Connecting the False Discovery Rate to shrunk estimates

N.W. Galwey,
3 February 2026

Top-level outline

- Review of the False Discovery Rate (FDR)
- Review of shrunk estimates
- The connection between the FDR and shrunk estimates
- Practical illustration of combined application
- Conclusions

Detailed Outline

Review of the False Discovery Rate (FDR)

- Specification of a significance test
- The multiple testing problem and the Replication Crisis
- The confusion matrix
- False-positive rate (*p*-value) vs. False Discovery Rate (FDR)
- The Benjamini-Hochberg FDR (BH-FDR) criterion: maximisation of statistical power
- Visualisation of the BH-FDR: individual *p*-value–FDR mapping
- The FDR as a predictor
- Bayesian interpretation of the FDR
- BH-FDR as an empirical-Bayes method: multiplicity not as the problem, but as part of the solution

cont'd.../

Detailed Outline/…cont'd.

Review of shrunk estimates

- Variation among group means: specification of model
- Specification of shrunk estimates
- Shrunk estimates as unbiased predictors
- Shrinkage of estimates as regression towards the mean
- Shrinkage of estimates as an empirical-Bayes method: multiplicity not as the problem, but as part of the solution

cont'd…/

Detailed Outline/…cont'd.

The connection between the FDR and shrunk estimates

- Common features of the FDR and shrunk estimates
- Specification of appropriate contexts in which to make the connection
- Application of a one-sided significance test to group means
- Bivariate distribution of observed test statistic (Z_{obs}) vs. true group means (G):
  the tool to connect FDR to shrinkage factor
- Parameters of the relationship:
  - overall mean (μ_G)
  - shrinkage factor $\left(\frac{\sigma_G^2}{\sigma_G^2+\sigma_E^2}\right)$
  - null hypothesis ($H_0: G \leq g_{H_0}$)
  - significance threshold (z_α)
- The bivariate distribution as a graphical representation of the confusion matrix
- Exploration of parameter-value combinations
- The FDR vs. significance threshold relationship

cont'd.../

Detailed Outline/…cont'd.

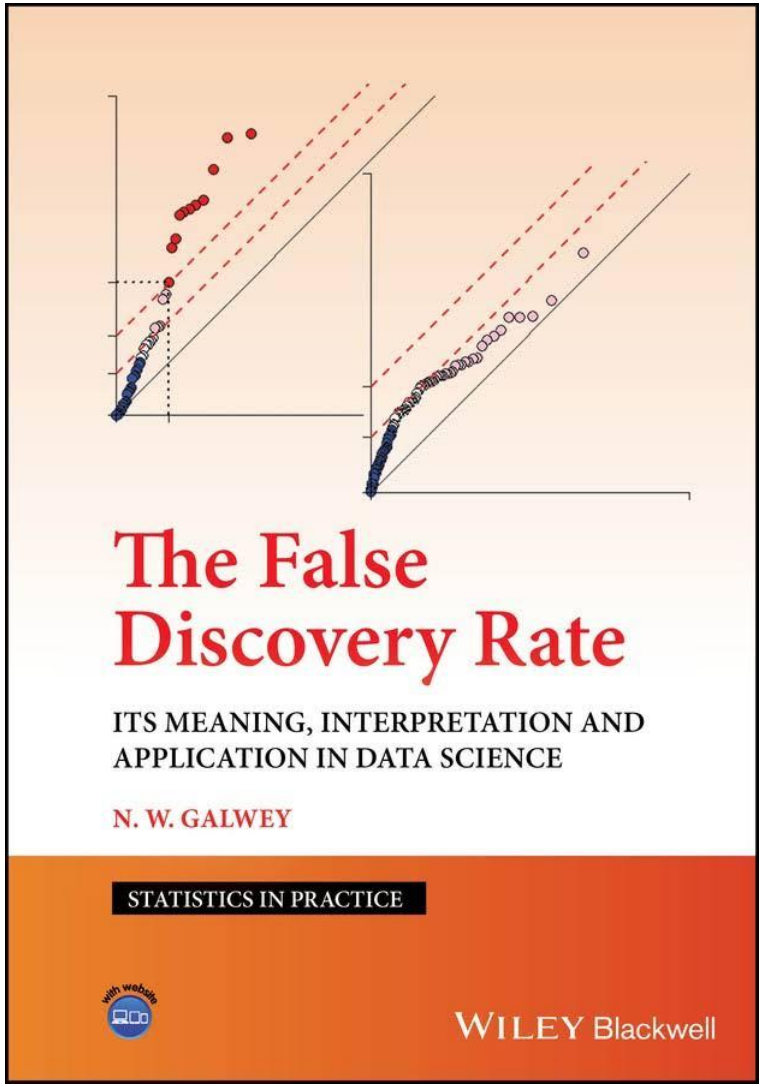
Practical illustration of combined application

- Application to published experimental data
- Application to simulated data: assessment of performance

Conclusions

Review of the False Discovery Rate (FDR)

Excellent reference covering this material



The specification of a significance test

G A quantitative effect of interest, e.g.

mean of group under consideration – mean of all groups

H_0 Null hypothesis, e.g. $G \leq 0$

H_1 Alternative hypothesis, e.g. $G > 0$. (N.B. One-sided test.)

Z A test statistic, a function of G with known distribution at the boundary of H_0,

e.g. $Z|(G = 0) \sim N(0, 1)$

z An observed value of Z, obtained from a significance test

The p-value is then defined as

$$p = P(Z \geq z|H_0)$$

α A significance threshold. A test result

$$p \leq \alpha$$

is then considered significant.

z_α The significance threshold of the Z statistic. That is,

$$P(Z \geq z_\alpha|H_0) = \alpha$$

For simplicity, when calculating p-value, assume $G = 0$ whenever H_0 is true. (For now.)

Graphical illustration of the significance threshold

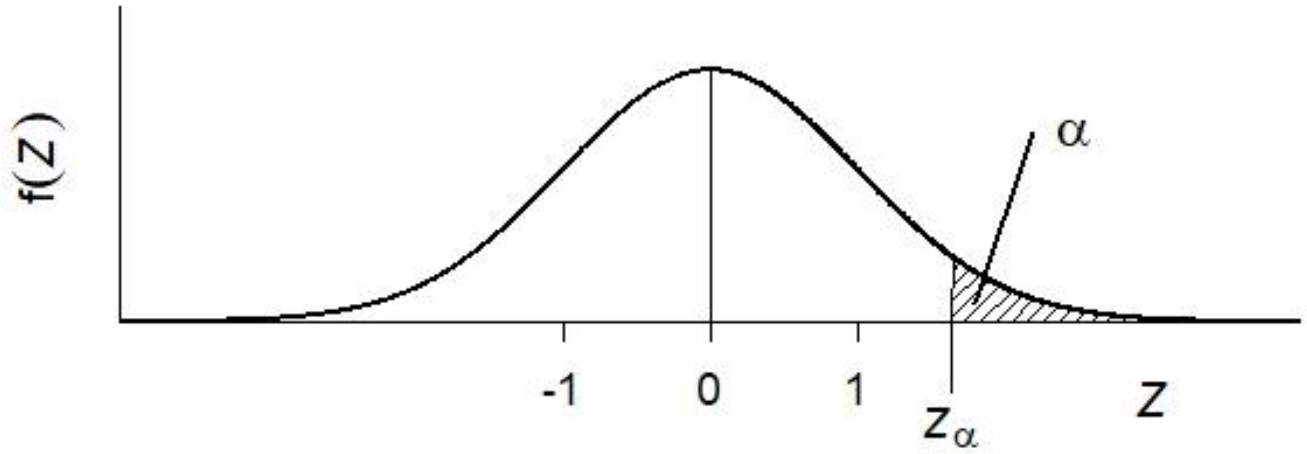
α A significance threshold. A test result

$$p \leq \alpha$$

is considered significant.

z_α The significance threshold of the Z statistic. That is,

$$P(Z \geq z_\alpha | H_0) = \alpha$$



The multiple testing problem

$$P(Z \geq z_\alpha | H_0) = \alpha$$

If one test is conducted and α is low, the probability of a false-positive result is low.

e.g. if

$$\alpha = 0.05$$

then

$$P(Z \geq z_{0.05} | H_0) = 0.05 .$$

However, consider the situation in which m_0 *independent* tests are conducted, of which k give significant results. If at least one test gives a significant result, then

$$k \geq 1 .$$

If H_0 is true in every case, the probability of at least one false-positive result is

$$P(k \geq 1 | \alpha, m_0, H_0) = 1 - (1 - \alpha)^{m_0}$$

e.g. if $\alpha = 0.05$, $m_0 = 30$, then

$$P(k \geq 1 | \alpha = 0.05, m_0 = 30, H_0) = 1 - (1 - 0.05)^{30} = 0.785 .$$

This has caused the 'Replication Crisis'

The replication crisis

- Observations of ‘high rate of nonreplication (lack of confirmation) of research discoveries’
- ‘…increasing concern that in modern research, false findings may be the majority or even the vast majority of published research claims.’
- Argument, on methological grounds, that this must indeed be the case.

Ioannidis, J.P.A. (2005) Why most published research findings are false. *PLoS Medicine* **19**:e1004085. https://doi.org/10.1371/journal.pmed.0020124

- ‘The replication crisis is frequently discussed in relation to psychology and medicine…Data strongly indicate that other natural and social sciences are affected as well.’

https://en.wikipedia.org/wiki/Replication_crisis , accessed 29 September 2025

The Bonferroni correction: an imperfect remedy

If H_0 is true in every case, the probability of at least one false-positive result is

$$\mathrm{P}(k \geq 1 | \alpha, m_0, H_0) = 1 - (1 - \alpha)^{m_0}$$

If m_0 is large and α is small,

$$1 - (1 - \alpha)^{m_0} \approx m_0 \alpha$$

$\Rightarrow$ Replace threshold α with $\frac{\alpha}{m_0}$ (Bonferroni, 1936) .

Then

$$1 - \left(1 - \frac{\alpha}{m_0}\right)^{m_0} \approx m_0 \frac{\alpha}{m_0} = \alpha \ .$$

Threshold is now very stringent. e.g. if α = 0.05, m_0 = 30, then

$$\frac{\alpha}{m_0} = \frac{0.05}{30} = 0.00167 \ .$$

$\Rightarrow$ Severe loss of statistical power.

The confusion matrix: definitions of values

True hypothesis	**Conclusion from test**	
	H_0	H_1
H_0	True negative	False positive (Type I error)
H_1	False negative (Type II error)	True positive

Definition of false-positive and false-discovery rates

Confusion matrix: value definitions

True hypothesis	Conclusion from test	
	H_0	H_1
H_0	True negative	False positive (Type I error)
H_1	False negative (Type II error)	True positive

Confusion matrix: expected counts

True hypothesis	Conclusion from test		
	H_0	H_1	
H_0	$m_0(1 - \alpha)$	$m_0\alpha$	m_0
H_1	$m_1 P(Z < z_\alpha \mid H_1)$	$m_1 \cdot P(Z \geq z_\alpha \mid H_1)$	m_1
	$m - k$	k	m

Specify:

m_0 = No. of tests for which H_0 is true

m_1 = No. of tests for which H_1 is true

$m = m_0 + m_1$

Expected values of rates:

False-positive rate = $\frac{m_0\alpha}{m_0\alpha + m_0(1-\alpha)} = \alpha$

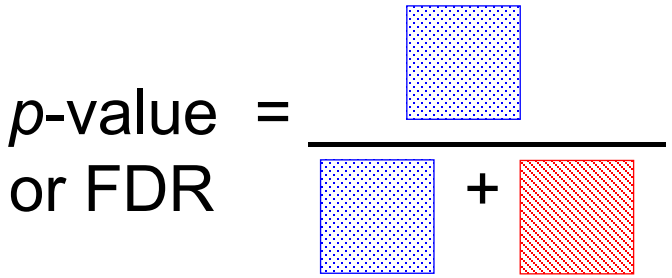
False discovery rate (FDR) = $\frac{m_0\alpha}{m_0\alpha + m_1 \cdot P(Z \geq z_\alpha \mid H_1)} = \frac{m_0\alpha}{k}$

when each significant test result is announced as a ‘discovery’.

