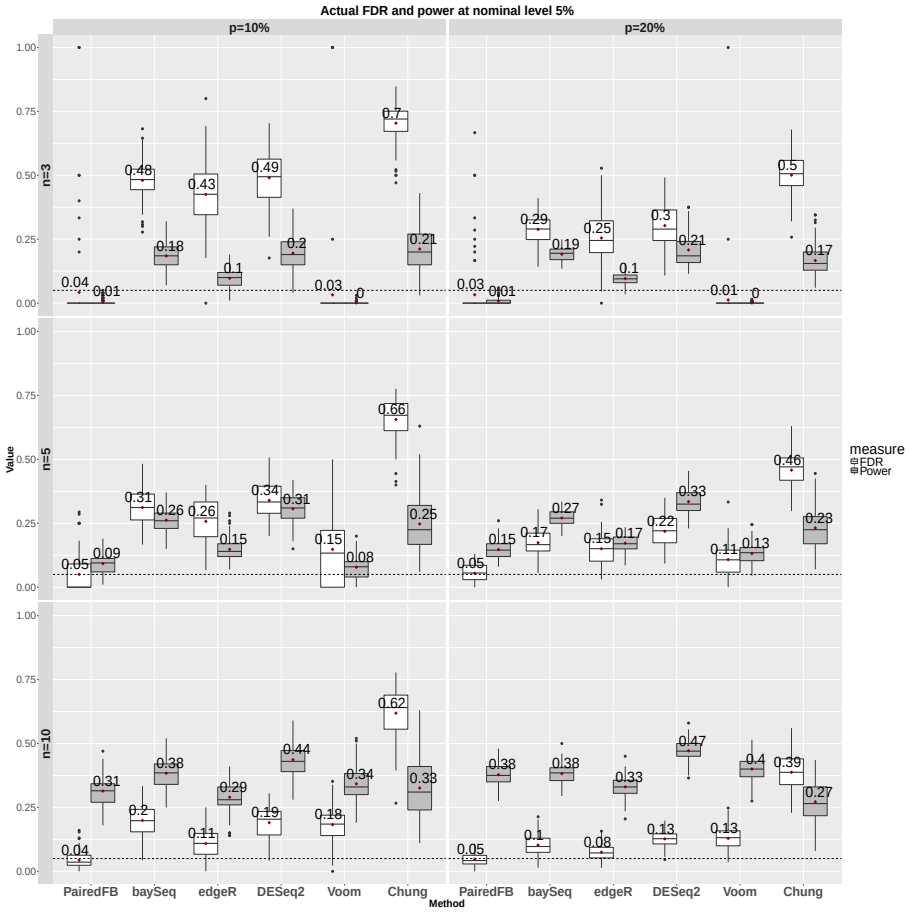


Figure S10: The box plots of the actual false discovery proportion and true detection rate over 100 simulations at nominal level 5% comparing all methods. Subplots in each row represent results for different sample sizes ($n = 3, 5, 10$) and column for different DE proportions ($p = 10\%, 20\%$). ϕ^g and π^{*g} of DE genes (π^{*g}_{DE}) are sampled under Setting E. The numbers besides each box indicate the mean value over 100 simulations. The horizontal black dashed line in each plot outlines the 5% nominal level.

Actual FDR and Power comparison at nominal level 10% under Setting E

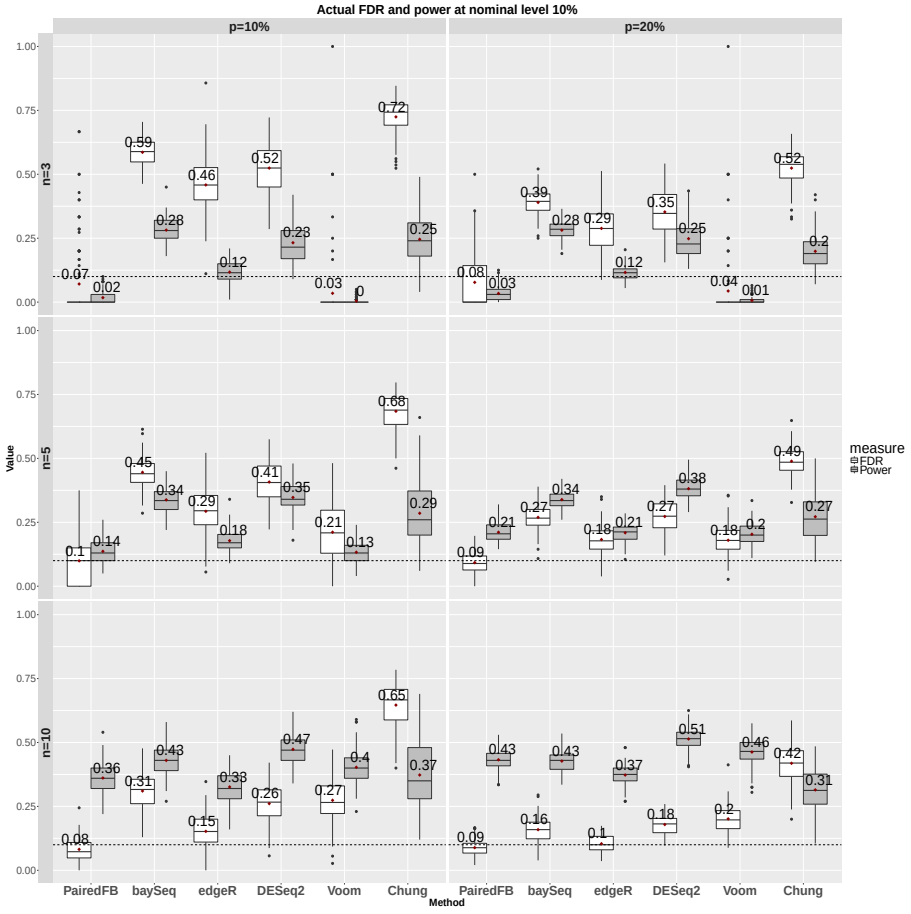


Figure S11: The box plots of the actual false discovery proportion and true detection rate over 100 simulations at nominal level 10% comparing all methods. Subplots in each row represent results for different sample sizes ($n = 3, 5, 10$) and column for different DE proportions ($p = 10\%, 20\%$). ϕ^g and π^{*g} of DE genes (π^{*g}_{DE}) are sampled under Setting E. The numbers besides each box indicate the mean value over 100 simulations. The horizontal black dashed line in each plot outlines the 10% nominal level.

2.2.4 Setting P1: $\pi_{DE}^{*g} \sim N(0, 5 \times 0.68^2)$, $\phi^g \sim beta(1, 12)$

Small DE proportion $p = 10\%$ under Setting P1

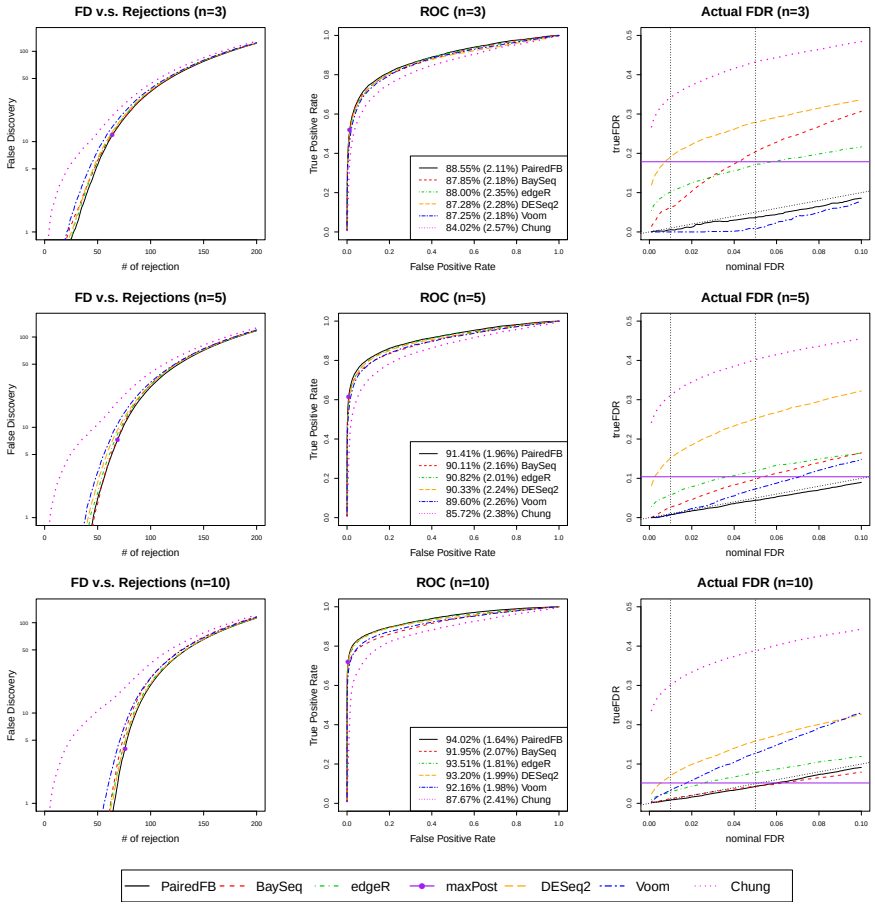


Figure S12: False discovery plots, ROC curves and the actual FDR curves comparing all methods under Setting P1 when DE proportion $p = 10\%$. In the third column, the black dotted line indicates the true nominal FDR level and two vertical dotted lines indicates 1% and 5% nominal level.

Large DE proportion $p = 20\%$ under Setting P1

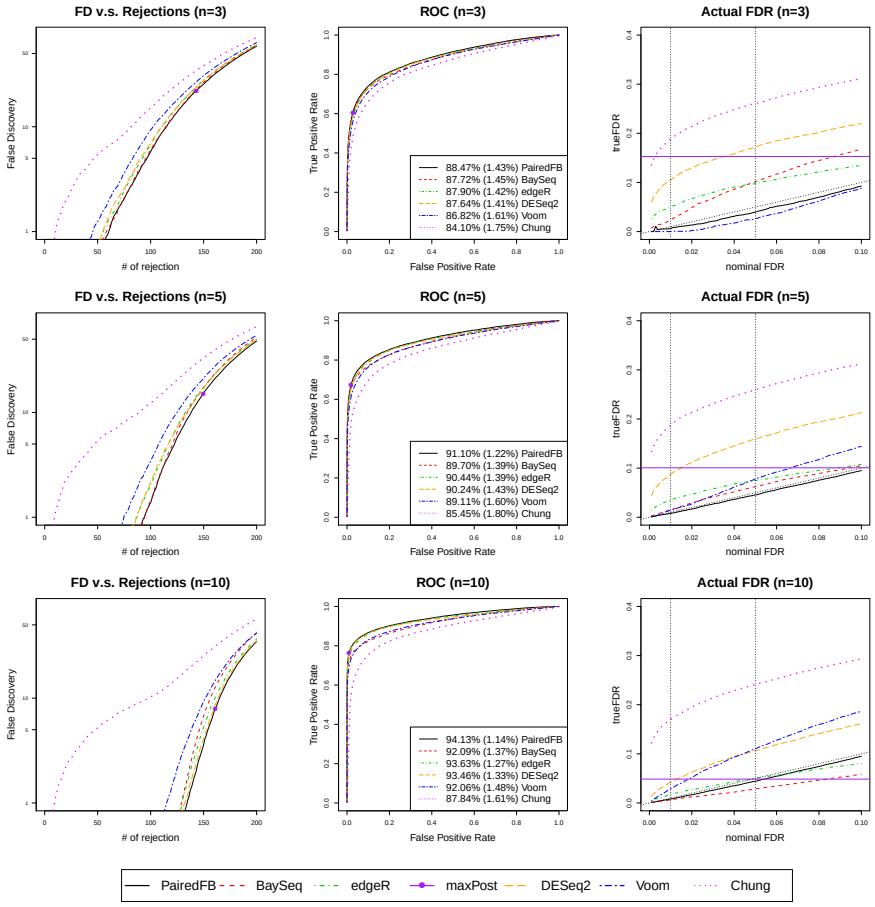


Figure S13: False discovery plots, ROC curves and the actual FDR curves comparing all methods under Setting P1 when DE proportion $p = 20\%$. In the third column, the black dotted line indicates the true nominal FDR level and two vertical dotted lines indicates 1% and 5% nominal level.

Actual FDR and Power comparison at nominal level 5% under Setting P1

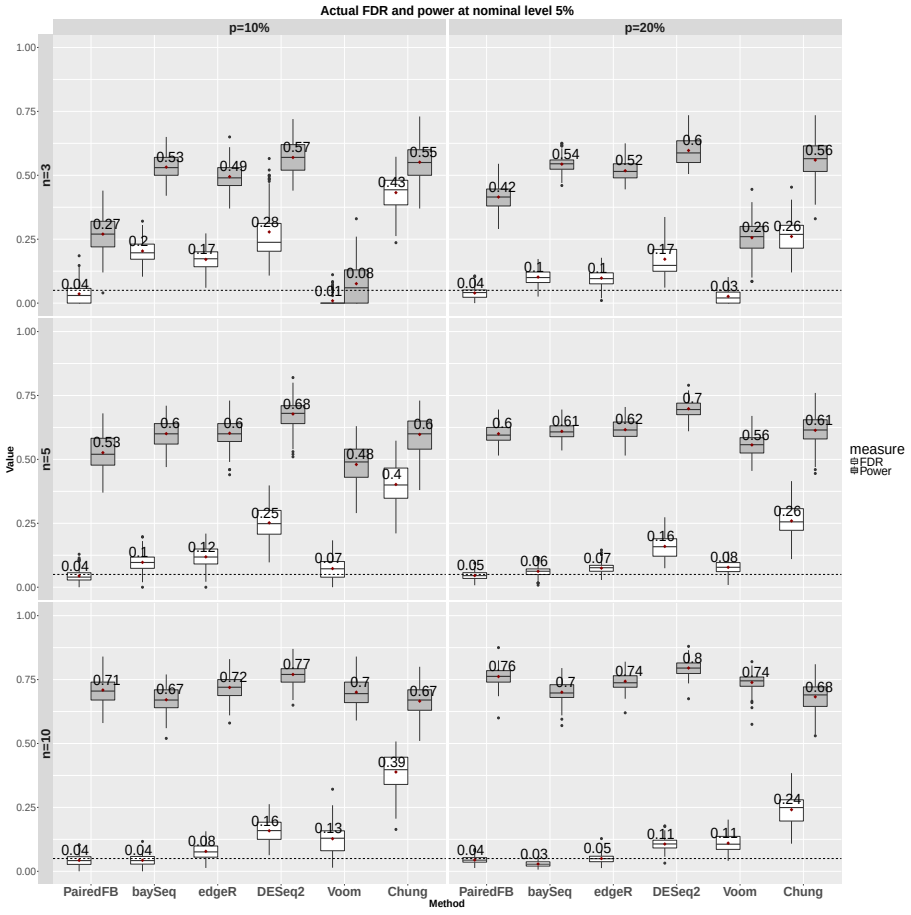


Figure S14: The box plots of the actual false discovery proportion and true detection rate over 100 simulations at nominal level 5% comparing all methods. Subplots in each row represent results for different sample sizes ($n = 3, 5, 10$) and column for different DE proportions ($p = 10\%, 20\%$). ϕ^g and π^{*g} of DE genes (π^{*g}_{DE}) are sampled under Setting P1. The numbers besides each box indicate the mean value over 100 simulations. The horizontal black dashed line in each plot outlines the 5% nominal level.

