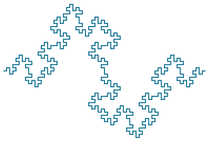
High-Throughput Sequencing Course

Multiple Testing

Biostatistics and Bioinformatics



Summer 2018

Duke University School of Medicine

DUKE BIOSTATISTICS & BIOINFORMATICS

Introduction

- You have previously considered the significance of a single gene

Introduction

- You have previously considered the significance of a single gene
- The analysis of high-dimensional data is concerned with assessing the significance of multiple loci/genes
  - Microarray : 20,000-50,000 probe sets
  - GWAS: 500,000-5,000,000 typed SNPs
  - RNA-Seq: 22,000 genes (humans)

Introduction

- You have previously considered the significance of a single gene
- The analysis of high-dimensional data is concerned with assessing the significance of multiple loci/genes
  - Microarray : 20,000-50,000 probe sets
  - GWAS: 500,000-5,000,000 typed SNPs
  - RNA-Seq: 22,000 genes (humans)
- This leads to the *Multiple Testing* problem

Introduction

- You have previously considered the significance of a single gene
- The analysis of high-dimensional data is concerned with assessing the significance of multiple loci/genes
  - Microarray : 20,000-50,000 probe sets
  - GWAS: 500,000-5,000,000 typed SNPs
  - RNA-Seq: 22,000 genes (humans)
- This leads to the *Multiple Testing* problem
- Before we address this problem, let's quickly review the single hypothesis case

Introduction: Type I and II Errors

Recall that in hypothesis testing with a single hypothesis (gene), errors can be classified as:

- Type I error - rejecting H_0 when it is true
- Type II error - not rejecting H_0 when it is false

Introduction: Type I and II Errors

Recall that in hypothesis testing with a single hypothesis (gene), errors can be classified as:

- Type I error - rejecting H_0 when it is true
- Type II error - not rejecting H_0 when it is false

We can define probabilities associated with each of these errors:

$$\begin{aligned} \alpha &= Pr(\text{reject } H_0 | H_0 \text{ true}) \\ \beta &= Pr(\text{do not reject } H_0 | H_0 \text{ false}) \end{aligned}$$

Introduction: Type I and II Errors

Decision	H_0 True	H_0 False
Do not reject H_0	$1 - \alpha$	β
Reject H_0	α	$1 - \beta$

Introduction: Type I and II Errors

- Always a compromise between type I and type II error

Introduction: Type I and II Errors

- Always a compromise between type I and type II error
- An α-level test has probability α of rejecting H_0 when H_0 is true

Introduction: Type I and II Errors

- Always a compromise between type I and type II error
- An α-level test has probability α of rejecting H_0 when H_0 is true
- Usually try to use the most powerful (smallest β) test for a given α-level

Introduction: Type I and II Errors

- Always a compromise between type I and type II error
- An α-level test has probability α of rejecting H_0 when H_0 is true
- Usually try to use the most powerful (smallest β) test for a given α-level → *control* type I error

Introduction: Multiple Tests

- Single-hypothesis type-I-error-control proves inadequate when experiment is characterized by a collection of hypotheses

Introduction: Multiple Tests

- Single-hypothesis type-I-error-control proves inadequate when experiment is characterized by a collection of hypotheses
  ← Genomics

Introduction: Multiple Tests

- Single-hypothesis type-I-error-control proves inadequate when experiment is characterized by a collection of hypotheses
  ← Genomics
- Why?

Introduction: Multiple Tests

- Assume we have m hypotheses we wish to test as part of a given experiment

Introduction: Multiple Tests

- Assume we have m hypotheses we wish to test as part of a given experiment
  - These hypotheses could correspond to m genes that we are investigating for differential expression between two groups

Introduction: Multiple Tests

- Assume we have m hypotheses we wish to test as part of a given experiment
  - These hypotheses could correspond to m genes that we are investigating for differential expression between two groups
- Assume that each hypothesis is tested using a α-level test

Introduction: Multiple Tests

- Assume we have m hypotheses we wish to test as part of a given experiment
  - These hypotheses could correspond to m genes that we are investigating for differential expression between two groups
- Assume that each hypothesis is tested using a α-level test
- Assume that the tests are *INDEPENDENT* and that the null is true for each of the hypothesis

Introduction: Multiple Tests

What is the probability of rejecting a null hypothesis even though all hypotheses are actually null?

