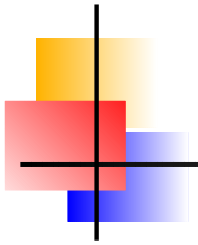


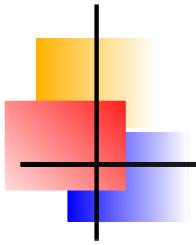
Advanced Statistical Methods for the Analysis of Gene Expression and Proteomics

Lecture 3 – Multiple Comparisons and False Discovery Rate (FDR); Phage analysis

Yuan Ji

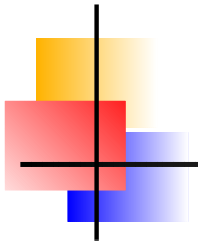
yuanji@mdanderson.org

Department of Bioinformatics and Computational Biology
University of Texas M.D. Anderson Cancer Center



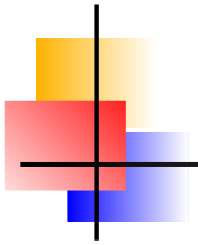
Outline

- Hypothesis test
- Multiple comparisons
- False discovery rate approaches
- Phage analysis – hierarchical modeling and controlling FDR



Hypothesis test

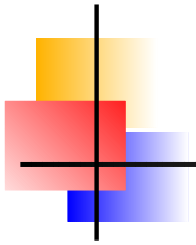
- A hypothesis is a statement or claim about some unknown aspect of the state of nature.
- A test of a hypothesis is a procedure, based on sample information, that culminates in an inferential statement about the hypothesis and possibly, in some situations, in a decision as to what action to take.
- The hypothesis being tested is called the null hypothesis, and the set of other possible claims is called the the alternative hypothesis.
- Typically, one put the desirable claims in the alternative hypothesis.
- Notation: H_0 for null, and H_1 or H_A for alternative.



Examples

- One-sample t-test
- Two-sample t-test
- F-test
- The likelihood ratio test
- χ^2-test
- The Fisher's exact test

All these known tests are based on theoretical proof.



General theory

- A set of data $\boldsymbol{Y}$ is observed.
- A probability model is assumed: $\boldsymbol{Y} \sim f(\boldsymbol{Y}|\boldsymbol{\theta})$.
- H_0 and H_1 are proposed as functions of $\boldsymbol{\theta}$.
- A pivotal statistics $T(\boldsymbol{Y})$ must be developed
  - $T(\boldsymbol{Y})$ is a function of the data $\boldsymbol{Y}$ only;
  - $T(\boldsymbol{Y})$ is "pivotal" – its distribution does not depend on θ.
- Plugging in the observed data values, we can compute an observed value of $T(\boldsymbol{Y}) = t_0$.
- The P-value corresponding to $T = t_0$ is probability (under the distribution of T) at and beyond t_0, in the direction of more extreme values.

Example – t-test

- Data: $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$
- Model: Y_i's are i.i.d. $N(\mu, \sigma^2)$. (so $\boldsymbol{\theta} = (\mu, \sigma^2)$)
- H_0: $\mu = \mu_0$ vs. H_1: $\mu \neq \mu_0$ (e.g., $\mu_0 = 0$)
- Pivotal statistics:

$$T(\boldsymbol{Y}) = \frac{\bar{Y} - \mu}{S/\sqrt{n}}$$

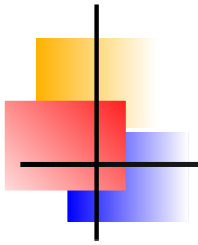
  where S is the sample standard deviation. $T(\boldsymbol{Y})$ follows a t-distribution with $n-1$ degrees of freedom.

- Under the H_0,

$$t_0(\boldsymbol{Y}) = \frac{\bar{Y} - \mu_0}{S/\sqrt{n}}$$

  .

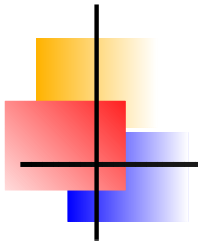
- P-value = $Pr(|T| > t_0)$ computed under the t-distribution with $n-1$ degrees of freedom



Hypothesis test

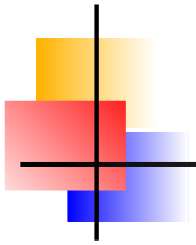
From a decision theoretic point of view, a hypothesis test is a decision rule that assignes one of the two actions, do not accept H_0 and accept H_0, based on the observed data $\boldsymbol{x}$.

- Suppose the observed data is $\boldsymbol{x} = (x_1, \ldots, x_n)$, and $E(x_i) = \mu$.
- $H_0 : \mu = 0$ vs. $H_1 : \mu = 1$.
- A test statistic is a function of the data: $T = t(x)$ (T does not depend on μ).
- A test is is function of T (and therefore of $\boldsymbol{x}$), $\phi(T(\boldsymbol{x}))$, which takes values 0 and 1.
- The level of a test is the probability $Pr(\phi = 1|H_0)$, which is equivalent to type I error rate.
- The power of a test is the probability $Pr(\phi = 1|H_1)$.
- The type II error rate is the probability $Pr(\phi = 0|H_1)$.



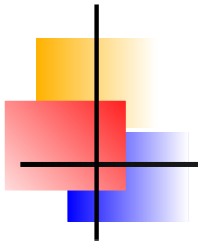
Level of a test

- Usually, we use α to denote the level, which controls the probability of falsely reject the null hypothesis.
- For example, if we reject the null in a comparison of a new drug vs. a standard drug and conclude that the new drug is more effective, we want to be very sure about our conclusion.
- This require our test has a low level α, e.g., 0.05.
- $\alpha = 0.05$ means that the probability of making a false conclusion that the new drug is more effective equals 0.05.
- For a single test, if we reject the null when p-value is less than 0.05, the test level $\alpha = 0.05$.



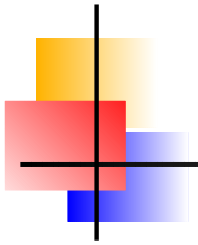
Multiple tests

- Question: if we have two tests, and each test has level α, what is the probability of falsely rejecting at least one null hypothesis?



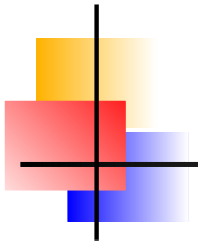
Multiple tests

- Question: if we have two tests, and each test has level α, what is the probability of falsely rejecting at least one null hypothesis?
- The answer is $1-(1-\alpha)^2$.
- When we have m tests, the probability of falsely rejecting at least one null hypothesis is $1-(1-\alpha)^m$.
- This quantity is called the familywise error rate (FWER).



Multiple tests

- Question: if we have two tests, and each test has level α, what is the probability of falsely rejecting at least one null hypothesis?
- The answer is $1-(1-\alpha)^2$.
- When we have m tests, the probability of falsely rejecting at least one null hypothesis is $1-(1-\alpha)^m$.
- This quantity is called the familywise error rate (FWER).
- Procedures that control the error rates of multiple tests are called multiple comparison procedures (MCPs).
- The most famous MCP is the Bonferroni procedure

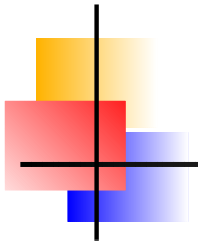


Bonferroni procedure

- Suppose each test has level of α_c.
- With m tests, the FWER is $1-(1-\alpha_c)^m$.
- If we want to control FWER at α, by solving

$$1-(1-\alpha_c)^m = \alpha,$$

  we have $\alpha_c = 1-(1-\alpha)^{1/m}$.
- Apply the Taylor expansion on $(1-\alpha)^{1/m}$ (assuming α is close to zero), we have $\alpha_c \approx \alpha/m$.
- Therefore, to control FWER at α, we reject each null when the p-value is less than α/m.



Multiple comparison in bioinformatics

Multiple comparisons are routinely encountered in Bioinformatics research.

- For each gene, we want to test the null hypothesis that the gene expression level is differentially expressed against the alternative hypothesis that the gene expression level is not.
- If we have 20,000 genes, we have 20,000 tests
- If we apply Bonferroni, we will reject each null when the p-value is smaller than $0.05/200000$ in order to maintain FWER at 0.05 level.
  - Very few null will be rejected
  - We will not have much power

  comparisons.

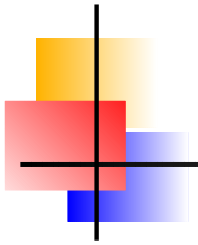
False Discovery Rate

Reference: "Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing" by Benjamini, Y. and Hochberg, Y.

Suppose we have m tests and m_0 null hypotheses are true.

	Not reject	Reject	Total
True null	$\boldsymbol{U}$	$\boldsymbol{V}$	m_0
True alternative	$\boldsymbol{S}$	$\boldsymbol{S}$	$m - m_0$
	$m - \boldsymbol{R}$	$\boldsymbol{R}$	m

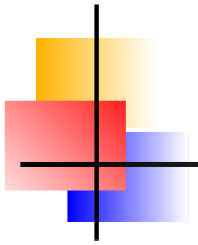
- FWER equals $Pr(\boldsymbol{V} \geq 1)$
- FDR equals $E(\boldsymbol{V}/\boldsymbol{R})$



FDR and FWER

FDR = $E(\boldsymbol{V}/\boldsymbol{R})$, FWER = $Pr(\boldsymbol{V} \geq 1)$.

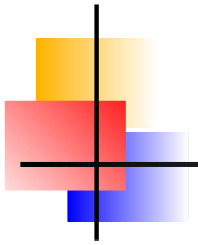
- If $\boldsymbol{R} = 0$, FDR $= 0$ by definition.
- Control of FDR implies control of FWER in the weak sense.
  - If $m_0 = m$, $\boldsymbol{S} = 0$, $\boldsymbol{V} = \boldsymbol{R}$. So $E(\boldsymbol{V}/\boldsymbol{R}) = 0Pr(\boldsymbol{V} = 0) + 1Pr(\boldsymbol{V} \geq 1)$.
- In general, controlling FWER implying controlling FDR
  - If $m_0 < m$ and $\boldsymbol{V} > 0$, then $\boldsymbol{V}/\boldsymbol{R} \leq 1$. Therefore, $1(\boldsymbol{V} \geq 1) = 1 \geq \boldsymbol{V}/\boldsymbol{R} = \boldsymbol{Q}$. Taking expectation of both sides we have $Pr(\boldsymbol{V} \geq 1) \geq E(\boldsymbol{V}/\boldsymbol{R})$.



Controlling FDR

Benjamini and Hochberg (1995) proposed the following procedure that will control the FDR at the level $\frac{m_0}{m}\alpha$

- For each test, obtain the p-value. We get $P_1, P_2, \ldots, P_m$.
- Let $\{P_{(1)}, P_{(2)}, \ldots, P_{(m)}\}$ be the set of ordered p-values. Denote $H_{(i)}$ the null hypothesis corresponding to $P_{(i)}$.
- Specify q^*, the desired FDR value.
- Let k be the largest i for which $P_{(i)} \leq \frac{i}{m}q^*$.
- Reject all $H_{(i)}$ $i = 1, 2, \ldots, k$.