Actual FDR and Power comparison at nominal level 10% under Setting P1

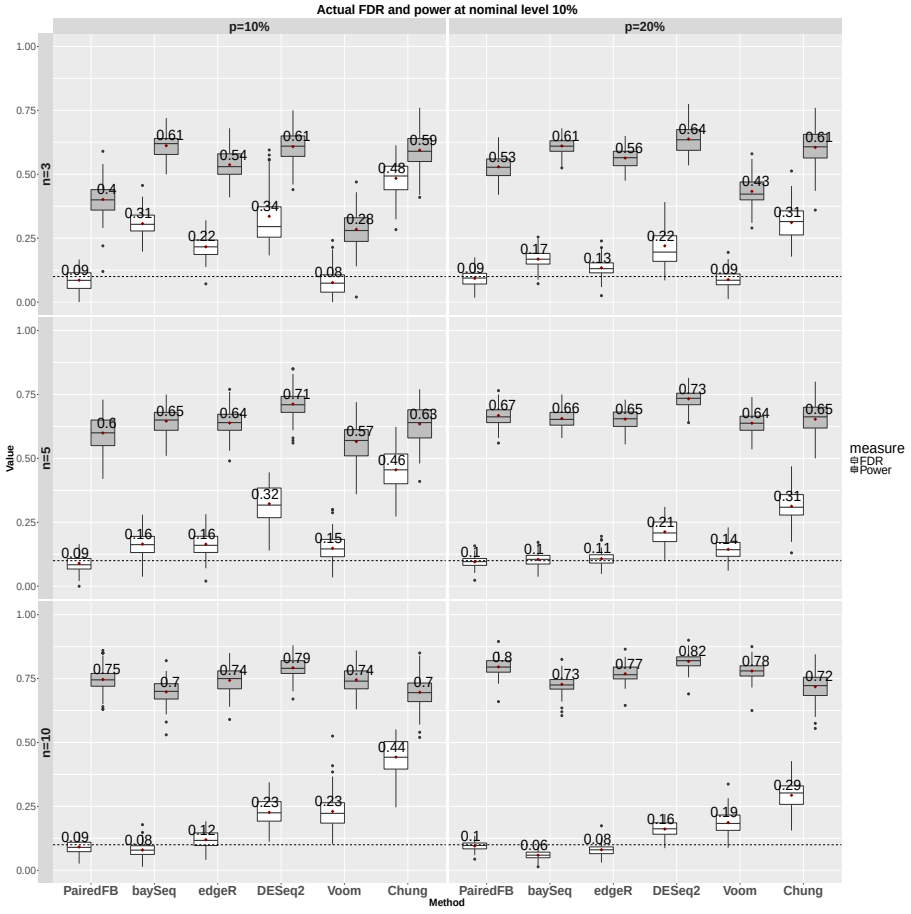


Figure S15: The box plots of the actual false discovery proportion and true detection rate over 100 simulations at nominal level 10% comparing all methods. Subplots in each row represent results for different sample sizes ($n = 3, 5, 10$) and column for different DE proportions ($p = 10\%, 20\%$). ϕ^g and π^{*g} of DE genes (π^{*g}_{DE}) are sampled under Setting P1. The numbers besides each box indicate the mean value over 100 simulations. The horizontal black dashed line in each plot outlines the 10% nominal level.

2.2.5 Setting P2: $\pi_{DE}^{*g} \sim N(0, 5 \times 0.68^2)$, $\phi^g \sim beta(1, 5)$

Small DE proportion $p = 10\%$

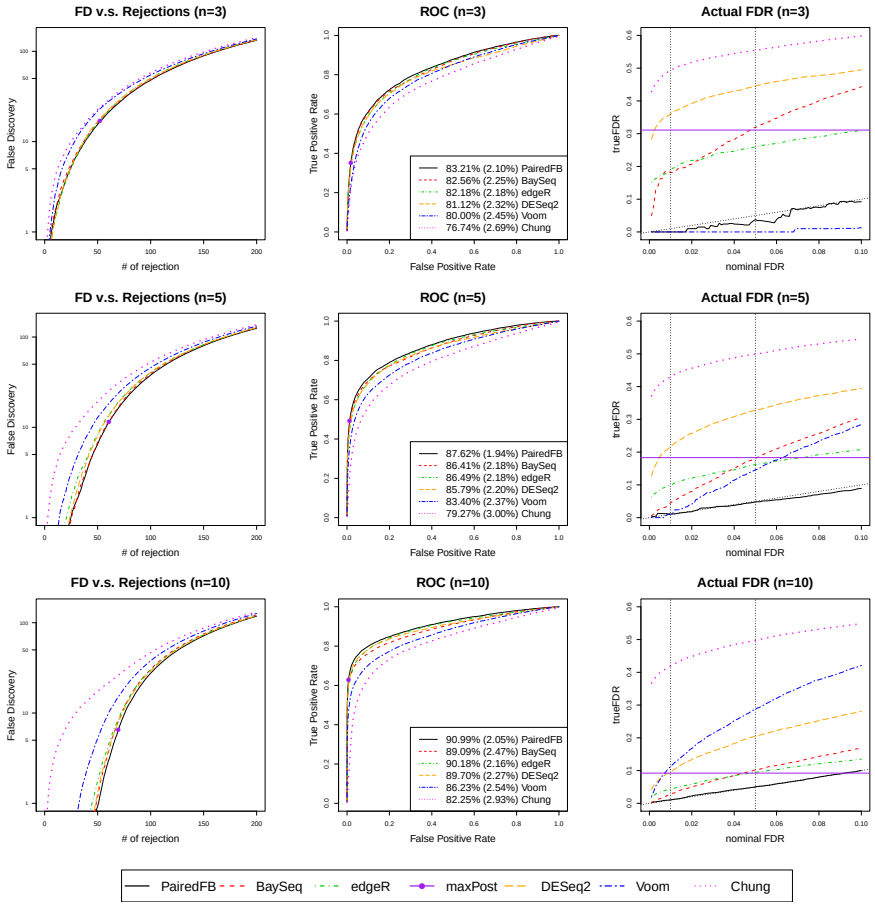


Figure S16: False discovery plots, ROC curves and the actual FDR curves comparing all methods under Setting P2 when DE proportion $p = 10\%$. In the third column, the black dotted line indicates the true nominal FDR level and two vertical dotted lines indicates 1% and 5% nominal level.

Large DE proportion $p = 20\%$

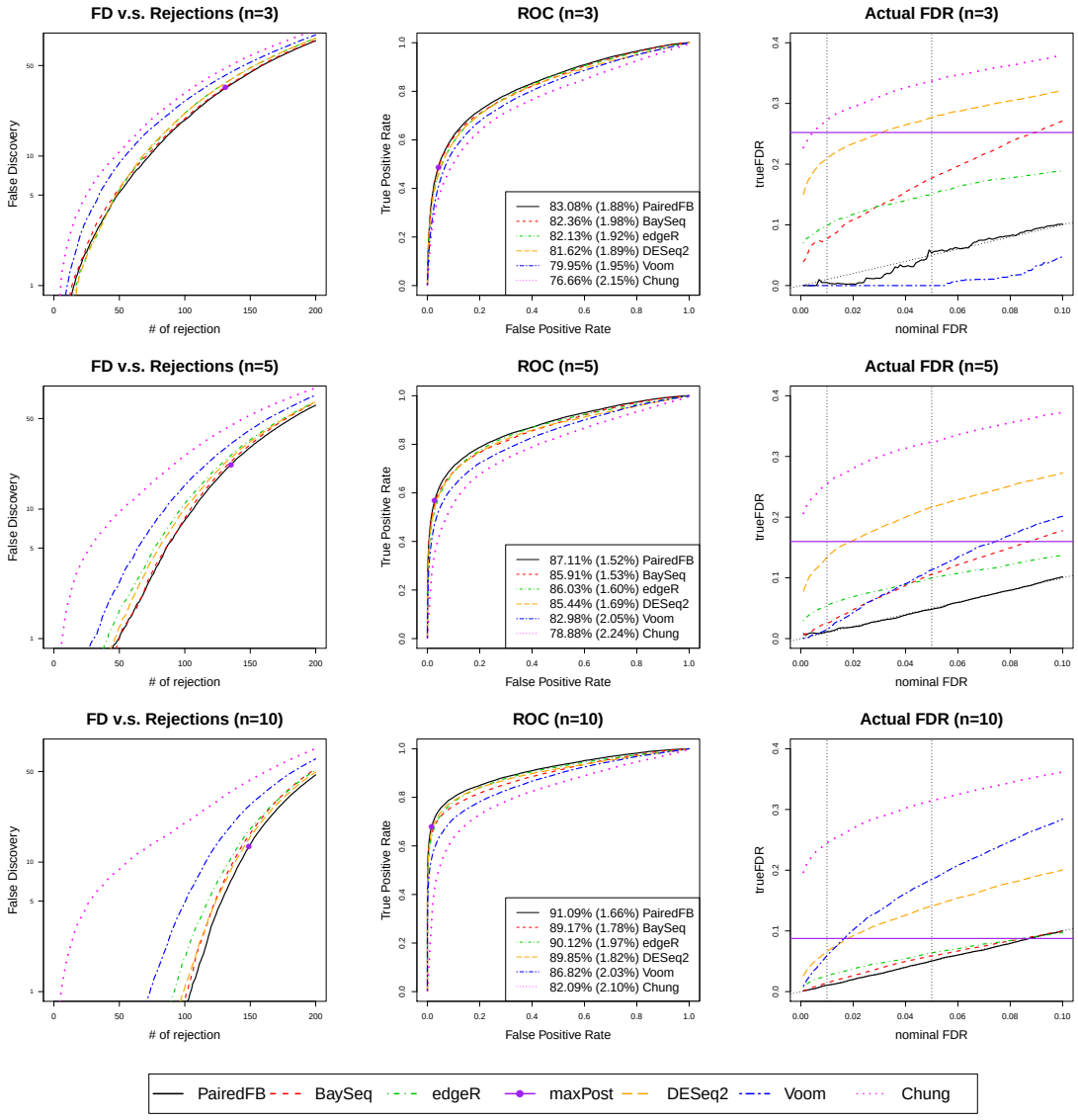


Figure S17: False discovery plots, ROC curves and the actual FDR curves comparing all methods under Setting P2 when DE proportion $p = 20\%$. In the third column, the black dotted line indicates the true nominal FDR level and two vertical dotted lines indicates 1% and 5% nominal level.

Actual FDR and Power comparison at nominal level 5% under Setting P2

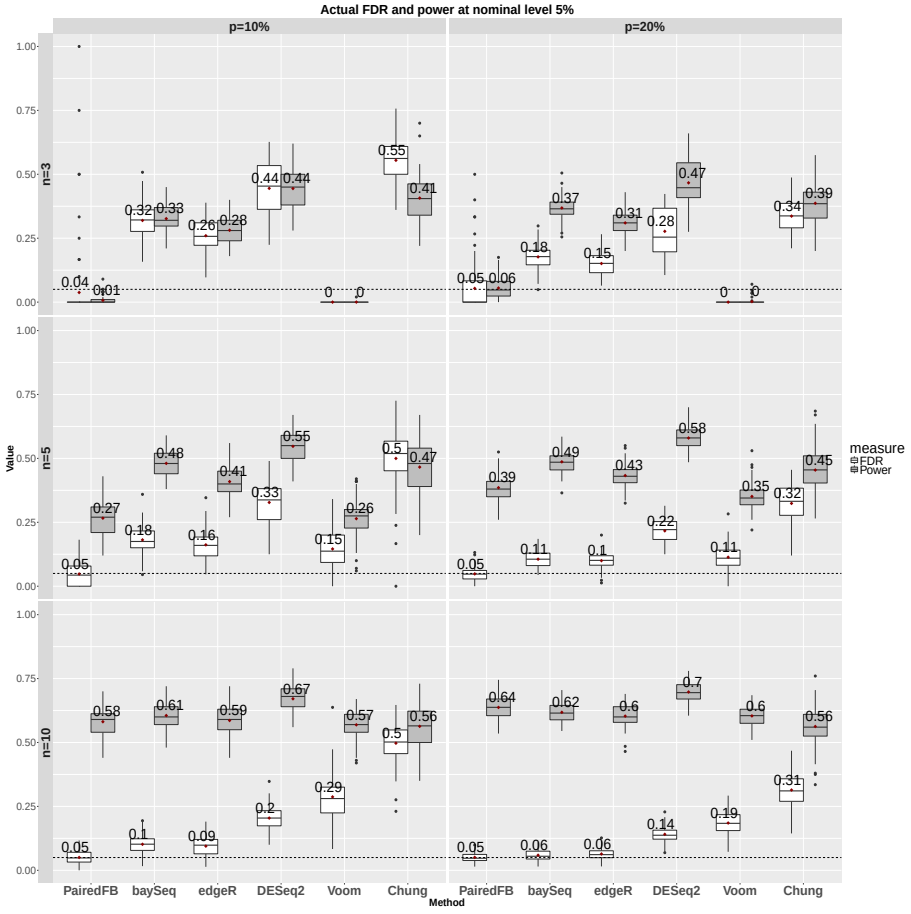


Figure S18: The box plots of the actual false discovery proportion and true detection rate over 100 simulations at nominal level 5% comparing all methods. Subplots in each row represent results for different sample sizes ($n = 3, 5, 10$) and column for different DE proportions ($p = 10\%, 20\%$). ϕ^g and π^{*g} of DE genes (π^{*g}_{DE}) are sampled under Setting P2. The numbers besides each box indicate the mean value over 100 simulations. The horizontal black dashed line in each plot outlines the 5% nominal level.

Actual FDR and Power comparison at nominal level 10% under Setting P2

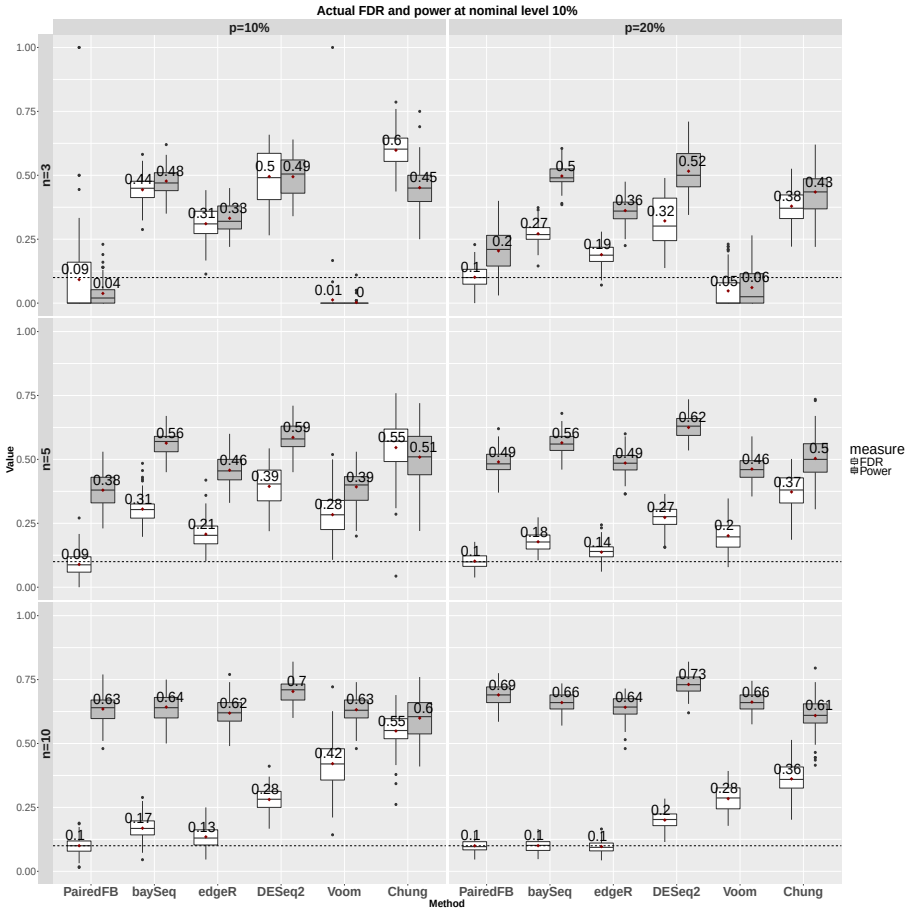


Figure S19: The box plots of the actual false discovery proportion and true detection rate over 100 simulations at nominal level 10% comparing all methods. Subplots in each row represent results for different sample sizes ($n = 3, 5, 10$) and column for different DE proportions ($p = 10\%, 20\%$). ϕ^g and π^{*g} of DE genes (π^{*g}_{DE}) are sampled under Setting P2. The numbers besides each box indicate the mean value over 100 simulations. The horizontal black dashed line in each plot outlines the 10% nominal level.