N.B. H_1 covers a range of values of Z.
Therefore there is no single value of $P(Z \geq z_\alpha \mid H_1)$

Relationship between *p*-value and FDR

True hypothesis	Conclusion from test	
	H_0	H_1
H_0	True negative	False pos.
H_1		



True hypothesis	Conclusion from test	
	H_0	H_1
H_0		False pos.
H_1		True positive

Consider *p* as an observation of a random variable *P*

$p\text{-value} = P(P \leq p | H_0)$

$FDR = P(H_0 | P \leq p)$

Estimation of the FDR from multiple-testing data

Confusion matrix expected counts

True hypothesis	**Conclusion from test**		
	H_0	H_1	
H_0	$m_0(1 - \alpha)$	$m_0\alpha$	m_0
H_1	$m_1 P(Z < z_\alpha \mid H_1)$	$m_1 \cdot P(Z \geq z_\alpha \mid H_1)$	m_1
	$m - k$	k	m

Following method is
conditional on the independence of the m_0 tests

$$\text{FDR} = \frac{m_0\alpha}{m_0\alpha + m_1 \cdot \mathrm{P}(Z \geq z_\alpha \mid H_1)} = \frac{m_0\alpha}{k}$$

m_0 is unknown.

But

$$m_0 \leq m$$

Hence

$$\text{FDR} \leq \frac{m\alpha}{k} = \frac{\alpha}{k/m}.$$

An upper boundary, obtainable from empirical data

$$\text{FDR} \leq \frac{\alpha}{k/m} = \frac{\text{significance threshold}}{\text{No.of significant test results/total No.of tests}}$$

Control of the FDR: the Benjamini-Hochberg step-down criterion (BH-FDR)

$$\text{FDR} \leq \frac{\alpha}{k/m} \qquad (1)$$

Consider results from a multiple-testing scenario,

p_i , $i = 1 \ldots m$.

Rank the *p*-values in ascending order,

$p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(m)}$.

Specify a value q^* at which the FDR is to be controlled, $0 \leq q^* \leq 1$.

Find the largest value *k* for which

$$\frac{p_{(k)}}{k/m} \leq q^* \; . \qquad (2)$$

Do this by a step-down process, evaluating $\frac{p_{(k)}}{k/m}$ for $k = m$, $k = m - 1$, etc., until the condition is satisfied.

Specify the significance threshold

$$\alpha = p_{(k)} \qquad (3)$$

Then combining (1), (2) and (3),

$$\text{FDR} \leq \frac{p_{(k)}}{k/m} \leq q^*$$

and the FDR is controlled at rate

BH-FDR = q^*

BH-FDR identifies largest set of 'discoveries' possible while controlling FDR.

$\Rightarrow$ BH-FDR maximises statistical power.

(Benjamini and Hochberg, 1995)

Control of the FDR: consequence of correlations between tests

Find the largest value *k* for which

$$\frac{p_{(k)}}{k/m} \leq q^* \ . \qquad (2)$$

Do this by a step-down process, evaluating $\frac{p_{(k)}}{k/m}$ for $k = m$, $k = m - 1$, etc.

- If positive correlations are present between the tests for which H_0 is true, effective number of tests < *m* .
- Value of *k* required to satisfy Inequality (2) tends to be reduced
- $\Rightarrow p_{(k)}$ delivered by step-down process becomes more stringent
- $\Rightarrow$ fewer test results are announced as 'discoveries'
- $\Rightarrow$ the BH-FDR is conservative
    - (like the Bonferroni correction).

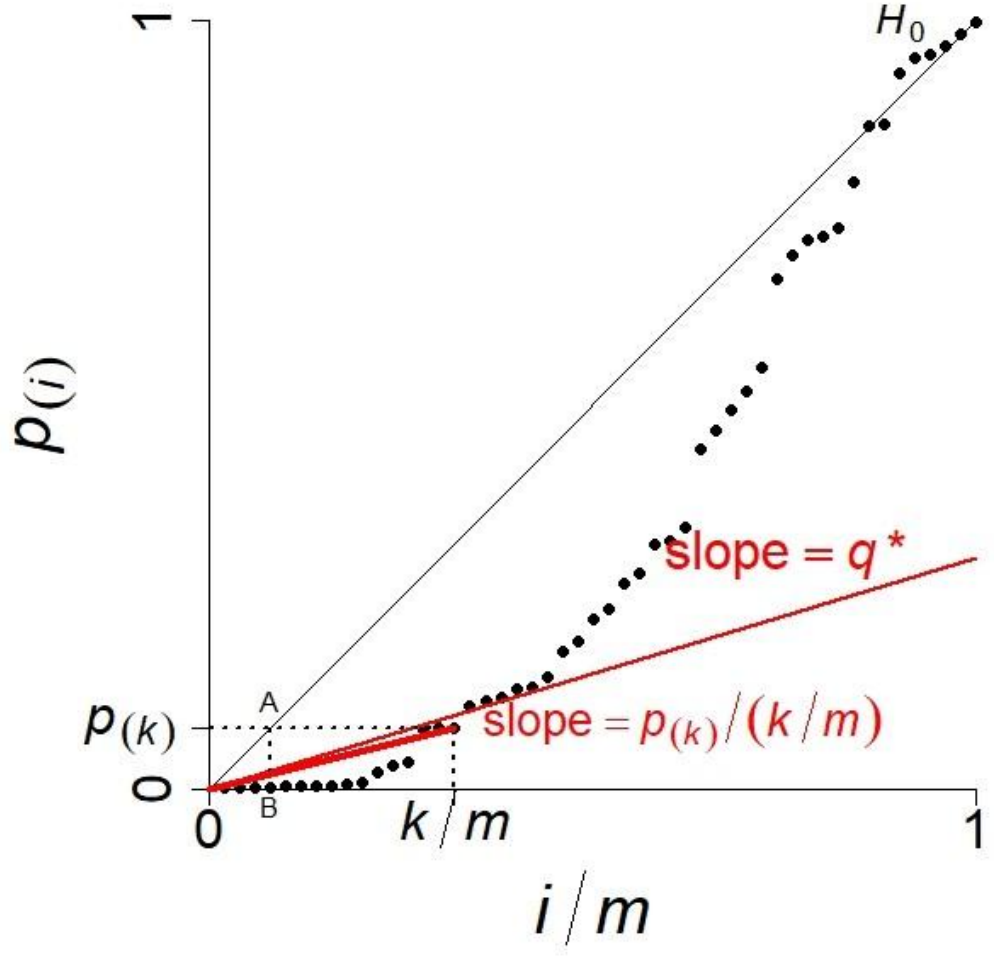
Illustration of the BH-FDR on a quantile-quantile (Q-Q) plot

Points below the line with slope q^* represent tests for which

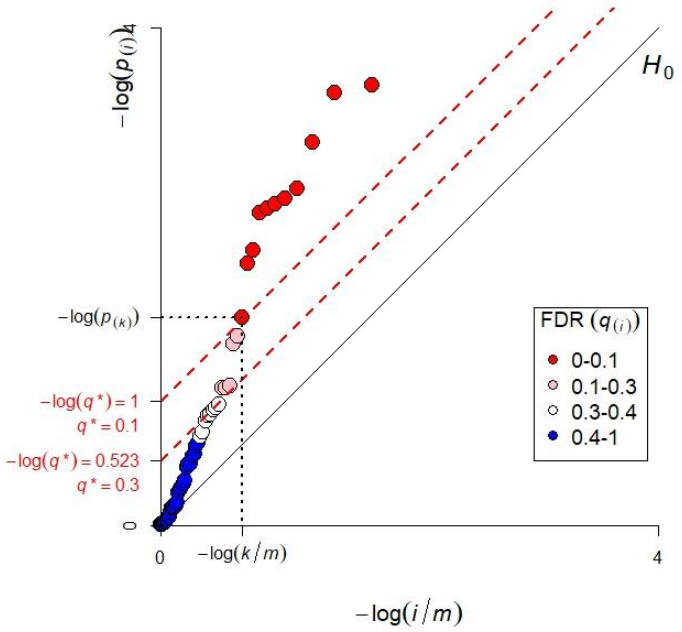
$$\frac{p_{(k)}}{k/m} \leq q^* \ .$$

Points to the left of the line AB indicates the number of significant results expected if H_0 is true for all tests.

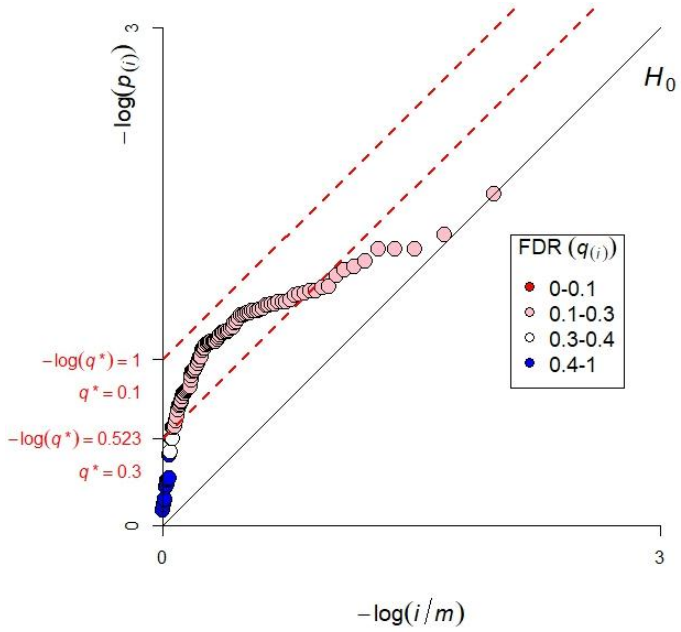
Points between AB and k/m indicate the additional number of additional significant results obtained



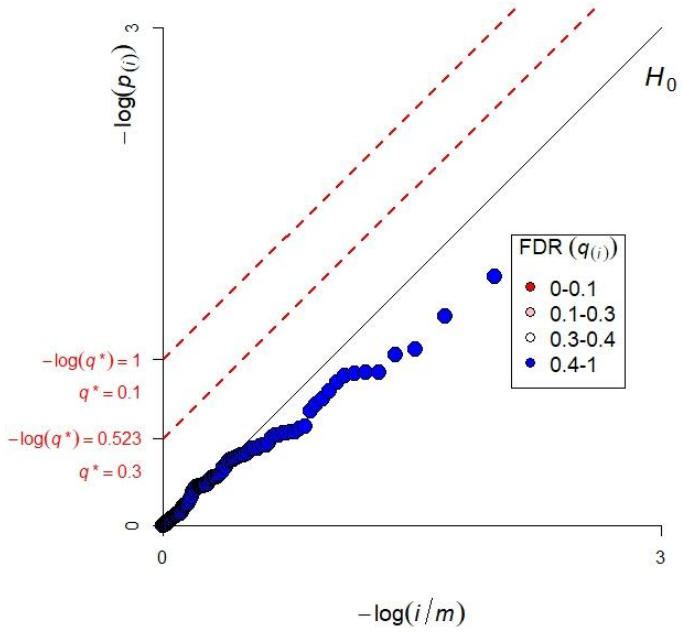
-$\log_{10}$ transformation and colour-coding of the Q-Q plot



H_1 true for some tests,
tests uncorrelated



H_1 true for some tests,
tests positively correlated



H_0 true for all tests,
tests positively correlated

$q_{(i)}$ = smallest value of q^* that causes $p_{(i)}$ to be significant (Storey, 2003).

Relationship between $p_{(i)}$ and $q_{(i)}$ is not strictly monotonic.

i.e. FDR is a feature of a ***set*** of tests.

See backup slides

The BH-FDR makes a prediction

$$\text{FDR} \leq \frac{p_{(k)}}{k/m} \leq q^*$$

If each of the *k* hypotheses that have given significant test results is thoroughly evaluated, the proportion that turn out to have been false positives will be no greater than q^*, on average.

A *p*-value does not intrinsically make a prediction.

Bayesian interpretation of the FDR

- Assertion H: H_0 is true
- Assertion D: $P \leq p_{(k)}$.