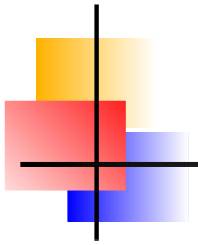


BH approach

The BH approach is a step-down procedure:

- Start from the largest p-value $P_{(m)}$.
- If $P_m > \alpha$, proceed to P_{m-1}; otherwise, all the null hypothese are rejected.
- Given $P_{(m)} > \alpha$, if $P_{(m-1)} > (m-1)\alpha/m$, proceed to $P_{(m_2)}$; otherwise, all the null hypotheses $H_{(1)}, \ldots, H_{(m-1)}$ are rejected.
- Continue on until the first time $P_{(k)} \leq k\alpha/m$ and reject all $H_{(i)}$ $i = 1, \ldots, k$.

The above procedure will control the FDR at α (in fact at $\frac{m_0}{m}\alpha$).



An example

- Suppose we have a set of P-values {.0001, .0004, .0019, .0095, .0201, .0278, .0298, .0344, .0459, .3240, .4262, .5719, .6528, .7590, 1.000}
- Controlling the FWER at 0.05, the Bonferroni approach would use 0.05/15=0.0033, and would reject three hypotheses.
- Controlling the FDR at 0.05, we would start at 1.000 and proceed using BH.
- Turns out $p_{(4)} = .0095 \leq 4/15 \times 0.05 = .013$ is the first time the condition is met. Therefore, the first four null hypotheses are rejected.

Bayesian hypothesis testing

- Let $\gamma = 0$ if H_0 is true and $\gamma = 1$ if H_1 is true.
- Assume $\gamma | p_0 \sim Bern(1 - p_0)$ – prior distribution
- Model:

$$\begin{aligned} \boldsymbol{Y} | \gamma = 0 &\sim f_0 \\ \boldsymbol{Y} | \gamma = 1 &\sim f_1 \end{aligned}$$

- Marginally, $\boldsymbol{Y} \sim f$ follows a mixture model:

$$f(\boldsymbol{y}) = p_0 f_0(\boldsymbol{y}) + (1 - p_0) f_1(\boldsymbol{y})$$

- The Bayes factor

$$B(\boldsymbol{Y}) = \frac{Pr(\gamma = 0 | \boldsymbol{Y}) / Pr(\gamma = 0)}{Pr(\gamma = 1 | \boldsymbol{Y}) / Pr(\gamma = 1)} = \frac{Pr(\gamma = 0 | \boldsymbol{Y}) / p_0}{Pr(\gamma = 1 | \boldsymbol{Y}) / (1 - p_0)}$$

- If $p_0 = 1/2$, then $B(\boldsymbol{Y})$ is decided by $Pr(\gamma = 0 | \boldsymbol{Y})$, the posterior probability of null.

Bayesian multiple hypothesis testing

- Let $\gamma_i = 0$ if H_{0i} is true and $\gamma_i = 1$ if H_{1i} is true.
- Assume $\gamma_i | p_0 \overset{i.i.d}{\sim} Bern(1 - p_0)$ – prior distribution (Note: marginally γ_i's are exchangeable but not independent)
- Model:

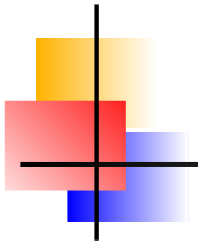
$$\begin{aligned} \boldsymbol{Y}_i | \gamma_i = 0 &\sim f_0 \\ \boldsymbol{Y}_i | \gamma_i = 1 &\sim f_1 \end{aligned}$$

- Marginally, $\boldsymbol{Y}_i \sim f$ follows a mixture model:

$$f(\boldsymbol{y}_i) = p_0 f_0(\boldsymbol{y}_i) + (1 - p_0) f_1(\boldsymbol{y}_i)$$

- The Bayes factor for the ith test

$$B_i(\boldsymbol{Y}_i) = \frac{Pr(\gamma_i = 0 | \boldsymbol{Y}_i)/Pr(\gamma_i = 0)}{Pr(\gamma_i = 1 | \boldsymbol{Y}_i)/Pr(\gamma_i = 1)} = \frac{Pr(\gamma_i = 0 | \boldsymbol{Y}_i)/p_0}{Pr(\gamma_i = 1 | \boldsymbol{Y}_i)/(1 - p_0)}$$



Bayesian multiple hypothesis testing

- If $p_0 = 1/2$, then $B_i(\boldsymbol{Y}_i)$ is decided by $Pr(\gamma_i = 0|\boldsymbol{Y}_i)$, the posterior probability of ith null H_{0i}.
- **Therefore, the important quantity is**
  $\pi_i = Pr(H_{0i} \text{ is ture}|\boldsymbol{Y}_i) = Pr(\gamma_i = 1|\boldsymbol{Y}_1)$.

Bayesian FDR

In Bayesian multiple hypothesis testing, reject the ith test is $\pi_i > \pi^*$. The problem is to specify π^* so that the FDR is controlled at a desirable level.

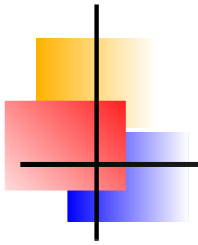
- Genovese and Waserman (02); Newton et al. (04); Bro et et al. (04)
- The posterior expected number of false discoveries

$$FD(\pi^*) = \sum_{i=1}^{m} \pi_i I(\pi_i < \pi^*)$$

(why – Homework 2)

- A Bayesian FDR procedure controls FDR at level α by rejecting H_{0i} if $\pi_i < \pi^*$ where

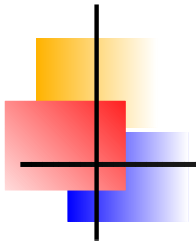
$$\pi^* = \max\{c : \frac{\sum_{i=1}^{m} \pi_i I(\pi_i \leq c)}{\sum_{i=1}^{m} I(\pi_i \leq c)} \leq \alpha\}$$



Bayesian FDR

The previous approach is a step-up procedure.

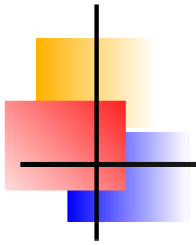
- Sort the marginal posterior probabilities to obtain $(\pi_{(1)}, \pi_{(2)}, \ldots, \pi_{(m)})$.
- Starting from the $\pi_{(1)}$. If $\pi_{(1)}/1 > \alpha$, then do not reject any null hypothesis.
- Otherwise, if $(\pi_{(1)} + \pi_{(2)})/2 > \alpha$, then reject $H_{(1)}$ only.
- Otherwise, if $(\pi_{(1)} + \pi_{(2)} + \pi_{(3)})/3 > \alpha$, then reject $H_{(1)}$ and $H_{(2)}$.
- Continue until the first time $\sum_{i=1}^{G} \pi_{(i)}/G > \alpha$, and reject $H_{(1)}, \ldots, H_{(G-1)}$.



Other approaches

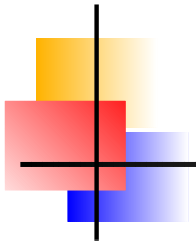
- BUM (Beta-Uniform Mixture) (Pound and Morris, 2003)
- pFDR (positive FDR) (Storey, 2003; Storey et al., 2004)
- Correlation and FDR (Efron, 2007)
- and MANY MANY others

Question: If there are 1,000 ordered test statistics, and I can only reject at most 10 tests, what should I do?



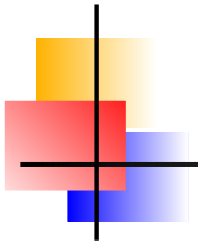
Two case studies

- Phase display experiments
- Bayesian FDR based on test statistics



Phage display

- A bacteriophage is a virus that ONLY infects bacteria (not human)
- By infecting bacteria, phage "kills" bacteria
- Phage provides important information on which proteins and peptides are potential drug candidates.