2.3 Simulation results for identifying biological significant DE genes

2.3.1 Setting SE1: π_{DE}^{*g} is drawn from empirical distribution scaled by $\sqrt{5}$ and ϕ^g is drawn from empirical distribution

Small DE proportion $p = 10\%$

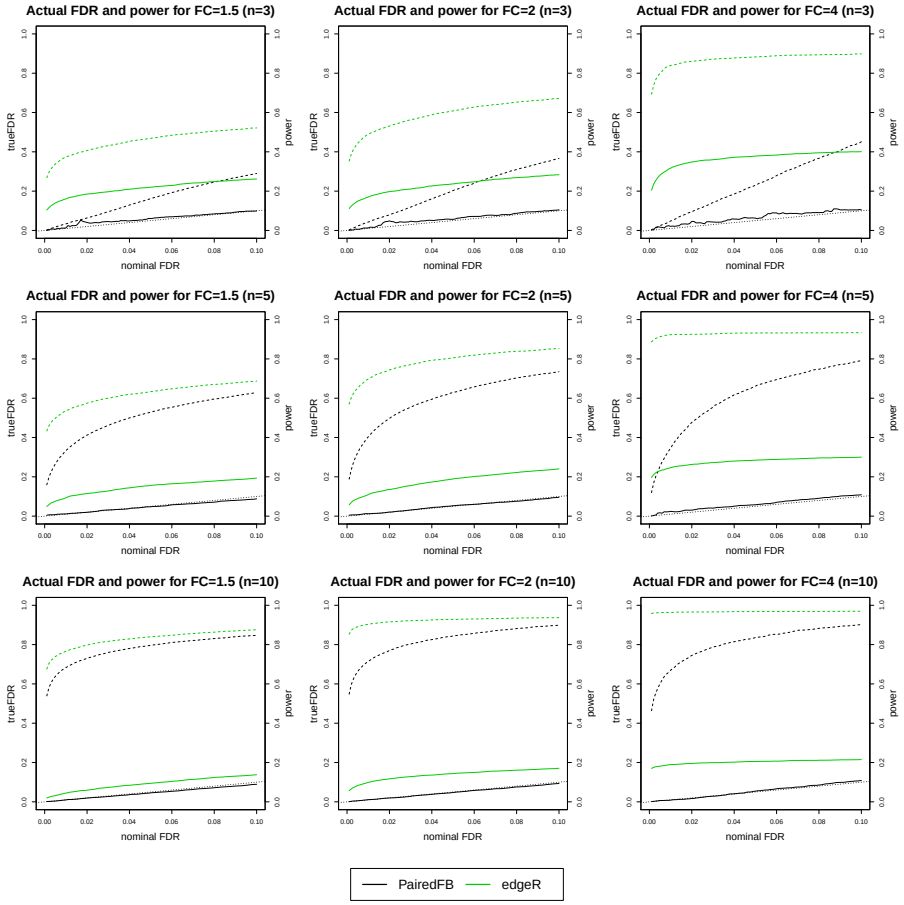


Figure S20: The actual FDR and power curves of the PairedFB and the two-stage edgeR procedure. Subplots in each row represent results for different sample sizes ($n = 3, 5, 10$) and column for different fold change thresholds to define the biological significance (FC= $1.5, 2, 4$). ϕ^g and π^{*g} of DE genes (π_{DE}^{*g}) are sampled under Setting SE1 and DE proportion $p = 10\%$.

Large DE proportion $p = 20\%$

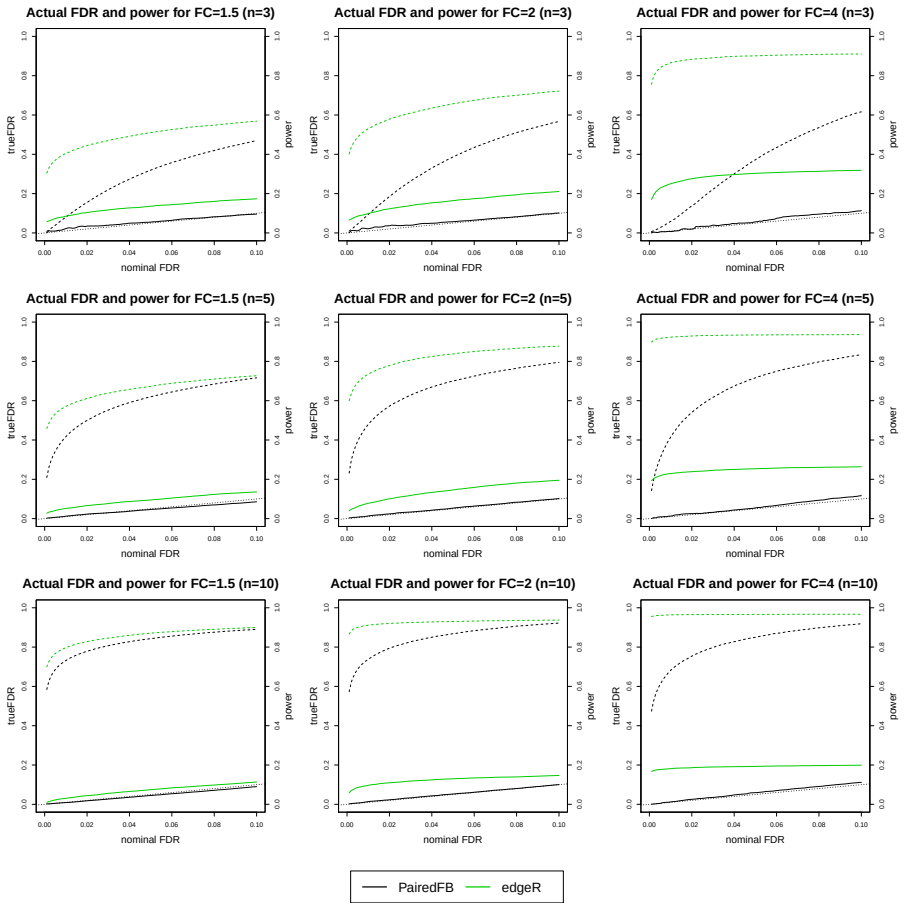


Figure S21: The actual FDR and power curves of the PairedFB and the two-stage edgeR procedure. Subplots in each row represent results for different sample sizes ($n = 3, 5, 10$) and column for different fold change thresholds to define the biological significance (FC= $1.5, 2, 4$). ϕ^g and π^{*g} of DE genes (π^{*g}_{DE}) are sampled under Setting SE1 and DE proportion $p = 20\%$.

2.3.2 Setting SE2: π_{DE}^{*g} is drawn from empirical distribution scaled by $\sqrt{10}$ and ϕ^g is drawn from empirical distribution

Small DE proportion $p = 10\%$

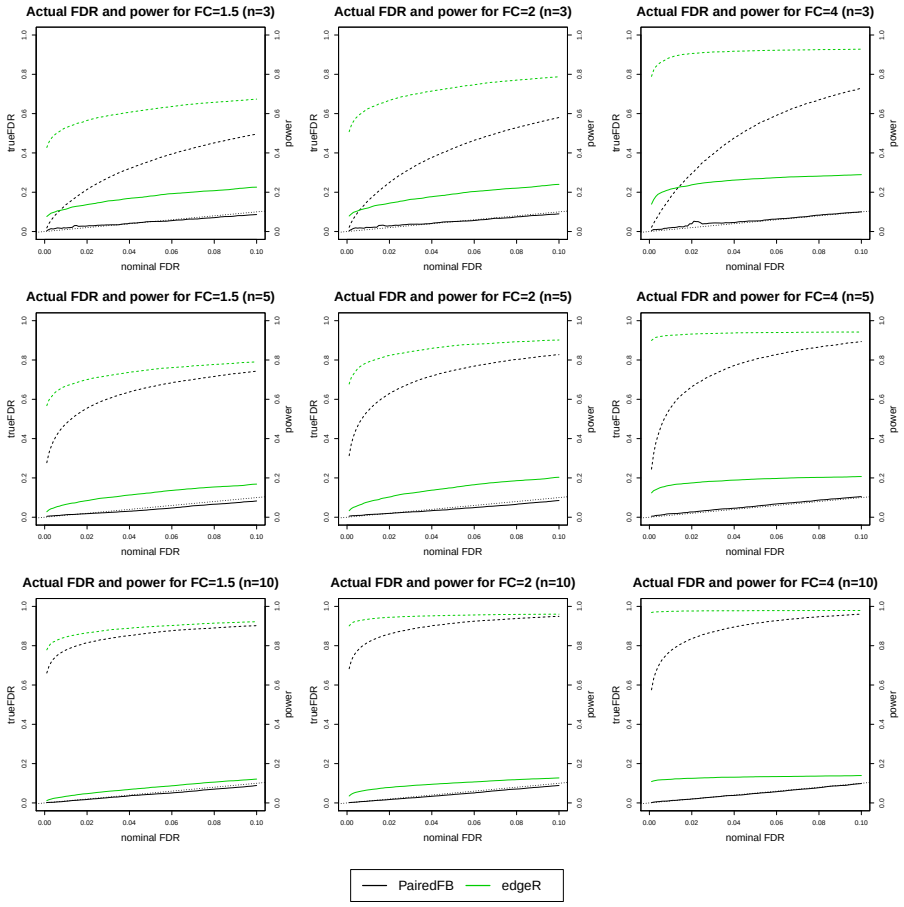


Figure S22: The actual FDR and power curves of the PairedFB and the two-stage edgeR procedure. Subplots in each row represent results for different sample sizes ($n = 3, 5, 10$) and column for different fold change thresholds to define the biological significance (FC= $1.5, 2, 4$). ϕ^g and π^{*g} of DE genes (π_{DE}^{*g}) are sampled under Setting SE2 and DE proportion $p = 10\%$.

Large DE proportion $p = 20\%$

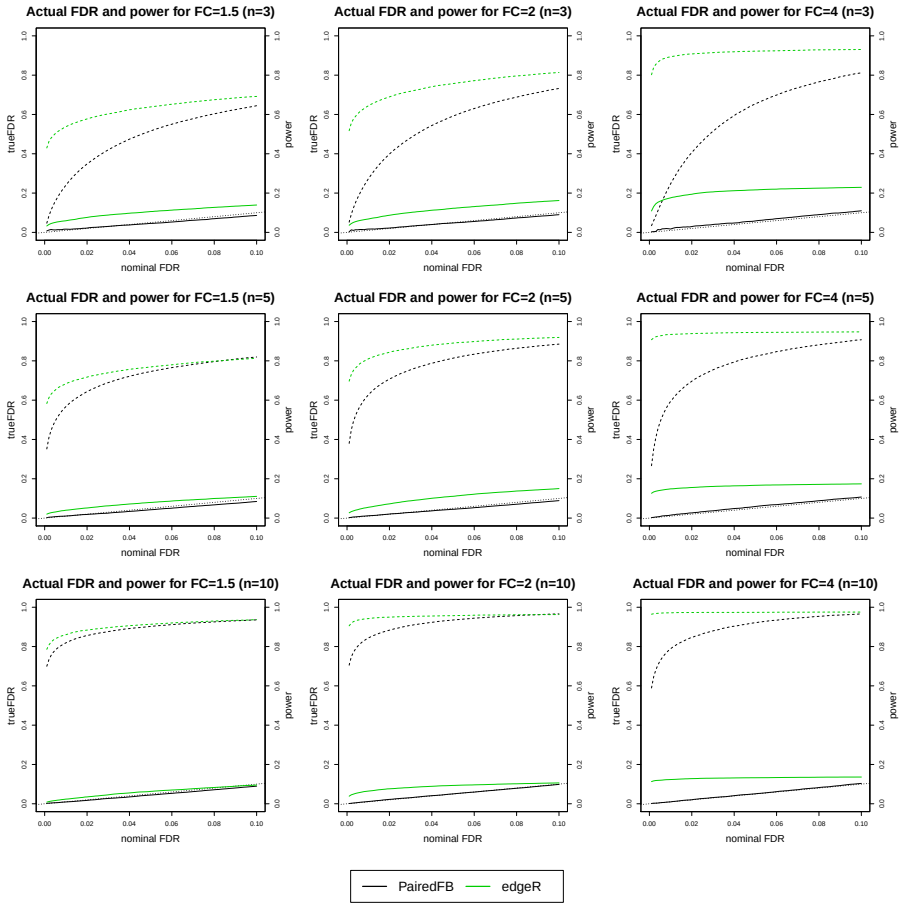


Figure S23: The actual FDR and power curves of the PairedFB and the two-stage edgeR procedure. Subplots in each row represent results for different sample sizes ($n = 3, 5, 10$) and column for different fold change thresholds to define the biological significance (FC= $1.5, 2, 4$). ϕ^g and π^{*g} of DE genes (π^{*g}_{DE}) are sampled under Setting SE2 and DE proportion $p = 20\%$.

Actual FDR and power comparison at nominal level 5% under Setting SE2

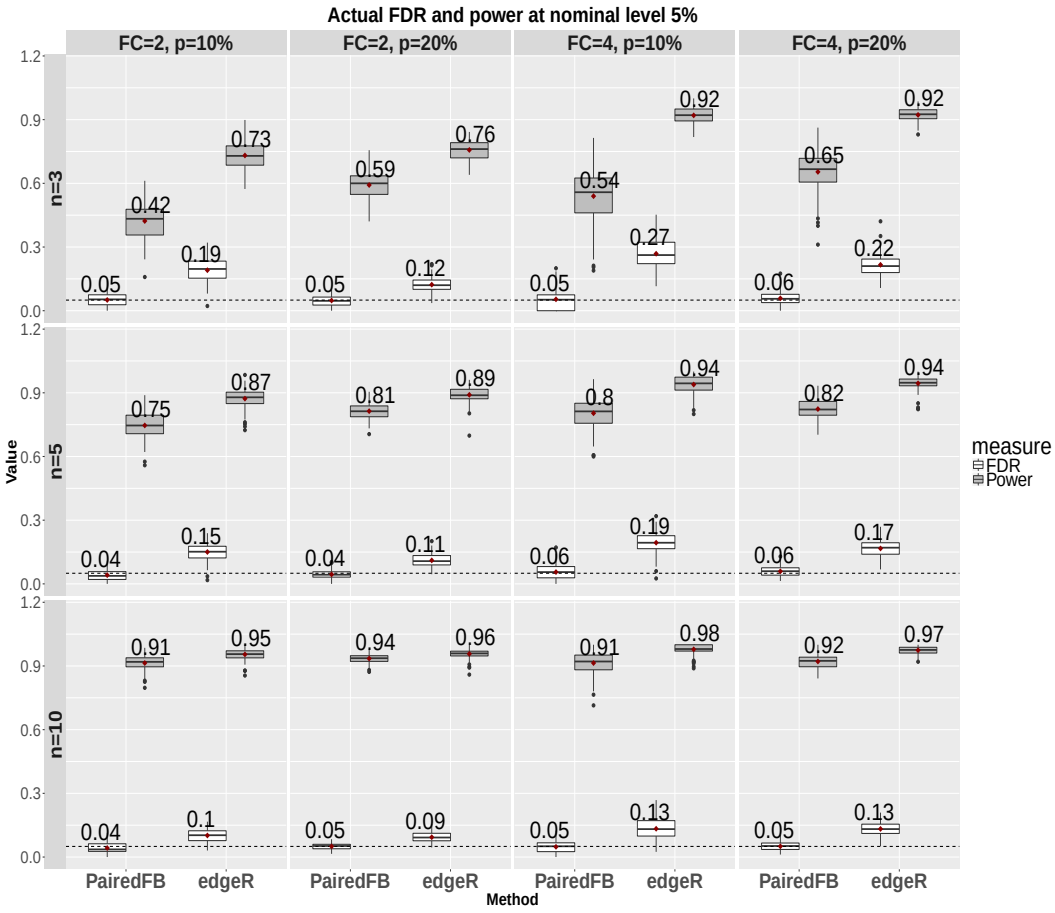


Figure S24: The box plots of the actual false discovery proportion and true detection rate over 100 simulations at nominal level 5% comparing PairedFB and the two-stage edgeR procedure. Subplots in each row represent results for different sample sizes ($n = 3, 5, 10$) and column for different DE proportions ($p = 10\%, 20\%$). ϕ^g and π^{*g} of DE genes (π^{*g}_{DE}) are sampled under Setting SE2. The numbers besides each box indicate the mean value over 100 simulations. The horizontal black dashed line in each plot outlines the 5% nominal level.

2.4 Simulation results for robustness test

2.4.1 Setting SE1: π_{DE}^{*g} is drawn from empirical distribution scaled by $\sqrt{5}$ and ϕ^g is drawn from empirical distribution.

Small DE proportion $p = 10\%$

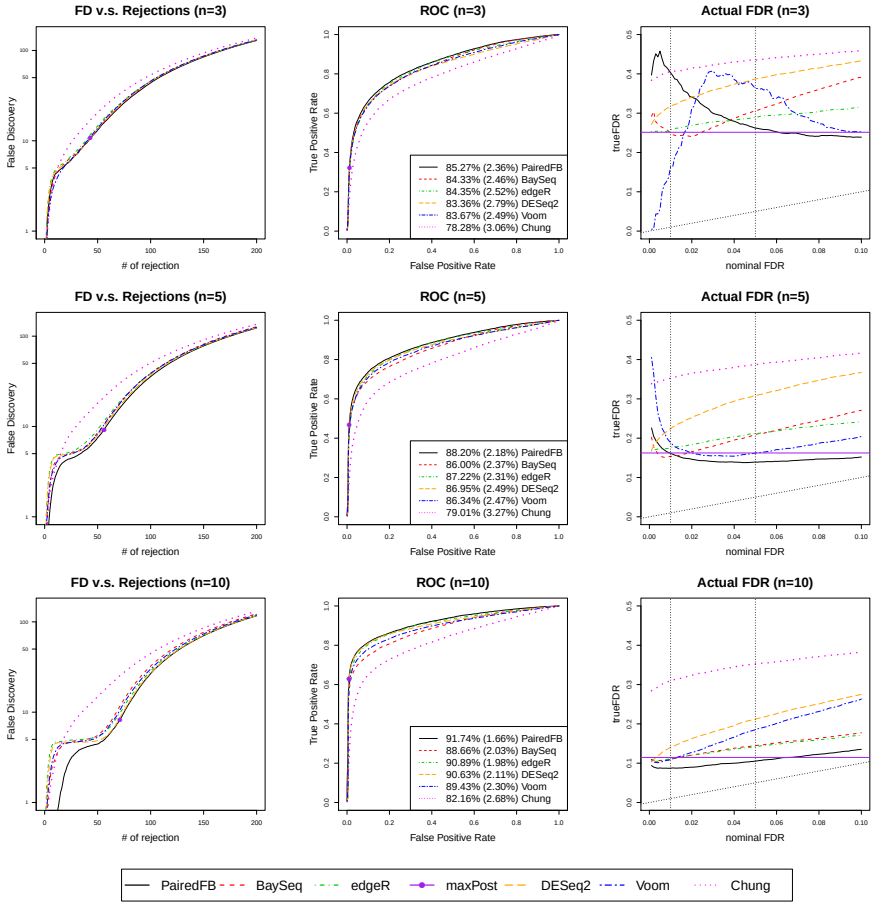


Figure S25: False discovery plots, ROC curves and the actual FDR curves when data is generated from Poisson-minimax mixture. Parameters of the minimax distribution match the ϕ^g and π_{DE}^{*g} sampled from empirical Setting SE1. DE proportion is $p = 10\%$. In the third column, black dotted line indicating the true nominal FDR level and two vertical dotted lines indicate 1% and 5% nominal level.