Consider p-value as an observation of a random variable P

Bayes theorem:

$$\mathrm{P}(H|D) = \frac{\mathrm{P}(D|H) \cdot \mathrm{P}(H)}{\mathrm{P}(D|H) \cdot \mathrm{P}(H) + \mathrm{P}(D|H') \cdot \mathrm{P}(H')} = \frac{\mathrm{P}(D|H) \cdot \mathrm{P}(H)}{\mathrm{P}(D)}$$

H' = Not H

Apply Bayes theorem to the assertions:

$$\mathrm{P}\left(H_0 \middle| P \leq p_{(k)}\right) = \frac{\mathrm{P}\left(P \leq p_{(k)} \middle| H_0\right) \cdot \mathrm{P}(H_0)}{\mathrm{P}\left(P \leq p_{(k)}\right)}$$

But $\mathrm{P}\left(P \leq p_{(k)} \middle| H_0\right) = p_{(k)}$, $\mathrm{P}\left(P \leq p_{(k)}\right) = k/m$ and $\mathrm{P}(H_0) \leq 1$

Hence

$$\mathrm{P}\left(H_0 \middle| P \leq p_{(k)}\right) \leq \frac{p_{(k)}}{k/m} \leq q^* \text{ =BH-FDR}$$

See backup slides

BH-FDR as an empirical-Bayes method

Bayes theorem:

$$\mathrm{P}(H|D) = \frac{\mathrm{P}(D|H) \cdot \mathrm{P}(H)}{\mathrm{P}(D)}$$

Bayesian formulation of FDR:

$$\mathrm{P}\left(H_0 \left| P \le p_{(k)}\right.\right) = \frac{\mathrm{P}\left(P \le p_{(k)} \middle| H_0\right) \cdot \mathrm{P}(H_0)}{\mathrm{P}\left(P \le p_{(k)}\right)}$$

Key obstacle to application of Bayesian methods is determination of the prior probability P(*H*), and hence P(*D*).

$$\mathrm{P}\left(P \le p_{(k)}\right) = k/m$$

Use only ranks of *p*-values. Consider tests as an exchangeable set.

The BH-FDR is an empirical-Bayes approach to obtaining P(*D*).

The set of *k* significant test results becomes the context for interpreting each individual test result.

Multiplicity ceases to be the problem: it becomes part of the solution.

Review of shrunk estimates

Variation among group means: specification of model

The model:

$Y = \mu + G + E$

where

Y = observable response variable

Y is recorded in each of r units in each of m groups

μ = overall mean
G = group effect
E = residual effect on unit within each group

$G \sim \mathrm{N}(0, \sigma_G)$
$E \sim \mathrm{N}(0, \sigma_E)$

$\bar{Y}_i$ = mean of observations on r units in Group i

Hence

$$\bar{Y}_i \sim \mathrm{N}\left(\mu + g_i, \frac{\sigma_E}{\sqrt{r}}\right)$$

where

g_i = realisation of G in Group i

Group effects, group means: specification of unshrunk and shrunk estimates

$Y = \mu + G + E$

$G \sim \mathrm{N}(0, \sigma_G)$
$E \sim \mathrm{N}(0, \sigma_E)$

$$\bar{Y}_i \sim \mathrm{N}\left(\mu + g_i, \frac{\sigma_E}{\sqrt{r}}\right)$$

where

g_i = realisation of G in Group i

$\bar{y}_{i,\,\mathrm{unshrunk}}$ = observation of $\bar{Y}_i$,
unshrunk estimate of $\mu + g_i$

$\hat{\mu}$ = overall mean of observations on r units in each of m groups

$\hat{g}_{i,\,\mathrm{unshrunk}} = \bar{y}_{i,\,\mathrm{unshrunk}} - \hat{\mu}$,
unshrunk estimate of g_i

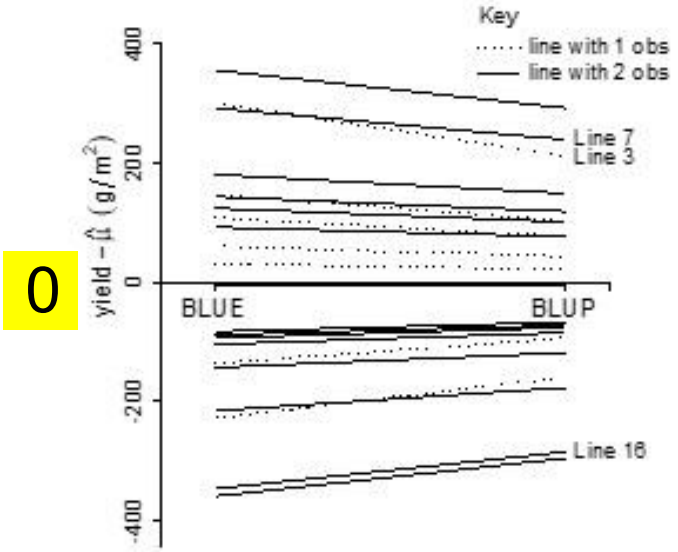
Shrinkage factor = $\dfrac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_E^2}{r}}$

$$\hat{g}_{i,\,\mathrm{shrunk}} = \left(\frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_E^2}{r}}\right) \cdot \hat{g}_{i,\,\mathrm{unshrunk}}$$

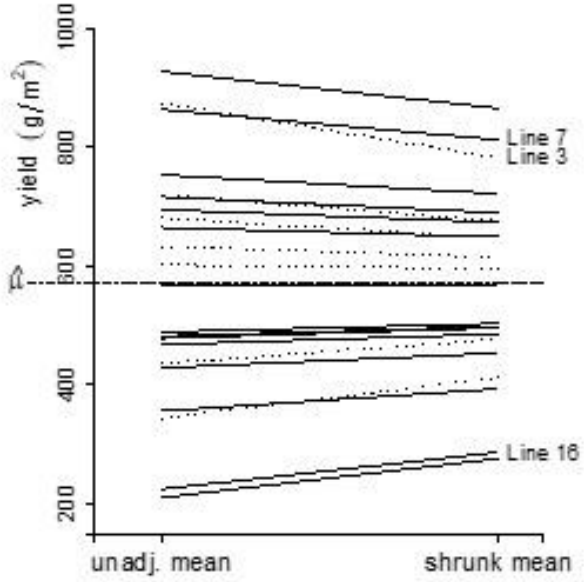
$$\bar{y}_{i,\,\mathrm{shrunk}} = \hat{\mu} + \left(\frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_E^2}{r}}\right) \cdot \hat{g}_{i,\,\mathrm{unshrunk}}$$

$$= \hat{\mu} + \left(\frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_E^2}{r}}\right) \cdot \left(\bar{y}_{i,\,\mathrm{unshrunk}} - \hat{\mu}\right)$$

Group effects, group means: visualisation of unshrunk and shrunk estimates



0



$\hat{\mu}$

$$\hat{g}_{i,\,\text{unshrunk}} = \bar{y}_{i,\,\text{unshrunk}} - \hat{\mu}$$

$$\hat{g}_{i,\,\text{shrunk}} = \left(\frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_E^2}{r}}\right) \cdot \hat{g}_{i,\,\text{unshrunk}}$$

$$\bar{y}_{i,\,\text{unshrunk}}$$

$$\bar{y}_{i,\,\text{shrunk}} = \hat{\mu} + \left(\frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_E^2}{r}}\right) \cdot \left(\bar{y}_{i,\,\text{unshrunk}} - \hat{\mu}\right)$$

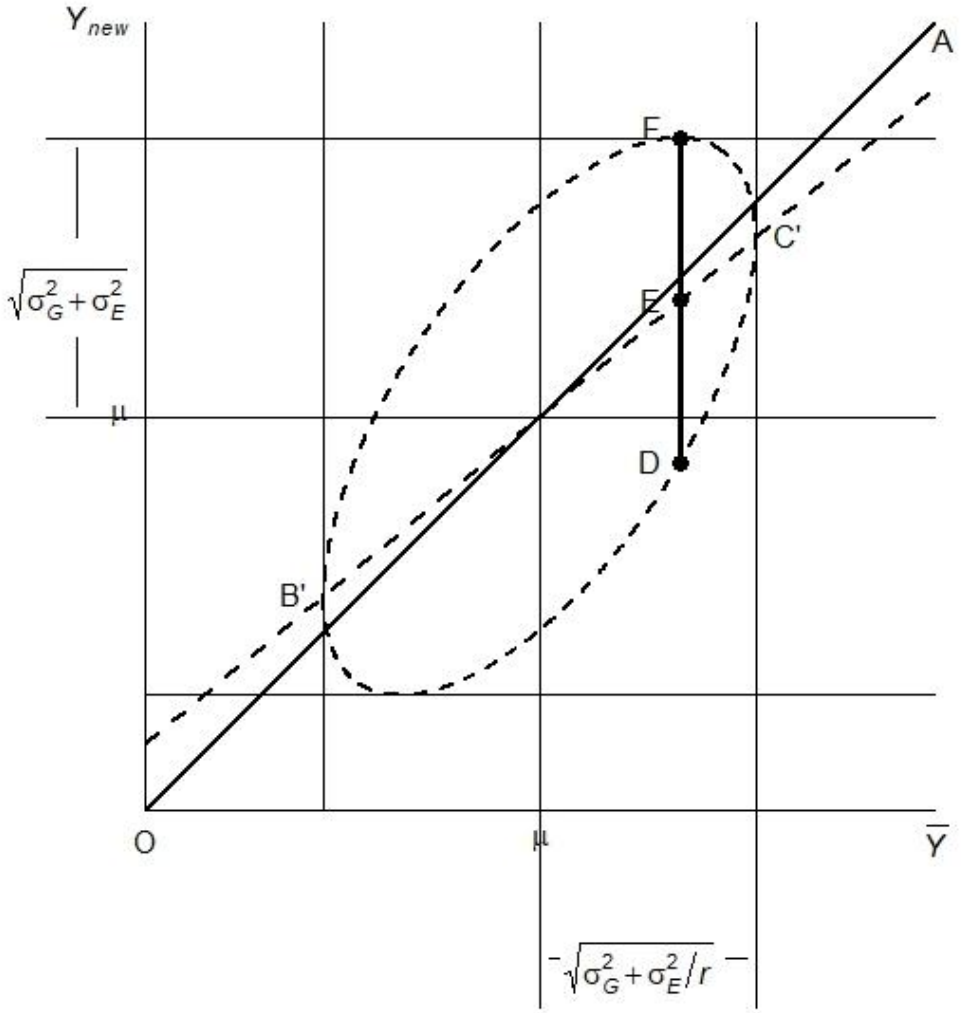
Best Linear Unbiased Estimate (BLUE)

Best Linear Unbiased Predictor (BLUP)

Unshrunk mean

Shrunk mean

$\overline{y_i}$ is a predictor of $y_{i,\mathbf{new}}$; shrinkage and regression towards the mean are equivalent



length(DE) = length(EF)

$\Rightarrow$ shrunk mean is an unbiased predictor of future observations

N.B. Unbiased, conditional on information from all groups observed $(\hat{\mu}, \widehat{\sigma_G^2}, \widehat{\sigma_E^2})$

See backup slides

Shrinkage of estimates as an empirical-Bayes method

Prior distribution of $\mu + g_i$:

$$\mu + g_i \sim \mathrm{N}(\mu, \sigma_G)$$

Distribution of unshrunk estimate of $\mu + g_i$:

$$\bar{Y}_i \sim \mathrm{N}\left(\mu + g_i, \frac{\sigma_E}{\sqrt{r}}\right)$$

Posterior distribution of $\mu + g_i$:

Define weights, $w_{\text{prior}} = \frac{1}{\sigma_G^2}$, $w_{\text{est}} = \frac{1}{\sigma_E^2/r}$

Then $\mu + g_i \sim \mathrm{N}\left(\frac{w_{\text{prior}} \cdot \mu + w_{\text{est}} \cdot \bar{Y}_i}{w_{\text{prior}} + w_{\text{est}}}, \sqrt{\frac{1}{w_{\text{prior}} + w_{\text{est}}}}\right) \leq \frac{\sigma_E}{\sqrt{r}}$

Rearranging the mean of this distribution,

$$\frac{w_{\text{prior}}\mu + w_{\text{est}}\bar{Y}_i}{w_{\text{prior}} + w_{\text{est}}} = \hat{\mu} + \left(\frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_E^2}{r}}\right) \cdot \left(\bar{y}_{i,\,\text{unshrunk}} - \hat{\mu}\right) = \bar{y}_{i,\,\text{shrunk}}$$

Use all $\bar{y}_{i,\,\text{unshrunk}}$, $i = 1 \ldots m$, to estimate μ and σ_G .

Shrunk estimates are an empirical-Bayes approach to obtaining unbiased predictions of group means.

The set of *m* group means becomes the prior context for interpreting each individual mean.

Multiplicity ceases to be the problem: it becomes part of the solution.