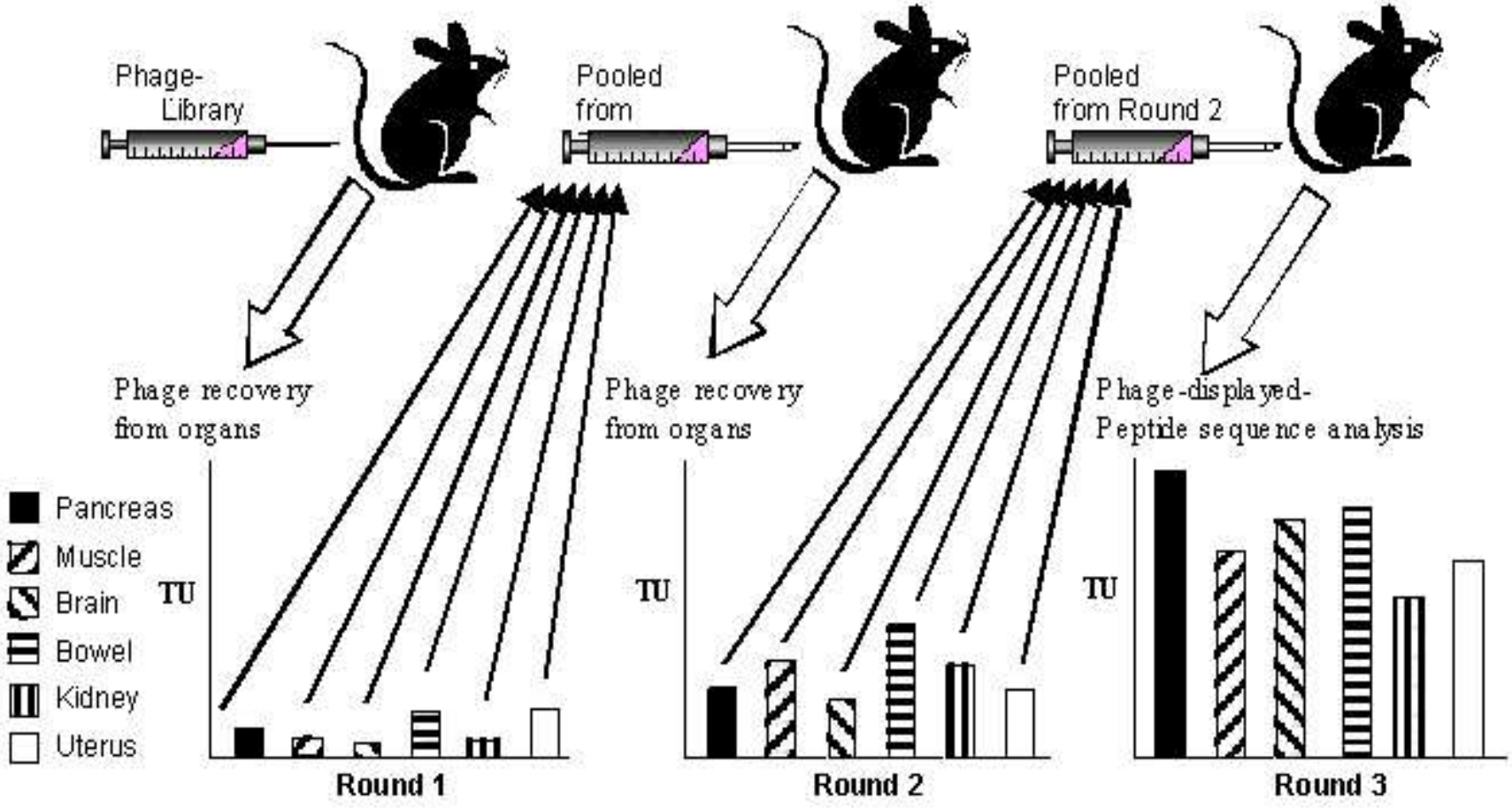


Phage display

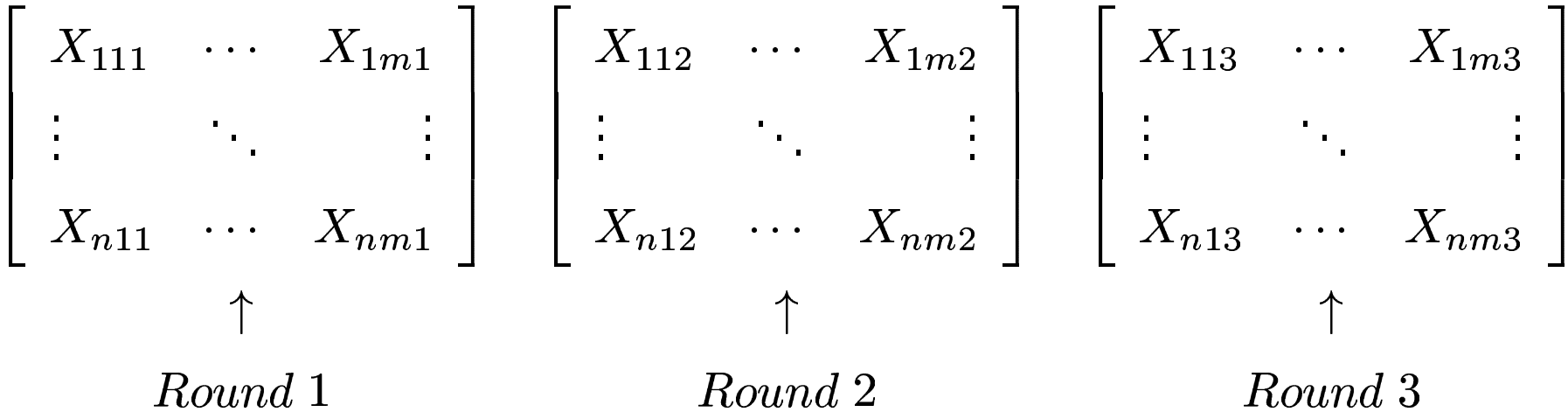
- A bacteriophage is a virus that ONLY infects bacteria (not human)
- By infecting bacteria, phage "kills" bacteria
- Phage provides important information on which proteins and peptides are potential drug candidates.
- Phage display is the process using a variety of phages in a phage library for peptide and protein screening
- The phage library is highly diversified. When exposed to a target tissue, some phage will bind with strong affinity
- If the phage binds to disease-causing molecules and changes their behavior, the peptide associated with the phage becomes a drug candidate

A nice introduction: http://www.dyax.com/phage/howitworks.asp

The mouse experiment



The count data

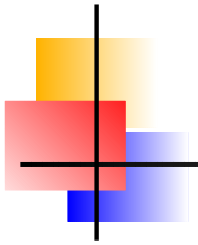


The count data

$$\begin{bmatrix} X_{111} & \cdots & X_{1m1} \\ \vdots & \ddots & \vdots \\ X_{n11} & \cdots & X_{nm1} \end{bmatrix} \quad \begin{bmatrix} X_{112} & \cdots & X_{1m2} \\ \vdots & \ddots & \vdots \\ X_{n12} & \cdots & X_{nm2} \end{bmatrix} \quad \begin{bmatrix} X_{113} & \cdots & X_{1m3} \\ \vdots & \ddots & \vdots \\ X_{n13} & \cdots & X_{nm3} \end{bmatrix}$$

$Round\ 1$ $Round\ 2$ $Round\ 3$

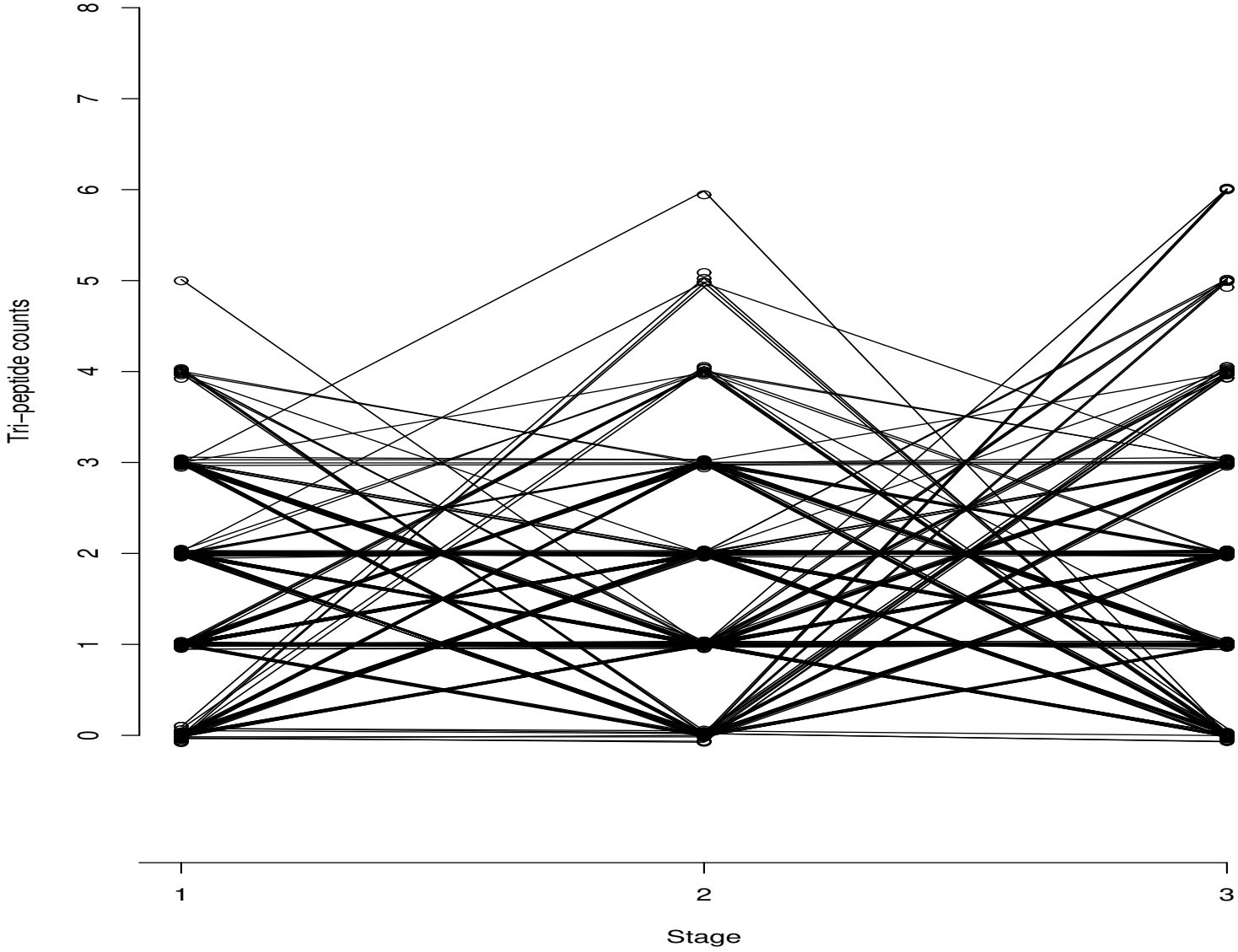
- At each round of the experiment, one data matrix obtained. Three in total
- n peptides measured for m tissues at K rounds
- $n = 4200$, $m = 6$, and $K = 3$
- X_{ijk} is the observed counts the peptide i for tissue j at round k.

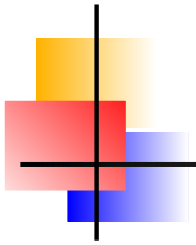


Data structure

- High dimensionality: n can be very large
- Complex correlations in the measurements
  - across tissues for the same peptide
  - across peptides for the same tissue
  - across rounds for the same pair of peptide and tissue
- Interested in the displaying patterns in the peptide counts across the three rounds

A visual display of the data

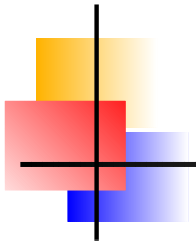




Main goal

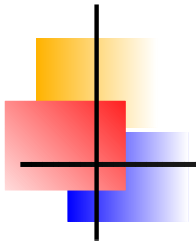
- If a peptide binds to a tissue strongly, the value of its count increase over the three rounds because of the enrichment – ascending pattern
- If a peptide does not bind to a tissue, the value of its count
  - decrease as it drops out of the selected peptide samples – descending pattern
  - oscillate due to sampling variation – oscillating pattern

Goal: To distinguish the three patterns



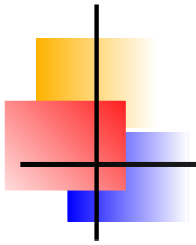
Challenges

- Mixture models are natural. Three patterns lead to three mixtures
- But mixtures of what?



Challenges

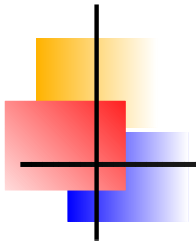
- Mixture models are natural. Three patterns lead to three mixtures
- But mixtures of what?
- A contingency table for each round
- However, three correlated tables – relationship between the tables is of major interest



The Bayesian model

- Model each cell count X_{ijk} as a Poisson random variable
- Treat the round id k as a covariate and regress the count X_{ijk} on k.

$$X_{ijk} \sim Poi(\mu_{ij} e^{k\beta_{ij}}) \tag{1}$$



The Bayesian model

- Model each cell count X_{ijk} as a Poisson random variable
- Treat the round id k as a covariate and regress the count X_{ijk} on k.

$$X_{ijk} \sim Poi(\mu_{ij} e^{k\beta_{ij}}) \tag{1}$$

- Mixtures on the distribution of the slopes

$$p(\beta_{ij}) = \pi_1 \phi(\beta_{ij}|s_1, \tau_1^2) + \pi_2 \phi(\beta_{ij}|s_2, \tau_2^2) + \pi_3 \phi(\beta_{ij}|s_3, \tau_3^2) \tag{2}$$

The Bayesian model

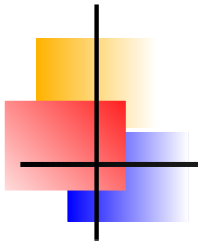
- Model each cell count X_{ijk} as a Poisson random variable
- Treat the round id k as a covariate and regress the count X_{ijk} on k.

$$X_{ijk} \sim Poi(\mu_{ij} e^{k\beta_{ij}}) \tag{1}$$

- Mixtures on the distribution of the slopes

$$p(\beta_{ij}) = \pi_1 \phi(\beta_{ij}|s_1, \tau_1^2) + \pi_2 \phi(\beta_{ij}|s_2, \tau_2^2) + \pi_3 \phi(\beta_{ij}|s_3, \tau_3^2) \tag{2}$$

- The prior of the s_1 centered at a negative value; fix $s_2 = 0$; and the prior of the s_3 centered at a positive value.