Large DE proportion $p = 20\%$.

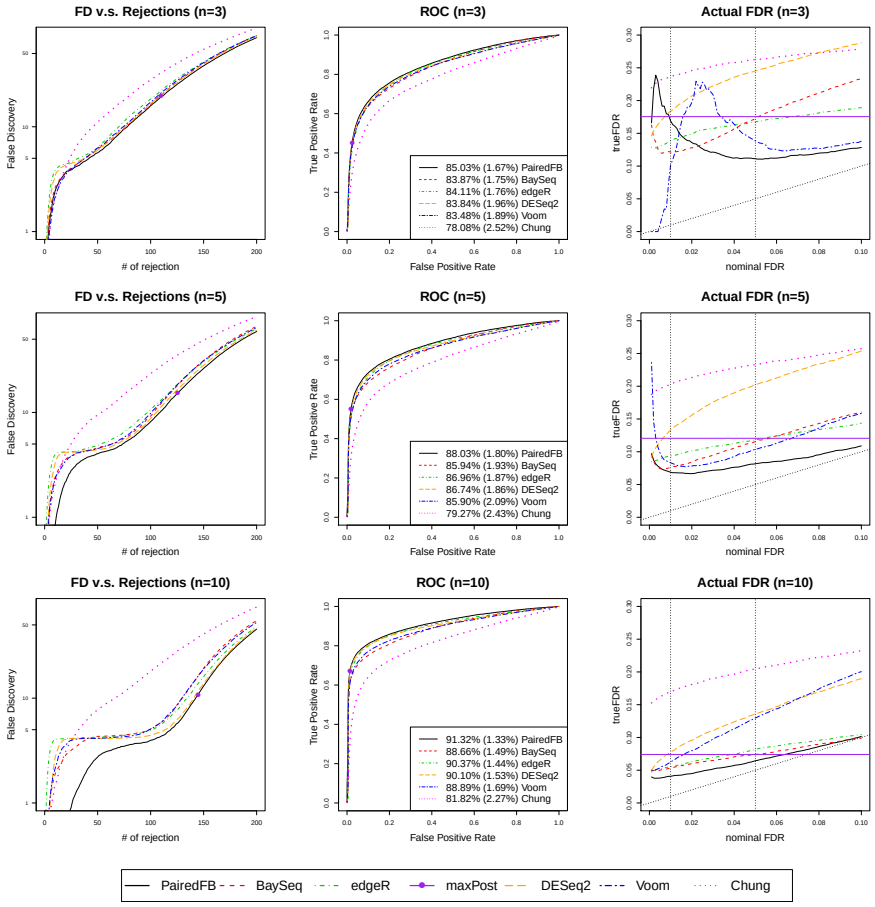


Figure S26: False discovery plots, ROC curves and the actual FDR curves when data is generated from Poisson-minimax mixture. Parameters of the minimax distribution match the ϕ^g and π_{DE}^{*g} sampled from empirical Setting SE1. DE proportion is $p = 20\%$. In the third column, black dotted line indicating the true nominal FDR level and two vertical dotted lines indicate 1% and 5% nominal level.

Actual FDR and Power comparison at nominal level 5% under Setting SE1

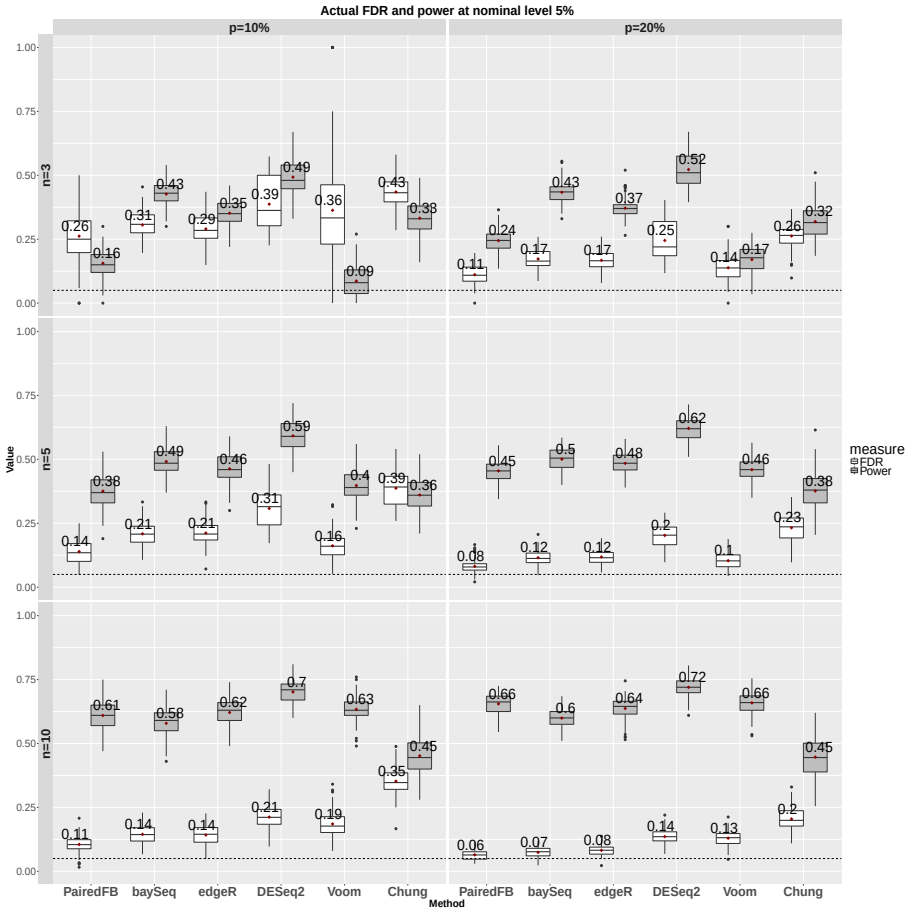


Figure S27: The box plots of the actual false discovery proportion and true detection rate over 100 simulations at nominal level 5% comparing all methods when data is generated from Poisson-minimax mixture. Subplots in each row represent results for different sample sizes ($n = 3, 5, 10$) and column for different DE proportions ($p = 5\%, 20\%$). Parameters of the minimax distribution match the ϕ^g and π^{*g} of DE genes (π^{*g}_{DE}) sampled from empirical setting I. The numbers besides each box indicate the mean value over 100 simulations. The horizontal black dashed line in each plot outlines the 5% nominal level.

Actual FDR and Power comparison at nominal level 10% under Setting SE1

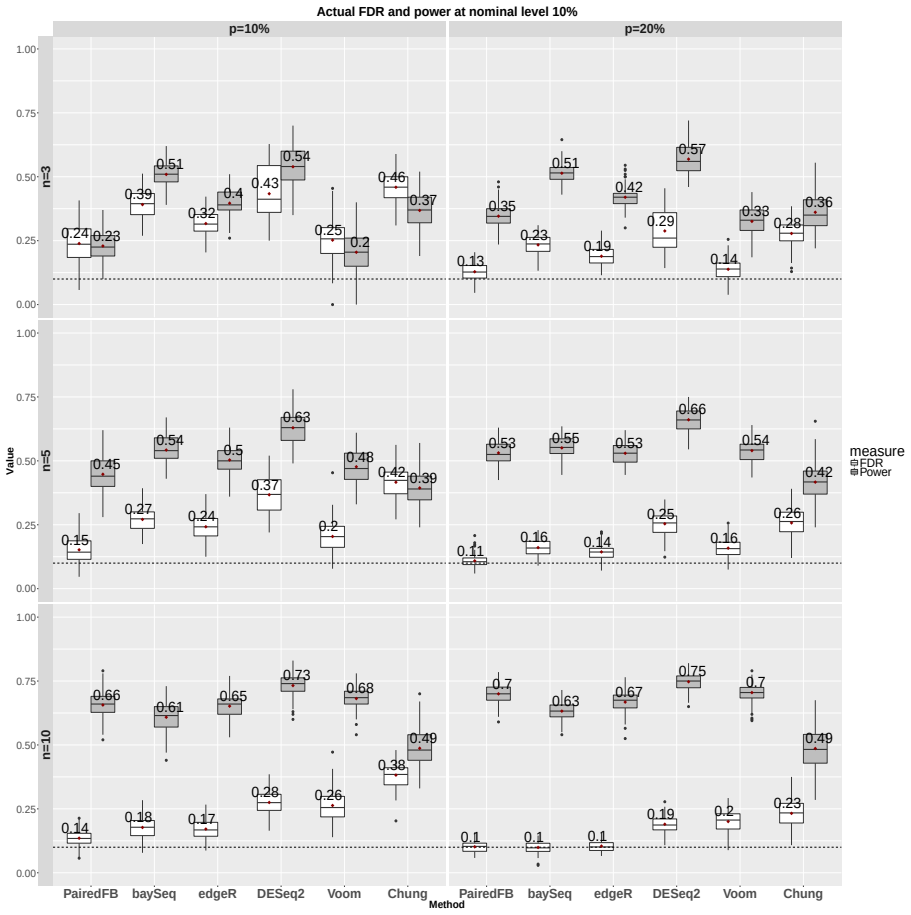


Figure S28: The box plots of the actual false discovery proportion and true detection rate over 100 simulations at nominal level 5% comparing all methods when data is generated from Poisson-minimax mixture. Subplots in each row represent results for different sample sizes ($n = 3, 5, 10$) and column for different DE proportions ($p = 5\%, 20\%$). Parameters of the minimax distribution match the ϕ^g and π^{*g} of DE genes (π^{*g}_{DE}) sampled from empirical setting I. The numbers besides each box indicate the mean value over 100 simulations. The horizontal black dashed line in each plot outlines the 10% nominal level.

2.4.2 Setting SE2: π_{DE}^{*g} is drawn from empirical distribution scaled by $\sqrt{10}$ and ϕ^g is drawn from empirical distribution.

Small DE proportion $p = 10\%$

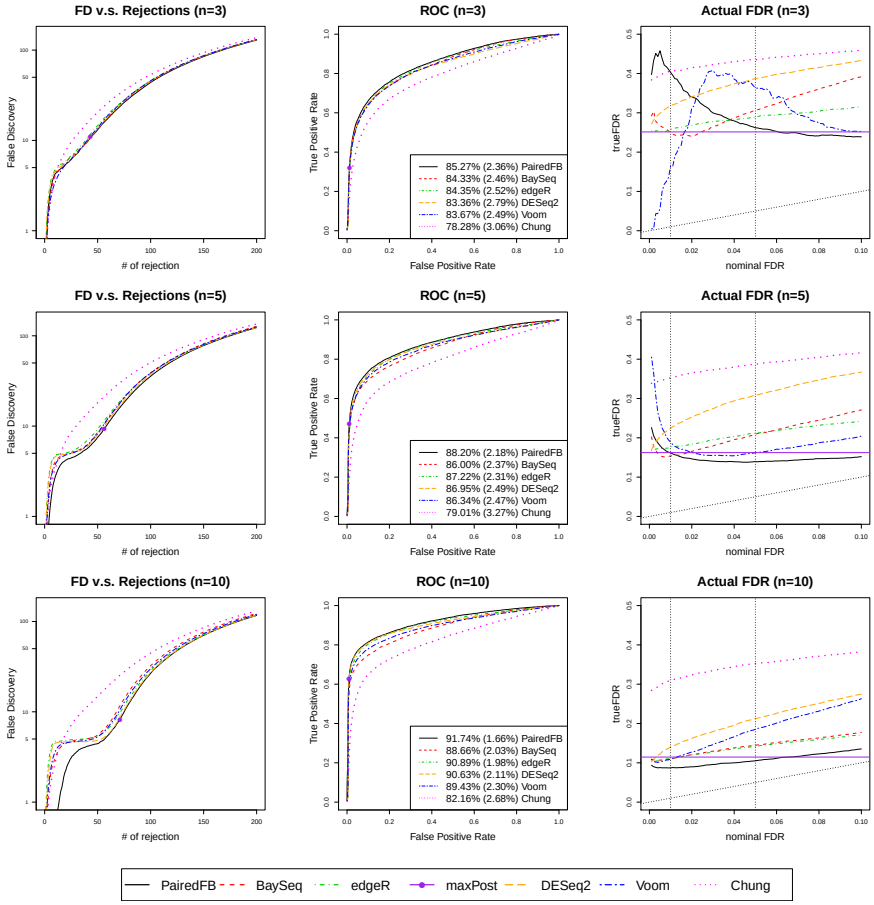


Figure S29: False discovery plots, ROC curves and the actual FDR curves when data is generated from Poisson-minimax mixture. Parameters of the minimax distribution match the ϕ^g and π_{DE}^{*g} sampled from empirical Setting SE2. DE proportion is $p = 10\%$. In the third column, black dotted line indicating the true nominal FDR level and two vertical dotted lines indicate 1% and 5% nominal level.

Large DE proportion $p = 20\%$

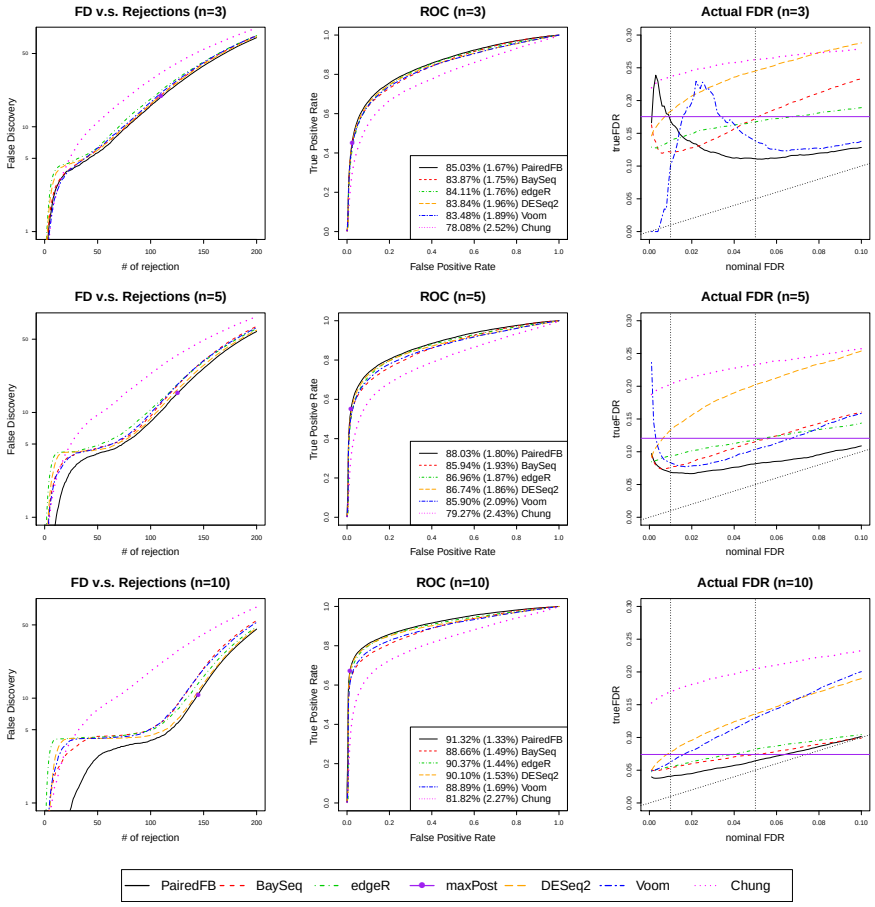


Figure S30: False discovery plots, ROC curves and the actual FDR curves when data is generated from Poisson-minimax mixture. Parameters of the minimax distribution match the ϕ^g and π_{DE}^{*g} sampled from empirical Setting SE1. DE proportion is $p = 20\%$. In the third column, black dotted line indicating the true nominal FDR level and two vertical dotted lines indicate 1% and 5% nominal level.