The connection between the FDR and shrunk estimates

FDR and shrunk estimates: common features

- Address the Reproducibility Crisis
- Multiplicity ceases to be the problem: becomes part of the solution
- Ask, ‘If we take away what could have happened by chance, what proportion are we left with?’
- Provide predictions/expectations
  - …whereas a *p*-value does ***not*** make a prediction
- Can be understood on an empirical-Bayesian basis
- The *m* tests or estimates provide the empirical prior distribution

See backup slides

Re-parameterise shrunk estimates model

The original model:

$Y_{\text{orig}} = \mu_{\text{orig}} + G_{\text{orig}} + E_{\text{orig}}$

$G_{\text{orig}} \sim \text{N}\left(0, \sigma_{G_{\text{orig}}}\right)$

$E_{\text{orig}} \sim \text{N}\left(0, \sigma_{E_{\text{orig}}}\right)$

Re-parameterised model:

$Z_{\text{obs}} = G + E$

$G \sim \text{N}(\mu_G, \sigma_G)$

$E \sim \text{N}(0,1)$

Hence

$G = \frac{G_{\text{orig}} + \mu_{\text{orig}}}{\sigma_{E_{\text{orig}}}/\sqrt{r}}$, $E = \frac{E_{\text{orig}}}{\sigma_{E_{\text{orig}}}/\sqrt{r}}$

and Z_{obs} is on the scale of the test statistic Z .

Shrinkage factor = $\frac{\sigma_G^2}{\sigma_G^2+1}$

See backup slides

Application of a significance test to group means

Re-parameterised model:

$Z_{obs} = G + E$

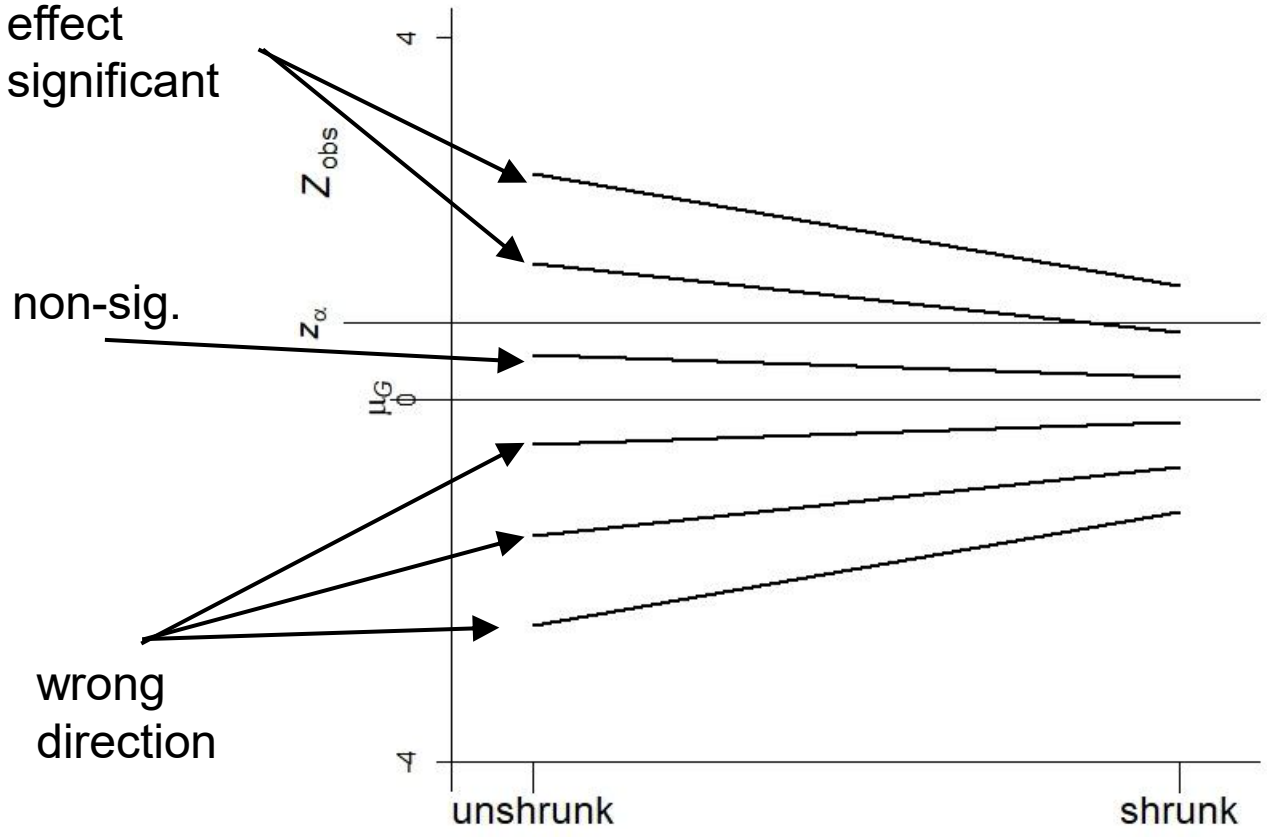
$G \sim \mathrm{N}(\mu_G, \sigma_G)$, $E \sim \mathrm{N}(0,1)$

An example:

$\mu_G = 0$, $\sigma_G = 1$, shrinkage factor $= \frac{1}{1^2+1} = \frac{1}{2}$

H_0: $G \leq 0$
H_1: $G > 0$
N.B. 1-sided test

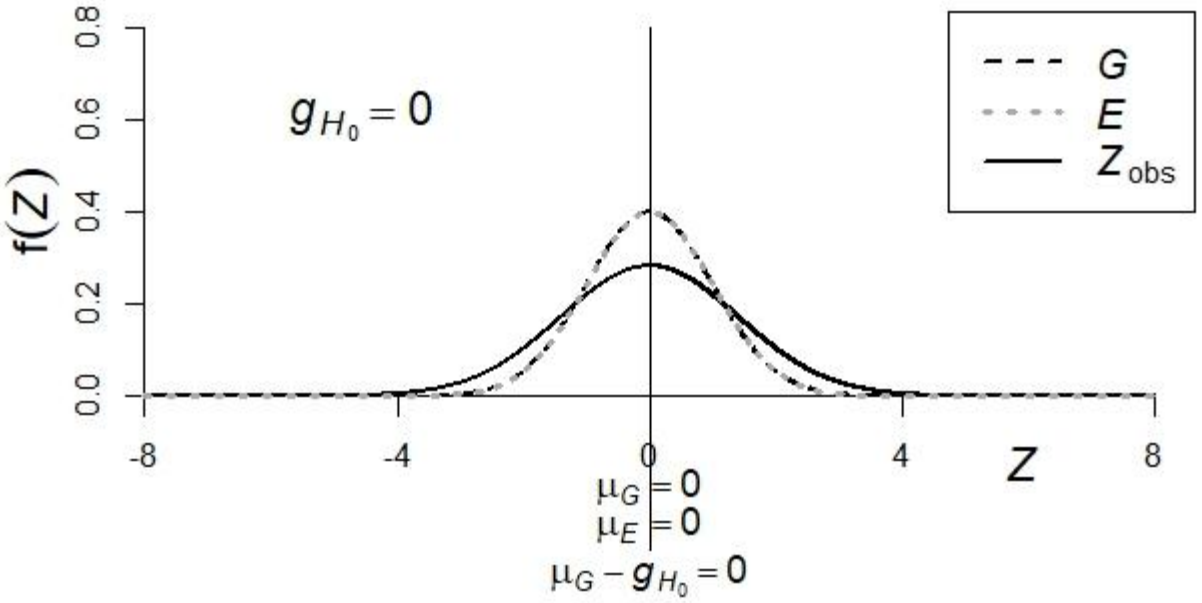
$\alpha = 0.20$ $z_\alpha = 0.842$
N.B. unrealistic, for illustration

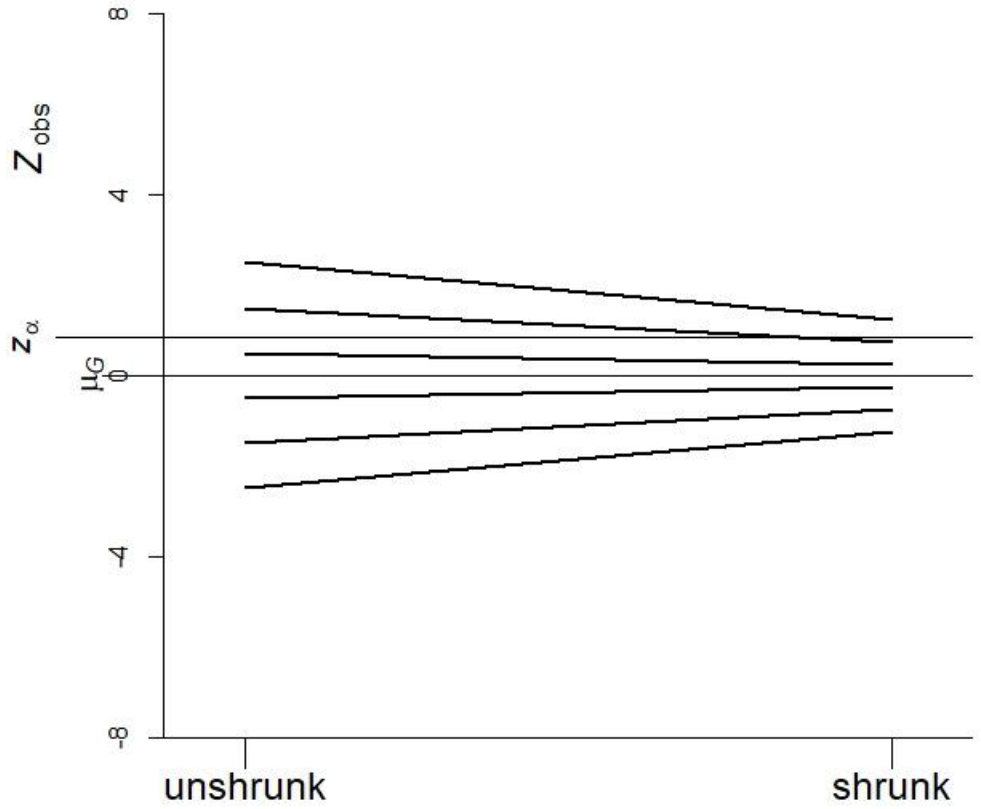


Connection of shrinkage factor to components of Z_{obs}

$\mu_G = 0$, $\sigma_G = 1$, $E \sim \mathrm{N}(0,1)$

shrinkage factor = $\frac{1}{2}$



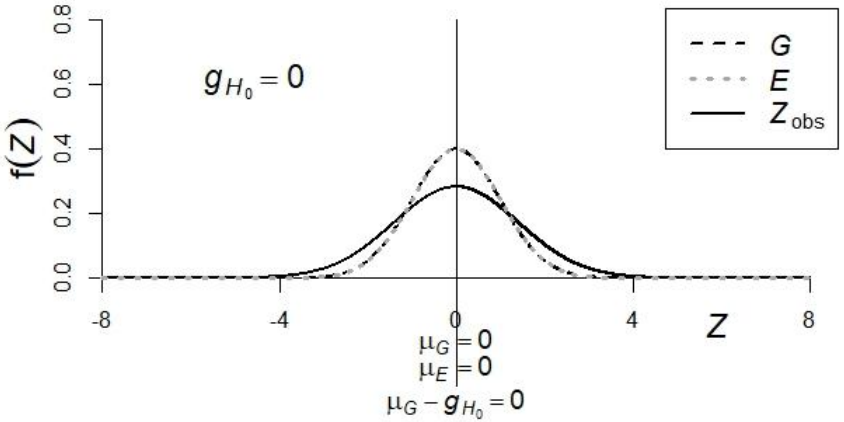


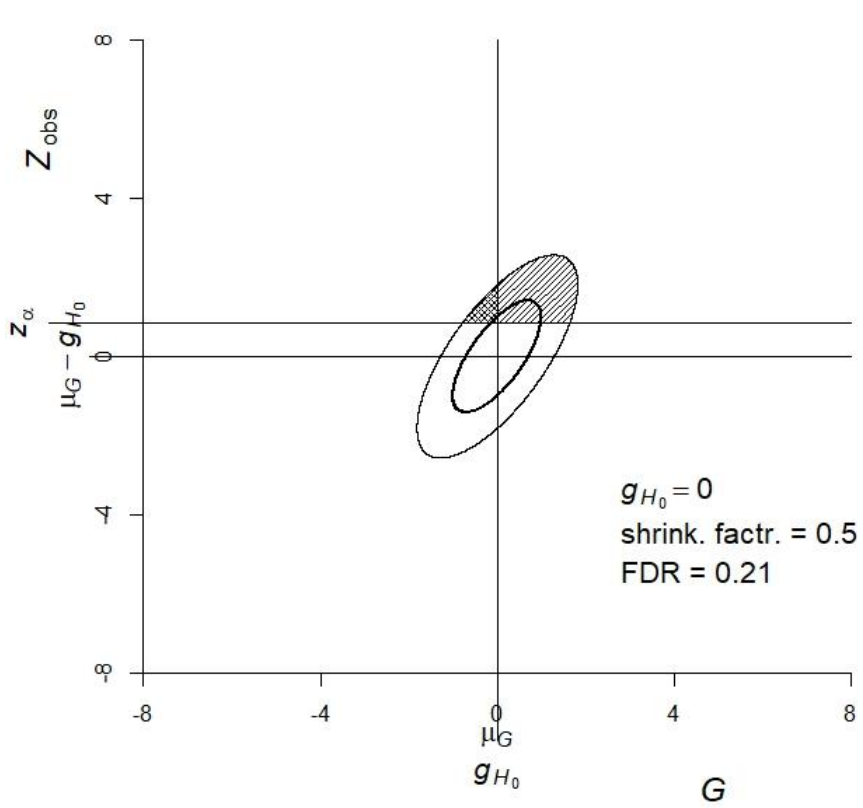
Connection of FDR to shrinkage factor

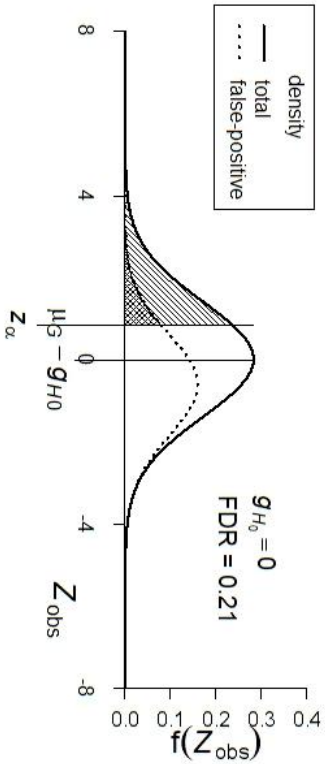
$\mu_G = 0$
$\sigma_G = 1$
shrinkage factor = $\frac{1}{2}$