Introduction: Multiple Tests

What is the probability of rejecting a null hypothesis even though all hypotheses are actually null?

- For a given test, we have an α chance of rejecting its null

Introduction: Multiple Tests

What is the probability of rejecting a null hypothesis even though all hypotheses are actually null?

- For a given test, we have an α chance of rejecting its null
- Therefore we have a $1 - \alpha$ chance of not rejecting

Introduction: Multiple Tests

What is the probability of rejecting a null hypothesis even though all hypotheses are actually null?

- For a given test, we have an α chance of rejecting its null
- Therefore we have a $1 - \alpha$ chance of not rejecting

- We have m independent α-level tests, so the chance of not rejecting any of them is:

$$(1-\alpha)^m$$

Introduction: Multiple Tests

What is the probability of rejecting a null hypothesis even though all hypotheses are actually null?

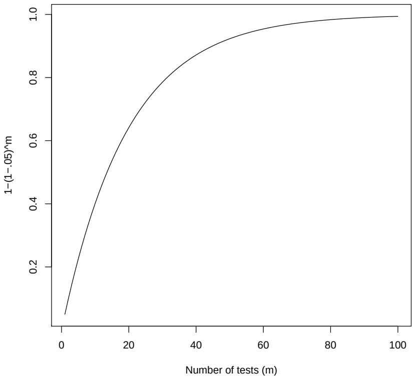
- For a given test, we have an α chance of rejecting its null
- Therefore we have a $1 - \alpha$ chance of not rejecting
- We have m independent α-level tests, so the chance of not rejecting any of them is:

$$(1-\alpha)^m$$

- Therefore, the chance of rejecting at least one hypothesis is:

$$1-(1-\alpha)^m$$

Introduction: Multiple Tests



Introduction: Notation

- Gene j (among the m genes) is either associated with the outcome or not
- The truth is unknown to us
- The null hypothesis for gene j is denoted by H_j
- H_j: gene j is not associated with the outcome of interest
- The alternative hypothesis is denoted by $\bar{H}_j$
- $\bar{H}_j$: gene j is associated with the outcome of interest
- H_j and $\bar{H}_j$ are called *marginal* or *local* hypotheses

Introduction: Marginal and Global Hypotheses

- H_j and $\bar{H}_j$ are called *marginal* or *local* hypotheses
- A global null hypothesis: None of the m genes is associated with the outcome
- A global alternative: At least one of the m genes is is associated with the outcome
- Using notation
  - Global Null: $\mathbb{H}_0 : H_1$ and H_2 and $\ldots H_m$
  - Global Alternative: $\mathbb{H}_1 : \bar{H}_1$ or $\bar{H}_2$ or $\ldots \bar{H}_m$

Introduction: Unadjusted vs Adjusted P-values

- Suppose that we only test a single gene, say gene j, among the m genes
- Let p_j (lower case p) denote P-value corresponding to H_j
- p_j is called the *marginal* or *unadjusted* P-value
- If m hypotheses are tested, inference on H_j on the basis of p_j is inappropriate
- The P-value for H_j has to account for testing the other $m - 1$ hypotheses
- We will denote the *adjusted* P-value by P_j (upper case P)
- When testing multiple genes, using the marginal P-value is inappropriate
- Why?

Introduction: More Notation

- Suppose that gene j is not associated with the outcome of interest (H_j is true)
- Then
  - Decision rule rejects → False-Positive (FP)
  - Decision rule fails to reject → True-Negative (TN)
- Suppose that gene j is associated with the outcome of interest (H_j is false)
  - Decision rule rejects → True-Positive (TP)
  - Decision rule fails to reject → False-Negative (FN)

Introduction: Summarizing a Multiple Testing Procedure

- The results from any multiple testing procedure can be summarized as the following table

	Accept	Reject	Total
Truth Null	A_0	R_0	m_0
Alt.	A_1	R_1	m_1
	A	R	m

- Notation:
  - m: Number of tests, m_0, m_1 number of null/true genes
  - R: Number of genes rejected according to the decision rule
  - A: Number of genes accepted according to the decision rule
  - R_0/R_1 number of TN/FP
  - A_0/A_1 number of FN/TP

Introduction: Example

- Results from an analysis based on $m = 10$ genes:

```
##        gene truth  pvalue
## 1   gene1     0 0.29070
## 2   gene2     1 0.61630
## 3   gene3     1 0.00320
## 4   gene4     0 0.01641
## 5   gene5     0 0.25150
## 6   gene6     0 0.58450
## 7   gene7     0 0.22890
## 8   gene8     1 0.12630
## 9   gene9     0 0.26080
## 10 gene10     0 0.04980
```

