

Multi-resolution hypothesis testing

Anirban Samaddar
Gustavo de los Campos

Michigan State University

August 28, 2020

1 Introduction

- An example of GWAS
- Multi-resolution hypothesis testing

2 FDR control strategy

- Our approach
- One existing method

3 Simulation

- Single chain Gibbs sampling
- Multiple chain Gibbs sampling

An example of GWAS

As a motivation of our work, we present an example of a GWAS using [mice](#) data from the [BGLR](#) package.

An example of GWAS

As a motivation of our work, we present an example of a GWAS using `mice` data from the `BGLR` package.

For this example we take the first $p = 500$ columns from `mice.X` matrix. These represent genotypical information in our example. The sample size is $n = 1814$.

An example of GWAS

As a motivation of our work, we present an example of a GWAS using `mice` data from the `BGLR` package.

For this example we take the first $p = 500$ columns from `mice.X` matrix. These represent genotypical information in our example. The sample size is $n = 1814$.

We generate Y or phenotype data by putting effects to 5 markers each placed significantly away from the others and adding random noise $E \sim N(0, \sigma_e^2 I)$. We generate the data in such a way that collectively the 5 SNPs explain 10% variance of Y .

An example of GWAS

We fit the following Bayesian linear regression model using **BGLR**:

$$Y = X\beta + E, \text{ where, } E \sim N(0, \sigma_e^2 I)$$

An example of GWAS

We fit the following Bayesian linear regression model using **BGLR**:

$$Y = X\beta + E, \text{ where, } E \sim N(0, \sigma_e^2 I)$$

We choose **model=BayesC** which assumes the variable selection or spike-slab prior on β :

$$p(\beta) = \prod_j [\pi N(\beta_j | 0, \sigma_\beta^2) + (1 - \pi) \mathbb{1}(\beta_j = 0)]$$

An example of GWAS

We fit the following Bayesian linear regression model using **BGLR**:

$$Y = X\beta + E, \text{ where, } E \sim N(0, \sigma_e^2 I)$$

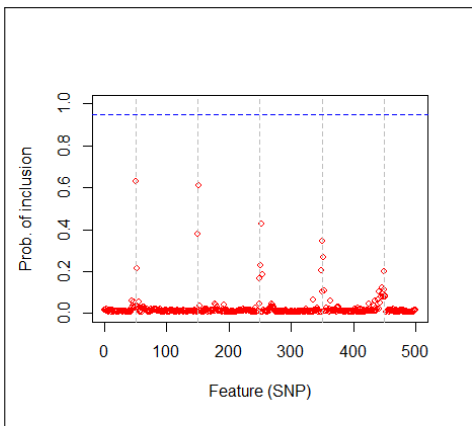
We choose **model=BayesC** which assumes the variable selection or spike-slab prior on β :

$$p(\beta) = \prod_j [\pi N(\beta_j | 0, \sigma_\beta^2) + (1 - \pi) \mathbb{1}(\beta_j = 0)]$$

Since the exact posterior is intractable, **BGLR** uses Gibbs sampling to sample from the exact posterior.

An example of GWAS

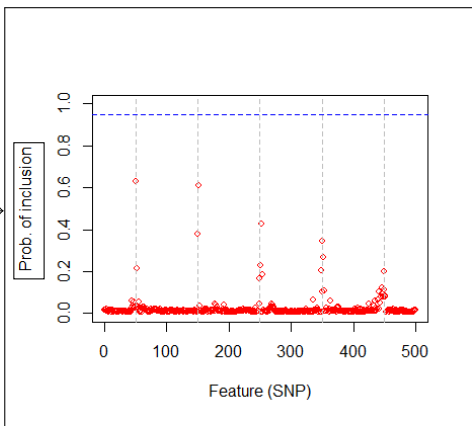
Let's look at the following Manhattan plot from our fitted model:



An example of GWAS

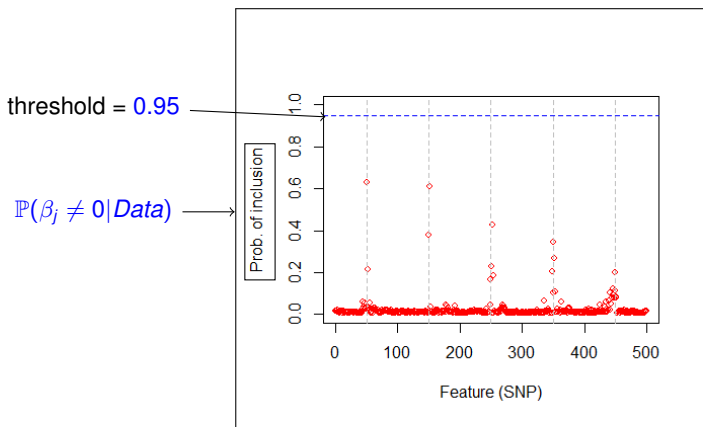
Let's look at the following Manhattan plot from our fitted model:

$$\mathbb{P}(\beta_j \neq 0 | \text{Data})$$



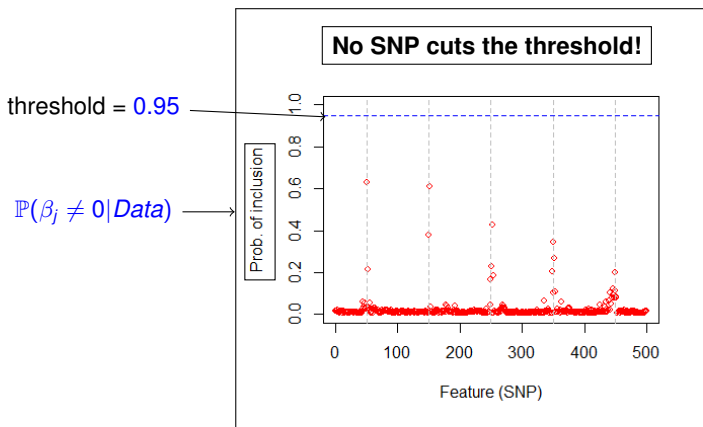
An example of GWAS

Let's look at the following Manhattan plot from our fitted model:



An example of GWAS

Let's look at the following Manhattan plot from our fitted model:



An example of GWAS

Objective: To find which of the 500 markers can be associated with the trait.

An example of GWAS

Objective: To find which of the 500 markers can be associated with the trait.

Formally we are interested in the following set of hypotheses:

$$\{H_{0j} : \beta_j = 0 \text{ vs } H_{0j}^c ; j = 1(1)500\}$$

An example of GWAS

Objective: To find which of the 500 markers can be associated with the trait.

Formally we are interested in the following set of hypotheses:

$$\{H_{0j} : \beta_j = 0 \text{ vs } H_{0j}^c ; j = 1(1)500\}$$

Decision Rule:

Let us fix a threshold 0.05 and we reject H_{0j} if,

$$\mathbb{P}(\beta_j = 0 | \text{Data}) \leq 0.05 \quad \forall j$$

$$\Rightarrow \text{local } fdr_j \leq 0.05$$

$$\Rightarrow \mathbb{P}(\beta_j \neq 0 | \text{Data}) \geq 0.95$$

An example of GWAS

Objective: To find which of the 500 markers can be associated with the trait.

Formally we are interested in the following set of hypotheses:

$$\{H_{0j} : \beta_j = 0 \text{ vs } H_{0j}^c ; j = 1(1)500\}$$

Decision Rule:

Let us fix a threshold 0.05 and we reject H_{0j} if,

$$\mathbb{P}(\beta_j = 0 | \text{Data}) \leq 0.05 \quad \forall j$$

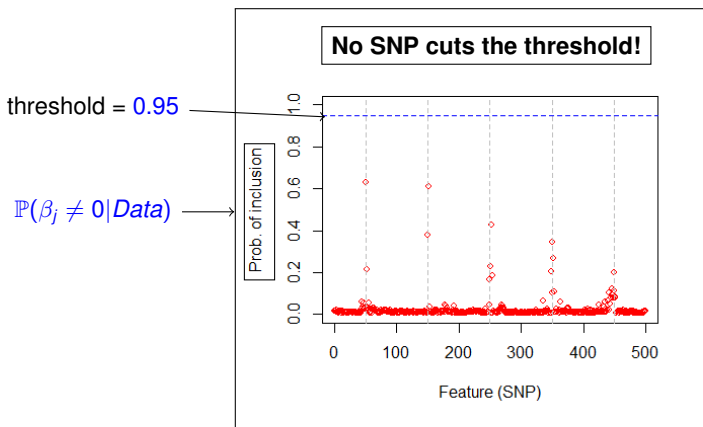
$$\Rightarrow \text{local } fdr_j \leq 0.05$$

$$\Rightarrow \mathbb{P}(\beta_j \neq 0 | \text{Data}) \geq 0.95$$

(Note that, if we reject H_{0j} then we are making a false detection at that marker with probability $\mathbb{P}(\beta_j = 0 | \text{Data})$. So it makes sense to define the decision rule in terms of $\mathbb{P}(\beta_j = 0 | \text{Data})$)