The Bayesian model II

Full Bayes hierarchical modeling

- μ_{ij} is the baseline count for peptide i for tissue j.
  $\mu_{ij} \sim \mu_0 G(\alpha, 1/\alpha)$
- Hierarchical priors on the hyperparameters
  - Dirichlet prior for (π_1, π_2, π_3)
  - Normal priors for s_1 and s_3 ($s_2 = 0$)
  - Inverse gamma priors for all the variance parameters
  - Inverse gamma prior for μ_0

Model fitting based on a hybrid of the Gibbs sampler and the Metropolis-Hastings algorithm.

$$[\mu_{ij} \mid \boldsymbol{N}, \text{ rest}] \sim G(\sum_{k=0}^{2} N_{ijk} + \alpha, \ \frac{1}{1 + e^{\beta_{ij}} + e^{2\beta_{ij}} + \alpha/\mu_0}),$$

and

$$[\mu_0 \mid \boldsymbol{N}, \text{ rest}] \sim IG(a_{\mu_0} + n, \ \frac{1}{1/b_{\mu_0} + \alpha \sum_{i=1}^{n} \sum_{j=1}^{m} \mu_{ij}})$$

$$[\boldsymbol{\lambda}_{ij} \mid \boldsymbol{N}, \text{ rest}] \sim Multi\left(1; \ \frac{\pi_1\phi_1}{\pi_1\phi_1 + \pi_2\phi_2 + \pi_3\phi_3}, \right.$$
$$\left. \frac{\pi_2\phi_2}{\pi_1\phi_1 + \pi_2\phi_2 + \pi_3\phi_3}, \ \frac{\pi_3\phi_3}{\pi_1\phi_1 + \pi_2\phi_2 + \pi_3\phi_3}\right)$$

and

$$[\pi \mid \boldsymbol{N}, \text{ rest}] \sim Dir(\pi_{1,0} + n_{\text{neg}}, \pi_{2,0} + n_{\text{zero}}, \pi_{3,0} + n_{\text{pos}}).$$

$$[\beta_{ij} \mid \boldsymbol{N}, \text{ rest}] \propto e^{-\mu_{ij}(1+e^{\beta_{ij}}+e^{2\beta_{ij}})} \mu_{ij}^{N_{ij0}+N_{ij1}+N_{ij2}} e^{\beta_{ij}(N_{ij1}+2N_{ij2})} \prod_{l=1}^{2} \phi^{\lambda_{ijl}}.$$

$$[s_1 \mid \boldsymbol{N}, \text{ rest}] \sim N(B_{\text{neg}}\bar{\beta}_{\text{neg}} + (1 - B_{\text{neg}})m_1, B_{\text{neg}}\tau_1^2/n_{\text{neg}})$$

and

$$[s_3 \mid \boldsymbol{N}, \text{ rest}] \sim N(B_{\text{pos}}\bar{\beta}_{\text{pos}} + (1 - B_{\text{pos}})m_3, B_{\text{pos}}\tau_3^2/n_{\text{pos}}),$$

where $B_{\text{neg}} = \frac{\eta_1^2}{\eta_1^2+\tau_1^2/n_{\text{neg}}}$ and $B_{\text{pos}} = \frac{\eta_3^2}{\eta_3^2+\tau_3^2/n_{\text{pos}}}$.

$$[\tau_1^2 \,|\boldsymbol{N},\ \text{rest}] \quad \sim \quad IG\left(a_\tau + \frac{n_{\text{neg}}}{2},\ \ \frac{1}{\frac{1}{b\,\tau} + \sum\limits_{(i,j)\in\Delta_{\text{neg}}} (\beta_{ij} - s_1)^2}\right)$$

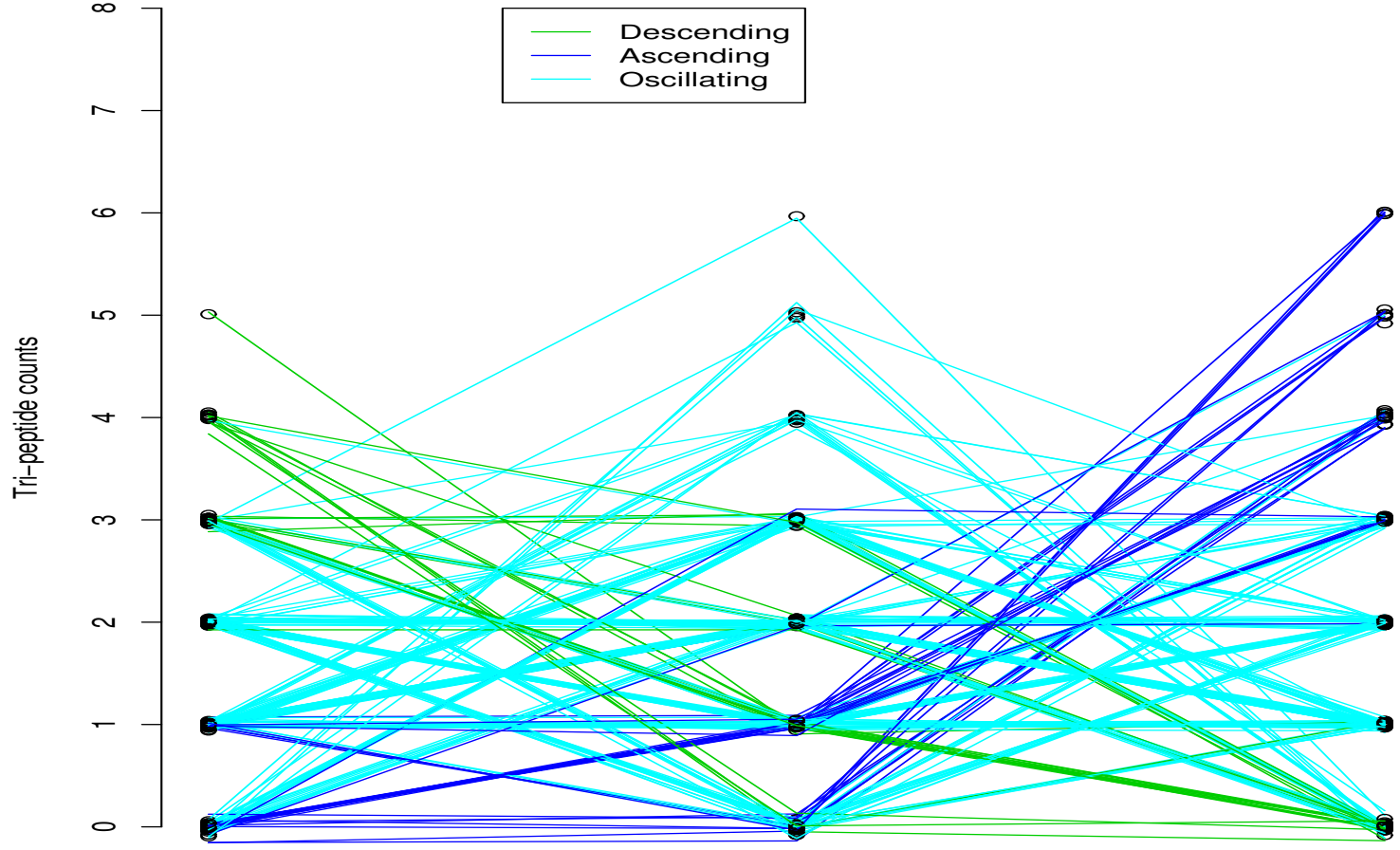
$$[\tau_2^2 \,|\boldsymbol{N},\ \text{rest}] \quad \sim \quad IG\left(a_\tau + \frac{n_{\text{zero}}}{2},\ \ \frac{1}{\frac{1}{b\,\tau} + \sum_{(i,j)\in\Delta_{\text{zero}}} \beta_{ij}^2}\right),$$

and

$$[\tau_3^2 \,|\boldsymbol{N},\ \text{rest}] \quad \sim \quad IG\left(a_\tau + \frac{n_{\text{pos}}}{2},\ \ \frac{1}{\frac{1}{b\,\tau} + \sum\limits_{(i,j)\in\Delta_{\text{pos}}} (\beta_{ij} - s_3)^2}\right)$$