- Investigator decides to use following decision rule: Any gene with a corresponding unadjusted P-value of less than 0.05 will be rejected.
- Reject H_j if $p_j < 0.05$ or accept H_j otherwise

Exercise: Fill in the 2x2 table

	Accept	Reject	Total
Truth Null	$A_0 =?$	$R_0 =?$	$m_0 =?$
Alt.	$A_1 =?$	$R_1 =?$	$m_1 =?$
	$A =?$	$R =?$	$m =?$

Example: Fill in the 2x2 table

	Accept	Reject	Total
Truth Null	$A_0 = 5$	$R_0 = 2$	$m_0 = 7$
Alt.	$A_1 = 2$	$R_1 = 1$	$m_1 = 3$
	$A = 7$	$R = 3$	$m = 10$

- $m_0 = 7$ and $m_1 = 3$
- $R = 3$ will be rejected based on the decision rule
- Consequently $A = m - R = 7$ will be accepted
- $R_0 = 2, R_1 = 1, A_0 = 5$ and $A_1 = 2$

The Truth

- What know or observe is this

```
##      gene  pvalue
## 1   gene1 0.29070
## 2   gene2 0.61630
## 3   gene3 0.00320
## 4   gene4 0.01641
## 5   gene5 0.25150
## 6   gene6 0.58450
## 7   gene7 0.22890
## 8   gene8 0.12630
## 9   gene9 0.26080
## 10 gene10 0.04980
```

- and not (truth colum is not known to us):

```
dat

##      gene truth  pvalue
## 1   gene1     0 0.29070
## 2   gene2     1 0.61630
## 3   gene3     1 0.00320
## 4   gene4     0 0.01641
## 5   gene5     0 0.25150
## 6   gene6     0 0.58450
## 7   gene7     0 0.22890
## 8   gene8     1 0.12630
## 9   gene9     0 0.26080
## 10 gene10     0 0.04980
```

Example: Fill in the 2x2 table (based on what we observe)

- We can only fill in the bottom row of the table

	Accept	Reject	Total
Truth Null	A_0	R_0	m_0
Alt.	A_1	R_1	m_1
	$A = 7$	$R = 3$	$m = 10$

- The remaining quantities are fixed unknown quantities or unobservable random variables.

Comments

	Accept	Reject	Total
Truth Null	A_0	R_0	m_0
Alt.	A_1	R_1	m_1
	A	R	m

Comments

	Accept	Reject	Total
Truth Null	A_0	R_0	m_0
Alt.	A_1	R_1	m_1
	A	R	m

- m is a known constant

Comments

	Accept	Reject	Total
Truth Null	A_0	R_0	m_0
Alt.	A_1	R_1	m_1
	A	R	m

- m is a known constant
- m_0 and m_1 are unknown constants

Comments

	Accept	Reject	Total
Truth Null	A_0	R_0	m_0
Alt.	A_1	R_1	m_1
	A	R	m

- m is a known constant
- m_0 and m_1 are unknown constants
- R and A are determined on the basis of applying the decision rule to the data

Comments

	Accept	Reject	Total
Truth Null	A_0	R_0	m_0
Alt.	A_1	R_1	m_1
	A	R	m

- m is a known constant
- m_0 and m_1 are unknown constants
- R and A are determined on the basis of applying the decision rule to the data
- They are *observable* random quantities

Comments

	Accept	Reject	Total
Truth Null	A_0	R_0	m_0
Alt.	A_1	R_1	m_1
	A	R	m

- m is a known constant
- m_0 and m_1 are unknown constants
- R and A are determined on the basis of applying the decision rule to the data
- They are *observable* random quantities
- The true states of the genes of the genes are unknown

Comments

	Accept	Reject	Total
Truth Null	A_0	R_0	m_0
Alt.	A_1	R_1	m_1
	A	R	m

- m is a known constant
- m_0 and m_1 are unknown constants
- R and A are determined on the basis of applying the decision rule to the data
- They are *observable* random quantities
- The true states of the genes of the genes are unknown
- A_0, A_1, R_0 and R_1 are *unobservable* random quantities

Introduction: Multiple Testing Problem

- Control error rate(s) in multiple testing context

Introduction: Multiple Testing Problem

- Control error rate(s) in multiple testing context
- Multiple testing methods are designed to control a particular error rate