H_0: $G \leq 0$
H_1: $G > 0$

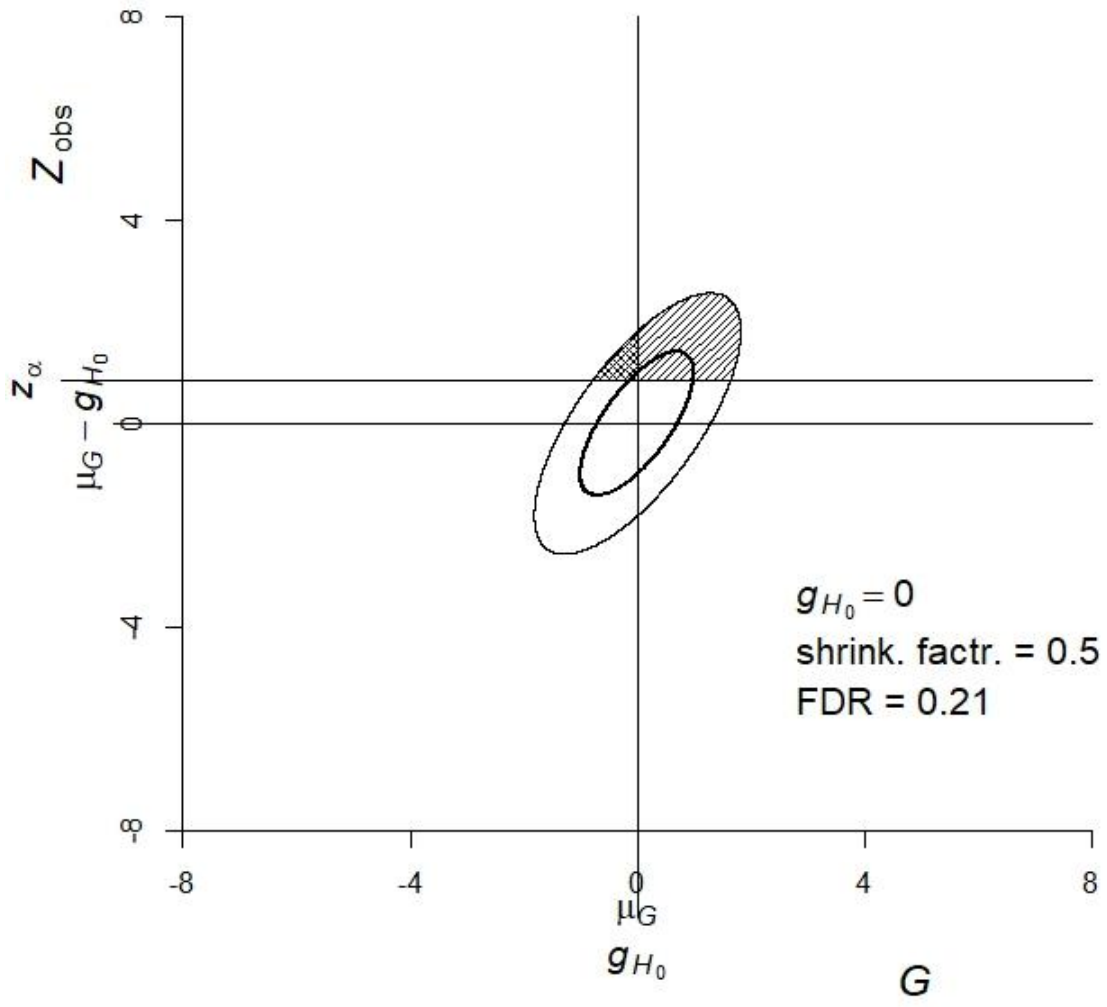
$\alpha = 0.20 \quad z_\alpha = 0.842$







The Z_{obs} vs. G bivariate distribution: the confusion matrix in graphical form

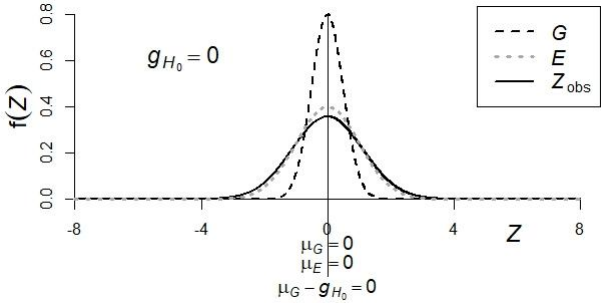


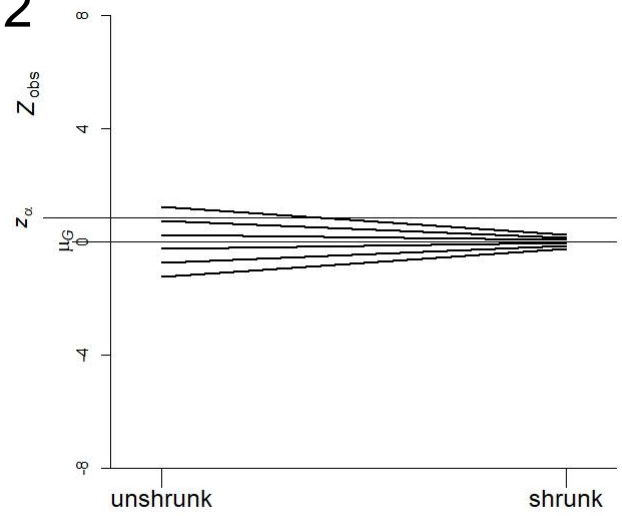
Conclusion from test	**True hypothesis**	
	H_0	H_1
H_1	False positive (Type I error)	True positive
H_0	True negative	False negative (Type II error)