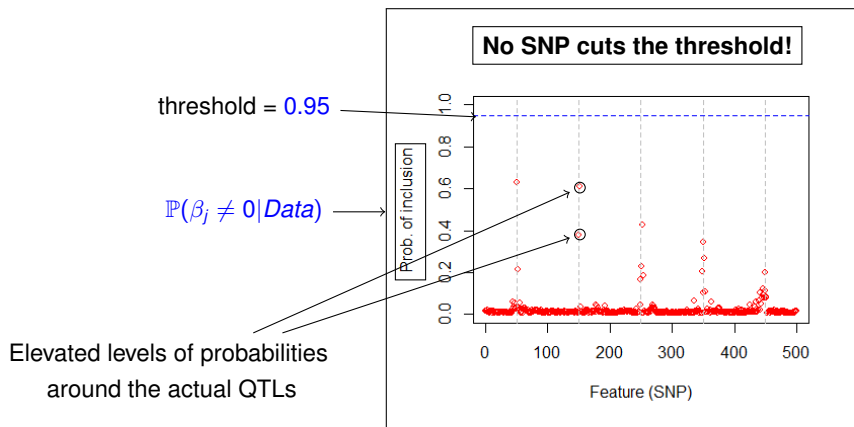
An example of GWAS

Let's look at the following Manhattan plot from our fitted model:



An example of GWAS

Let's look at the following Manhattan plot from our fitted model:

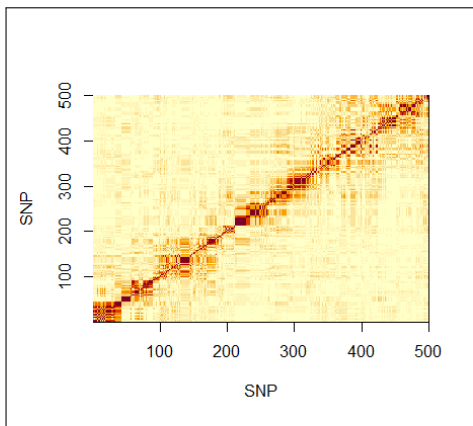


Correlation structure

We investigate this behavior by looking at the correlation structure between the columns of X :

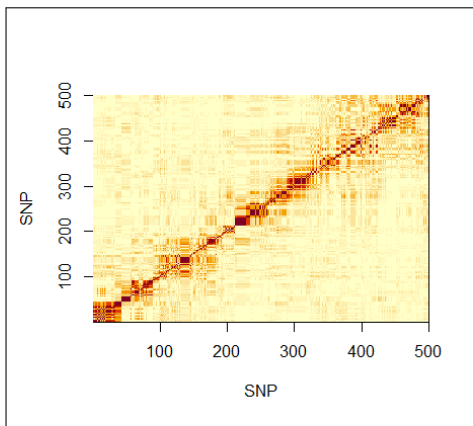
Correlation structure

We investigate this behavior by looking at the correlation structure between the columns of X :



Correlation structure

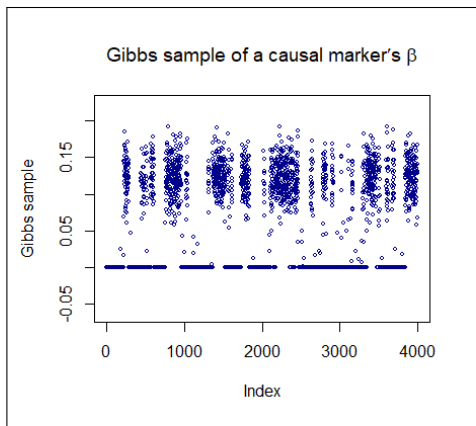
We investigate this behavior by looking at the correlation structure between the columns of X :



There is high LD between an SNP and its neighbors

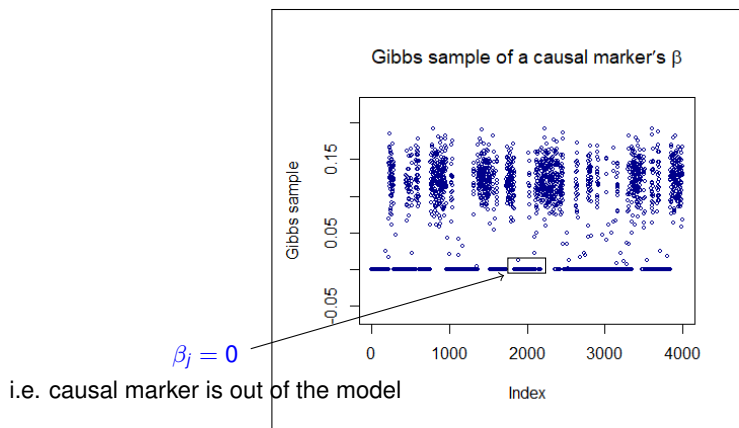
Trace plot

Due to the high LD between consecutive SNPs, often the Gibbs sampler fails to include the causal markers in the model and includes nearby markers. This is evident from the following trace plot of the Gibbs samples:



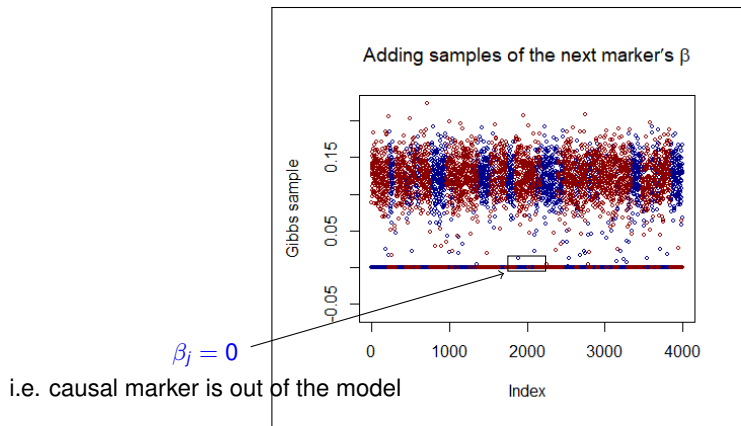
Trace plot

Due to the high LD between consecutive SNPs, often the Gibbs sampler fails to include the causal markers in the model and includes nearby markers. This is evident from the following trace plot of the Gibbs samples:



Trace plot

Due to the high LD between consecutive SNPs, often the Gibbs sampler fails to include the causal markers in the model and includes nearby markers. This is evident from the following trace plot of the Gibbs samples:



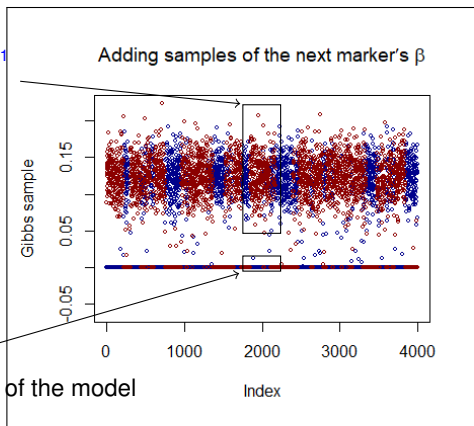
Trace plot

Due to the high LD between consecutive SNPs, often the Gibbs sampler fails to include the causal markers in the model and includes nearby markers. This is evident from the following trace plot of the Gibbs samples:

Correlated neighbor β_{j+1}
included in the model

$\beta_j = 0$

i.e. causal marker is out of the model



Summary of the GWAS example

When there is high LD, very few SNPs make it to the discovery set that has a small FDR threshold. In the example, no SNP gets included into the discovery set with 0.05 FDR threshold.

Summary of the GWAS example

When there is high LD, very few SNPs make it to the discovery set that has a small FDR threshold. In the example, no SNP gets included into the discovery set with 0.05 FDR threshold.

Hence we ask whether it is possible to discover **clusters** of SNPs where FDR is controlled. For example, if we consider $\{\beta_j, \beta_{j+1}\}$ together then they have $\mathbb{P}(\beta_j \neq 0 \text{ OR } \beta_{j+1} \neq 0 | \text{Data}) \geq 0.95$. Hence, together they get included in the discovery set.

Summary of the GWAS example

When there is high LD, very few SNPs make it to the discovery set that has a small FDR threshold. In the example, no SNP gets included into the discovery set with 0.05 FDR threshold.

Hence we ask whether it is possible to discover **clusters** of SNPs where FDR is controlled. For example, if we consider $\{\beta_j, \beta_{j+1}\}$ together then they have $\mathbb{P}(\beta_j \neq 0 \text{ OR } \beta_{j+1} \neq 0 | \text{Data}) \geq 0.95$. Hence, together they get included in the discovery set.

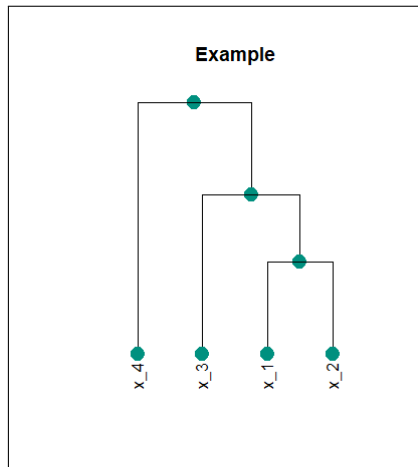
This should lead us to more discoveries than before.

Method

Let us consider a model with 4 SNPs and following is the hierarchical clustering structure between the columns of X :

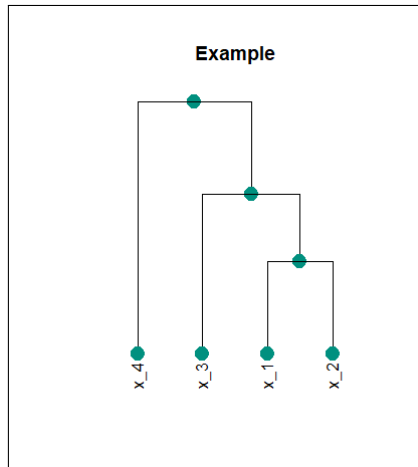
Method

Let us consider a model with 4 SNPs and following is the hierarchical clustering structure between the columns of X :



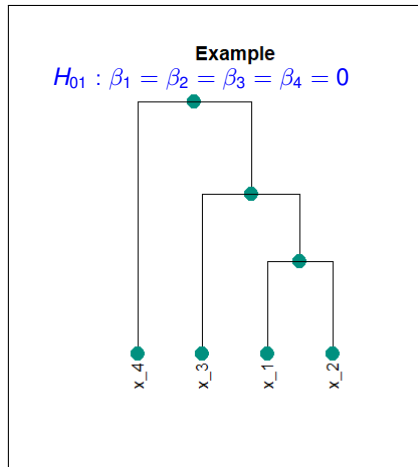
Method

We define the null hypotheses corresponding to each cluster in the hierarchy. At each cluster we are interested in testing H_{0i} vs $H_{1i} = H_{0i}^c$.