Actual FDR and Power comparison at nominal level 5% under Setting SE2

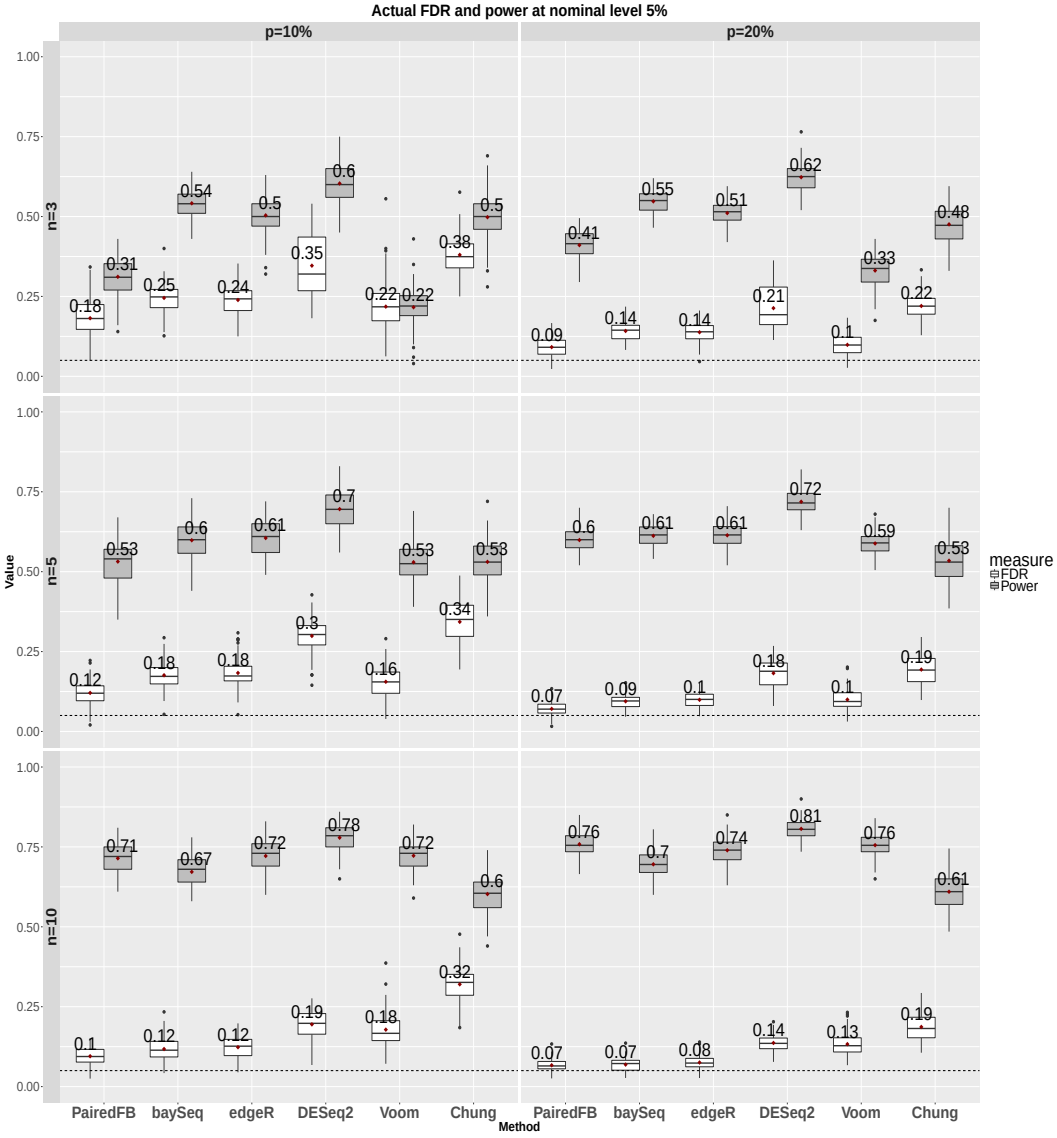


Figure S31: The box plots of the actual false discovery proportion and true detection rate over 100 simulations at nominal level 5% comparing all methods when data is generated from Poisson-minimax mixture. Subplots in each row represent results for different sample sizes ($n = 3, 5, 10$) and column for different DE proportions ($p = 5\%, 20\%$). Parameters of the minimax distribution match the ϕ^g and π^{*g} of DE genes (π^{*g}_{DE}) sampled from empirical setting II. The numbers besides each box indicate the mean value over 100 simulations. The horizontal black dashed line in each plot outlines the 5% nominal level.

Actual FDR and Power comparison at nominal level 10% under Setting SE2

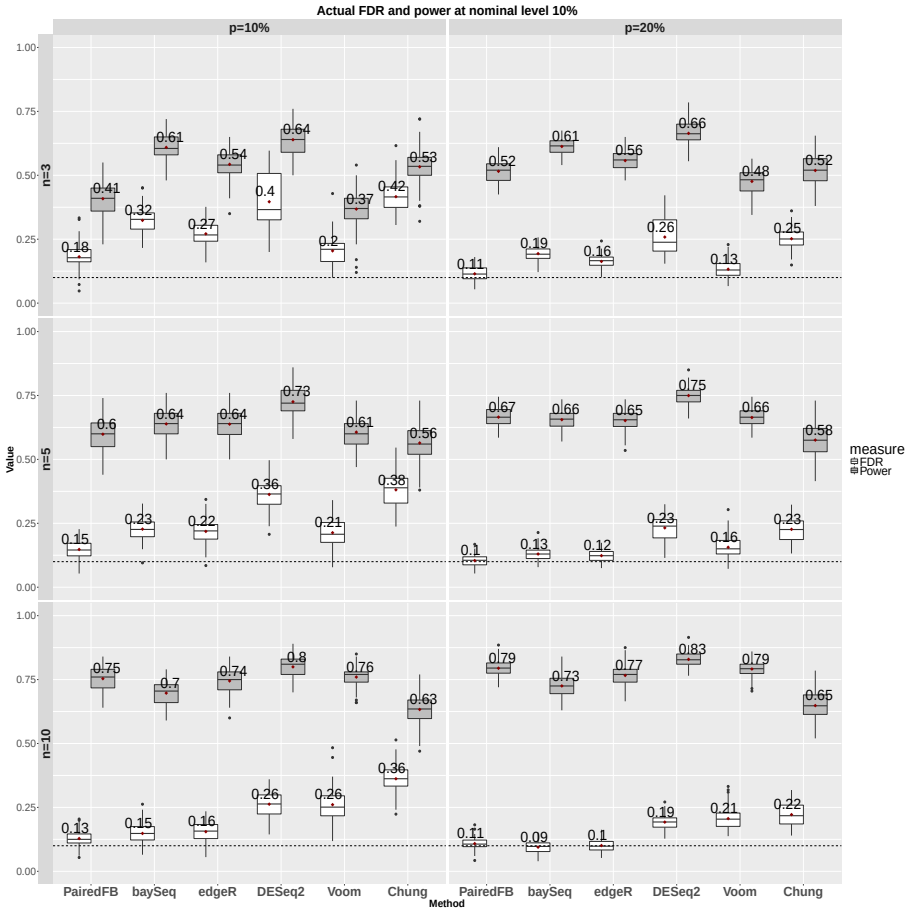


Figure S32: The box plots of the actual false discovery proportion and true detection rate over 100 simulations at nominal level 5% comparing all methods when data is generated from Poisson-minimax mixture. Subplots in each row represent results for different sample sizes ($n = 3, 5, 10$) and column for different DE proportions ($p = 5\%, 20\%$). Parameters of the minimax distribution match the ϕ^g and π^{*g} of DE genes (π^{*g}_{DE}) sampled from empirical setting I. The numbers besides each box indicate the mean value over 100 simulations. The horizontal black dashed line in each plot outlines the 10% nominal level.

2.5 Results when running longer MCMC for Chung's methods

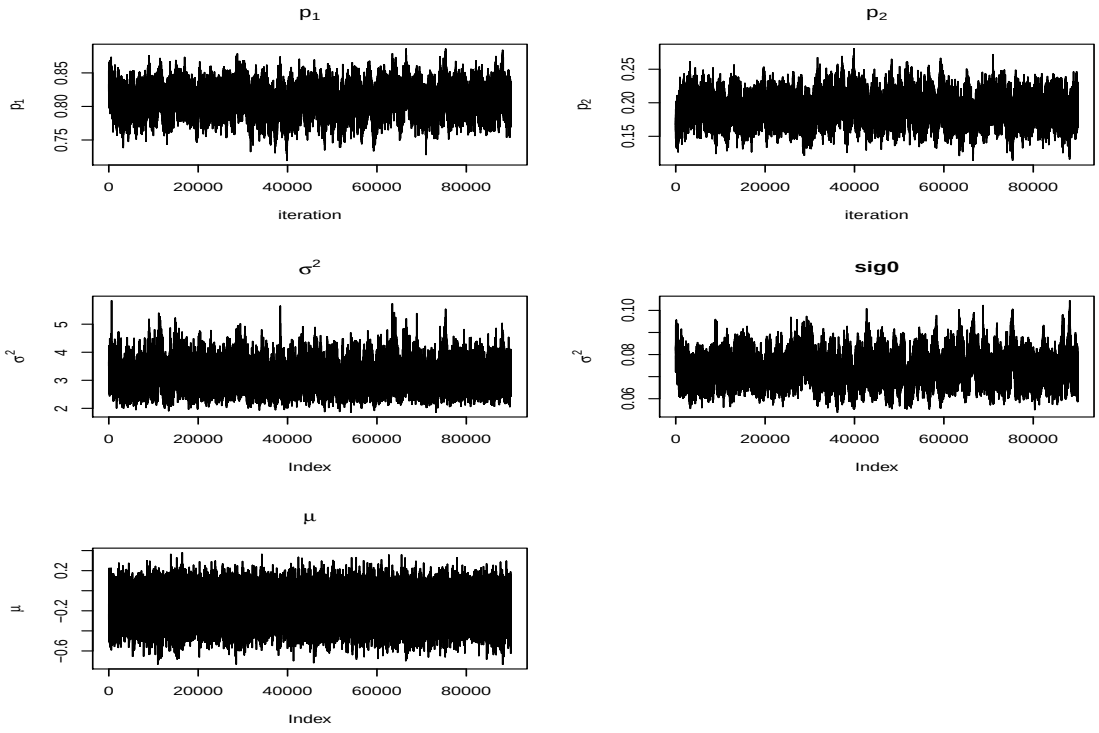


Figure S33: Trace plots (90000 iterations after discarding 10000 burn-in samples) for hyper-parameters $(1-p), p, \sigma^2, \sigma_0^2, \mu$ (from left to right and top to bottom) in the full Bayesian model of Chung *et al.* (2013) from one simulation under Setting SE1 when $p = 20\%$ and $n = 10$.

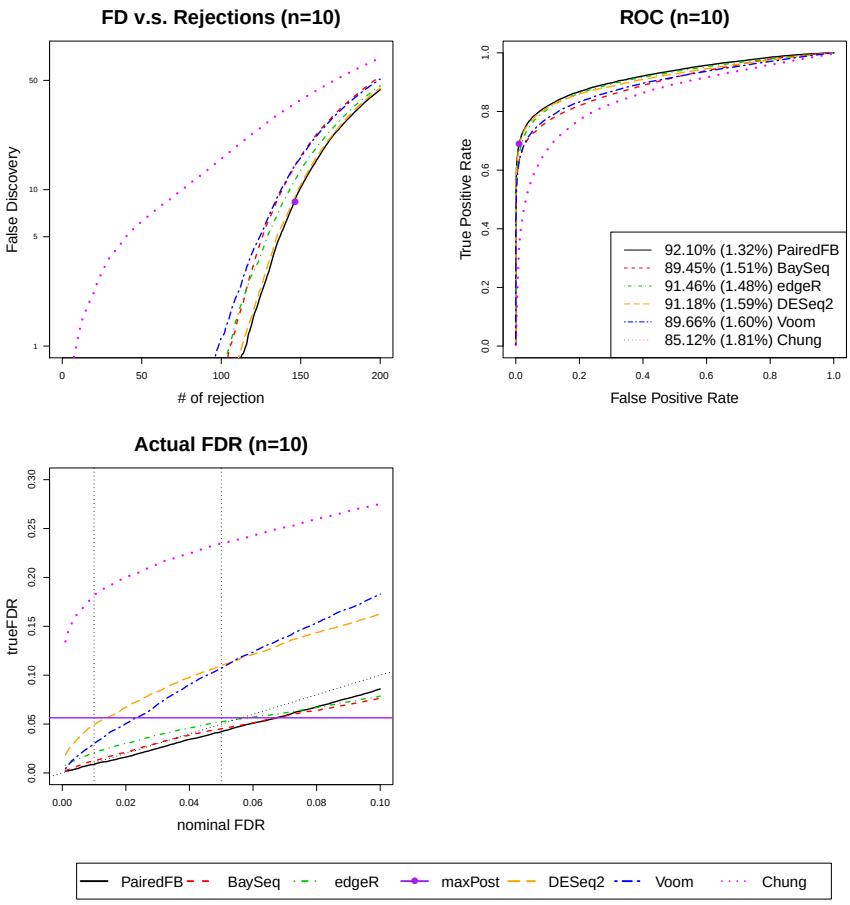


Figure S34: False discovery plots, ROC curves and the actual FDR curves comparing all methods under Setting SE1 when $n = 10$ and DE proportion $p = 20\%$ when running longer MCMC for Chung's method. Here we increase the iterations of MCMC for Chung's full Bayesian model to $100,000$ and the burn-in samples are $10,000$.

2.6 Real Data Analysis

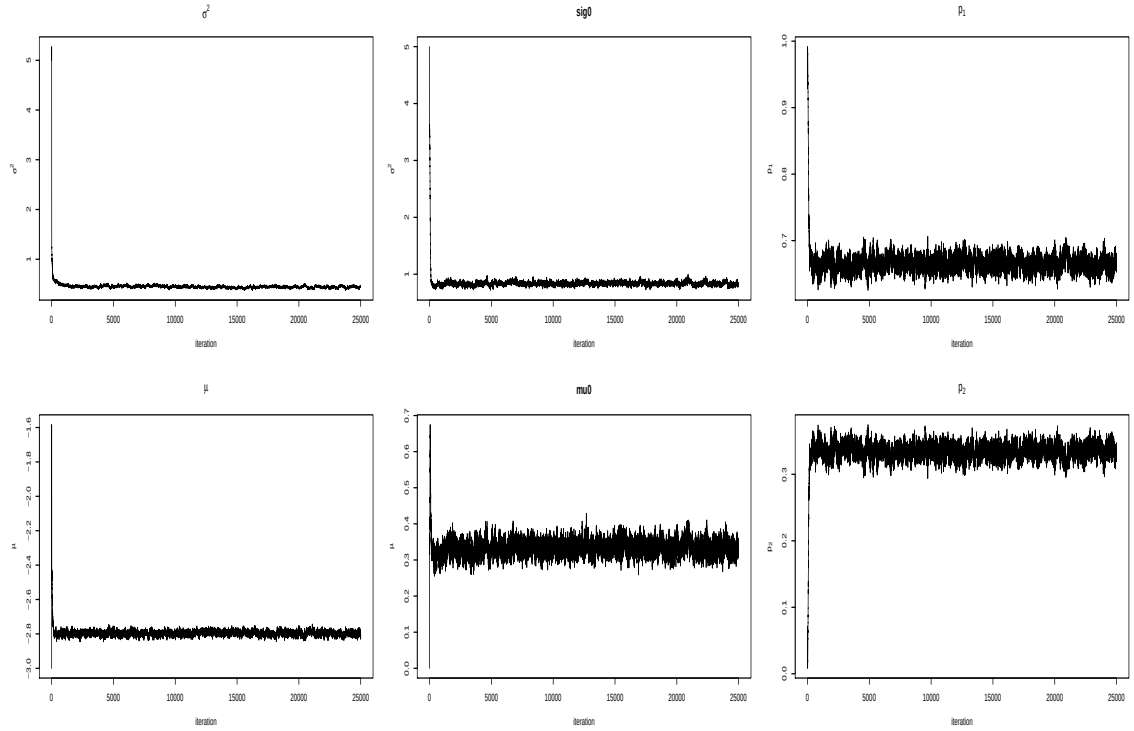


Figure S35: Trace plots for hyper-parameters $\sigma^2, \mu, \sigma_0^2, \mu_0, (1-p), p$ (from top to bottom and left to right) in the real data analysis.