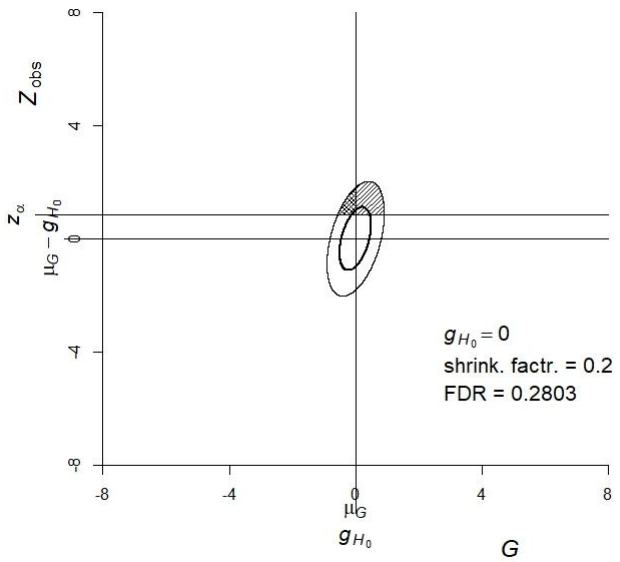
- matrix is rotated
- rows are permuted

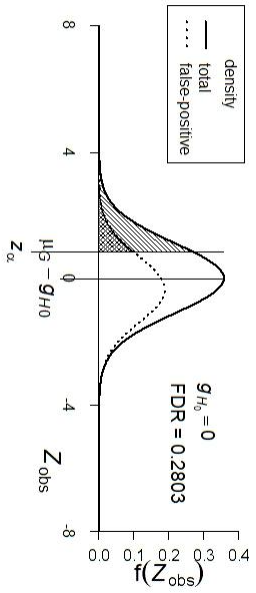
Explore a range of shrinkage factors

σ_G = 0.5, shrinkage factor = 0.2

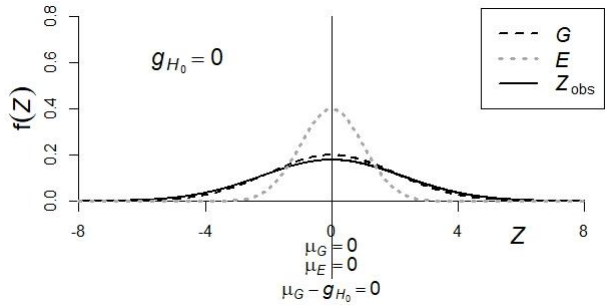


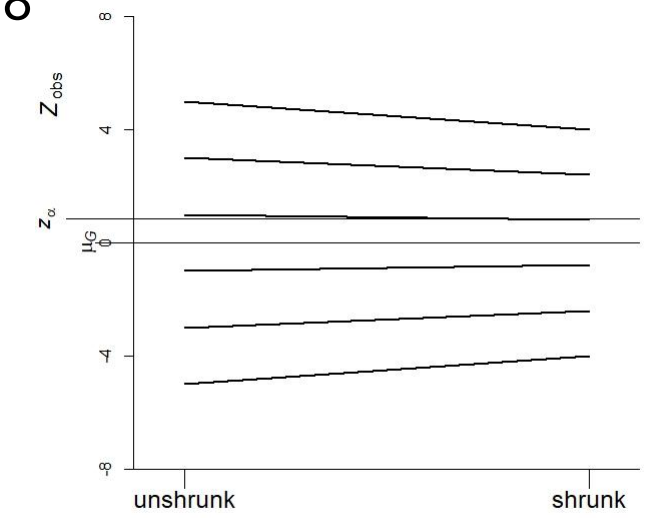


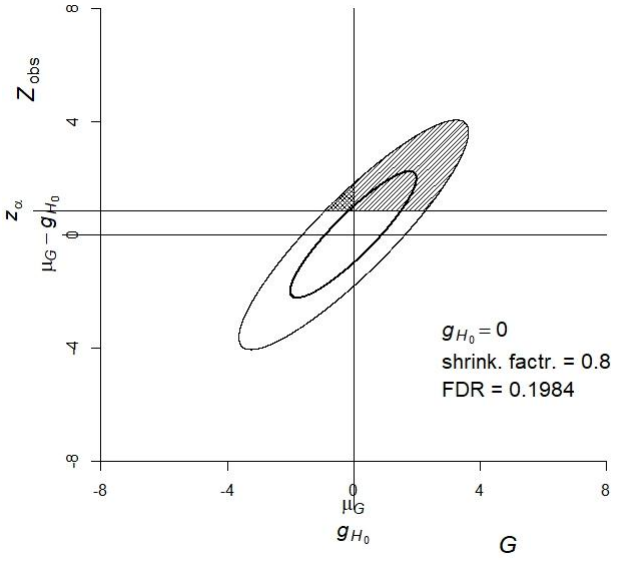


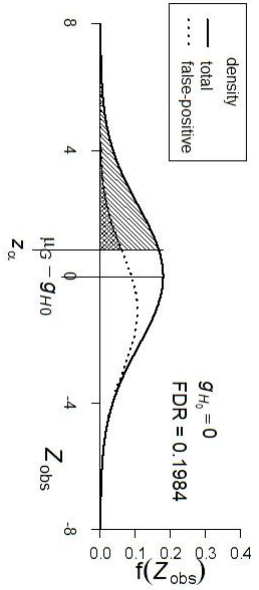


σ_G = 2, shrinkage factor = 0.8



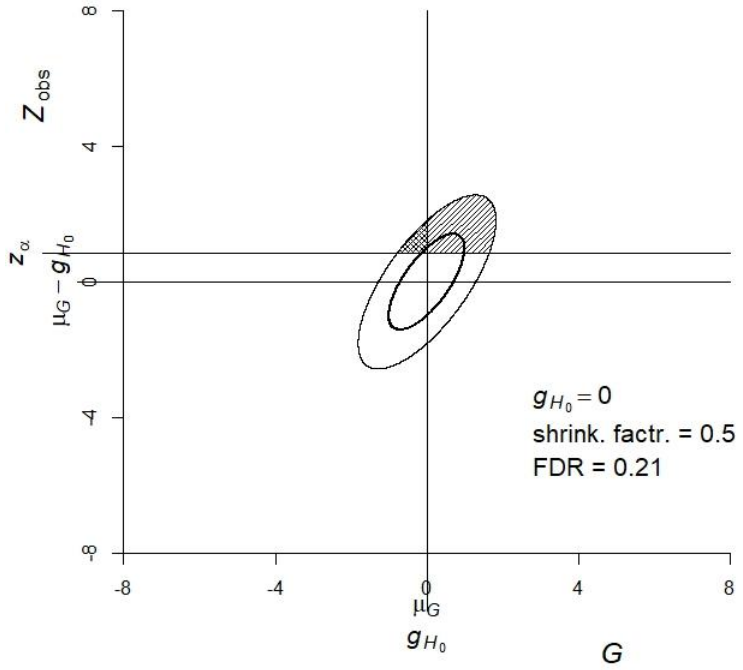


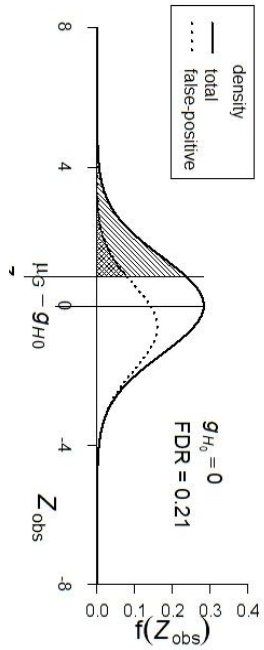


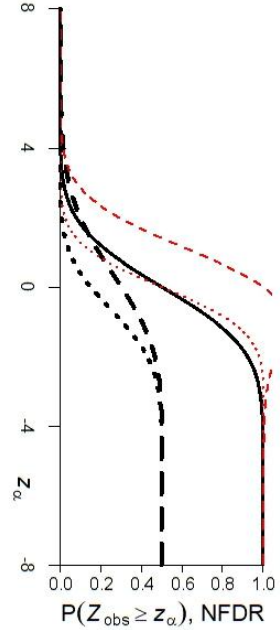


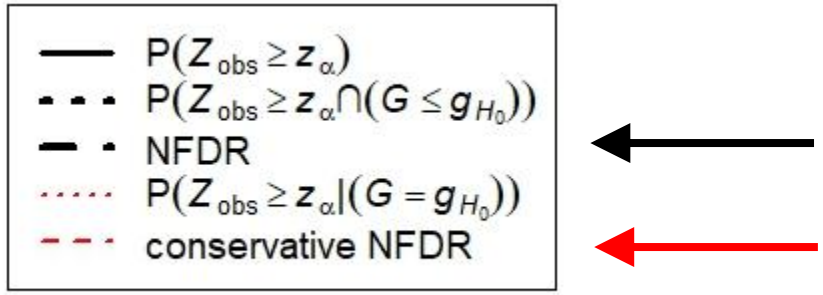
See backup slides

Relationship between FDR and significance threshold









LFDR = 'local FDR', conditional on $Z = z_\alpha$.

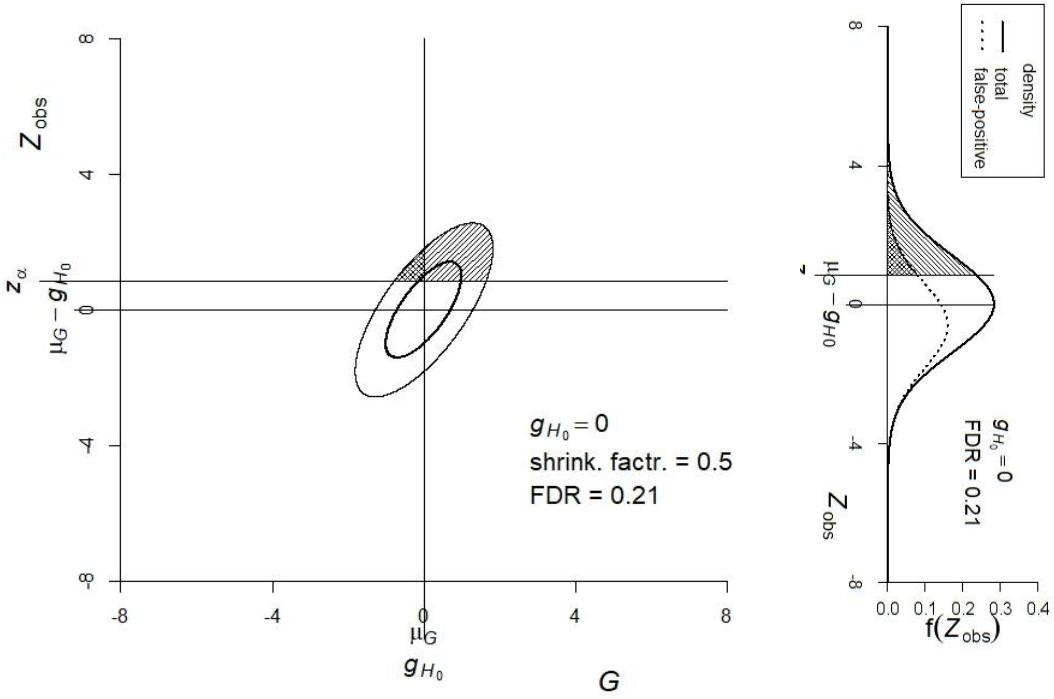
NFDR = 'non-local FDR', over the range $Z \geq z_\alpha$. (Bickel, 2020)

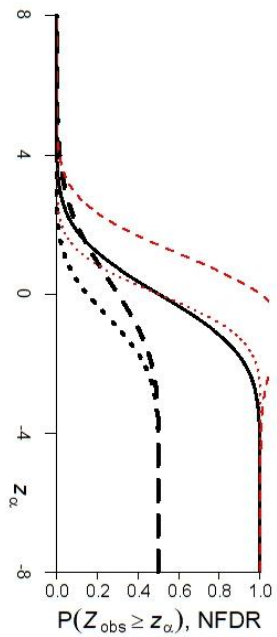
m_0/m is known. H_0: $G \leq g_{H_0}$

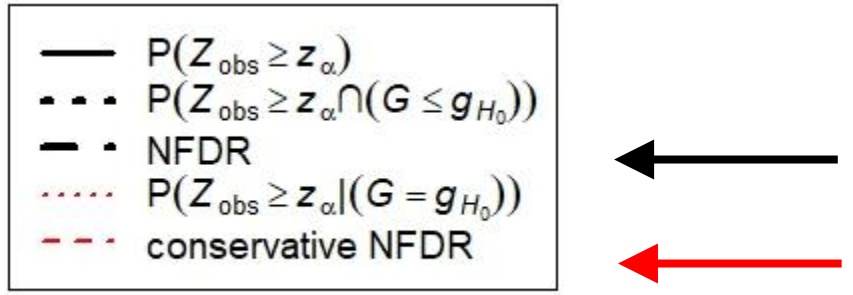
conservative NFDR .
Set m_0/m = 1, H_0: $G = g_{H_0}$.

BH-FDR is an empirical implementation of the conservative NFDR

FDR vs. sig. threshold: dependence on shrinkage factor



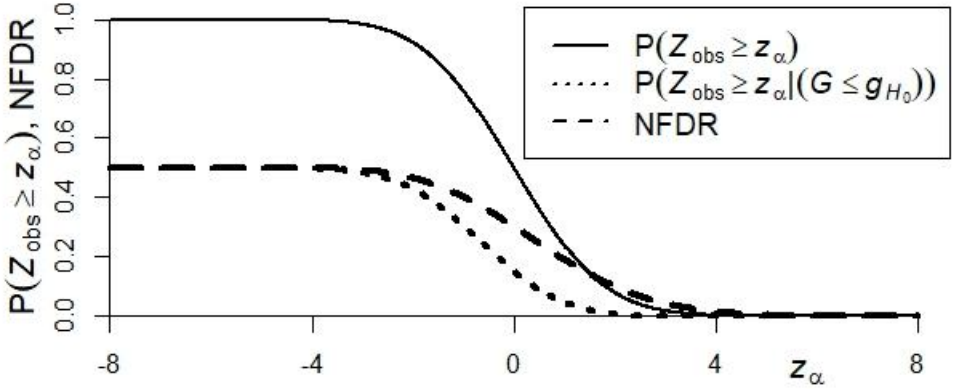




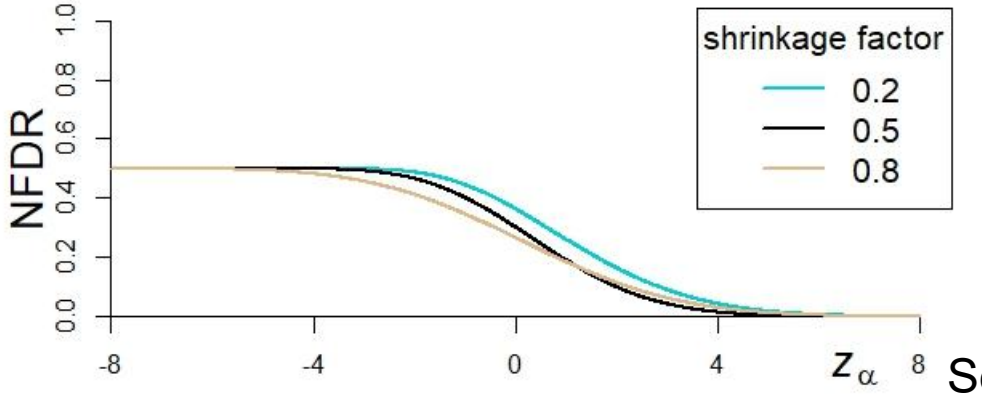
NFDR = 'non-local FDR',
over the range $Z \geq z_\alpha$.
m_0/m is known. H_0: $G \leq g_{H_0}$

conservative NFDR .
Set m_0/m = 1 , H_0: $G = g_{H_0}$.

Rotate and flip to plot *Z* on horizontal axis



Extend to other values of shrinkage factor



See backup slides

Combined application of FDR and shrunk estimates: practical illustration

Application to published experimental data

Previous example	**This example**
m groups of *r* sampling units	single small set of sampling units
single response variable	*m* response variables

See backup slides

A publication reporting gene expression results

Deng J-T, Wang X-L, Chen Y-X, O'Brien ER, Gui Y, Walsh MP (2015) The Effects of Knockdown of Rho-Associated Kinase 1 and Zipper-Interacting Protein Kinase on Gene Expression and Function in Cultured Human Arterial Smooth Muscle Cells. PLoS ONE 10(2): e0116969. doi:10.1371/journal.pone.0116969

Data Availability Statement: The raw data sets for array comparisons have been deposited in the Gene Expression Omnibus website: www.ncbi.nlm.nih.gov/geo/ (accession number: GSE56819).

NCBI Resources How To Reviewer access Sign Out

GEO Home | Documentation | Query & Browse | Email GEO

Gene Expression Omnibus

GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles.