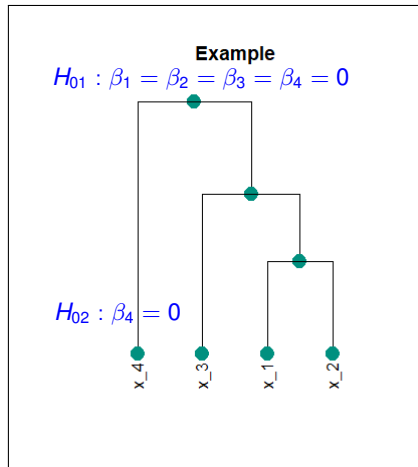
Method

We define the null hypotheses corresponding to each cluster in the hierarchy. At each cluster we are interested in testing H_{0i} vs $H_{1i} = H_{0i}^c$.



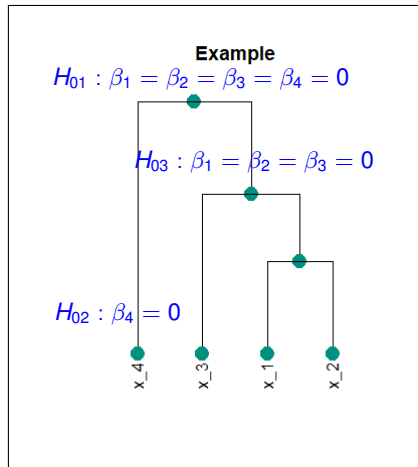
Method

We define the null hypotheses corresponding to each cluster in the hierarchy. At each cluster we are interested in testing H_{0i} vs $H_{1i} = H_{0i}^c$.



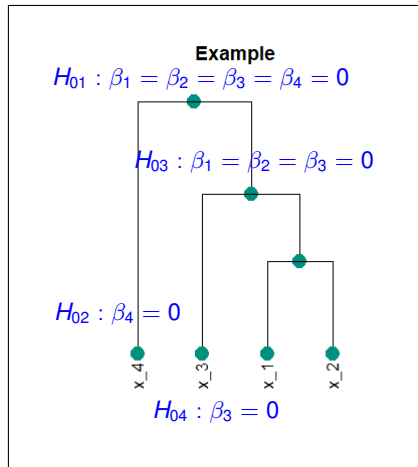
Method

We define the null hypotheses corresponding to each cluster in the hierarchy. At each cluster we are interested in testing H_{0i} vs $H_{1i} = H_{0i}^c$.



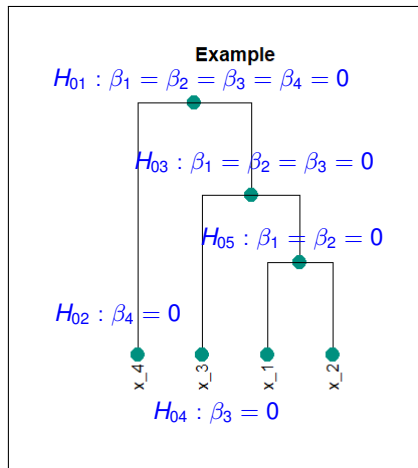
Method

We define the null hypotheses corresponding to each cluster in the hierarchy. At each cluster we are interested in testing H_{0i} vs $H_{1i} = H_{0i}^c$.



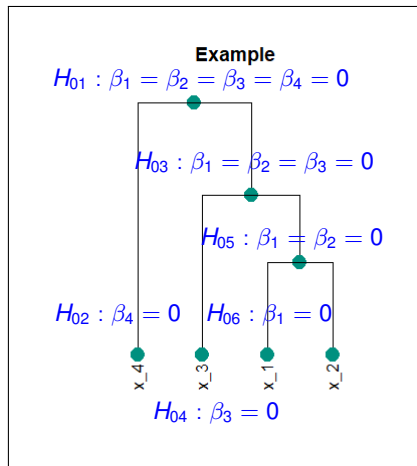
Method

We define the null hypotheses corresponding to each cluster in the hierarchy. At each cluster we are interested in testing H_{0i} vs $H_{1i} = H_{0i}^c$.



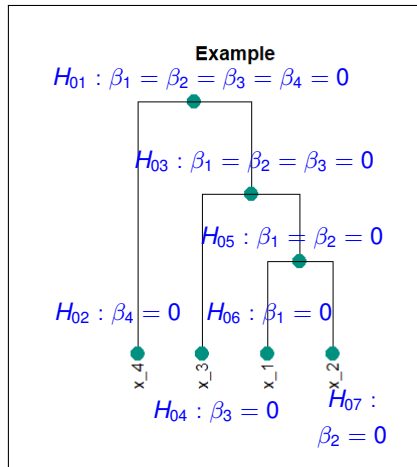
Method

We define the null hypotheses corresponding to each cluster in the hierarchy. At each cluster we are interested in testing H_{0i} vs $H_{1i} = H_{0i}^c$.



Method

We define the null hypotheses corresponding to each cluster in the hierarchy. At each cluster we are interested in testing H_{0i} vs $H_{1i} = H_{0i}^c$.

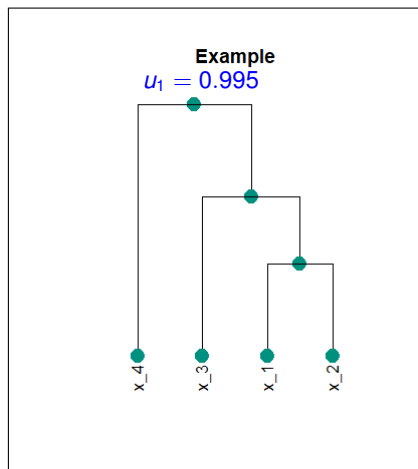


Method

The inclusion probability of at least one of the SNPs in a cluster be: $u_i = \mathbb{P}(H_{1i} | \text{Data})$. Let us assume the following values of u_i in the hierarchy:

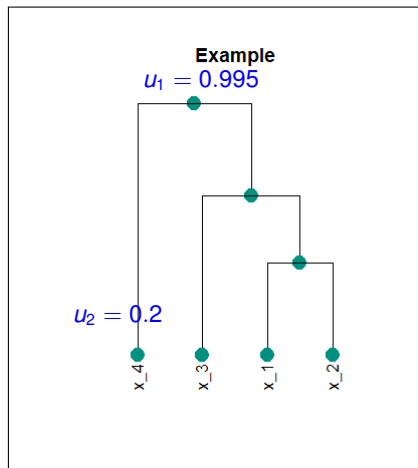
Method

The inclusion probability of at least one of the SNPs in a cluster be:
 $u_i = \mathbb{P}(H_{1i} | \text{Data})$. Let us assume the following values of u_i in the hierarchy:



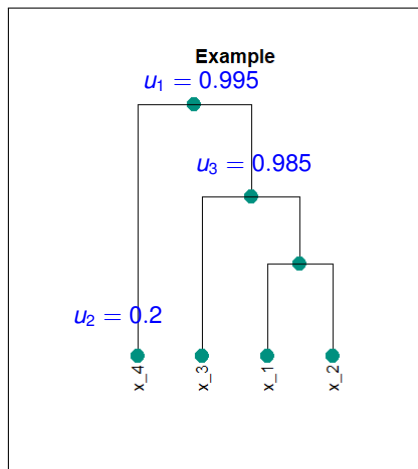
Method

The inclusion probability of at least one of the SNPs in a cluster be:
 $u_i = \mathbb{P}(H_{1i} | \text{Data})$. Let us assume the following values of u_i in the hierarchy:



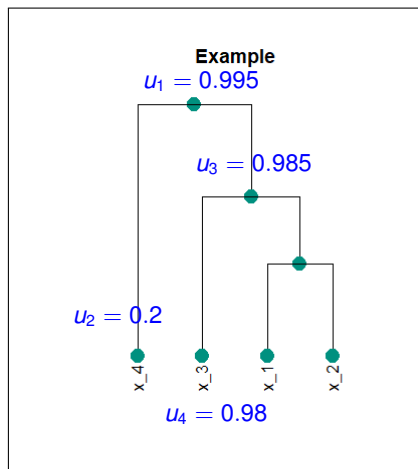
Method

The inclusion probability of at least one of the SNPs in a cluster be:
 $u_i = \mathbb{P}(H_{1i} | \text{Data})$. Let us assume the following values of u_i in the hierarchy:



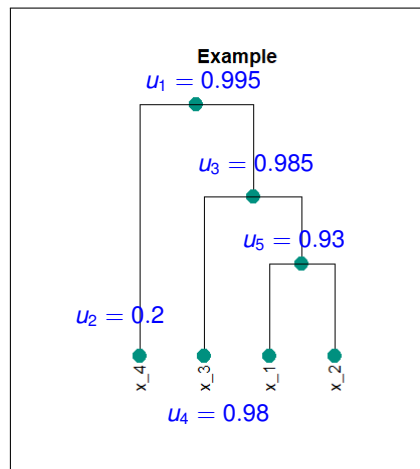
Method

The inclusion probability of at least one of the SNPs in a cluster be:
 $u_i = \mathbb{P}(H_{1i} | \text{Data})$. Let us assume the following values of u_i in the hierarchy:



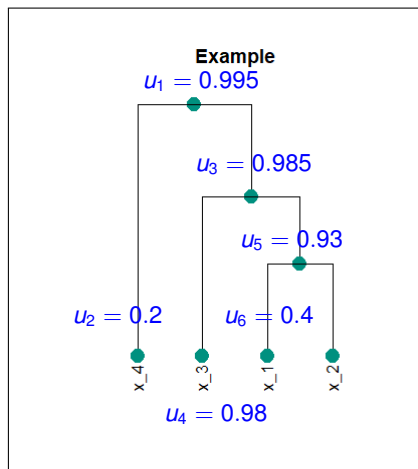
Method

The inclusion probability of at least one of the SNPs in a cluster be: $u_i = \mathbb{P}(H_{1i} | \text{Data})$. Let us assume the following values of u_i in the hierarchy:



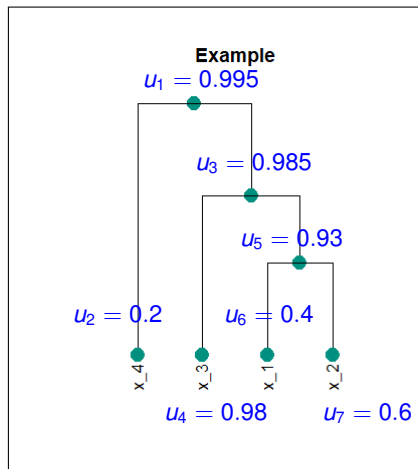
Method

The inclusion probability of at least one of the SNPs in a cluster be:
 $u_i = \mathbb{P}(H_{1i} | \text{Data})$. Let us assume the following values of u_i in the hierarchy:



Method

The inclusion probability of at least one of the SNPs in a cluster be: $u_i = \mathbb{P}(H_{1i} | \text{Data})$. Let us assume the following values of u_i in the hierarchy:

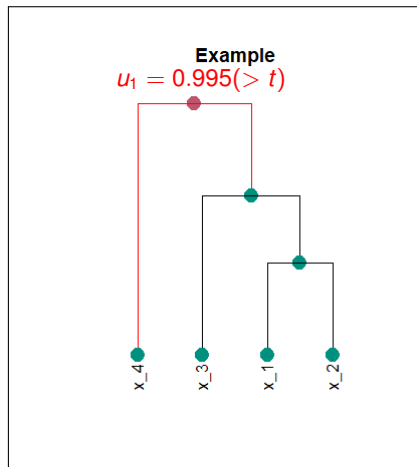


Method

We proceed with hierarchically testing each hypothesis starting from the top of the tree. Let us fix the threshold $t = (1 - t^*) = 0.9$.

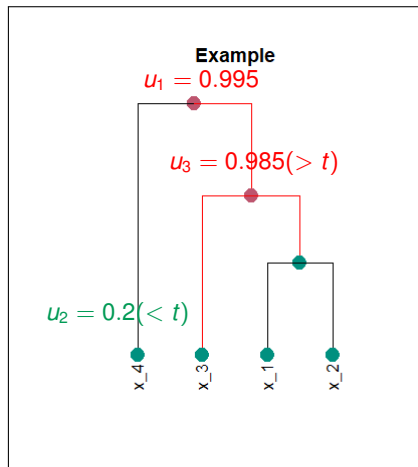
Method

We proceed with hierarchically testing each hypothesis starting from the top of the tree. Let us fix the threshold $t = (1 - t^*) = 0.9$.



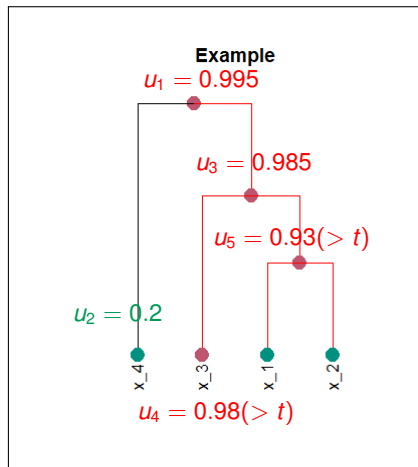
Method

We proceed with hierarchically testing each hypothesis starting from the top of the tree. Let us fix the threshold $t = (1 - t^*) = 0.9$.



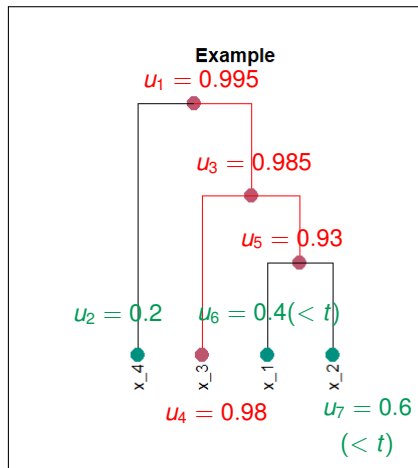
Method

We proceed with hierarchically testing each hypothesis starting from the top of the tree. Let us fix the threshold $t = (1 - t^*) = 0.9$.



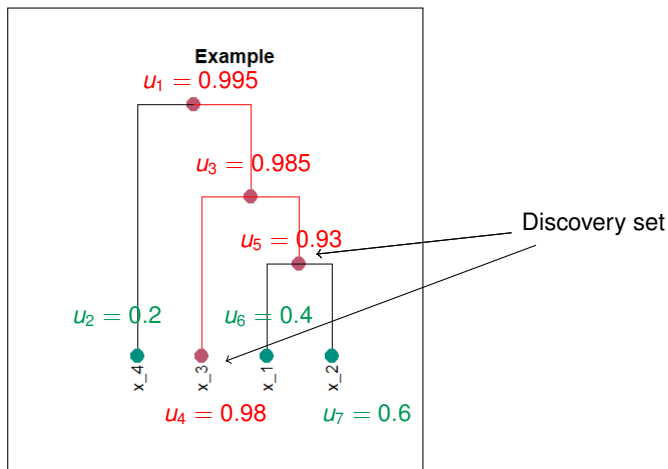
Method

We proceed with hierarchically testing each hypothesis starting from the top of the tree. Let us fix the threshold $t = (1 - t^*) = 0.9$.



Method

So the discovery set is $\{\{\beta_3\}, \{\beta_1, \beta_2\}\}$



If we perform the set of tests: $\{H_{0i} : \beta_i = 0 \text{ vs } H_{1i} = H_{0i}^c ; i = 1(1)4\}$,
i.e. test at the individual SNP level, then we discover only $\{\beta_3\}$.

If we perform the set of tests: $\{H_{0i} : \beta_i = 0 \text{ vs } H_{1i} = H_{0i}^c ; i = 1(1)4\}$, i.e. test at the individual SNP level, then we discover only $\{\beta_3\}$.

Since we have exploited the clustering structure of the columns of \mathbf{X} and performed a hierarchical testing, we have discovered $\{\{\beta_3\}, \{\beta_1, \beta_2\}\}$.

If we perform the set of tests: $\{H_{0i} : \beta_i = 0 \text{ vs } H_{1i} = H_{0i}^c ; i = 1(1)4\}$, i.e. test at the individual SNP level, then we discover only $\{\beta_3\}$.

Since we have exploited the clustering structure of the columns of X and performed a hierarchical testing, we have discovered $\{\{\beta_3\}, \{\beta_1, \beta_2\}\}$.

Note that, the discovery set contains clusters of β s. The interpretation of a discovery set cluster $\{\beta_1, \beta_2\}$ is: at least one $\beta \in \{\beta_1, \beta_2\}$ is non-zero.

In general we can expect this method to provide us:

In general we can expect this method to provide us:

i) More discovery than testing at the individual SNP level. Hence we can expect increase in power.

In general we can expect this method to provide us:

- i) More discovery than testing at the individual SNP level. Hence we can expect increase in power.
- ii) Optimal resolution level where discovery is possible without making significant false discoveries.

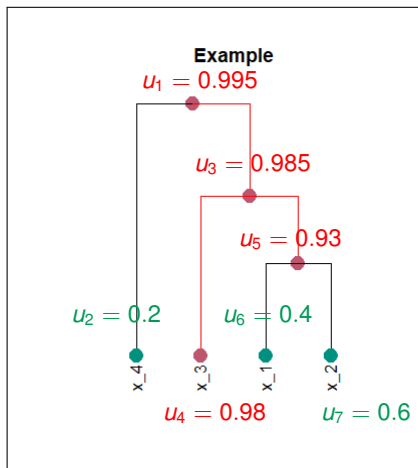
In general we can expect this method to provide us:

- i) More discovery than testing at the individual SNP level. Hence we can expect increase in power.
- ii) Optimal resolution level where discovery is possible without making significant false discoveries.

Note that, in general, more discovery doesn't necessarily increase power if we don't control any false discovery error rate. In the next section we discuss one such rate - Bayesian FDR or **BFDR**

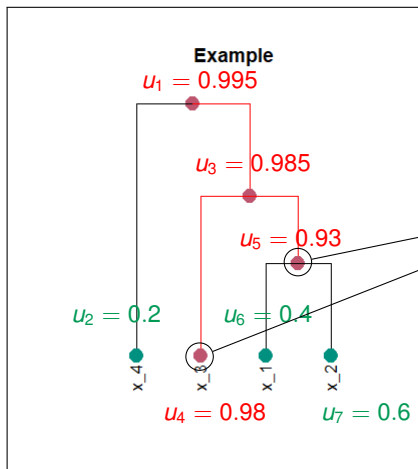
FDR control strategy - example

So in our example where $t = 0.9$:



FDR control strategy - example

So in our example where $t = 0.9$:



Discovery set

$$BFDR(t) = \frac{((1-u_4) + (1-u_5))}{2} = 0.045$$

FDR control strategy

Our interest is to control FDR at the discovery set level.