Introduction: Multiple Testing Problem

- Control error rate(s) in multiple testing context
- Multiple testing methods are designed to control a particular error rate
- Multiple error rates exist

Introduction: Multiple Testing Problem

- Control error rate(s) in multiple testing context
- Multiple testing methods are designed to control a particular error rate
- Multiple error rates exist → need to chose error rate to control and then method to control it

Introduction: Error Rates

- **Family-wise error rate** (FWER): the probability of at least one type I error

Introduction: Error Rates

- **Family-wise error rate** (FWER): the probability of at least one type I error
- **False discovery rate** (FDR): the expected proportion of type I errors among the rejected hypotheses.

Family-wise Error Rate (FWER)

- Probability of committing at least one false-rejection (among m) given that *all* genes are null

Family-wise Error Rate (FWER)

- Probability of committing at least one false-rejection (among m) given that *all* genes are null
- FWER $= P(R \geq 1|m = m0)$

Family-wise Error Rate (FWER)

- Probability of committing at least one false-rejection (among m) given that *all* genes are null
- FWER $= P(R \geq 1|m = m0)$
- Note that when $m = 1$ (single gene), this definition is identical to the type I error we have previously considered

Controlling FWER: Sidak's method

Recall that we showed that with m independent α-level tests:

$$\text{FWER} = 1 - (1 - \alpha)^m$$

Controlling FWER: Sidak's method

Recall that we showed that with m independent α-level tests:

$$\text{FWER} = 1 - (1 - \alpha)^m$$

Therefore,

$$\begin{aligned} 1 - \text{FWER} &= (1 - \alpha)^m \\ (1 - \text{FWER})^{1/m} &= 1 - \alpha \\ 1 - (1 - \text{FWER})^{1/m} &= \alpha \end{aligned}$$

Controlling FWER: Sidak's method

Recall that we showed that with m independent α-level tests:

$$\text{FWER} = 1 - (1 - \alpha)^m$$

Therefore,

$$\begin{aligned} 1 - \text{FWER} &= (1 - \alpha)^m \\ (1 - \text{FWER})^{1/m} &= 1 - \alpha \\ 1 - (1 - \text{FWER})^{1/m} &= \alpha \end{aligned}$$

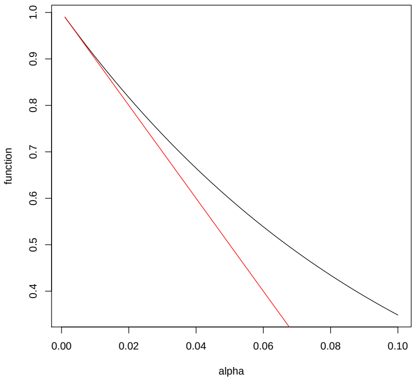
This suggests that we can control FWER by choosing α for each individual test to be $1 - (1 - \text{FWER})^{1/m}$

Controlling FWER: Bonferroni

Note that $(1-x)^m \approx 1 - mx$ (for x close to zero)

Controlling FWER: Bonferroni

Note that $(1-x)^m \approx 1 - mx$ (for x close to zero)



Controlling FWER: Bonferroni

Note that $(1-x)^m \approx 1 - mx$ (for x close to zero)

Therefore

$$\begin{aligned} 1 - \text{FWER} &= (1-\alpha)^m \\ 1 - \text{FWER} &\approx 1 - m\alpha \\ \text{FWER} &\approx m\alpha \end{aligned}$$

Controlling FWER: Bonferroni

Note that $(1-x)^m \approx 1 - mx$ (for x close to zero)

Therefore

$$\begin{aligned} 1 - \text{FWER} &= (1-\alpha)^m \\ 1 - \text{FWER} &\approx 1 - m\alpha \\ \text{FWER} &\approx m\alpha \end{aligned}$$

This suggests that we can control FWER by choosing α for each individual test to be $\frac{\text{FWER}}{m}$

Controlling FWER: Bonferroni

- The Bonferroni adjusted *P-value* is defined as

$$P_j = m \times p_j$$

- Technical note: P_j is defined above could be larger than 1 so a more technically rigorous definition is

$$P_j = \min\{m \times p_j, 1\}$$

- In other words, if $m \times p_j$ is larger than 1, then truncate P_j at 1.