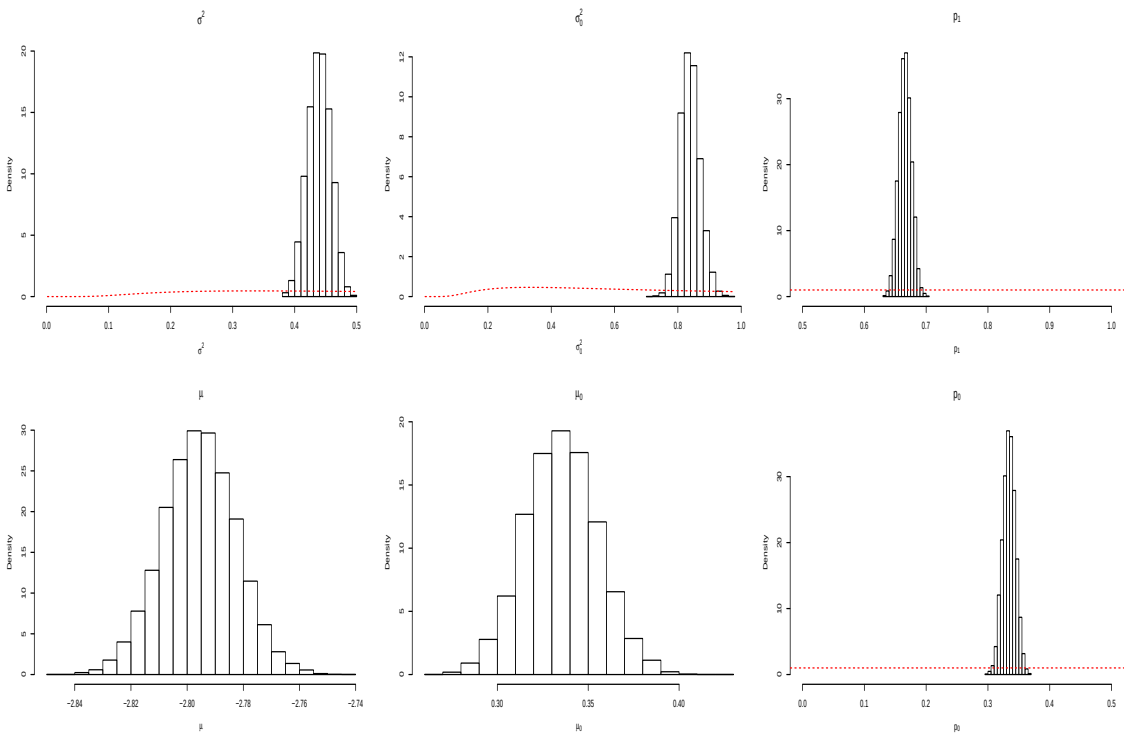


Figure S36: Histograms of posterior samples of hyper-parameters $\sigma^2, \mu, \sigma_0^2, \mu_0, p_1, p_0$ (from top to bottom and left to right) in the real data analysis. The red dashed line indicates the proper prior density.

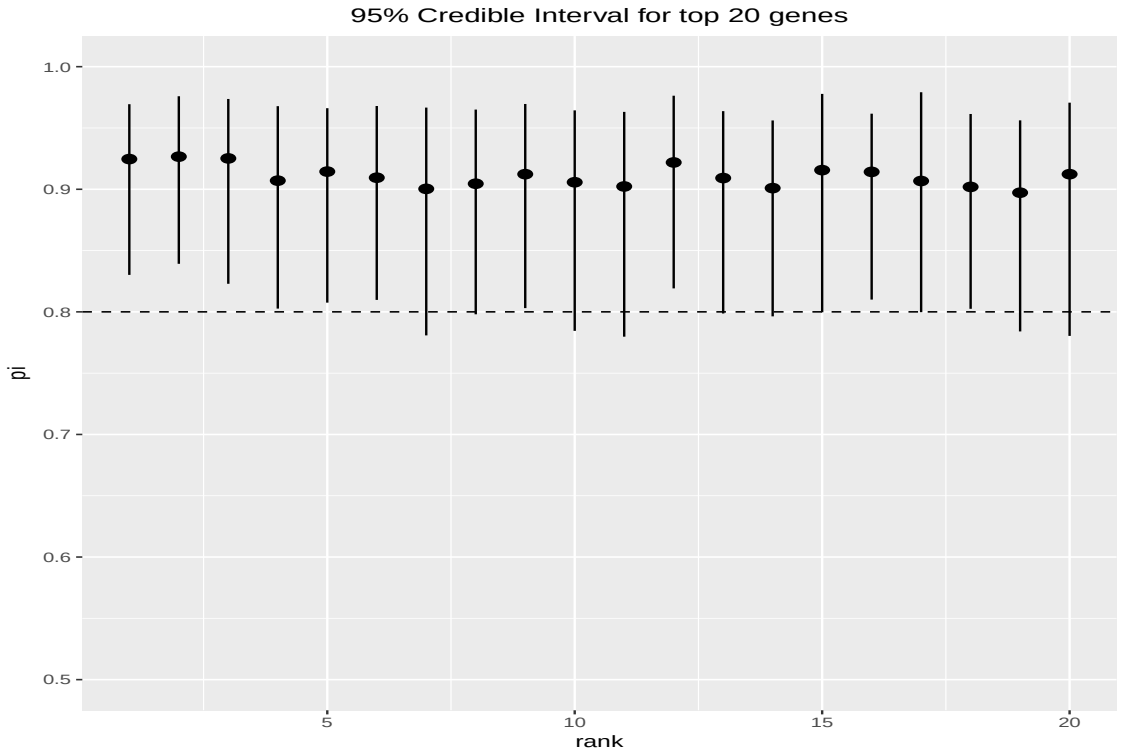


Figure S37: The 95% credible interval of posterior samples of π^g for the top 20 ranked genes in Table S1 identified by PairedFB method. The dashed line indicates FC=4.

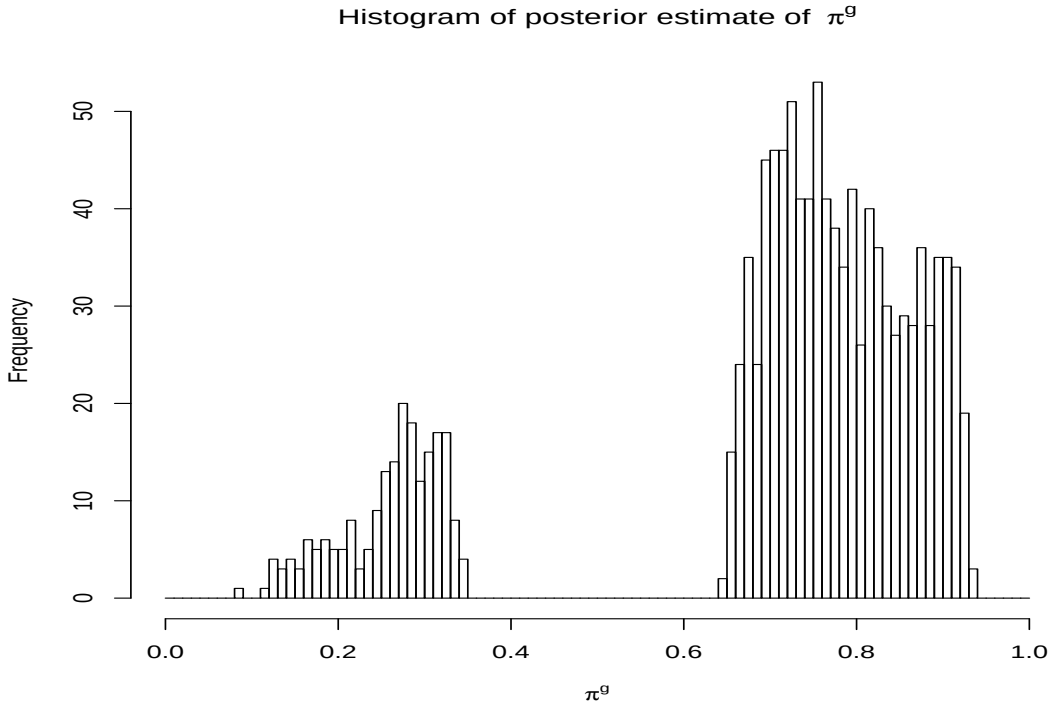


Figure S38: The histogram of posterior mean of π^g for the 1190 DEGs identified by PairedFB when FDR is controlled at 5%.

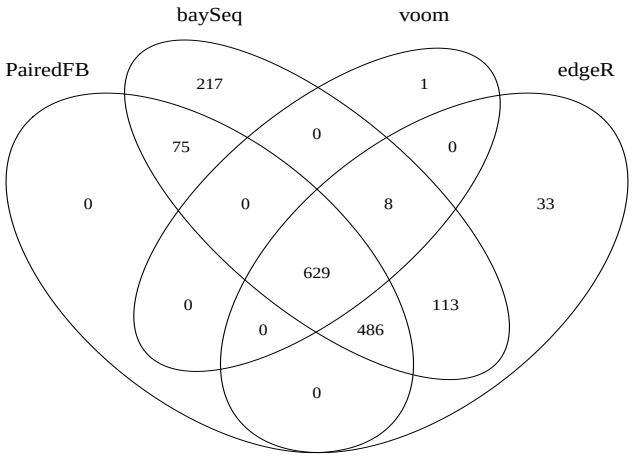


Figure S39: Venn Diagram of DE genes detected by different methods for the carcinoma data by Tuch *et al.* (2010) at nominal level 5%.

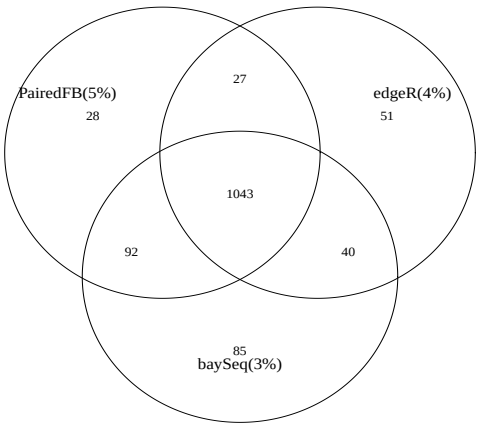
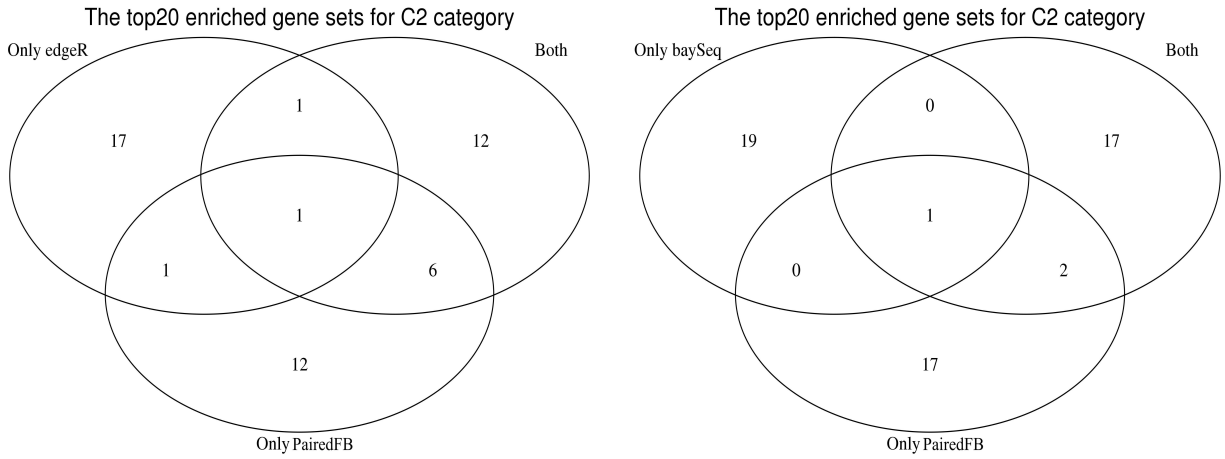


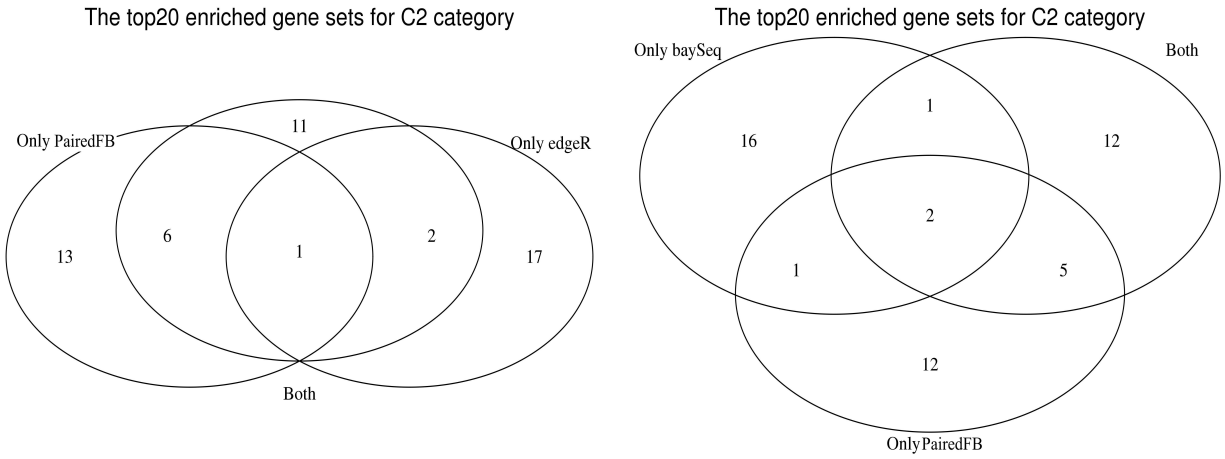
Figure S40: Venn Diagram of DE genes detected by different methods for the carcinoma data by Tuch *et al.* (2010) at different nominal levels. The nominal level is controlled at 5% for PairedFB method and 4% for the edgeR and 3% for paired baySeq method.



(a) Compare to edgeR

(b) Compare to baySeq

Figure S41: Venn diagram of top 20 enriched gene sets for C2 category of DE genes detected by both PairedFB and edgeR (or baySeq), PairedFB only and edgeR (or baySeq) only at comparable nominal levels: 5% for PairedFB, 4% for edgeR and 3% for paired baySeq.



(a) Compare to edgeR

(b) Compare to baySeq

Figure S42: Venn diagram of top 20 enriched gene sets for C2 category of DE genes detected by both PairedFB and edgeR (or baySeq), PairedFB only and edgeR (or baySeq) only in the top 500 DEGs list.

3 Tables

Gene	X8N	X8T	X33N	X33T	X51N	X51T	log2FC	pp	pp_2FC	pp_4FC
TMPRSS11B	286	0	744	24	153	1	6.29	1.000	1.000	0.988
MAL	297	0	376	20	79	0	5.87	1.000	1.000	0.998
CEACAM7	21	0	10	1	14	0	5.00	1.000	1.000	0.995
LDB3	165	0	124	2	124	13	4.71	1.000	0.999	0.980
ACTA1	368	1	305	2	619	59	4.92	1.000	1.000	0.978
DES	1057	12	639	10	3791	371	4.64	1.000	1.000	0.985
KLHL40	12	0	14	0	24	4	4.32	1.000	1.000	0.967
HSPB7	67	0	55	0	130	22	4.30	1.000	1.000	0.962
MYOZ3	28	0	21	0	25	5	4.09	1.000	1.000	0.979
TMOD1	24	0	24	0	79	12	4.44	1.000	0.999	0.969
ENO3	120	1	123	5	220	12	4.87	1.000	0.999	0.952
CASQ1	57	0	82	0	38	5	4.63	1.000	1.000	0.996
PTGFR	50	1	27	1	79	3	4.98	1.000	1.000	0.973
LAMB4	54	1	124	10	13	1	4.11	1.000	1.000	0.964
TRIM54	24	0	12	0	43	5	4.80	1.000	1.000	0.982
PYGM	154	3	121	1	99	6	5.11	1.000	1.000	0.978
ASB5	47	0	47	1	194	21	4.60	1.000	1.000	0.976
GPX3	194	2	141	12	226	14	4.28	1.000	0.999	0.974
SRL	91	1	62	3	95	11	4.14	1.000	0.997	0.961
SMYD1	54	0	49	1	64	8	4.45	1.000	1.000	0.970

Table S1: The raw count data (showing counts per million) and other statistics of the top 20 DE genes identified by the PairedFB from the analysis of the carcinoma data by Tuch *et al.* (2010). The last column gives the rank of each gene based on the PairedFB method and the pp column gives the posterior probability of being DE. pp_2FC and pp_4FC correspond to the posterior probability of each gene having mean fold change exceeding 2 and 4 respectively.