Results

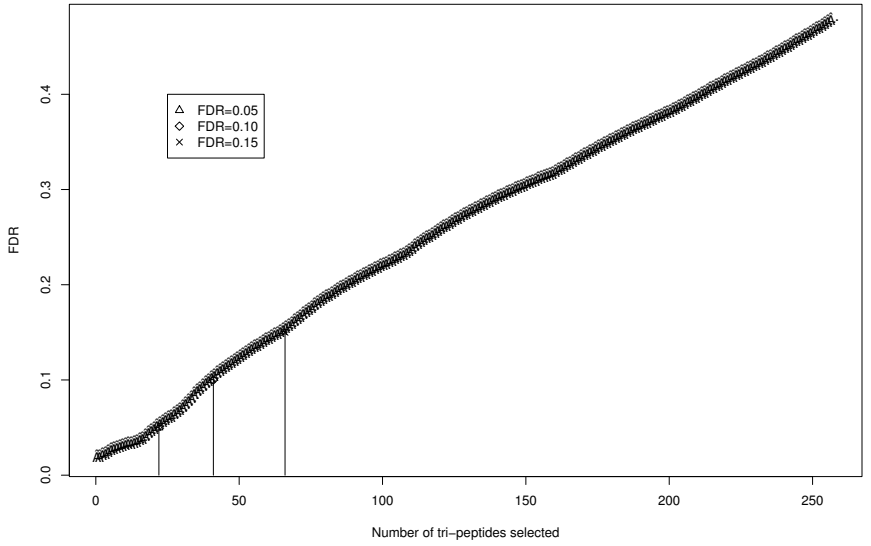
Three display patterns identified



Results II

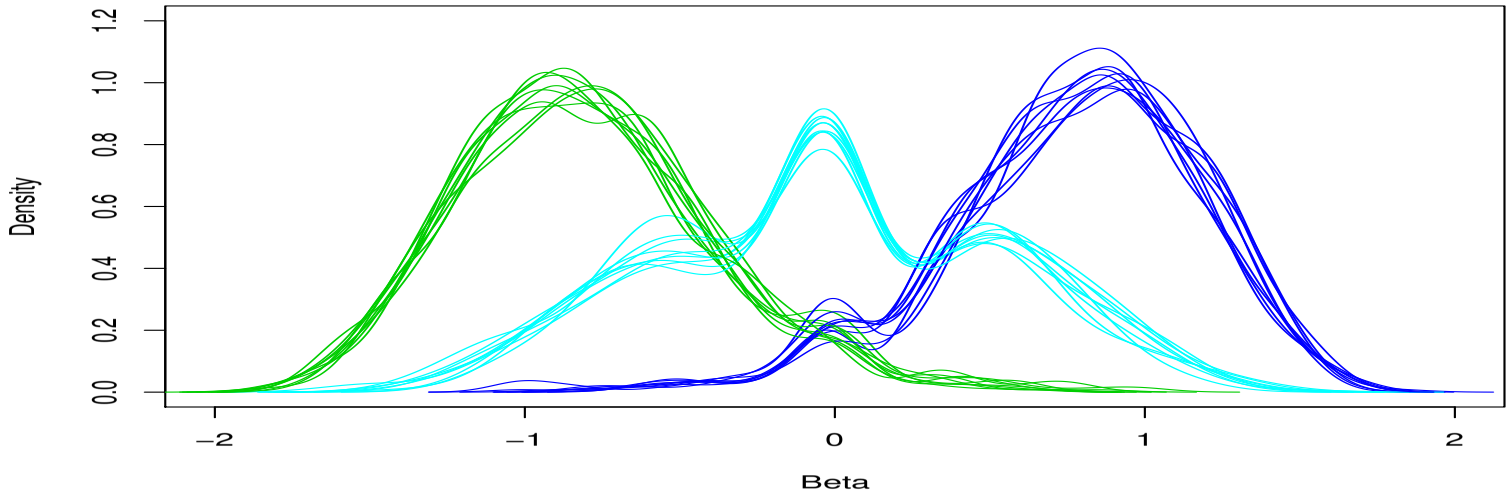
Particularly interested in the blue group, which indicate that the peptide bind strongly the the corresponding tissue

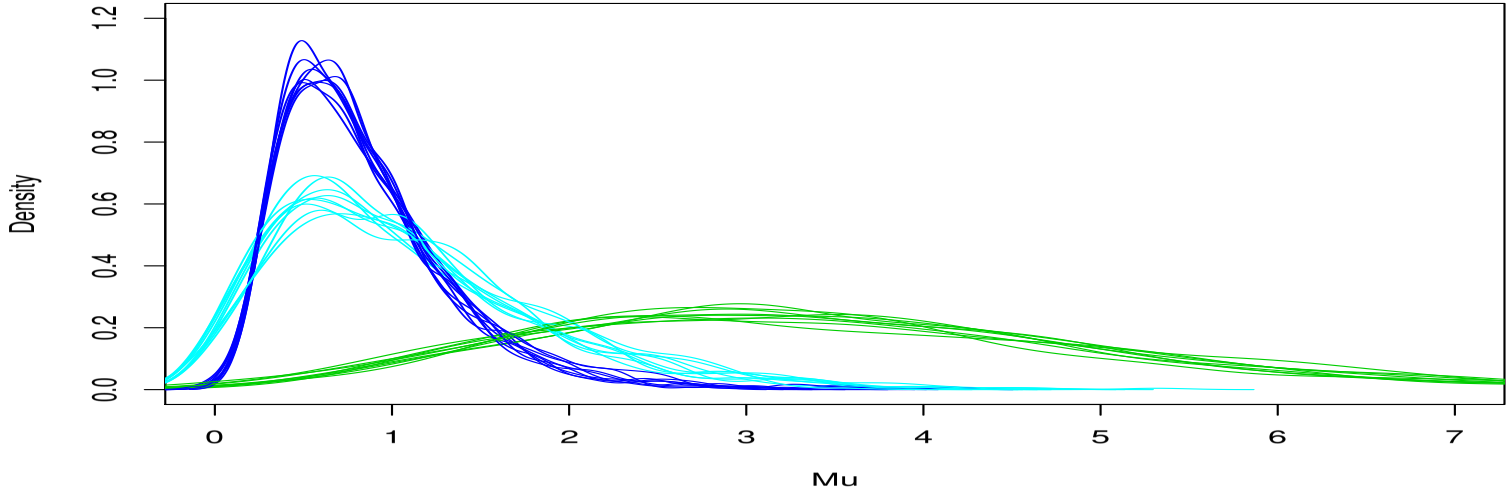
- Compute the posterior probability $P(\beta_{ij} > 0|\text{Data})$
- FDR based on the posterior probabilities, Newton *et al.* (2004)

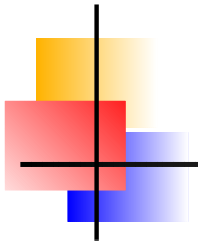


Organ	Tri-peptide	Counts (estimated Poisson me		
A	GGL	1 (*0.81*)	0 (*2.14*)	6 (*5.66*
B	DRW	0 (*0.42*)	0 (*1.26*)	4 (*3.77*
B	AGV	0 (*0.38*)	0 (*1.12*)	4 (*3.67*
B	FGG	0 (*0.39*)	0 (*1.21*)	4 (*3.73*
B	GGR	1 (*0.85*)	0 (*2.19*)	6 (*5.74*
B	GLL	0 (*0.62*)	1 (*1.38*)	3 (*3.06*
K	LRV	0 (*0.63*)	1 (*1.62*)	4 (*4.20*
K	LGS	1 (*1.46*)	2 (*2.71*)	5 (*5.02*
M	GGT	0 (*0.38*)	0 (*1.34*)	5 (*4.68*
M	FSG	0 (*0.62*)	1 (*1.80*)	5 (*5.25*
M	AGS	0 (*0.61*)	1 (*1.79*)	5 (*5.26*
M	IGS	0 (*0.60*)	1 (*1.77*)	5 (*5.22*
M	AIG	0 (*0.41*)	0 (*1.23*)	4 (*3.70*
M	IAY	0 (*0.42*)	0 (*1.26*)	4 (*3.77*
M	DFS	0 (*0.42*)	0 (*1.26*)	4 (*3.77*
M	RRS	0 (*0.58*)	1 (*1.56*)	4 (*4.16*
M	FRS	0 (*0.64*)	1 (*1.42*)	3 (*3.10*
M	SGV	0 (*0.61*)	1 (*1.38*)	3 (*3.11*
P	SSV	1 (*0.82*)	0 (*2.17*)	6 (*5.74*
P	SSV	0 (*0.62*)	1 (*1.37*)	3 (*3.14*
P	GWR	0 (*0.62*)	1 (*1.39*)	3 (*3.06*
U	AAG	0 (*0.63*)	1 (*1.70*)	4 (*4.21*

simulation


Density
0.0
0.2
0.4
0.6
0.8
1.0
1.2
−2
−1
0
1
2
Beta


Density
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0
1
2
3
4
5
6
7
Mu



Discussions

- Poisson assumption vs. Multinomial assumption for the counts
- Mixture of Normals vs. others, e.g.
  $P(\beta_{ij}) = -\pi_1 G(\beta_{ij}|g_1, h_1) + \pi_2 N(\beta_{ij}|s_2, \tau_2^2) + \pi_3 G(\beta_{ij}|g_2, h_2)$
- Different baseline count μ_{ij} vs. one baseline μ for all the (i, j)'s – simulation
- Functional data analysis if the covariate is time

Bayesian Multiple Testing Based on Test Statistics

Yuan Ji

January 28, 2008

Outline

1. The problem of multiple testing
2. The proposed approach
   1. A hierarchical modeling approach for multiple testing
   2. An illustrative example – F-tests
   3. A model assessment tool
3. Application
4. A brief discussion

A Bayesian framework

- Suppose a sequence of m null hypotheses H_{0i} is tested against a corresponding sequence of alternative H_{1i} for $i = 1, \ldots, m$.
- A Bayesian procedure for this problem:
  - Construct a latent indicator $J_i = 0$ if H_{0i} is true and $J_i = 1$ if H_{1i} is true.
  - Compute the marginal posterior probability $\Pr(J_i = 1|\text{data})$ based on some appropriate models.
  - Adjust for multiplicity using the marginal posterior probabilities.

Bayesian modeling

For test i, observed data y_i. A Bayesian hierarchical model consists of

- Probability distribution $p(y_i|J_i = k) = p_k(y_i; \boldsymbol{\theta_k})$, $k = 0, 1$.
- The likelihood function:

$$p_0(y_i; \boldsymbol{\theta_0})^{1-J_i} p_1(y_i; \boldsymbol{\theta_1})^{J_i}.$$

- Priors for $\boldsymbol{\theta_k}$ is $f_k(\boldsymbol{\theta_k})$; prior $\Pr(J_i = 1) = \pi$.
- Hyperpriors for the parameters in the priors (e.g., π).

Compute

$$r_i = \Pr(J_i = 1|y_1, \ldots, y_m)$$

the marginal posterior probability that H_{1i} is true.

Multiplicity

- Probabilities r_i adjust for multiplicities automatically as long as
  - $\Pr(J_i = 1) > 0$ for all $i = 1, \ldots, m$;
  - $\pi \sim p(\pi)$, rather than fixed.
  - *Ref. Scott and Berger (2003); Müller et al. (2006)*
- Optimal decision (Müller et al., 2004) is

$$I(r_i > t),$$

  to reject all the null hypotheses with $r_i > t$ for some fixed value t.
- Choice of t depends on choice of loss functions.

Motivation

- Construction of appropriate Bayesian models can be difficult. (e.g., construction of priors for $\boldsymbol{\theta_k}$).
- Values of posterior probabilities r_i are often sensitive to the prior densities.
- MCMC computation can be intensive, especially for high-dimensional data (e.g., genomics/proteomics data).

Hierarchical model based on test statistics

- Johnson (2005) proposed computing posterior probabilities r_j based on test statistics.
- Main idea:
  - Base the models on the sampling distributions of test statistics.
  - The null distributions are often completely specified – no need for prior specification.
  - The alternative distributions of test statistics can often be described with a parsimonious parametrization.

Hierarchical model based on test statistics (cont)

Therefore,

- Models under the null $p_0(y_i)$ are free of parameters.
- Models under the alternative $p_1(y_i; \boldsymbol{\theta_1})$ depend on few parameters (often just one).
- $\Pr(J_i = 1 | y_1, \ldots, y_m; \boldsymbol{\theta_1})$ has a closed-form solution – easy to sample.