Gene Expression Omnibus

Keyword or GEO Accession Search

Getting Started

- Overview
- FAQ
- About GEO DataSets
- About GEO Profiles
- About GEO2R Analysis
- How to Construct a Query
- How to Download Data

Tools

- Search for Studies at GEO DataSets
- Search for Gene Expression at GEO Profiles
- Search GEO Documentation
- Analyze a Study with GEO2R
- Studies with Genome Data Viewer Tracks
- Programmatic Access
- FTP Site
- ENCODE Data Listings and Tracks

Browse Content

- Repository Browser

DataSets:	4348
Series:	263097
Platforms:	27703
Samples:	8025984

Information for Submitters

- Login to Submit
- Submission Guidelines
- Update Guidelines
- MIAME Standards
- Citing and Linking to GEO
- Guidelines for Reviewers
- GEO Publications

Sample specifications

Sample	Treatment	Title
GSM1369856	Control	siRNA 1
GSM1369857	Control	siRNA 2
GSM1369858	Control	siRNA 3
GSM1369859	ROCK1	siRNA 1
GSM1369860	ROCK1	siRNA 2
GSM1369861	ROCK1	siRNA 3
GSM1369862	ZIPK	siRNA 1
GSM1369863	ZIPK	siRNA 2
GSM1369864	ZIPK	siRNA 3

Expression values

GSM1369856

#ID_REF =	
#VALUE = RMA signal	
ID_REF	VALUE
7892501	7.545518
7892502	4.32842
7892503	4.168302
7892504	9.103029
7892505	4.599297

33,297 rows

GSM1369857

#ID_REF =	
#VALUE = RMA signal	
ID_REF	VALUE
7892501	5.080564
7892502	3.893911
7892503	4.329082
7892504	9.094555
7892505	4.345451

9 samples

...

GSM1369864

#ID_REF =	
#VALUE = RMA signal	
ID_REF	VALUE
7892501	6.65136
7892502	4.722325
7892503	3.762711
7892504	9.089062
7892505	3.735544

N.B. Probeset IDs are not collated

Probeset specifications

ID	adj.P.Val	P.Value	t	B	logFC	Gene.symbol	Gene.title
8027448	0.000553	1.66E-08	-3.48E+01	7.76898	-2.67	DPY19L3	dpy-19 like 3 (C. elegans)
8085033	0.00068	4.08E-08	-3.02E+01	7.466	-2.94	LMLN	leishmanolysin like peptidase
8022441	0.000882	7.95E-08	2.72E+01	7.20076	3.54	ROCK1	Rho associated coiled-coil containing protein kinase 1

33,297 rows

ID_REF	GSM1369856	GSM1369857	GSM1369858	GSM1369859	GSM1369860	GSM1369861	GSM1369862	GSM1369863	GSM1369864
	Control	Control	Control	ROCK1	ROCK1	ROCK1	ZIPK	ZIPK	ZIPK
7892501	7.545518	5.080564	6.675656	6.233815	7.140542	6.609992	6.174588	6.321959	6.65136
7892502	4.32842	3.893911	4.077985	4.617835	4.629821	4.277309	5.117608	4.241955	4.722325
7892503	4.168302	4.329082	4.180284	4.179679	3.866266	3.937363	4.059075	3.995274	3.762711
7892504	9.103029	9.094555	9.264637	8.849144	9.211445	8.83327	8.898712	9.165777	9.089062
7892505	4.599297	4.345451	3.511323	3.0886	4.118563	4.001799	4.393135	3.600587	3.735544

33,297 rows

For further details of data extraction and collation, see backup slides

Analysis of a representative probeset

Probeset specification

ID	Gene.symbol	Gene.title
7984779	PML	promyelocytic leukemia

Expression values

	Control			ROCK1			ZIPK		
Probeset	GSM1369856	GSM1369857	GSM1369858	GSM1369859	GSM1369860	GSM1369861	GSM1369862	GSM1369863	GSM1369864
7984779	9.450409	9.348948	9.393126	9.775353	9.603642	9.920068	9.146811	9.627286	9.3111

Anova

Source of variation	DF	MS	*F*	*p*
Treatments	2	0.1505	5.1718	0.0495
Residuals	6	0.0291		

$r = 3$

Means

Treatments	Mean	SE
Control	9.3975	0.0985
ROCK1	9.7664	0.0985
ZIPK	9.3617	0.0985
Grand mean	9.5085	

$$\mathrm{SE}_{\mathrm{mean}} = \sqrt{\frac{\widehat{\sigma_E^2}}{r}} = \sqrt{\frac{\mathrm{MS}_{\mathrm{Resid}}}{r}} = \sqrt{\frac{0.0291}{3}} = 0.0985$$

Contrasts

Contrast	Estimate	SE	DF	*t*	*p*(two-sided)
ROCK1.v.control	0.3689	0.1393	6	2.648	0.0381
ZIPK.v.control	-0.0358	0.1393	6	-0.260	0.8060
ROCK1.v.ZIPK	0.4046	0.1393	6	2.905	0.0272

$$\mathrm{SE}_{\mathrm{cont}} = \sqrt{\frac{2\widehat{\sigma_E^2}}{r}} = \sqrt{\frac{2 \times \mathrm{MS}_{\mathrm{Resid}}}{r}} = \sqrt{\frac{2 \times 0.0291}{3}} = 0.1393$$

ROCK1 vs. ZIPK. δ = true difference. One-sided tests.

$$t = \frac{0.4046}{0.1393} = 2.905$$

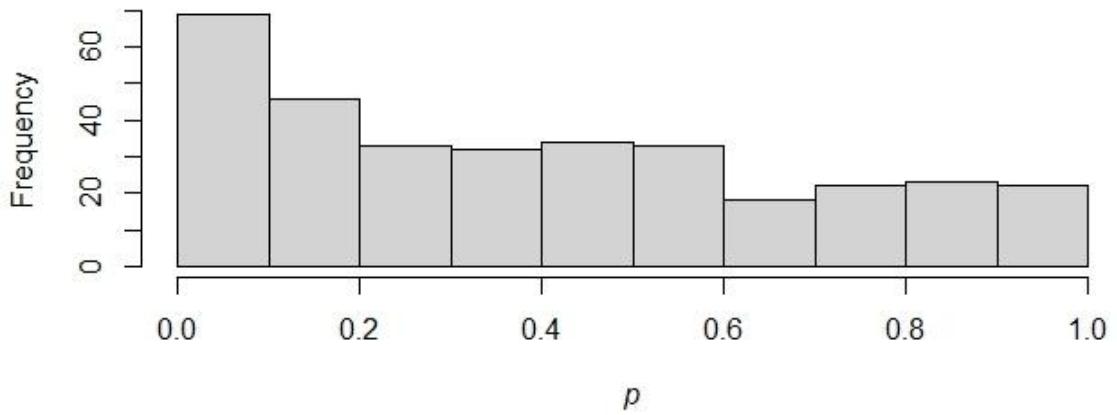
DF = 6

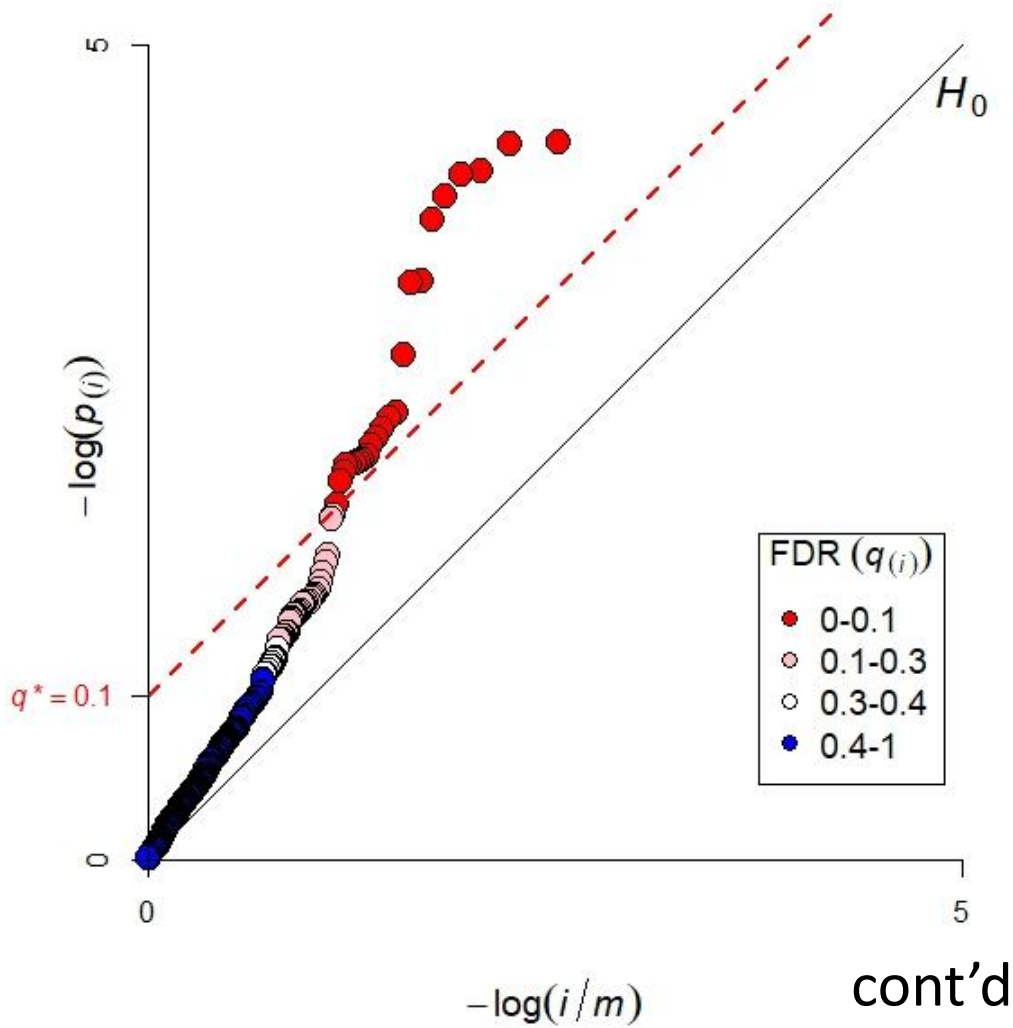
H_0: $\delta \leq 0$; H_1: $\delta > 0$. $p = 0.01358$

H_0: $\delta \geq 0$; H_1: $\delta < 0$. $p = 0.98642$

Analysis of every 100th probeset: 2-sided test

m = 332 representative probesets

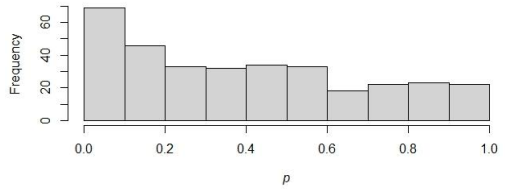


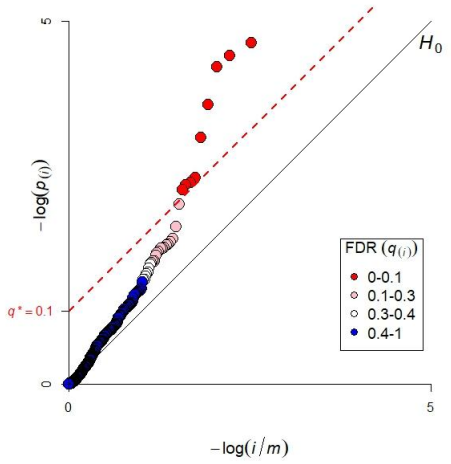


cont'd…/

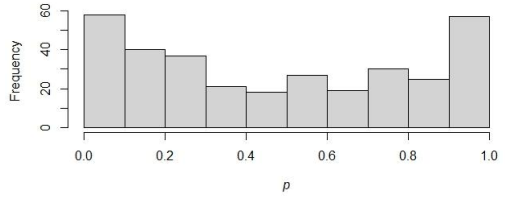
Analysis of every 100th probeset:
two 1-sided tests

2-sided



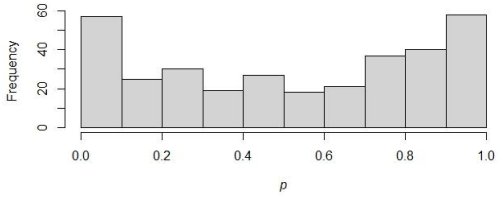


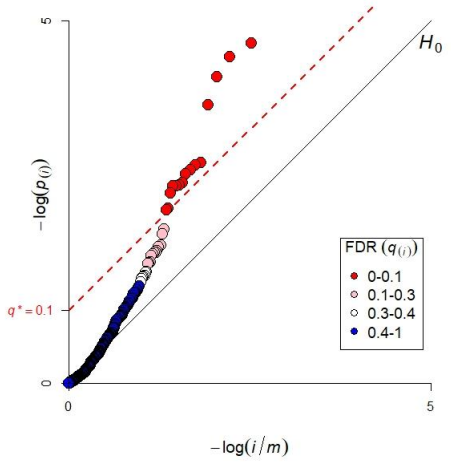
Upper tail



Two 1-sided

Lower tail





Obtain effect estimates (unshrunk and shrunk) from each set of 1-sided tests.

cont'd…/

Application of shrinkage to multiple tests of a contrast

For 1-sided upper-tail test of *i*th hypothesis,

$$Z|H_0 \sim \mathrm{N}(0,1)$$

$$p_i = \mathrm{P}\left(Z \middle| H_0 > z_i\right)$$

Assume $\mathrm{E}(Z) = 0$ whenever H_0 is true

where

p_i = observed *p*-value

z_i = value of Z_{obs}

for i^{th} test.