Gene	X8N	X8T	X33N	X33T	X51N	X51T	log2FC	pp	pp_2FC	pp_4FC	rank
MAL	297	0	376	20	79	0	5.87	1.000	1.000	0.998	2
SPP1	2	48	3	634	7	94	-4.66	0.999	0.983	0.917	74
MMP12	1	35	3	79	1	6	-3.75	0.999	0.988	0.875	142
KRT13	43428	290	34287	7836	9622	102	3.78	0.999	0.995	0.943	143
TGM3	2849	190	2213	331	580	4	3.81	0.999	0.991	0.921	152
LAMC2	35	567	36	336	63	1110	-3.74	0.999	0.986	0.908	163
KRT4	8432	45	9445	1526	2153	2	4.30	0.999	0.996	0.936	170
EMP1	1366	291	2111	142	767	126	2.79	0.999	0.991	0.796	171
ECM1	331	28	694	104	114	36	2.51	0.996	0.980	0.693	261
MMP1	5	158	6	116	395	1164	-3.00	0.951	0.964	0.676	706
MMP10	1	10	5	6	13	342	-2.05	0.950	0.861	0.474	722
PLAU	27	81	27	96	88	250	-1.64	0.918	0.768	0.047	879
KRT17	241	2202	267	702	547	1293	-1.79	0.887	0.781	0.173	994
POSTN	20	112	51	96	150	453	-1.59	0.876	0.713	0.073	1025
ITGA6	262	940	192	330	206	1545	-1.71	0.874	0.737	0.116	1036
MMP3	5	28	4	40	73	85	-1.70	0.797	0.618	0.130	1280
CXCL8	3	54	9	85	51	43	-1.74	0.718	0.605	0.158	1505
IFI6	49	72	17	29	15	108	-1.23	0.641	0.355	0.017	1755
TNC	239	457	147	934	1284	1317	-1.06	0.484	0.184	0.004	2373
FN1	211	542	235	1234	1937	1593	-1.02	0.429	0.157	0.004	2657
COL1A2	351	2681	825	978	5845	2317	-0.41	0.203	0.029	0.001	5232
COL5A2	144	707	200	265	1635	847	-0.47	0.202	0.027	0.000	5261
COL4A1	206	1026	354	350	1941	586	-0.12	0.164	0.013	0.000	6376
SPARC	245	445	286	222	1483	746	0.16	0.140	0.006	0.000	7575
KRT5	10445	7358	9335	5948	4091	10548	-0.04	0.136	0.005	0.000	7834

Table S2: The raw count data (showing counts per million) and other statistics based on the PairedFB for the important head and neck squamous cell carcinoma 25 genes as identified by Yu *et al.* (2008). The bolded 15 genes are detected by PairedFB at 5% nominal FDR level. The last column gives the rank of each gene based on the PairedFB method and the pp column gives the posterior probability of being DE. pp_2FC and pp_4FC correspond to the posterior probability of each gene having mean fold change exceeding 2 and 4 respectively.

Gene	X8N	X8T	X33N	X33T	X51N	X51T	log2FC	pp	pp_2FC	pp_4FC	rank
CASQ1	57	0	82	0	38	5	4.63	1.000	1.000	0.996	12
INHBA	3	80	4	77	33	119	-3.16	0.985	0.965	0.794	456
WIF1	9	5	20	1	3	0	2.68	0.984	0.967	0.737	471
HMGA2	1	181	2	30	4	8	-2.69	0.981	0.928	0.635	511
SHANK2	3	151	5	17	3	15	-2.65	0.955	0.886	0.580	689
MMP1	5	158	6	116	395	1164	-3.00	0.951	0.964	0.676	706
CCND1	152	1931	191	138	72	140	-1.01	0.467	0.254	0.021	2446
CTTN	166	2316	152	113	150	166	-0.76	0.354	0.128	0.005	3212
CDKN2B	61	0	79	74	30	241	0.24	0.317	0.157	0.027	3568

Table S3: The raw count data (showing counts per million) and other statistics based on the PairedFB for the 9 genes of interest identified by Tuch *et al.* (2010). The bolded 6 genes are identified by PairedFB at 5% nominal FDR level. The last column gives the rank of each gene based on the PairedFB method and the pp column gives the posterior probability of being DE. pp_2FC and pp_4FC correspond to the posterior probability of each gene having mean fold change exceeding 2 and 4 respectively.

GOID	Term	Ontology	P-Value	FDR
GO:0003012	muscle system process	BP	1.53E-33	2.87E-29
GO:0006936	muscle contraction	BP	1.06E-32	9.89E-29
GO:0043292	contractile fiber	CC	2.63E-31	1.64E-27
GO:0044449	contractile fiber part	CC	6.84E-31	3.20E-27
GO:0030016	myofibril	CC	1.79E-30	6.70E-27
GO:0003008	system process	BP	3.13E-30	9.75E-27
GO:0030017	sarcomere	CC	1.23E-29	3.28E-26
GO:0032501	multicellular organismal process	BP	2.04E-28	4.76E-25
GO:0071944	cell periphery	CC	1.38E-26	2.87E-23
GO:0005886	plasma membrane	CC	1.54E-24	2.89E-21
GO:0005578	proteinaceous extracellular matrix	CC	1.99E-24	3.38E-21
GO:0009888	tissue development	BP	7.85E-24	1.22E-20
GO:0005576	extracellular region	CC	1.33E-22	1.91E-19
GO:0044459	plasma membrane part	CC	4.35E-22	5.81E-19
GO:0044707	single-multicellular organism process	BP	4.39E-21	5.47E-18
GO:0051239	regulation of multicellular organismal process	BP	1.78E-20	2.08E-17
GO:0005615	extracellular space	CC	2.09E-20	2.30E-17
GO:0031012	extracellular matrix	CC	3.03E-20	3.15E-17
GO:0030049	muscle filament sliding	BP	6.97E-20	6.51E-17
GO:0033275	actin-myosin filament sliding	BP	6.97E-20	6.51E-17
GO:0003013	circulatory system process	BP	1.84E-19	1.64E-16
GO:0008015	blood circulation	BP	2.75E-19	2.33E-16
GO:0044057	regulation of system process	BP	1.13E-18	9.09E-16
GO:0055001	muscle cell development	BP	1.17E-18	9.09E-16
GO:0048856	anatomical structure development	BP	2.93E-18	2.19E-15
GO:0005887	integral component of plasma membrane	CC	6.05E-18	4.35E-15
GO:0048731	system development	BP	7.38E-18	5.11E-15
GO:0031226	intrinsic component of plasma membrane	CC	8.71E-18	5.82E-15
GO:0048513	animal organ development	BP	1.97E-17	1.27E-14
GO:0044421	extracellular region part	CC	4.03E-17	2.51E-14
GO:0061061	muscle structure development	BP	5.05E-17	3.04E-14
GO:0006928	movement of cell or subcellular component	BP	5.37E-17	3.14E-14
GO:0030029	actin filament-based process	BP	7.68E-17	4.35E-14
GO:0030154	cell differentiation	BP	8.80E-17	4.84E-14
GO:0031674	I band	CC	1.48E-16	7.91E-14
GO:0055002	striated muscle cell development	BP	2.44E-16	1.27E-13
GO:0009653	anatomical structure morphogenesis	BP	3.45E-16	1.74E-13
GO:0005539	glycosaminoglycan binding	MF	3.62E-16	1.78E-13
GO:0007275	multicellular organism development	BP	1.04E-15	5.01E-13
GO:0040011	locomotion	BP	1.94E-15	8.88E-13
GO:0032502	developmental process	BP	1.95E-15	8.88E-13
GO:0016477	cell migration	BP	2.10E-15	9.32E-13
GO:0070252	actin-mediated cell contraction	BP	2.61E-15	1.13E-12
GO:0048870	cell motility	BP	3.05E-15	1.27E-12
GO:0051674	localization of cell	BP	3.05E-15	1.27E-12
GO:0090257	regulation of muscle system process	BP	4.28E-15	1.74E-12
GO:0004872	receptor activity	MF	4.89E-15	1.90E-12
GO:0060089	molecular transducer activity	MF	4.89E-15	1.90E-12
GO:0030198	extracellular matrix organization	BP	7.37E-15	2.75E-12
GO:0043062	extracellular structure organization	BP	7.37E-15	2.75E-12

Table S4: Top 50 GO terms of DE results from PairedFB at nominal level 5% in the analysis of the carcinoma data by Tuch *et al.* (2010).

(a) Compared with edgeR

	Overlap		Only edgeR		Only PairedFB	
	Gene Set Name	p-Value	Gene Set Name	p-Value	Gene Set Name	p-Value
Top1	**SMID BREAST CANCER NORMAL LIKE UP**	1.68E-87	**SMID BREAST CANCER BASAL UP**	1.99E-13	ZWANG TRANSIENTLY UP BY 1ST EGF PULSE ONLY	1.03E-11
Top2	**SWEET LUNG CANCER KRAS DN**	2.25E-86	ALFANO MYC TARGETS	5.33E-11	ENK UV RESPONSE EPIDERMIS DN	1.05E-08
Top3	LEE BMP2 TARGETS UP	5.96E-86	KAN RESPONSE TO ARSENIC TRIOXIDE	1.48E-10	WONG ADULT TISSUE STEM MODULE	5.90E-08
Top4	BOQUEST STEM CELL CULTURED VS FRESH UP	1.67E-81	MEISSNER BRAIN HCP WITH H3K4ME3 AND H3K27ME3	3.35E-10	**WALLACE PROSTATE CANCER RACE UP**	1.03E-07
Top5	MEISSNER BRAIN HCP WITH H3K4ME3 AND H3K27ME3	2.01E-73	**VECCHI GASTRIC CANCER EARLY UP**	1.03E-09	MEISSNER BRAIN HCP WITH H3K4ME3 AND H3K27ME3	1.40E-07
Top6	CUI TCF21 TARGETS 2 DN	9.59E-68	**SMID BREAST CANCER LUMINAL B DN**	1.36E-09	LEE BMP2 TARGETS UP	6.45E-07
Top7	DODD NASOPHARYNGEAL CARCINOMA UP	3.33E-67	DOUGLAS BMI1 TARGETS UP	1.42E-09	**ACEVEDO LIVER TUMOR VS NORMAL ADJACENT TISSUE DN**	6.81E-07
Top8	**LIU PROSTATE CANCER DN**	3.91E-64	BLUM RESPONSE TO SALIRASIB DN	1.72E-09	PICCALUGA ANGIOIMMUNOBLASTIC LYMPHOMA UP	1.23E-06
Top9	**RICKMAN HEAD AND NECK CANCER F**	1.13E-62	KINSEY TARGETS OF EWSR1 FLII FUSION UP	4.29E-09	MIKKELSEN NPC HCP WITH H3K4ME3 AND H3K27ME3	1.45E-06
Top10	WONG ADULT TISSUE STEM MODULE	1.14E-61	PEREZ TP53 TARGETS	1.06E-08	**ACEVEDO LIVER CANCER DN**	1.69E-06
Top11	NABA MATRISOME	9.55E-61	GRAESSMANN APOPTOSIS BY DOXORUBICIN DN	1.14E-08	RUIZ TNC TARGETS DN	2.14E-06
Top12	**SMID BREAST CANCER LUMINAL B DN**	3.10E-57	DUTERTRE ESTRADIOL RESPONSE 24HR UP	1.82E-08	PEREZ TP53 TARGETS	2.53E-06
Top13	LIM MAMMARY STEM CELL UP	4.50E-57	**FARMER BREAST CANCER BASAL VS LULMINAL**	2.13E-08	NABA MATRISOME	3.14E-06
Top14	CHICAS RB1 TARGETS CONFLUENT	5.11E-56	**CHARAFE BREAST CANCER LUMINAL VS BASAL DN**	2.58E-08	BENPORATH SUZ12 TARGETS	3.49E-06
Top15	PICCALUGA ANGIOIMMUNOBLASTIC LYMPHOMA UP	1.19E-54	MARTINEZ TP53 TARGETS DN	2.81E-08	CHEN METABOLIC SYNDROM NETWORK	3.58E-06
Top16	BOQUEST STEM CELL UP	1.64E-54	ONDER CDH1 TARGETS 2 DN	3.10E-08	**LIU PROSTATE CANCER DN**	5.24E-06
Top17	**LINDGREN BLADDER CANCER CLUSTER 2B**	5.84E-54	LI WILMS TUMOR VS FETAL KIDNEY 1 DN	3.91E-08	BOQUEST STEM CELL UP	5.91E-06
Top18	ONDER CDH1 TARGETS 2 UP	1.41E-52	MASSARWEH TAMOXIFEN RESISTANCE DN	5.03E-08	BURTON ADIPOGENESIS 4	7.87E-06
Top19	**SCHUETZ BREAST CANCER DUCTAL INVASIVE UP**	1.99E-52	BUYTAERT PHOTODYNAMIC THERAPY STRESS DN	5.78E-08	BASAKI YBX1 TARGETS DN	8.22E-06
Top20	**SMID BREAST CANCER BASAL DN**	6.68E-51	ONKEN UVEAL MELANOMA DN	9.94E-08	WANG SMARCE1 TARGETS UP	9.56E-06

(b) Compared with baySeq

	Overlap		Only baySeq		Only PairedFB	
	Gene Set Name	p-Value	Gene Set Name	p-Value	Gene Set Name	p-Value
Top1	**SMID BREAST CANCER NORMAL LIKE UP**	5.23E-88	BERENJENO TRANSFORMED BY RHOA UP	9.12E-15	YANG BCL3 TARGETS UP	2.70E-07
Top2	LEE BMP2 TARGETS UP	1.47E-86	KINSEY TARGETS OF EWSR1 FLII FUSION UP	6.89E-13	YOSHIMURA MAPK8 TARGETS UP	3.53E-07
Top3	**SWEET LUNG CANCER KRAS DN**	1.83E-83	PUJANA BRCA1 PCC NETWORK	2.67E-12	LEE BMP2 TARGETS UP	2.86E-06
Top4	BOQUEST STEM CELL CULTURED VS FRESH UP	2.63E-82	BLALOCK ALZHEIMERS DISEASE DN	2.92E-12	NABA MATRISOME	3.42E-06
Top5	MEISSNER BRAIN HCP WITH H3K4ME3 AND H3K27ME3	4.16E-75	HELLER HDAC TARGETS SILENCED BY METHYLATION UP	3.82E-11	BENPORATH EED TARGETS	4.45E-06
Top6	**DODD NASOPHARYNGEAL CARCINOMA UP**	2.46E-70	REACTOME CELL CYCLE MITOTIC	1.69E-09	MEISSNER BRAIN HCP WITH H3K4ME3 AND H3K27ME3	4.69E-06
Top7	CUI TCF21 TARGETS 2 DN	2.11E-67	KAN RESPONSE TO ARSENIC TRIOXIDE	1.89E-09	KIM WT1 TARGETS UP	6.01E-06
Top8	**LIU PROSTATE CANCER DN**	4.96E-67	REACTOME CELL CYCLE	2.07E-09	BASAKI YBX1 TARGETS DN	6.72E-06
Top9	WONG ADULT TISSUE STEM MODULE	1.93E-66	NUYTTEN EZH2 TARGETS DN	3.03E-09	BENPORATH ES WITH H3K27ME3	6.73E-06
Top10	PICCALUGA ANGIOIMMUNOBLASTIC LYMPHOMA UP	2.72E-62	FEVR CTNNB1 TARGETS DN	4.33E-09	**RICKMAN HEAD AND NECK CANCER C**	1.07E-05
Top11	**RICKMAN HEAD AND NECK CANCER F**	1.57E-61	MEISSNER BRAIN HCP WITH H3K4ME3 AND H3K27ME3	5.72E-09	CHICAS RB1 TARGETS GROWING	1.11E-05
Top12	NABA MATRISOME	5.96E-61	GOBERT OLIGODENDROCYTE DIFFERENTIATION DN	6.65E-09	FEVR CTNNB1 TARGETS UP	1.68E-05
Top13	BOQUEST STEM CELL UP	1.28E-57	BASAKI YBX1 TARGETS UP	8.03E-09	ONDER CDH1 TARGETS 2 DN	1.96E-05
Top14	LIM MAMMARY STEM CELL UP	7.79E-57	**PATIL LIVER CANCER**	1.85E-08	MCDOWELL ACUTE LUNG INJURY UP	2.22E-05
Top15	**SMID BREAST CANCER LUMINAL B DN**	8.68E-57	DUTERTRE ESTRADIOL RESPONSE 24HR UP	2.29E-08	TRAYNOR RETT SYNDROM UP	2.22E-05
Top16	**LINDGREN BLADDER CANCER CLUSTER 2B**	3.78E-55	PENG RAPAMYCIN RESPONSE DN	2.68E-08	CREIGHTON ENDOCRINE THERAPY RESISTANCE 5	2.42E-05
Top17	CHICAS RB1 TARGETS CONFLUENT	1.28E-54	**SMID BREAST CANCER BASAL UP**	2.80E-08	SABATES COLORECTAL ADENOMA DN	2.64E-05
Top18	ONDER CDH1 TARGETS 2 UP	1.76E-54	PUJANA CHEK2 PCC NETWORK	3.12E-08	BENPORATH SUZ12 TARGETS	3.14E-05
Top19	**SCHUETZ BREAST CANCER DUCTAL INVASIVE UP**	8.27E-54	KOBAYASHI EGFR SIGNALING 24HR DN	3.30E-08	GROSS HYPOXIA VIA ELK3 DN	3.81E-05
Top20	**SMID BREAST CANCER BASAL DN**	3.83E-51	**CHIANG LIVER CANCER SUBCLASS PROLIFERATION UP**	3.46E-08	RAMJAUN APOPTOSIS BY TGFB1 VIA MAPK1 DN	3.92E-05

Table S5: Top 20 enriched gene sets of DE genes based on the overlapping genes of both PairedFB and edgeR (or baySeq), edgeR (or baySeq) only and PairedFB only when the nominal level is controlled at 5% for PairedFB method and 3% for edgeR (or baySeq). Bolded gene sets are cancer related or HNSCC related. Underline is used to indicate the common enriched gene sets between results from DEG list of uniquely and commonly identified

Top	PairedFB & edgeR	PairedFB & baySeq	edgeR & baySeq
Top 100	50	57	61
Top 200	141	150	141
Top 500	434	434	410

Table S6: Number of overlapping DE genes in different top ranked gene lists between any two methods from PairedFB, edgeR and baySeq for the carcinoma data by Tuch *et al.* (2010).