Controlling FWER: Holm's

- Sidak and Bonferroni are one-step approaches

Controlling FWER: Holm's

- Sidak and Bonferroni are one-step approaches
- Some power can be gained by sequential methods

Controlling FWER: Holm's

- Sidak and Bonferroni are one-step approaches
- Some power can be gained by sequential methods

- The simplest sequential method is Holm's method:

Controlling FWER: Holm's

- Sidak and Bonferroni are one-step approaches
- Some power can be gained by sequential methods
- The simplest sequential method is Holm's method:
  - Order the unadjusted P-values such that $p_1 \leq p_2 \leq ... \leq p_m$

Controlling FWER: Holm's

- Sidak and Bonferroni are one-step approaches
- Some power can be gained by sequential methods
- The simplest sequential method is Holm's method:
  - Order the unadjusted P-values such that $p_1 \leq p_2 \leq ... \leq p_m$
  - To control FWER, the step-down Holm adjusted P-values are

$$P_j = \min\{(m - j + 1) \times p_j, 1\}$$

Controlling FWER: Holm's

- Sidak and Bonferroni are one-step approaches
- Some power can be gained by sequential methods
- The simplest sequential method is Holm's method:
  - Order the unadjusted P-values such that $p_1 \leq p_2 \leq ... \leq p_m$
  - To control FWER, the step-down Holm adjusted P-values are

$$P_j = \min\{(m - j + 1) \times p_j, 1\}$$

  - Note that every unadjusted P-value is not multiplied by same factor

Controlling FWER: Permutation

- These methods work well when tests are independent

Controlling FWER: Permutation

- These methods work well when tests are independent (not generally the case in genomics experiments)

Controlling FWER: Permutation

- These methods work well when tests are independent (not generally the case in genomics experiments)
- When tests are correlated, these methods are conservative

Controlling FWER: Permutation

- These methods work well when tests are independent (not generally the case in genomics experiments)
- When tests are correlated, these methods are conservative
- Permutation approaches are useful in this context

Controlling FWER: Permutation

Assume we are interested in assessing differential expression between 2 groups :

1. Compute minimum unadjusted P-value for all genes from the observed data (call it p_1)
2. Randomly permute the group labels
   - *Breaks* relationship between group and expression
   - Reflects sample from global null hypothesis
3. Compute minimum P-value from data set generated in 2 (call it p_1^1)
4. Repeat 2 and 3 B times to get $p_1^1, p_1^2, ..., p_1^B$
5. Compute the proportion of $p_1^1, p_1^2, ..., p_1^B$ that are $\leq p_1$
6. This proportion is the permutation adjusted P_1

Controlling FWER: Permutation

- Can get $P_2, ..., P_m$ in a similar way (a Holm's-like permutation proceedure)

Controlling FWER: Permutation

- Can get $P_2, ..., P_m$ in a similar way (a Holm's-like permutation proceedure)

- Correlation among the genes is accounted for

False Discovery Rate (FDR)

- Consider the quantity $\frac{R_0}{R}$
- This is the proportion of of false discoveries among the genes rejected
- This is an *unobservable* random quantity (R_0 is not observable)
- In the FDR framework is based on controlling the *expected* value of this ratio
- FDR $\equiv E[\frac{R_0}{R}]$
  - Expectation is set to zero if $R = 0$, therefore FDR $= E[\frac{R_0}{R}|R > 0]Pr(R > 0)$
- Note that when $m_0 = m$ (none of the genes are true), FWER=FDR

Controlling FDR: Benjamini and Hochberg

1. Order the unadjusted P-values $p_1 \leq p_2 \leq ... \leq p_m$

Controlling FDR: Benjamini and Hochberg

1. Order the unadjusted P-values $p_1 \leq p_2 \leq ... \leq p_m$

2. Find the largest j such that $p_j \leq j\alpha/m$

Controlling FDR: Benjamini and Hochberg

1. Order the unadjusted P-values $p_1 \leq p_2 \leq ... \leq p_m$
2. Find the largest j such that $p_j \leq j\alpha/m$
3. Reject null hypotheses affiliated with $p_1, p_2, ..., p_j$

Controlling FDR: Benjamini and Hochberg

1. Order the unadjusted P-values $p_1 \leq p_2 \leq ... \leq p_m$
2. Find the largest j such that $p_j \leq j\alpha/m$
3. Reject null hypotheses affiliated with $p_1, p_2, ..., p_j$

This proceedure will control FDR at $\alpha m_0/m$ when the tests are independent and continuous