FDR control strategy

Our interest is to control FDR at the discovery set level.

If our discovery set is $\mathcal{A}(t)$ for a fixed threshold $t \in (0, 1)$, then the Bayesian FDR or $BFDR(t) = \frac{\sum_{i \in I(\mathcal{A}(t))} (1 - u_i)}{|I(\mathcal{A}(t))|}$, where, $I(\mathcal{A}(t))$ is the index set of the clusters in $\mathcal{A}(t)$ and $|\cdot|$ is the cardinality of a set.

FDR control strategy

Our interest is to control FDR at the discovery set level.

If our discovery set is $\mathcal{A}(t)$ for a fixed threshold $t \in (0, 1)$, then the Bayesian FDR or $BFDR(t) = \frac{\sum_{i \in I(\mathcal{A}(t))} (1 - u_i)}{|I(\mathcal{A}(t))|}$, where, $I(\mathcal{A}(t))$ is the index set of the clusters in $\mathcal{A}(t)$ and $|\cdot|$ is the cardinality of a set.

As $(1 - u_i) = \text{local } fdr_i$, the $BFDR(t)$ is the average of *local fdrs* in the discovery set.

FDR control strategy

So the next question is - given an $\alpha \in (0, 1)$ can we find an optimum threshold t such that the discovery set $BFDR(t) \leq \alpha$?

FDR control strategy

So the next question is - given an $\alpha \in (0, 1)$ can we find an optimum threshold t such that the discovery set $BFDR(t) \leq \alpha$?

Formally, given an $\alpha \in (0, 1)$, we want to find -
 $t_{min} = \min\{t : BFDR(t) \leq \alpha\}$

FDR control strategy

So the next question is - given an $\alpha \in (0, 1)$ can we find an optimum threshold t such that the discovery set $BFDR(t) \leq \alpha$?

Formally, given an $\alpha \in (0, 1)$, we want to find -
 $t_{min} = \min\{t : BFDR(t) \leq \alpha\}$

So the optimum decision rule:

$$u_i \geq t_{min} \quad \forall i$$

FDR control strategy

So the next question is - given an $\alpha \in (0, 1)$ can we find an optimum threshold t such that the discovery set $BFDR(t) \leq \alpha$?

Formally, given an $\alpha \in (0, 1)$, we want to find -
 $t_{min} = \min\{t : BFDR(t) \leq \alpha\}$

So the optimum decision rule:

$$u_i \geq t_{min} \quad \forall i$$

And based on the above decision rule we find the optimum discovery set $\mathcal{A}(t_{min})$.

FDR control strategy

So the next question is - given an $\alpha \in (0, 1)$ can we find an optimum threshold t such that the discovery set $BFDR(t) \leq \alpha$?

Formally, given an $\alpha \in (0, 1)$, we want to find -
 $t_{min} = \min\{t : BFDR(t) \leq \alpha\}$

So the optimum decision rule:

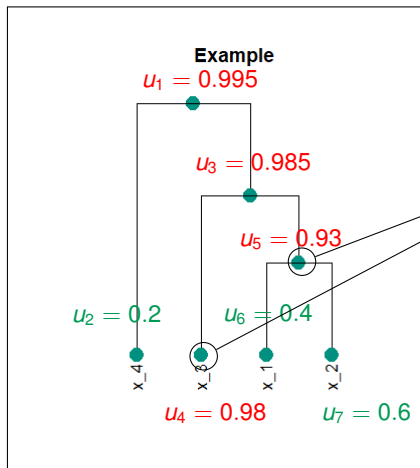
$$u_i \geq t_{min} \quad \forall i$$

And based on the above decision rule we find the optimum discovery set $\mathcal{A}(t_{min})$.

This way of controlling Bayesian FDR has been discussed in Genovese and Wasserman (2002).

FDR control strategy - example

So when $\alpha = 0.05$ the optimal threshold $t_{min} = 0.93$ and $BFDR = 0.045$.

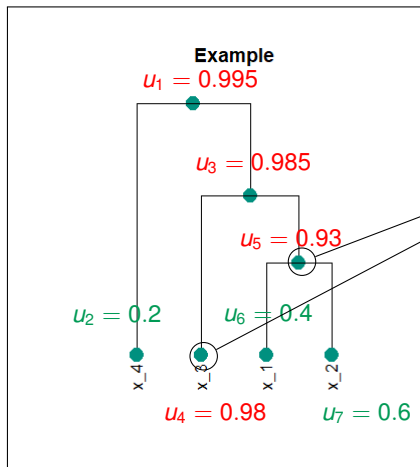


$$t_{min} = u_5$$

Discovery set
 $BFDR = 0.045$
 < 0.05

FDR control strategy - example

So when $\alpha = 0.05$ the optimal threshold $t_{min} = 0.93$ and $BFDR = 0.045$.



$$t_{min} = u_5$$

Discovery set
 $BFDR = 0.045$
 < 0.05

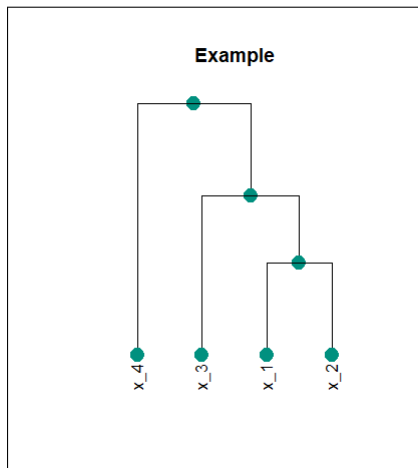
Note that,
even if $BFDR$
threshold is 0.05
the optimal
local fdr threshold is
 $(1 - t_{min}) = 0.07 > 0.05$

Existing method

Daniel Yekutieli (2008) explored a hierarchical testing procedure where he focused on controlling FDR for disjoint subfamilies of hypotheses when the subfamilies are arranged in a tree. So in our example:

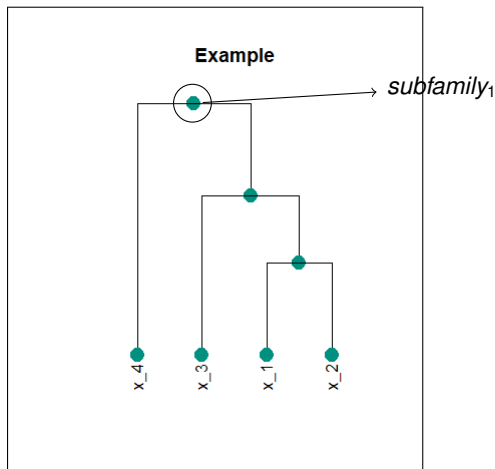
Existing method

Daniel Yekutieli (2008) explored a hierarchical testing procedure where he focused on controlling FDR for disjoint subfamilies of hypotheses when the subfamilies are arranged in a tree. So in our example:



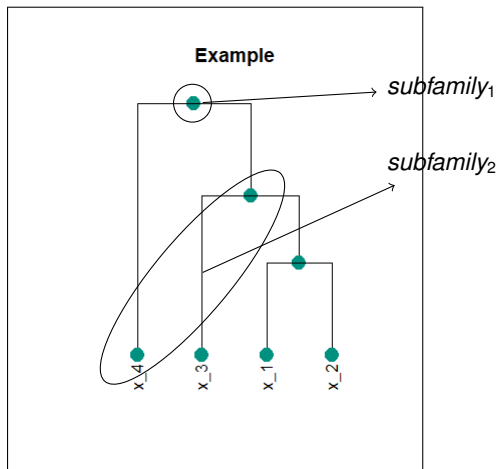
Existing method

Daniel Yekutieli (2008) explored a hierarchical testing procedure where he focused on controlling FDR for disjoint subfamilies of hypotheses when the subfamilies are arranged in a tree. So in our example:



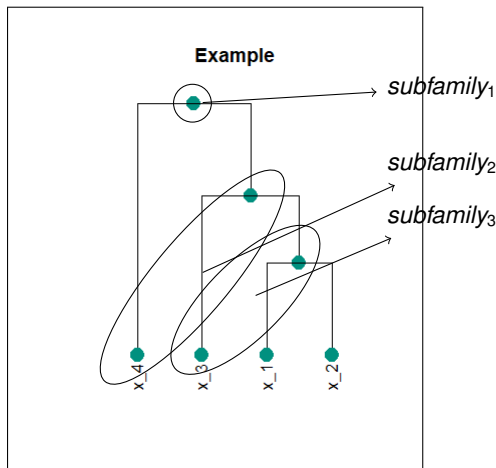
Existing method

Daniel Yekutieli (2008) explored a hierarchical testing procedure where he focused on controlling FDR for disjoint subfamilies of hypotheses when the subfamilies are arranged in a tree. So in our example:



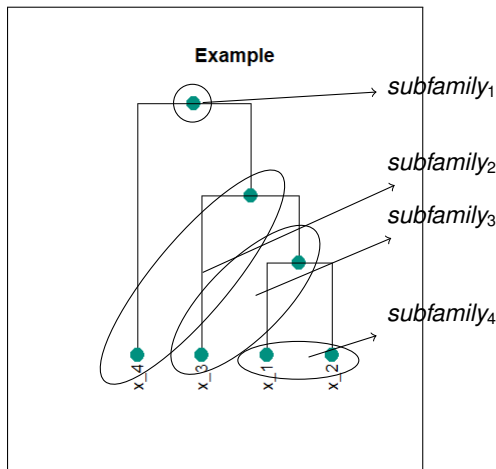
Existing method

Daniel Yekutieli (2008) explored a hierarchical testing procedure where he focused on controlling FDR for disjoint subfamilies of hypotheses when the subfamilies are arranged in a tree. So in our example:



Existing method

Daniel Yekutieli (2008) explored a hierarchical testing procedure where he focused on controlling FDR for disjoint subfamilies of hypotheses when the subfamilies are arranged in a tree. So in our example:



Existing method - comparison

Our method is different because -

i) In the paper, he uses B-H procedure (Benjamini and Hochberg (1995)) to control the expected FDR of the subfamilies. Whereas we use a Bayesian approach and find an optimum decision rule based on posterior probability of inclusion to control the Bayesian FDR of the discovery set.