Probability model

Let f_i be the test statistic (e.g., χ^2-, $F-$, $t-$ or $z-$statistic) for null H_{0i} vs. H_{1i};

- Likelihood $p(f_i|J_i,\tau) = p_0(f_i)^{1-J_i} p_1(f_i|\tau)^{J_i}$;
- Prior of $J_i \sim \text{Bin}(1,\pi)$;
  - Hyperprior of $\pi \sim \text{Beta}(p_0,(1-p_0))$, where p_0 is fixed.
- Prior of $1/\tau \sim \text{Gamma}(1,2)$;

MCMC

MCMC algorithm for $\{\pi, \tau, J_1, \ldots, J_m\}$

- Full conditional

$$\Pr(J_i = 1 | f_1, \ldots, f_m, \tau, \pi) = \frac{p_1(f_i|\tau)\pi}{p_1(f_i|\tau)\pi + p_0(f_i)(1-\pi)}$$

- $\pi | J_1, \ldots, J_m \sim \text{Beta}(p_0 + \sum J_i, (1 - p_0) + m - \sum J_i)$.
- Sample τ, e.g., using random-walk Metropolis-Hastings.

A special case – $F-$tests

Suppose

$$\mathbf{y}_i|\beta_i, \sigma_i^2 \sim N_n(\mathbf{X}_i\beta_i, \sigma_i^2\mathbf{I}).$$

For testing the validity of linear constraint $H_{0i} : \mathbf{Q}'\beta_i = \xi$, the classical $F-$ statistic f_i is the ratio of average sums of squares.

- $p_0(f_i)$ is a central $F-$ distribution;
- Suppose alternative H_{1i} assumes that

$$\beta_i \sim N(\beta_i^*, \tau\sigma_i^2(\mathbf{X}_i'\mathbf{X}_i)^{-1})$$

  where β_i^* is a value satisfying H_{0i},
- then $p_1(f_i|\tau) \sim (1+\tau)p_0(f_i)$.

Decision rules

- Posterior probability $r_i = \Pr(J_i = 1 | f_1, \ldots, f_m)$ is computed using the MCMC sample.
- Reject H_{0i} if $r_i > t$ for some value of t (more discussion later)

qqplot

A quantile-quantile plot is proposed to check the model fitting.

- Suppose $\{\tau^1, \ldots, \tau^B\}$ is the MCMC sample.
- Randomly draw τ^s.
- Obtain the corresponding posterior sample $\{J_1^s, \ldots, J_m^s\}$ from the s^{th} iteration of the MCMC.
- Assign the test statistics f_i to the null group if $J_i^s = 0$, and to the alternative group if $J_i^s = 1$.

qqplot continued

- Plot the sample quantiles of f_i in the null group against the theoretical quantiles based on the distribution $p_0(f_i)$;
- Plot the sample quantiles of f_i in the alternative group against the theoretical quantiles based on the distribution $p_1(f_i|\tau^k)$;
- Compare the curves with the 45 degree line.

This procedure only works for quantities of which the sampling distributions are free of parameters – such as the F-statistics (its distribution only depends on two degrees of freedom).

Simulation 1

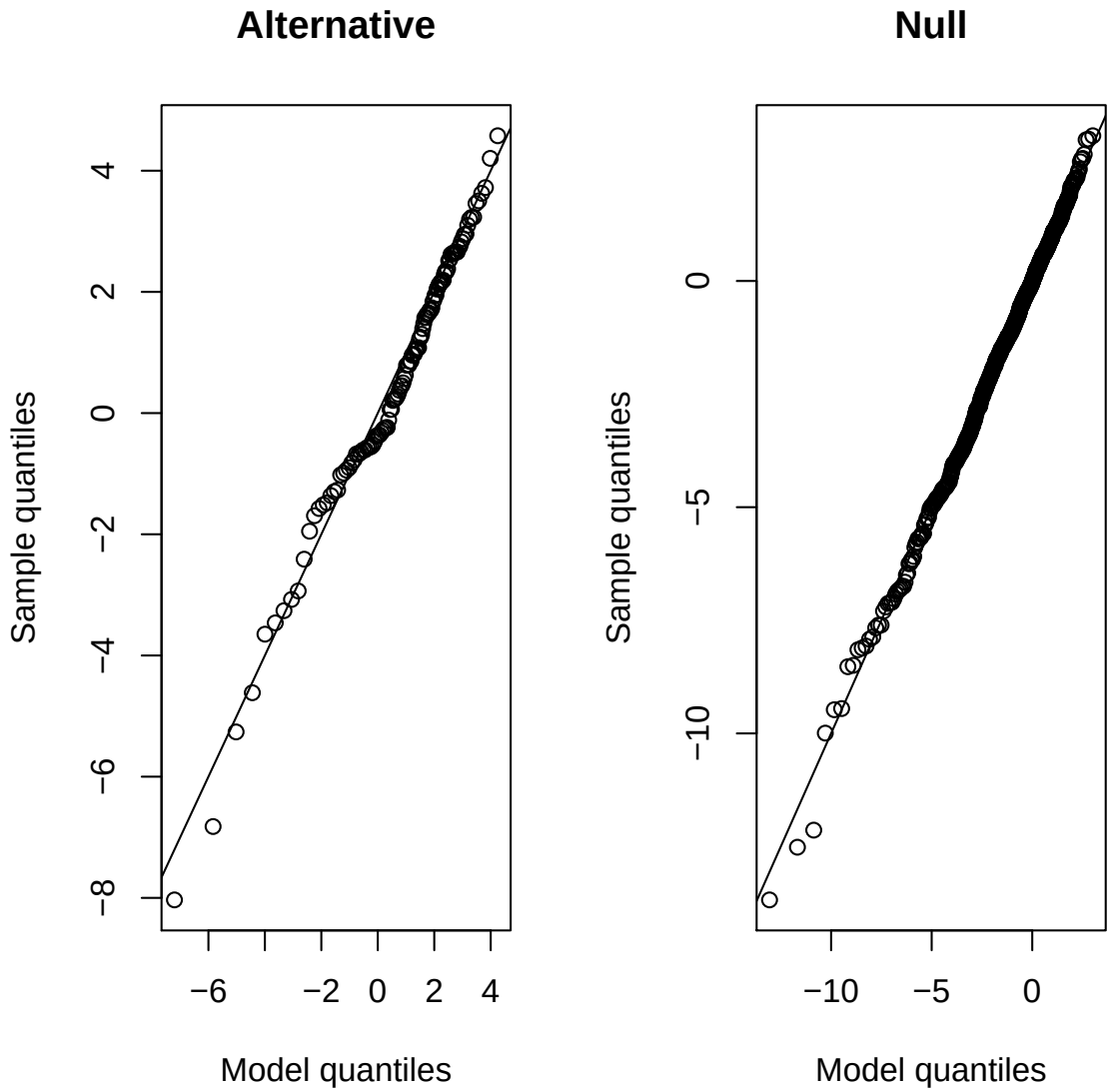
Consider one-sample t-tests $H_{0i} : \mu_i = 0$, $i = 1, \ldots, m$.

Observed data for test i are samples $\{y_{i1}, \ldots, y_{in}\}$. The $F-$statistic f_i is the square of the one-sample $t-$statistic.

- We generated $m = 1000$ tests.
- Sample sizes per test $n = 11$.
- Under H_{0i}, $f_i \sim F_{1,10}$ and under alternative $f_i \sim (1 + \tau)F_{1,10}$.

Simulation scheme consists of sampling τ, π, $J_i|\pi$, and $f_i|J_i, \pi$ (in this order), from their true distributions under the proposed model.

qq-plots 1


Alternative
Sample quantiles
Model quantiles
−8
−6
−4
−2
0
2
4
−6
−2
0
2
4
Null
Sample quantiles
Model quantiles
−10
−5
0
−10
−5
0

Simulation 2

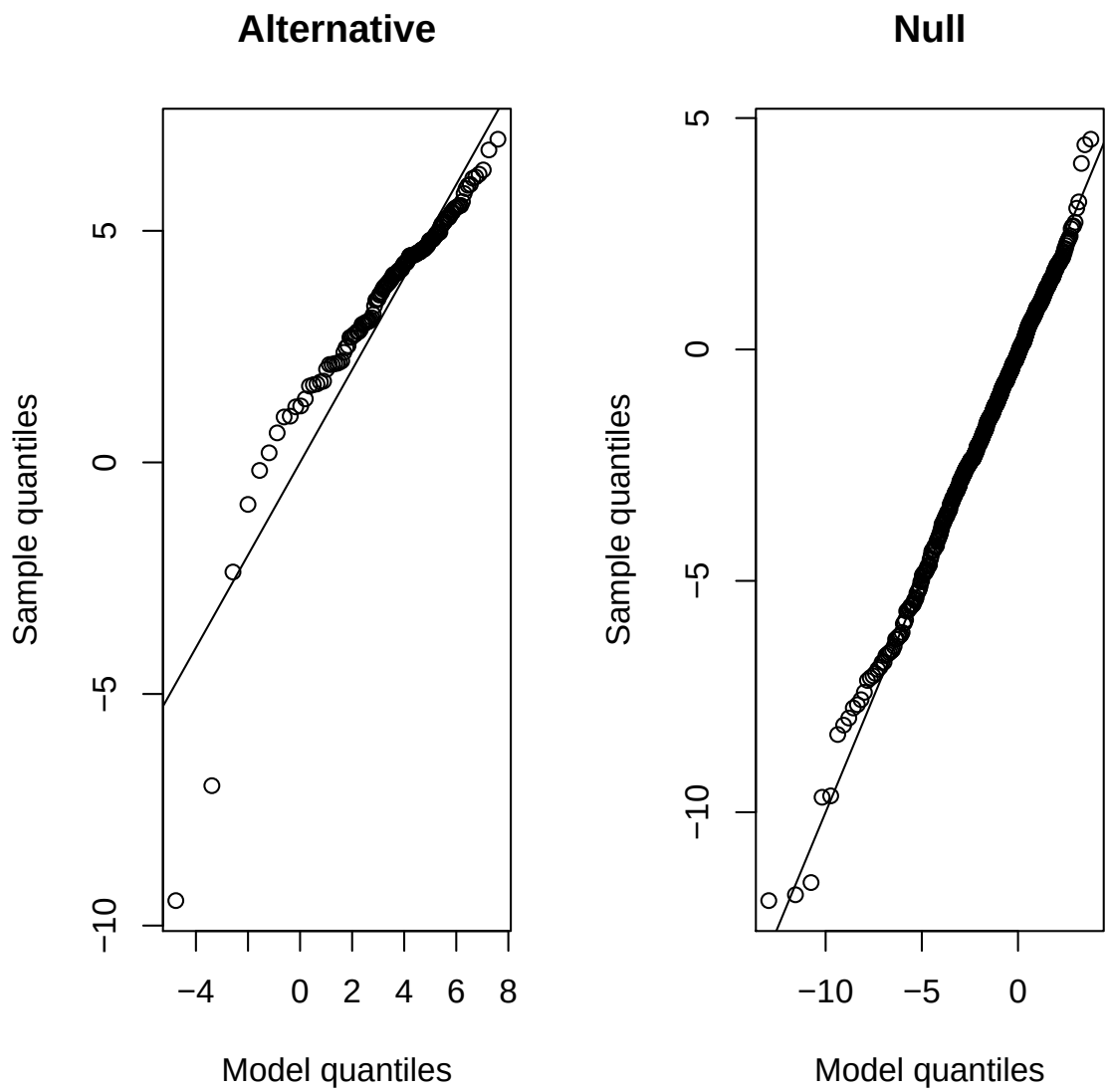
- Sample $y_{i1}, \ldots, y_{in} \overset{iid}{\sim} N(3,1)$ for $i = 1, \ldots, 100$;
- Sample $y_{i1}, \ldots, y_{in} \overset{iid}{\sim} N(0,1)$ for $i = 101, \ldots, 1000$;
- $H_{0i} : \mu_i = 0$
- Compute

$$t_i = \frac{\bar{y}_{i.}}{\hat{\sigma}/\sqrt{n}}$$

where $\bar{y}_{i.}$ is the sample mean and $\hat{\sigma}$ is the sample standard deviation.