Controlling FDR: Benjamini and Hochberg

1. Order the unadjusted P-values $p_1 \leq p_2 \leq ... \leq p_m$
2. Find the largest j such that $p_j \leq j\alpha/m$
3. Reject null hypotheses affiliated with $p_1, p_2, ..., p_j$

This proceedure will control FDR at $\alpha m_0/m$ when the tests are independent and continuous

Benjamini and Yekutieli (2001) proposed a modification of the BH procedure that always controls FDR (no larger than $\alpha m_0/m$)

Controlling FDR: Benjamini and Hochberg

1. Order the unadjusted P-values $p_1 \leq p_2 \leq ... \leq p_m$
2. Find the largest j such that $p_j \leq j\alpha/m$
3. Reject null hypotheses affiliated with $p_1, p_2, ..., p_j$

This proceedure will control FDR at $\alpha m_0/m$ when the tests are independent and continuous

Benjamini and Yekutieli (2001) proposed a modification of the BH procedure that always controls FDR (no larger than $\alpha m_0/m$)
← can be quite conservative

Controlling FDR: Benjamini and Hochberg

1. Order the unadjusted P-values $p_1 \leq p_2 \leq ... \leq p_m$
2. Find the largest j such that $p_j \leq j\alpha/m$
3. Reject null hypotheses affiliated with $p_1, p_2, ..., p_j$

This proceedure will control FDR at $\alpha m_0/m$ when the tests are independent and continuous

Benjamini and Yekutieli (2001) proposed a modification of the BH procedure that always controls FDR (no larger than $\alpha m_0/m$)
← can be quite conservative

Note that when $m_0 = m$ (i.e., all hypotheses are null), these procedures maintain FWER at α

Controlling pFDR: q-values

$$\text{FDR} = E[\tfrac{R_0}{R}|R > 0]Pr(R > 0)$$

Controlling pFDR: q-values

$$\text{FDR} = E[\tfrac{R_0}{R}|R > 0]Pr(R > 0)$$

$$\text{pFDR} = E[\tfrac{R_0}{R}|R > 0] \leftarrow \textit{positive}\ \text{FDR}$$

Controlling pFDR: q-values

$$\text{FDR} = E[\tfrac{R_0}{R}|R > 0]Pr(R > 0)$$

$$\text{pFDR} = E[\tfrac{R_0}{R}|R > 0] \leftarrow \textit{positive}\ \text{FDR}$$

- Since $Pr(R > 0)$ is often ~ 1 in most genomics experiments, FDR and pFDR are ver similar

Controlling pFDR: q-values

- q-value is the minimum pFDR for which the affiliated hypothesis is rejected

Controlling pFDR: q-values

- q-value is the minimum pFDR for which the affiliated hypothesis is rejected
- q-value can be interpreted as the expected proportion of false positives incurred when calling that test significant

Genome-wide Significance

- In GWAS papers, $\alpha = 5 \times 10^{-8}$ is typically considered the threshold for genome-wide significance
- It is based on a Bonferroni correction: If you consider testing $m = 1,000,000$ SNPs at the FWER level of 0.05, then each SNP should be tested at the
$$\alpha = \frac{0.05}{1,000,000} = 5 \times 10^{-8},$$
level
- Suppose that the unadjusted P=value for a SNP is 5×10^{-7}
- Is this ”reaching” genome-wide significance?
- The term ”suggestive” is also used

”Reaching” Genome-wide Significance

- Suppose that your $m = 1,000,000$ SNPs are independent
- The adjusted P-value is
$$P = 5 \times 10^{-7} \times m = 5 \times 10^{-7} \times 10^{6} = 0.5,$$
- This is off by an order of magnitude ($0.5 = 0.05 \times 10$)
- It is not ”reaching”
- Note: Due to linkage disequiblirium among SNPs the adjusted P-value is likely to be smaller than 0.5
- The point is that while 5×10^{-7} is small number, it may not be small enough when tesing a large number of hypotheses

Conclusions

- Multiple testing *must* be accounted for when testing for associations in the context of high-dimensional data
- FWER and FDR are the two common frameworks for quantifying error
- Error rate estimates can be used to compute 'adjusted' p-values
- Resampling-based methods can increase power in controlling error when sample sizes are sufficient for their use.
- When large-scale patterns of differential expression are observed, it is important to consider if such effects are biologically reasonable, and if technical factors can be attributed to the variation.