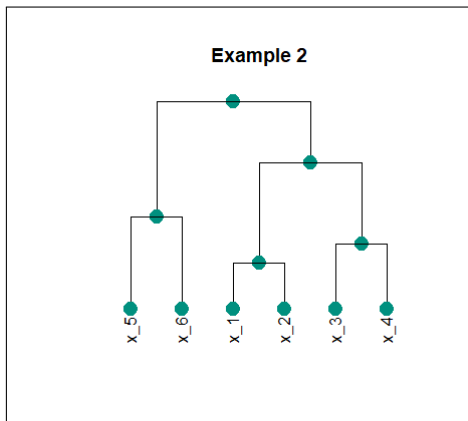
Existing method - comparison

Our method is different because -

- i) In the paper, he uses B-H procedure (Benjamini and Hochberg (1995)) to control the expected FDR of the subfamilies. Whereas we use a Bayesian approach and find an optimum decision rule based on posterior probability of inclusion to control the Bayesian FDR of the discovery set.
- ii) If we try to control the Bayesian FDR of the subfamilies, in situation, we may get smaller final discovery set. This is discussed in the following example.

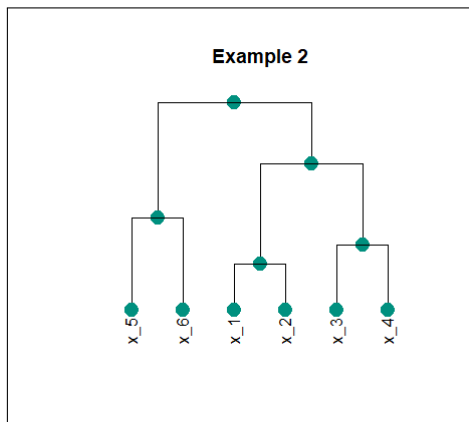
Existing method

Let us consider a model with 6 SNPs and following is the hierarchical clustering structure between the columns of X :



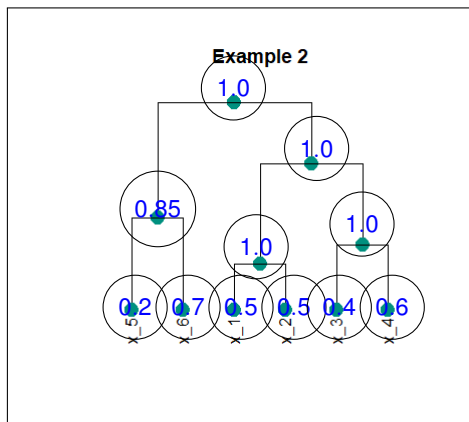
Existing method - comparison via example

We fit the **BGLR** and get the following $u_i = \mathbb{P}(H_{1i}|\text{Data})$ in the hierarchy:



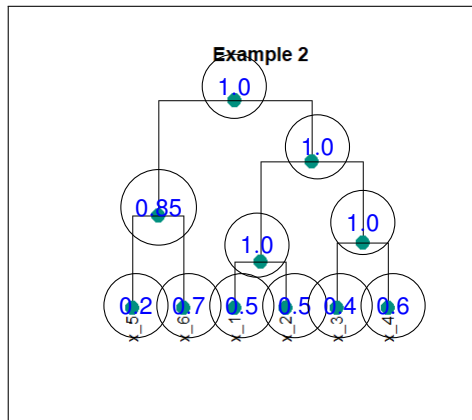
Existing method - comparison via example

We fit the **BGLR** and get the following $u_i = \mathbb{P}(H_{1i} | \text{Data})$ in the hierarchy:



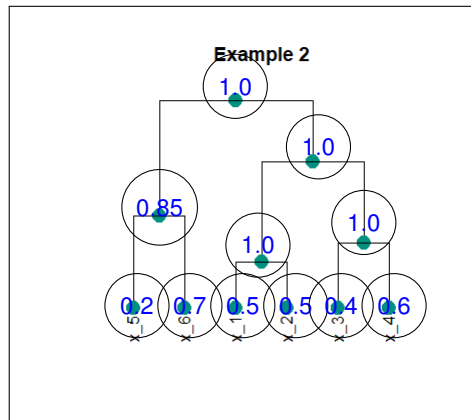
Existing method - comparison via example

When $\alpha = 0.05$, our method obtains the following:



Existing method - comparison via example

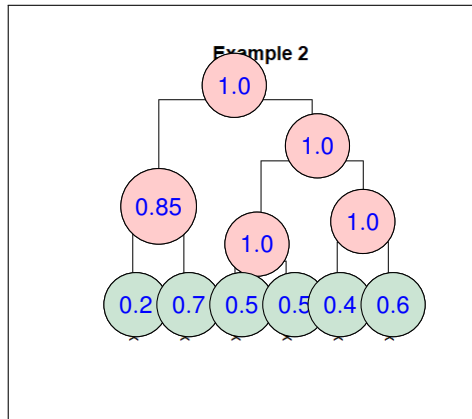
When $\alpha = 0.05$, our method obtains the following:



$$t_{min} = 0.85$$

Existing method - comparison via example

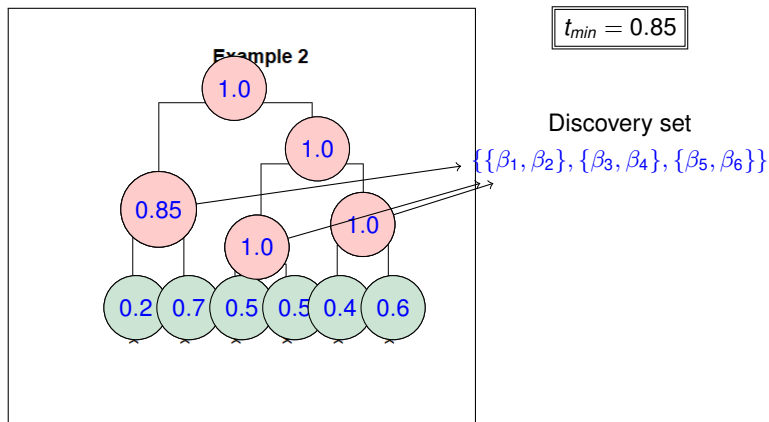
When $\alpha = 0.05$, our method obtains the following:



$$t_{min} = 0.85$$

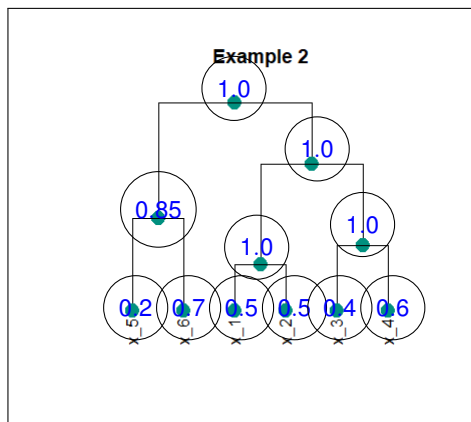
Existing method - comparison via example

When $\alpha = 0.05$, our method obtains the following:



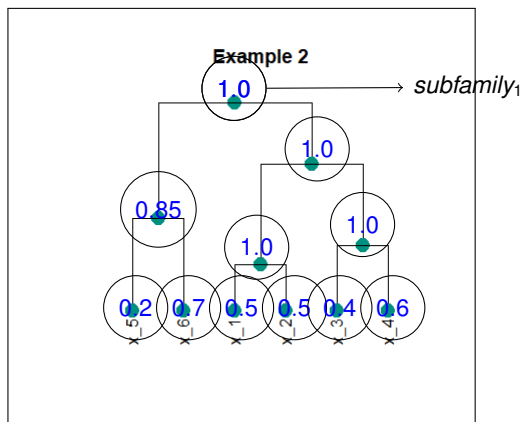
Existing method - comparison via example

Now if we control **BFDR** subfamily-wise then we get:



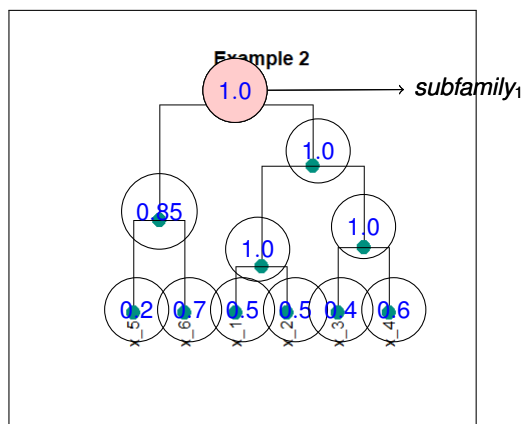
Existing method - comparison via example

Now if we control **BFDR** subfamily-wise then we get:



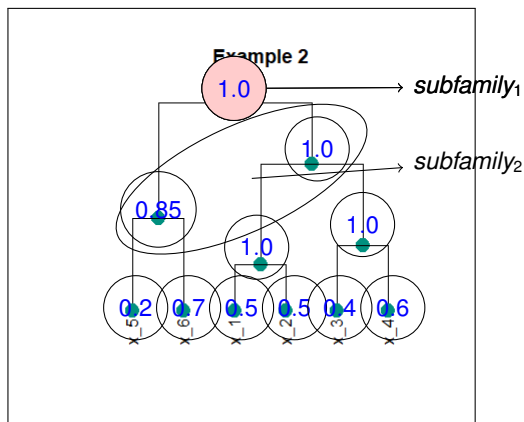
Existing method - comparison via example

Now if we control **BFDR** subfamily-wise then we get:



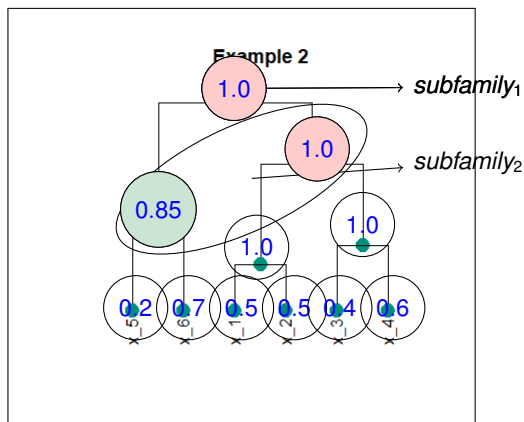
Existing method - comparison via example

Now if we control **BFDR** subfamily-wise then we get:



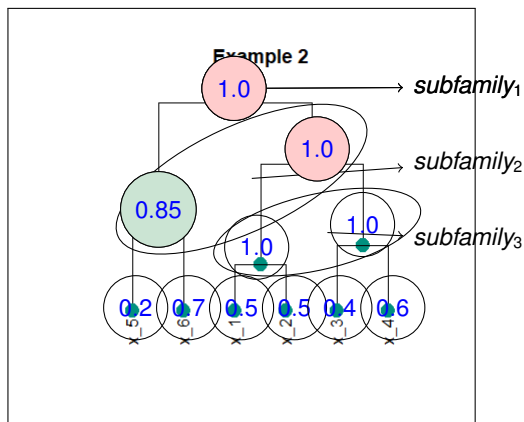
Existing method - comparison via example

Now if we control **BFDR** subfamily-wise then we get:



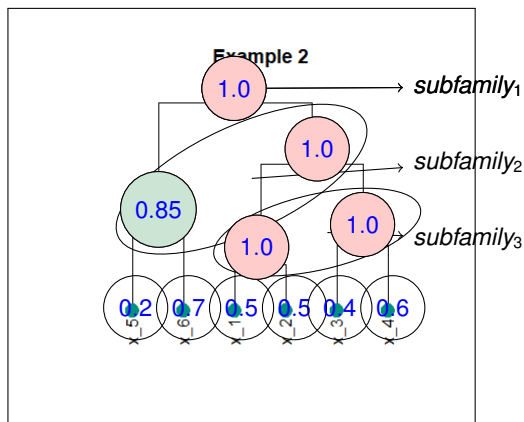
Existing method - comparison via example

Now if we control **BFDR** subfamily-wise then we get:



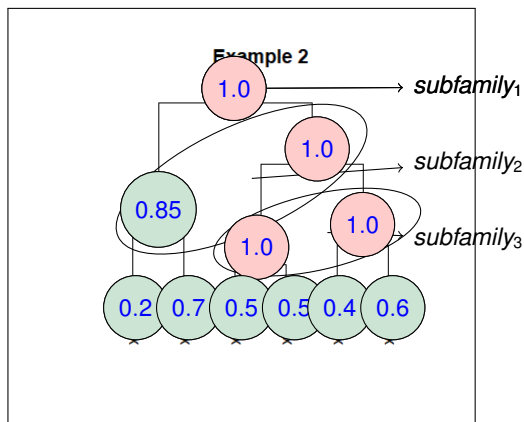
Existing method - comparison via example

Now if we control **BFDR** subfamily-wise then we get:



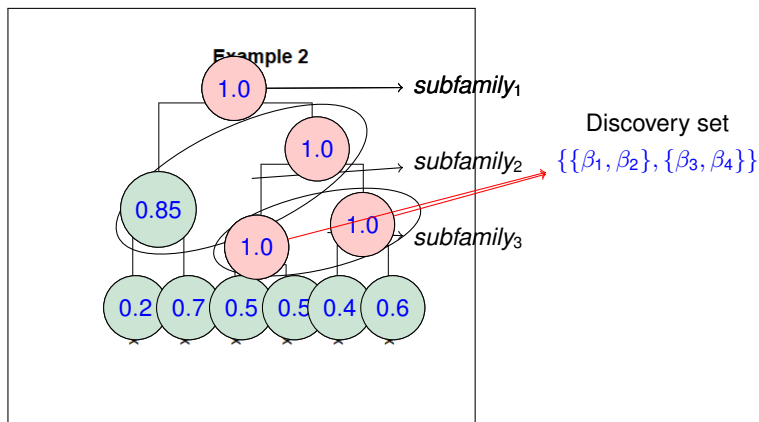
Existing method - comparison via example

Now if we control **BFDR** subfamily-wise then we get:



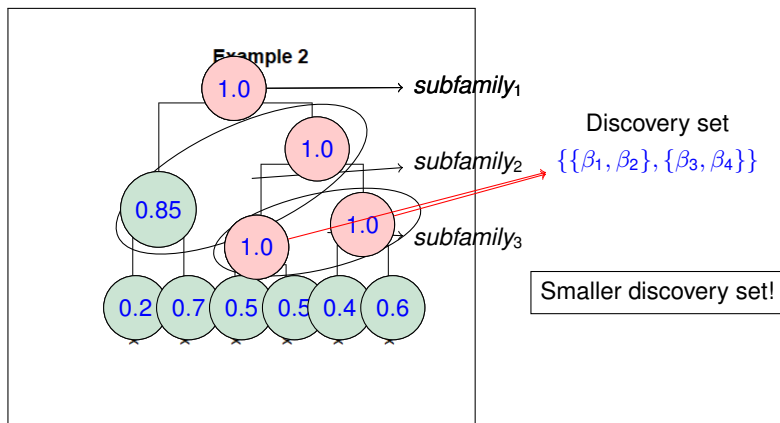
Existing method - comparison via example

Now if we control **BFDR** subfamily-wise then we get:



Existing method - comparison via example

Now if we control **BFDR** subfamily-wise then we get:



Simulated data

n represents the sample size in all the simulation results.

Simulated data

n represents the sample size in all the simulation results.

We generate 100 columns of X with LD and they represent the genotypical information in our simulation.

Simulated data

n represents the sample size in all the simulation results.

We generate 100 columns of X with LD and they represent the genotypical information in our simulation.

We generate Y by adding non-zero weights to 4 markers positioned at (30,40,60,70) and adding random noise $E \sim N(0, \sigma_e^2 I)$.

Simulated data

n represents the sample size in all the simulation results.

We generate 100 columns of X with LD and they represent the genotypical information in our simulation.

We generate Y by adding non-zero weights to 4 markers positioned at (30,40,60,70) and adding random noise $E \sim N(0, \sigma_e^2 I)$.

Together the 4 markers explains 1% of the variance in Y .

Simulated data

n represents the sample size in all the simulation results.

We generate 100 columns of X with LD and they represent the genotypical information in our simulation.

We generate Y by adding non-zero weights to 4 markers positioned at (30,40,60,70) and adding random noise $E \sim N(0, \sigma_e^2 I)$.

Together the 4 markers explains 1% of the variance in Y .

As a measure of LD around the causal markers, we use R-sq = average of r^2 values between a causal marker and it's immediate neighbors left and right. r = Pearson correlation coefficient.

Simulation results

We run BGLR model with the generated data. All the following results are obtained from running a single chain Gibbs sampler with 2000 burn-in samples and 8000 samples post burn-in with the thinning of 5 samples. So effectively we have $\frac{8000}{5} = 1400$ samples from the posterior.

Simulation results: notations

In the simulation results following notations have been used:

Simulation results: notations

In the simulation results following notations have been used:

FDR Level or α = The level at which we control the **BFDR**.

Simulation results: notations

In the simulation results following notations have been used:

FDR Level or α = The level at which we control the **BFDR**.

Actual FDR = The actual value of the FDR of the discovery set.

Simulation results: notations

In the simulation results following notations have been used:

FDR Level or α = The level at which we control the **BFDR**.

Actual FDR = The actual value of the FDR of the discovery set.

Predicted BFDR = The **BFDR** of the discovery set. Since in reality the position of the causal markers are unknown this provides an estimate of the **Actual FDR**.

Simulation results: notations

In the simulation results following notations have been used:

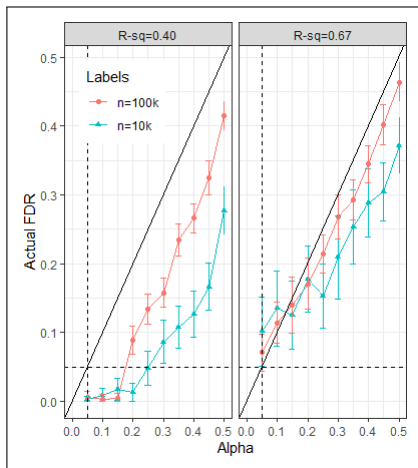
FDR Level or α = The level at which we control the **BFDR**.

Actual FDR = The actual value of the FDR of the discovery set.