After applying the proposed method,

qqplots 2

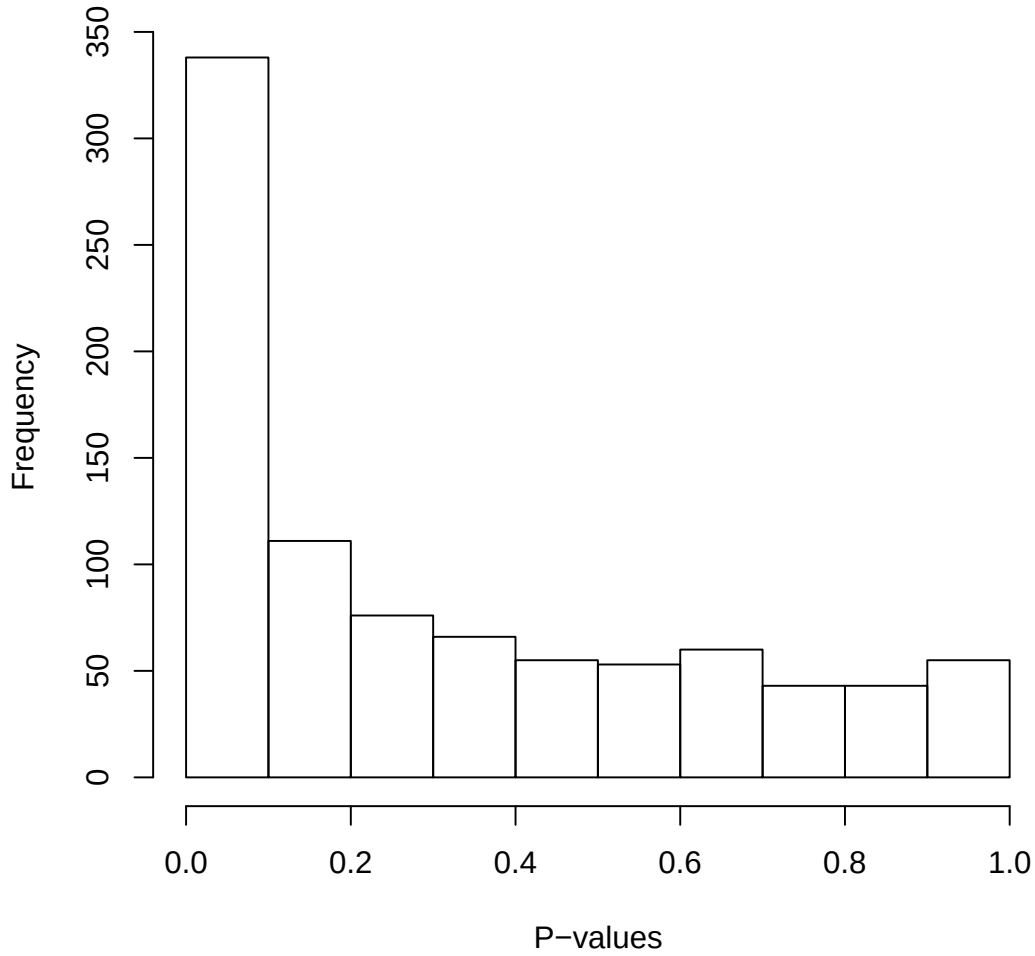

Alternative
Sample quantiles
Model quantiles
−10
−5
0
5
−4
0
2
4
6
8
Null
Sample quantiles
Model quantiles
−10
−5
0
5
−10
−5
0

siRNA screening

An siRNA screening experiment conducted by Gordon Mills and his lab.

- A kinase library of about 900 siRNA's are screened for their silencing properties.
- A functional silencing siRNA significantly reduced cell viability (measured as a continuous variable).
- Using 96-well plates, the library is screened with 30 plates in triplicates.
- F-statistics f_i are computed for all 900 siRNA's with degrees of freedom $(1, 4)$.

Histogram of p-values

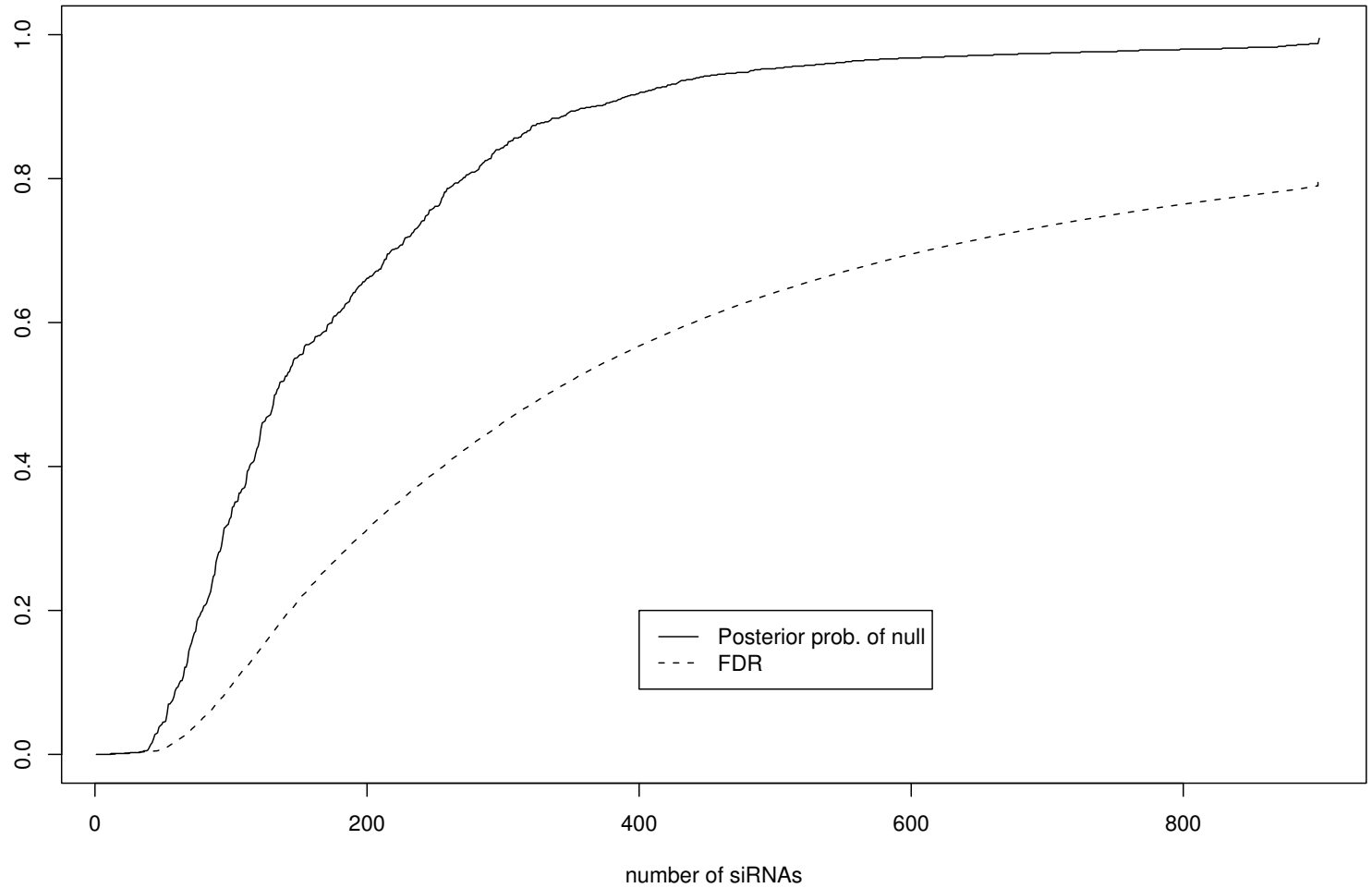


Results using the Bayesian procedure

We applied the proposed method for the 900 F-statistics f_i.

- Assume $\pi \sim \text{Beta}(0.5, 0.5)$.
- Assume $1/\tau \sim \text{Gamma}(1, 2)$.
- Under null, $f_i \sim F(1, 4)$.
- Under alternative, $f_i \sim (1 + \tau)F(1, 4)$.

Posterior probability and FDR



Decision rules

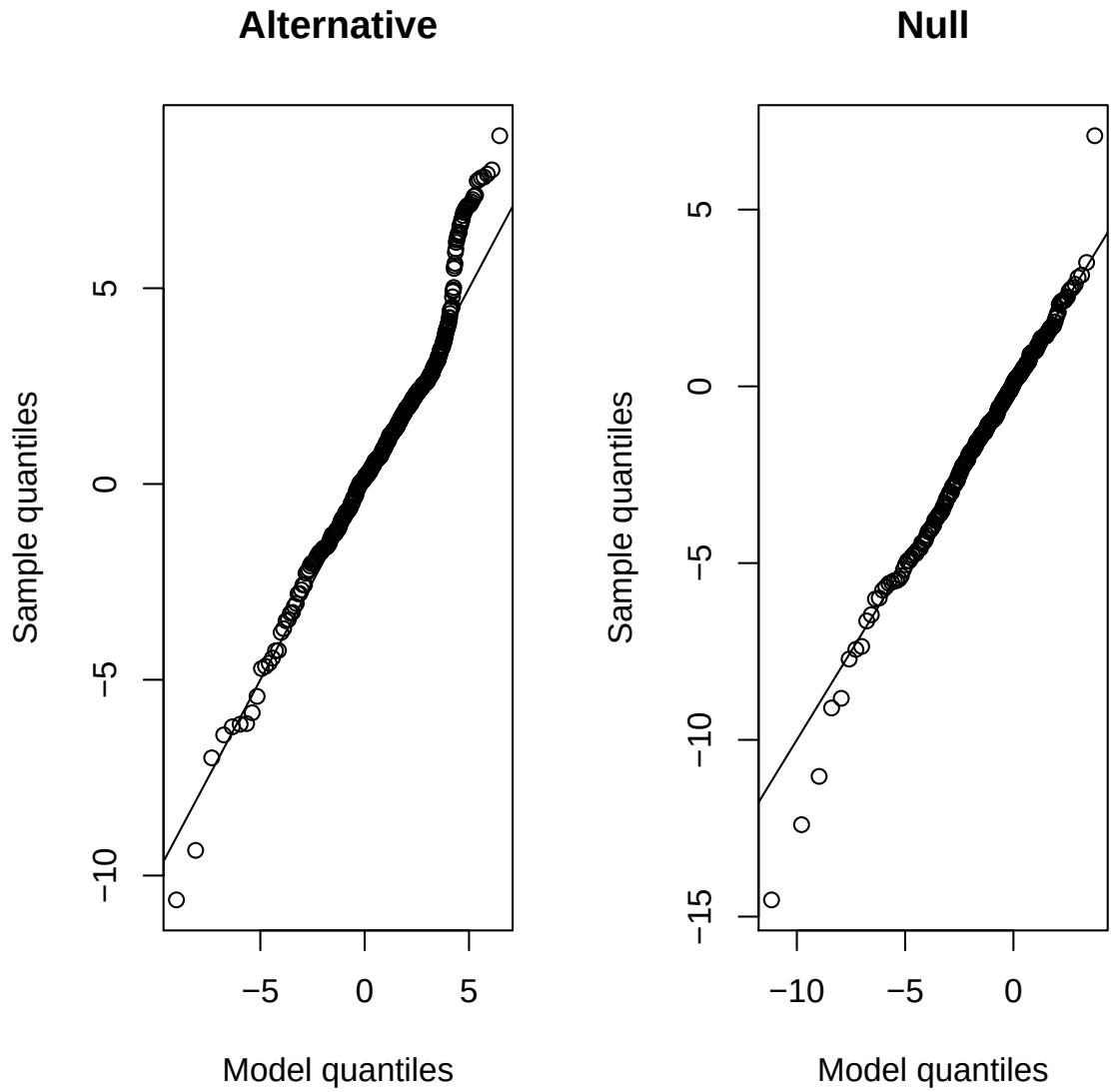
The optimal decision takes the form (M̈(u)ller et al., 2004)