(a) Compared with edgeR

	Overlap		Only edgeR		Only PairedFB	
	Gene Set Name	p-Value	Gene Set Name	p-Value	Gene Set Name	p-Value
Top1	**RICKMAN HEAD AND NECK CANCER F**	3.68E-75	NABA MATRISOME	2.23E-09	**DODD NASOPHARYNGEAL CARCINOMA UP**	5.83E-13
Top2	**SMID BREAST CANCER NORMAL LIKE UP**	5.64E-51	BERENJENO TRANSFORMED BY RHOA DN	5.50E-09	**SMID BREAST CANCER NORMAL LIKE UP**	8.44E-11
Top3	**SWEET LUNG CANCER KRAS DN**	7.72E-51	PID SYNDECAN 1 PATHWAY	6.86E-09	**WEST ADRENOCORTICAL TUMOR DN**	3.59E-10
Top4	KUNINGER IGF1 VS PDGFB TARGETS UP	2.61E-45	BOQUEST STEM CELL CULTURED VS FRESH UP	1.06E-08	BOQUEST STEM CELL DN	9.50E-10
Top5	BOQUEST STEM CELL CULTURED VS FRESH UP	2.30E-41	MARTINEZ RB1 TARGETS UP	4.28E-08	**DELYS THYROID CANCER DN**	1.67E-09
Top6	BOQUEST STEM CELL UP	5.20E-36	HAN SATB1 TARGETS UP	1.03E-07	**RICKMAN TUMOR DIFFERENTIATED WELL VS MODERATELY UP**	1.25E-08
Top7	LEE BMP2 TARGETS UP	4.34E-35	VART KSHV INFECTION ANGIOGENIC MARKERS UP	1.49E-07	**SWEET LUNG CANCER KRAS DN**	1.29E-08
Top8	NABA MATRISOME	7.96E-35	BASAKI YBX1 TARGETS UP	2.10E-07	CUI TCF21 TARGETS 2 DN	2.72E-08
Top9	**DODD NASOPHARYNGEAL CARCINOMA UP**	7.89E-34	HAN SATB1 TARGETS DN	2.42E-07	ZHENG BOUND BY FOXP3	3.64E-08
Top10	**VECCHI GASTRIC CANCER EARLY DN**	1.63E-33	**DELYS THYROID CANCER UP**	2.47E-07	ONDER CDH1 TARGETS 2 UP	9.02E-08
Top11	**ACEVEDO FGFR1 TARGETS IN PROSTATE CANCER MODEL DN**	5.10E-33	KOINUMA TARGETS OF SMAD2 OR SMAD3	2.77E-07	BOQUEST STEM CELL UP	1.00E-07
Top12	EBAUER MYOGENIC TARGETS OF PAX3 FOXO1 FUSION	1.16E-31	PID INTEGRIN3 PATHWAY	4.59E-07	NABA CORE MATRISOME	1.47E-07
Top13	SABATES COLORECTAL ADENOMA DN	1.66E-31	KIM WT1 TARGETS 12HR DN	5.96E-07	WANG SMARCE1 TARGETS UP	1.66E-07
Top14	**SMID BREAST CANCER LUMINAL B DN**	7.19E-31	BLALOCK ALZHEIMERS DISEASE UP	7.18E-07	NABA MATRISOME	2.33E-07
Top15	**LIU PROSTATE CANCER DN**	9.52E-31	**SCHUETZ BREAST CANCER DUCTAL INVASIVE UP**	7.56E-07	**ACEVEDO LIVER CANCER WITH H3K27ME3 UP**	2.36E-07
Top16	**SMID BREAST CANCER BASAL DN**	1.74E-30	ROY WOUND BLOOD VESSEL UP	8.50E-07	MEISSNER BRAIN HCP WITH H3K4ME3 AND H3K27ME3	3.42E-07
Top17	CROMER TUMORIGENESIS DN	1.63E-29	NABA MATRISOME ASSOCIATED	1.31E-06	**BONOME OVARIAN CANCER SURVIVAL SUBOPTIMAL DEBULKING**	7.12E-07
Top18	WONG ADULT TISSUE STEM MODULE	5.47E-28	**SMID BREAST CANCER LUMINAL B DN**	1.51E-06	MARKEY RB1 CHRONIC LOF DN	8.08E-07
Top19	BOSCO EPITHELIAL DIFFERENTIATION MODULE	6.52E-28	KRIEG HYPOXIA NOT VIA KDM3A	1.57E-06	WONG ADULT TISSUE STEM MODULE	9.14E-07
Top20	EBAUER TARGETS OF PAX3 FOXO1 FUSION UP	9.05E-28	PID INTEGRIN5 PATHWAY	1.90E-06	LEE BMP2 TARGETS UP	1.20E-06

(b) Compared with baySeq

	Overlap		Only baySeq		Only PairedFB	
	Gene Set Name	p-Value	Gene Set Name	p-Value	Gene Set Name	p-Value
Top1	**RICKMAN HEAD AND NECK CANCER F**	3.68E-75	LEE BMP2 TARGETS UP	6.68E-10	**DODD NASOPHARYNGEAL CARCINOMA UP**	5.83E-13
Top2	**SMID BREAST CANCER NORMAL LIKE UP**	1.62E-53	CAIRO HEPATOBLASTOMA CLASSES DN	2.33E-08	NABA MATRISOME	1.89E-10
Top3	**SWEET LUNG CANCER KRAS DN**	3.93E-52	BOQUEST STEM CELL DN	2.83E-08	HAN SATB1 TARGETS UP	2.70E-10
Top4	KUNINGER IGF1 VS PDGFB TARGETS UP	3.31E-47	VART KSHV INFECTION ANGIOGENIC MARKERS UP	1.49E-07	WONG ADULT TISSUE STEM MODULE	4.61E-10
Top5	BOQUEST STEM CELL UP	2.04E-41	GOZGIT ESR1 TARGETS DN	1.70E-07	BOQUEST STEM CELL CULTURED VS FRESH UP	5.48E-10
Top6	BOQUEST STEM CELL CULTURED VS FRESH UP	1.44E-36	NABA MATRISOME	2.33E-07	LEE BMP2 TARGETS UP	9.03E-09
Top7	**DODD NASOPHARYNGEAL CARCINOMA UP**	7.89E-34	PID INTEGRIN3 PATHWAY	4.59E-07	GOZGIT ESR1 TARGETS DN	1.46E-08
Top8	LEE BMP2 TARGETS UP	4.29E-33	KINSEY TARGETS OF EWSR1 FLII FUSION DN	4.90E-07	**SMID BREAST CANCER NORMAL LIKE UP**	2.79E-08
Top9	**SMID BREAST CANCER LUMINAL B DN**	5.70E-33	**SCHUETZ BREAST CANCER DUCTAL INVASIVE UP**	7.56E-07	BRUINS UVC RESPONSE VIA TP53 GROUP A	6.02E-08
Top10	**DELYS THYROID CANCER DN**	1.21E-32	PEREZ TP53 TARGETS	8.58E-07	FLECHNER BIOPSY KIDNEY TRANSPLANT REJECTED VS OK DN	8.96E-08
Top11	NABA MATRISOME	4.30E-32	DAVICIONI TARGETS OF PAX FOXO1 FUSIONS UP	1.89E-06	BLALOCK ALZHEIMERS DISEASE UP	1.02E-07
Top12	EBAUER MYOGENIC TARGETS OF PAX3 FOXO1 FUSION	1.16E-31	KYNG DNA DAMAGE BY 4NQO OR UV	2.17E-06	**SWEET LUNG CANCER KRAS DN**	2.15E-07
Top13	**VECCHI GASTRIC CANCER EARLY DN**	3.71E-31	CUI TCF21 TARGETS 2 DN	2.89E-06	RODRIGUES THYROID CARCINOMA POORLY DIFFERENTIATED DN	2.24E-07
Top14	**ACEVEDO FGFR1 TARGETS IN PROSTATE CANCER MODEL DN**	1.44E-30	GEORGES TARGETS OF MIR192 AND MIR215	5.22E-06	MEISSNER BRAIN HCP WITH H3K4ME3 AND H3K27ME3	3.42E-07
Top15	EBAUER TARGETS OF PAX3 FOXO1 FUSION UP	2.07E-30	LINDVALL IMMORTALIZED BY TERT DN	5.65E-06	LEI MYB TARGETS	3.90E-07
Top16	SABATES COLORECTAL ADENOMA DN	4.70E-29	SCHAEFFER PROSTATE DEVELOPMENT 48HR UP	6.53E-06	**LIU PROSTATE CANCER DN**	4.59E-07
Top17	**LIU PROSTATE CANCER DN**	1.32E-28	**SWEET LUNG CANCER KRAS UP**	6.88E-06	LIM MAMMARY STEM CELL UP	5.19E-07
Top18	YAUCH HEDGEHOG SIGNALING PARACRINE DN	4.30E-28	**SMID BREAST CANCER BASAL DN**	7.42E-06	PICCALUGA ANGIOIMMUNOBLASTIC LYMPHOMA UP	5.32E-07
Top19	BOSCO EPITHELIAL DIFFERENTIATION MODULE	6.52E-28	MARTORIATI MDM4 TARGETS FETAL LIVER DN	9.26E-06	TAKEDA TARGETS OF NUP98 HOXA9 FUSION 8D DN	5.32E-07
Top20	**SMID BREAST CANCER BASAL DN**	1.35E-27	ZWANG CLASS 1 TRANSIENTLY INDUCED BY EGF	9.49E-06	PILON KLF1 TARGETS UP	6.51E-07

Table S7: Top 20 enriched gene sets of DE genes based on the overlapping genes of both PairedFB and edgeR (or baySeq), edgeR (or baySeq) only and PairedFB only when the top 500 DEGs are considered for each method. Bolded gene sets are cancer related or HNSCC related. Underline is used to indicate the common enriched gene sets between results from DEG list of uniquely and commonly identified.

	Gene Set Name	p-Value	FDR
Top1	**RICKMAN HEAD AND NECK CANCER F**	2.88E-72	1.36E-68
Top2	KUNINGER IGF1 VS PDGFB TARGETS UP	2.50E-47	5.91E-44
Top3	YAUCH HEDGEHOG SIGNALING PARACRINE DN	9.79E-32	1.54E-28
Top4	CHEMELLO SOLEUS VS EDL MYOFIBERS UP	2.58E-25	3.05E-22
Top5	EBAUER MYOGENIC TARGETS OF PAX3 FOXO1 FUSION	8.62E-25	8.15E-22
Top6	BOSCO EPITHELIAL DIFFERENTIATION MODULE	2.04E-22	1.61E-19
Top7	REACTOME STRIATED MUSCLE CONTRACTION	3.01E-20	2.04E-17
Top8	REN ALVEOLAR RHABDOMYOSARCOMA UP	5.95E-20	3.52E-17
Top9	SWEET LUNG CANCER KRAS DN	2.69E-19	1.41E-16
Top10	EBAUER TARGETS OF PAX3 FOXO1 FUSION UP	3.70E-19	1.75E-16
Top11	**RICKMAN HEAD AND NECK CANCER E**	7.07E-19	3.04E-16
Top12	HUMMERICH SKIN CANCER PROGRESSION DN	3.94E-18	1.55E-15
Top13	ACEVEDO FGFR1 TARGETS IN PROSTATE CANCER MODEL DN	2.35E-17	8.56E-15
Top14	REACTOME MUSCLE CONTRACTION	4.87E-17	1.64E-14
Top15	**CROMER TUMORIGENESIS DN**	1.02E-16	3.20E-14
Top16	SABATES COLORECTAL ADENOMA DN	2.96E-15	8.74E-13
Top17	DODD NASOPHARYNGEAL CARCINOMA UP	5.62E-15	1.56E-12
Top18	MARTINEZ TP53 TARGETS UP	3.96E-14	1.04E-11
Top19	DELASERNA MYOD TARGETS UP	6.85E-14	1.70E-11
Top20	BOQUEST STEM CELL UP	1.60E-13	3.79E-11

Table S8: Top 20 enriched gene sets of the biologically significant DEG list (180 genes) from PairedFB with mean fold change exceeding 4 and nominal FDR controlled at 0.05. Bolded gene sets are OSCC or HNSCC related. P-value is calculated for the Fisher's exact test for association and FDR adjusts for multiple testing for all the curated gene sets considered.

	Gene Set Name	p-Value	FDR
Top1	**RICKMAN HEAD AND NECK CANCER F**	2.02E-70	9.55E-67
Top2	SMID BREAST CANCER NORMAL LIKE UP	3.09E-65	7.30E-62
Top3	SWEET LUNG CANCER KRAS DN	1.62E-63	2.56E-60
Top4	BOQUEST STEM CELL CULTURED VS FRESH UP	8.05E-62	9.52E-59
Top5	NABA MATRISOME	3.08E-49	2.91E-46
Top6	LIU PROSTATE CANCER DN	3.42E-47	2.70E-44
Top7	BOQUEST STEM CELL UP	1.63E-45	1.10E-42
Top8	DODD NASOPHARYNGEAL CARCINOMA UP	2.60E-45	1.53E-42
Top9	SMID BREAST CANCER LUMINAL B DN	1.20E-43	6.31E-41
Top10	KUNINGER IGF1 VS PDGFB TARGETS UP	1.63E-41	7.71E-39
Top11	VECCHI GASTRIC CANCER EARLY DN	8.22E-41	3.53E-38
Top12	SMID BREAST CANCER BASAL DN	2.45E-40	9.66E-38
Top13	WONG ADULT TISSUE STEM MODULE	1.62E-39	5.89E-37
Top14	LEE BMP2 TARGETS UP	1.44E-38	4.86E-36
Top15	SABATES COLORECTAL ADENOMA DN	9.66E-37	3.04E-34
Top16	ACEVEDO FGFR1 TARGETS IN PROSTATE CANCER MODEL DN	1.04E-36	3.08E-34
Top17	DELYS THYROID CANCER DN	1.68E-36	4.66E-34
Top18	NABA CORE MATRISOME	2.38E-34	6.25E-32
Top19	CUI TCF21 TARGETS 2 DN	7.11E-33	1.77E-30
Top20	MEISSNER BRAIN HCP WITH H3K4ME3 AND H3K27ME3	2.01E-32	4.76E-30

Table S9: Top 20 enriched gene sets of the biologically significant genes (634 genes) from edgeR with mean fold change exceeding 4 among DEG list at nominal level 0.05. Bolded gene sets are HNSCC related. P-value is calculated for the Fisher's exact test for association and FDR adjusts for multiple testing for all the curated gene sets considered.

References

Abramowitz, M. and Stegun, I. A. (1964). Hypergeometric functions. In *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55, chapter 15, pages 555–566. Courier Corporation.

Chung, L. M., Ferguson, J. P., Zheng, W., Qian, F., Bruno, V., Montgomery, R. R., and Zhao, H. (2013). Differential expression analysis for paired rna-seq data. *BMC bioinformatics*, **14**(1), 110.

Hardcastle, T. J. and Kelly, K. A. (2013). Empirical bayesian analysis of paired high-throughput sequencing data with a beta-binomial distribution. *BMC bioinformatics*, **14**(1), 135.

Jones, M. (2004). The minimax distribution: a beta-type distribution with some tractability advantages.

Robinson, M. D. and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of rna-seq data. *Genome biology*, **11**(3), 1.

Tuch, B. B., Laborde, R. R., Xu, X., Gu, J., Chung, C. B., Monighetti, C. K., Stanley, S. J., Olsen, K. D., Kasperbauer, J. L., Moore, E. J., *et al.* (2010). Tumor transcriptome sequencing reveals allelic expression imbalances associated with copy number alterations. *PloS one*, **5**(2), e9317.

Van De Wiel, M. A., Leday, G. G., Pardo, L., Rue, H., Van Der Vaart, A. W., and Van Wieringen, W. N. (2013). Bayesian analysis of rna sequencing data by estimating multiple shrinkage priors. *Biostatistics*, **14**(1), 113–128.

Yu, Y.-H., Kuo, H.-K., and Chang, K.-W. (2008). The evolving transcriptome of head and neck squamous cell carcinoma: a systematic review. *PLoS One*, **3**(9), e3215.