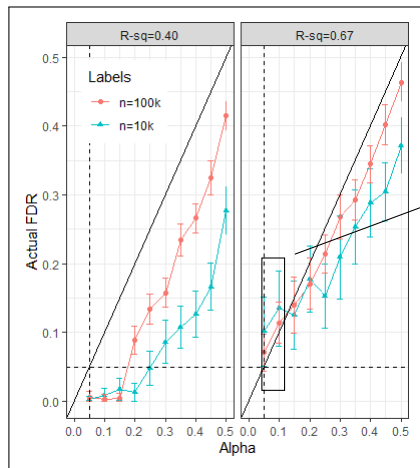
Predicted BFDR = The **BFDR** of the discovery set. Since in reality the position of the causal markers are unknown this provides an estimate of the **Actual FDR**.

Power = The power of the discovery set.

Actual FDR vs FDR Level(α)

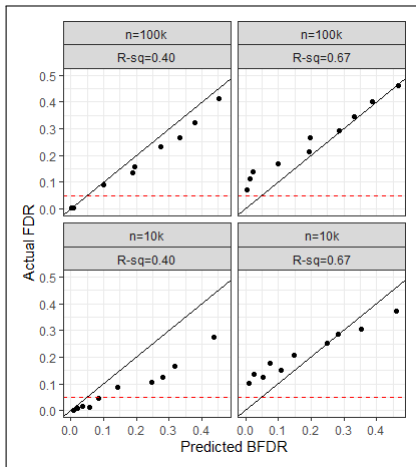


Actual FDR vs FDR Level(α)

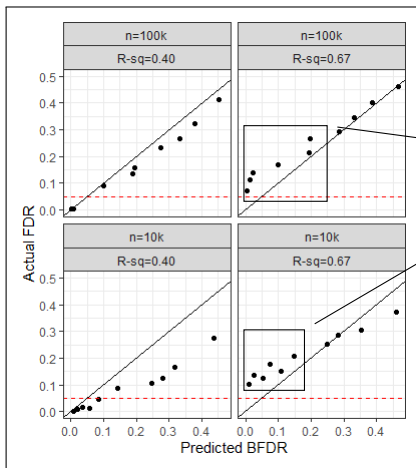


For small α ,
when LD is high
FDR control breaks down

Actual FDR vs Predicted FDR



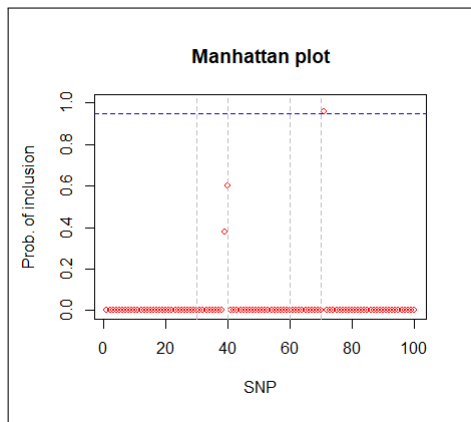
Actual FDR vs Predicted FDR



When LD is high
the **Predicted BFDR**
is underestimating
the **Actual FDR**

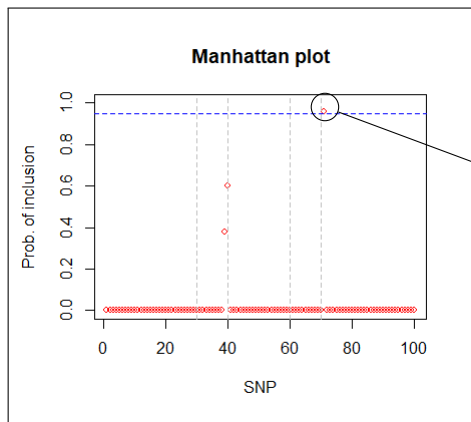
High LD scenario

In case of high LD, we observe that FDR control breaks down. This is largely due to the cases like the following:



High LD scenario

In case of high LD, we observe that FDR control breaks down. This is largely due to the cases like the following:



Error: Marker 71 which is correlated with causal marker 70 gets included in the model with 0.95 inclusion prob. So in our method marker 71 will get included in the discovery set by itself.

Summary - Single chain simulation

To summarize -

Summary - Single chain simulation

To summarize -

- For high LD the FDR control breaks down

Summary - Single chain simulation

To summarize -

- For high LD the FDR control breaks down
- Erroneous cases are those where the model is always including a marker which is not causal but highly correlated with a nearby causal marker. So this "wrong" marker has very high inclusion probability.

Summary - Single chain simulation

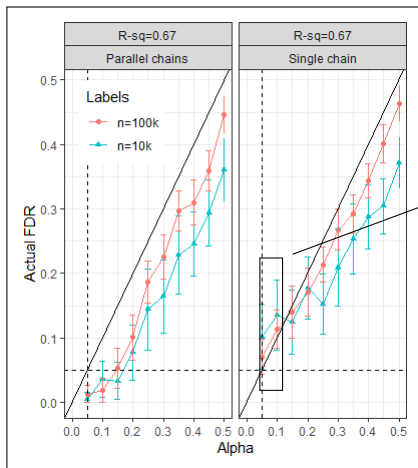
To summarize -

- For high LD the FDR control breaks down
- Erroneous cases are those where the model is always including a marker which is not causal but highly correlated with a nearby causal marker. So this "wrong" marker has very high inclusion probability.

So if we run multiple chains of Gibbs sampler then we should get a good mixing between the inclusion of causal marker and it's highly correlated neighbor.

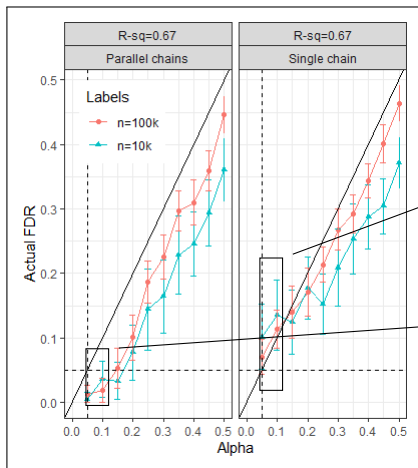
We fit the **BGLR** model to the previously generated data. Now we run **4** parallel chains each with **2000** burn-in and **8000** post burn-in samples with a thinning of **5** samples.

Actual FDR vs FDR level(α)



For small α ,
when LD is high
FDR control breaks down

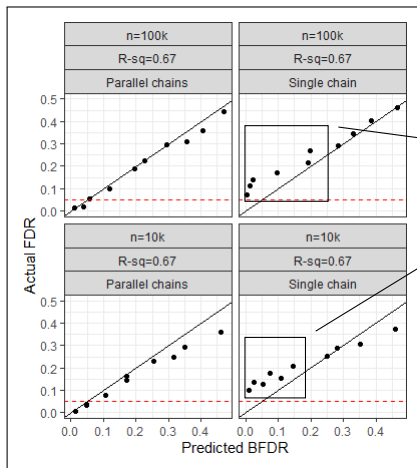
Actual FDR vs FDR level(α)



For small α ,
when LD is high
FDR control breaks down

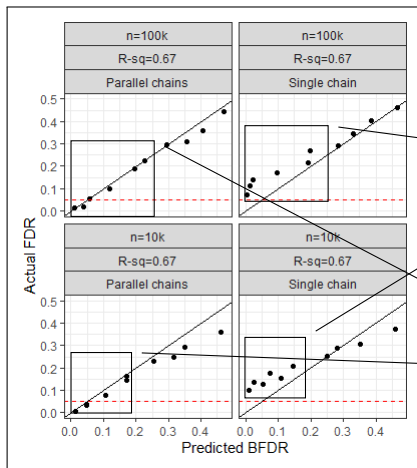
The actual FDR
is still controlled

Actual FDR vs Predicted FDR

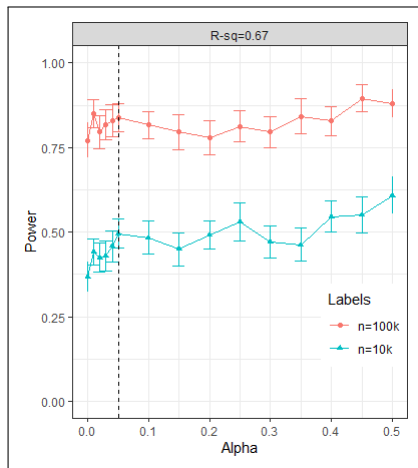


When LD is high
the **Predicted BFDR**
is underestimating
the **Actual FDR**

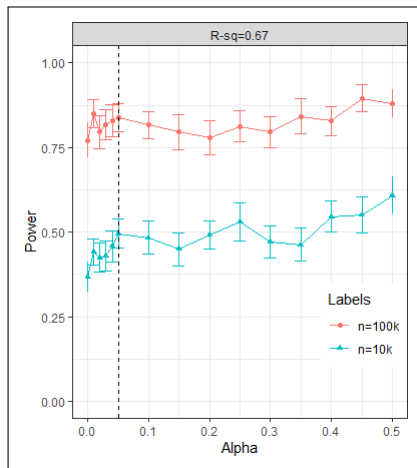
Actual FDR vs Predicted FDR



Power vs FDR level(α)



Power vs FDR level(α)



Observation:
Power at 0% FDR
is still high.



Benjamini, Y., and Hochberg, Y. (1995)
Controlling the false discovery rate: A practical and powerful
approach to multiple testing
J. Roy. Statist. Soc. Ser. B 57 289–300.



Yekutieli, D. (2008)
Hierarchical false discovery rate–controlling methodology
J. Amer. Statist. Assoc.



Genovese, C., and Wasserman, L. (2003)
Bayesian and Frequentist Multiple Testing
Bayesian Statistics 7.

Questions?