$$d_i = I(r_i \leq t)$$

- If the "goal" (loss function) is to minimize FNR subject to FDR $\leq \alpha$, then t equals the largest r_i such that the corresponding posterior expected FDR (by rejecting all the $r_j \leq r_i$) is $\leq \alpha$.

- In the above plot, draw a horizontal line at y-axis $= 0.2$. Draw a vertical line at the intersection between the horizontal line and the dotted curve. The intersection between the vertical line and the solid curve is the optimal t value in d_i.

Model assessment



A gene expression experiment

Khodarev et al. (2005) studied the association between progression of Barrett's Metaplasia to Adenocarcinoma and gene expression levels. Three conditions are examined:

- Normal esophageal epithelium
- Premalignant Barrett's metaplasi,
- Esophageal adenocarcinoma

For each condition, $n = 8$ Affymetrix U133A arrays were produced from 8 different patients with the same condition. After normalization using dChip (Li and Wong, 2001), we obtained $m = 16384$ genes, each with 24 measurements.

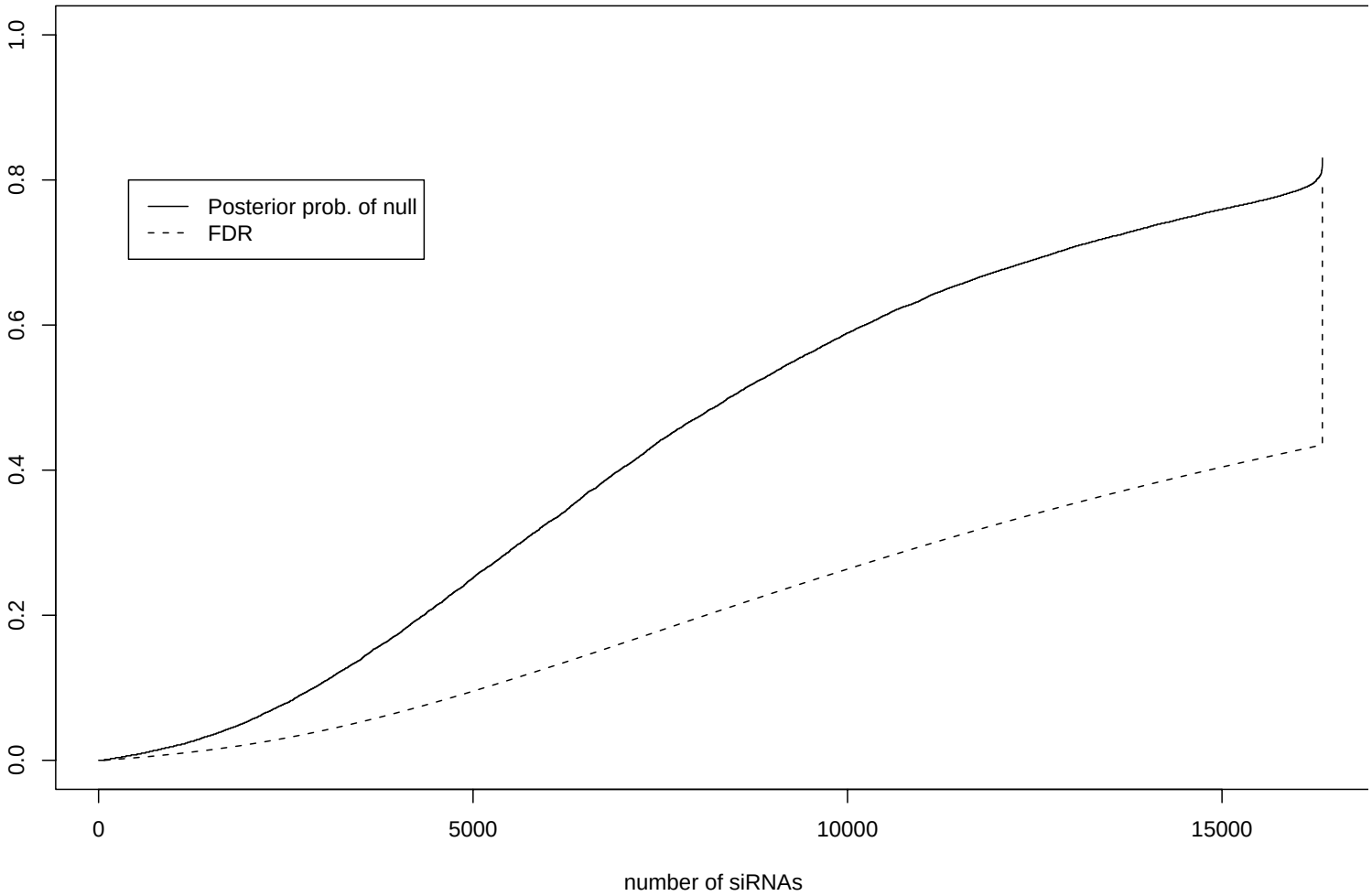
A gene expression experiment (cont)

For each gene, we performed a one-way ANOVA using the three conditions as a factor. We obtained $m = 16384$ $F-$statistics with degrees of freedom $(2, 21)$. Therefore,

- $p_0(f_i)$ follows $F_{2,21}$
- $p_1(f_i|\tau)$ follows $(1+\tau)F_{2,21}$.

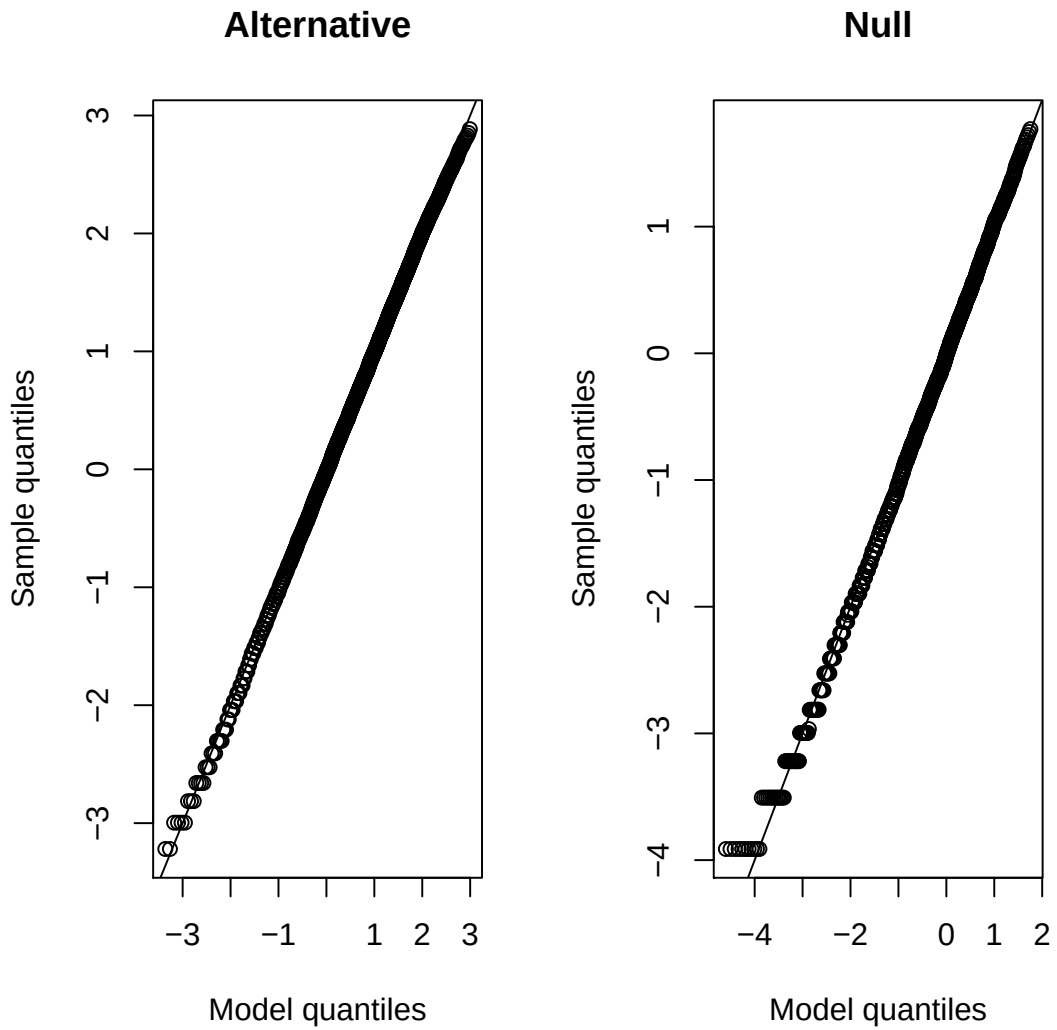
We applied the proposed method and computed $r_i = \Pr(J_i = 1|f_1, \ldots, f_m)$ for each gene i.

Posterior probability and FDR



Model assessment

We let $\pi \sim \text{Beta}(.5, .5)$.



Conclusions

- The proposed model simplifies the process of specifying prior distributions for unknown parameters, which can be tricky.
- Only one parameter needs to be sampled using M-H; others are sampled directly from Bernoulli distributions.
- Information across all the tests is used in the decision making for each single test – through the common parameter τ.
- We provide a simple model-assessment tool to check the model fitting.
- Additional research is needed to explore more general assumptions under the alternative when model does not fit.