Obtain z_i by back-transformation from p_i .

If H_0 is true for all tests,

$$Z_{\mathrm{obs}} \sim \mathrm{N}(0,1) \, .$$

whence the appropriate shrinkage factor is

$$\mathrm{S.F.} = \frac{\widehat{\mathrm{var}}(Z_{\mathrm{obs}}) - 1}{\widehat{\mathrm{var}}(Z_{\mathrm{obs}})}$$

where $\widehat{\mathrm{var}}(Z_{\mathrm{obs}})$ is obtained over all z_i , $i = 1 \ldots m$.

Apply S.F. to each z_i, then multiply by $\mathrm{SE}_{\mathrm{cont},\, i}$ to back-transform from the scale of *Z* to the scale of the effect estimate:

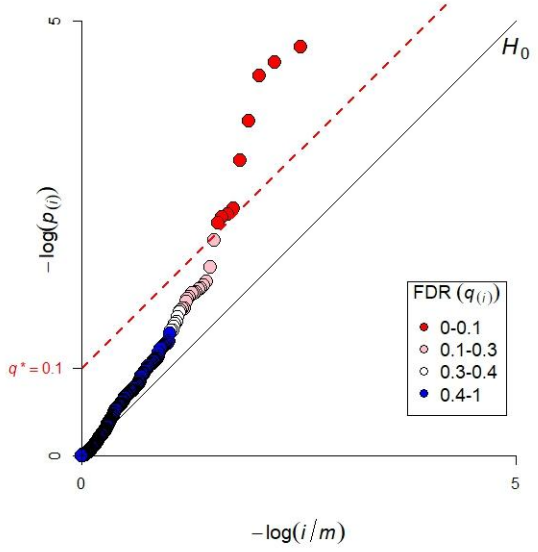
$$\mathrm{Est}_{\mathrm{shrunk},i} = \mathrm{S.F.} \times z_i \times \mathrm{SE}_{\mathrm{cont},i}$$

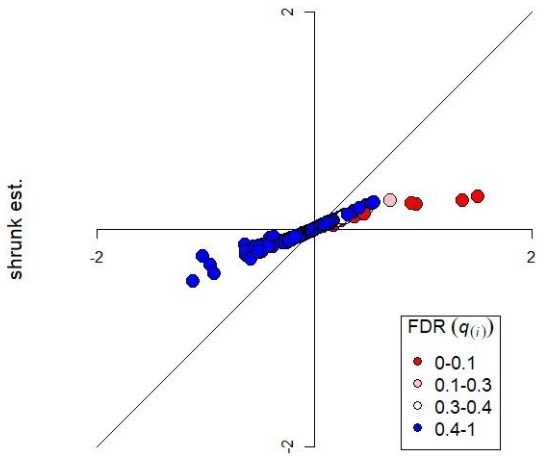
??

Analysis of every 100th probeset:

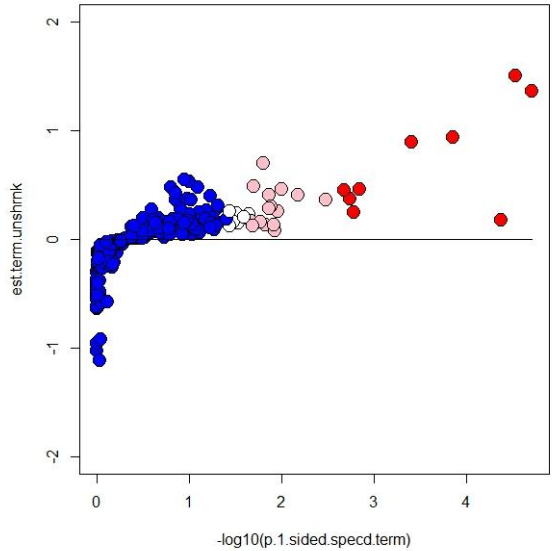
two 1-sided tests and shrunk estimate

Upper tail

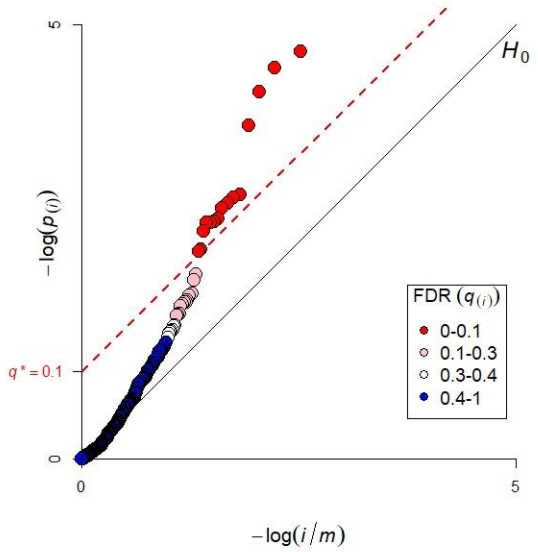


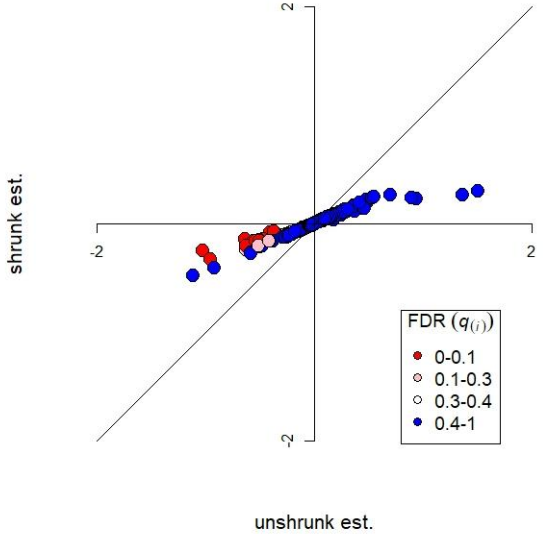


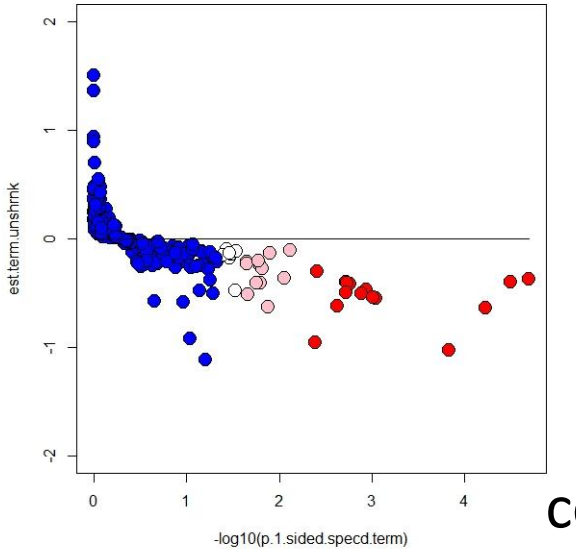
$r_{\text{shrunk. unshrunk}} = 0.962$



Lower tail





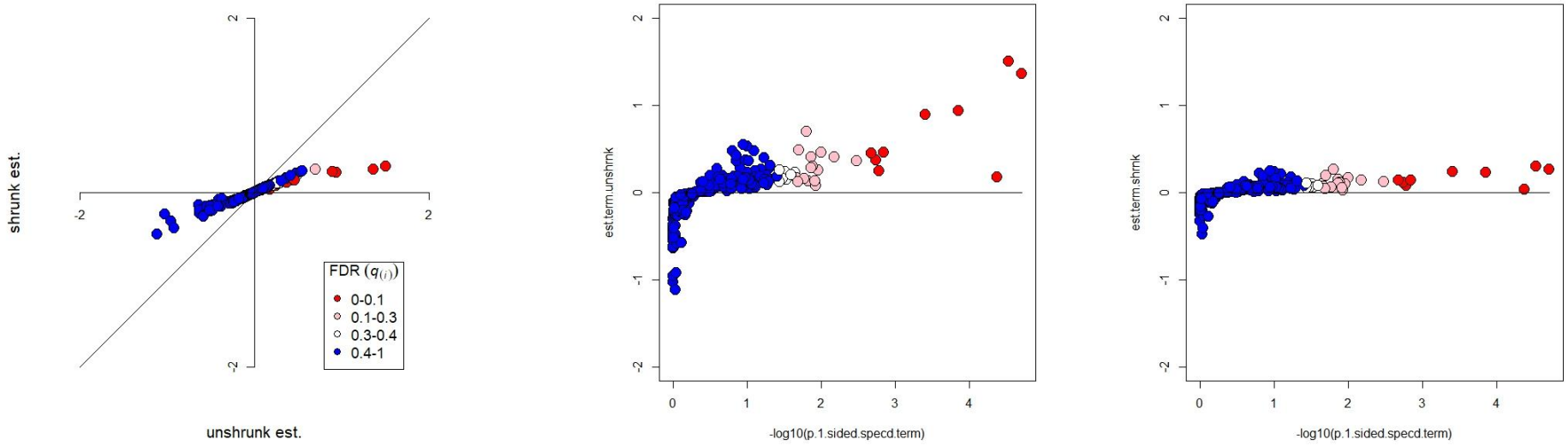


cont'd.../

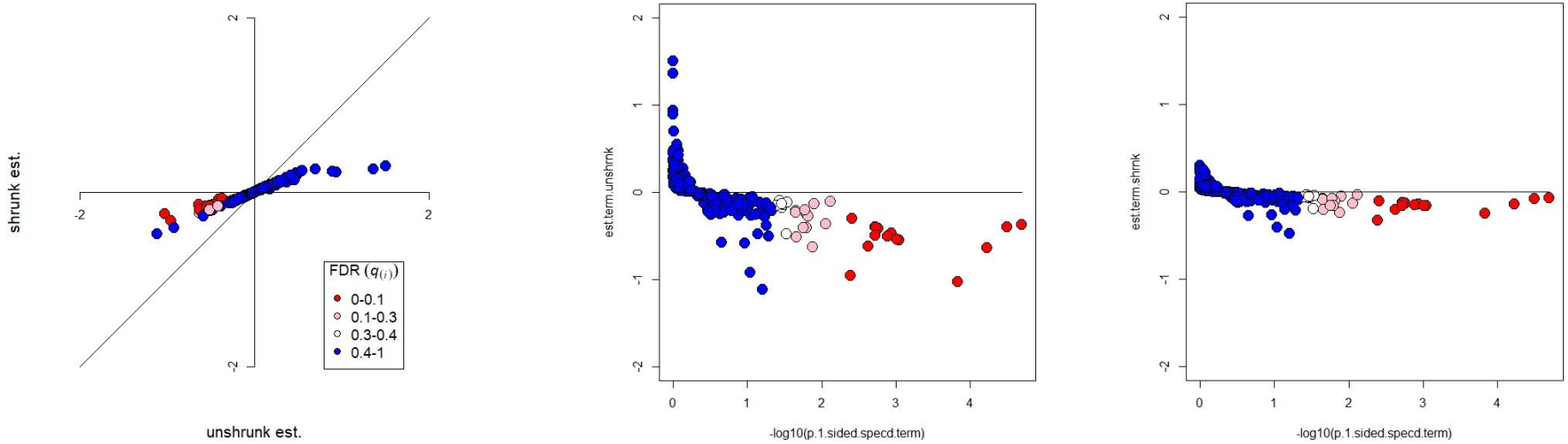
Analysis of every 100th probeset/…cont'd.

Two 1-sided tests

Upper tail



Lower tail



Application to simulated data

- Shrunk estimate achieves reduced MSE (when σ_G^2 is small) by 'borrowing' information between tests.
- But this assumes that σ_E^2 is the same for all tests (no heteroskedasticity).
- Methods taking account of heteroskedasticity have been developed, e.g. function `voom()` in R package `"limma"` (Law et al., 2014).
- The approach presented here can be applied to effect estimates and *p*-values obtained from such methods.

See backup slides

Conclusions

Conclusions

- You have to live with the culture of a discipline.
- A research community may be committed to significance tests when shrunk estimates would be more appropriate.
- A move from Bonferroni-corrected *p*-values to BH-FDR values will then bring them closer to an optimum interpretation of their data…
- …while keeping them in a hypothesis-testing framework.
- But why not present shrunk estimates too?
- Shrunk estimates and the FDR have shared merits:
    - they address the Replication Crisis
    - multiplicity becomes part of the solution
    - they make predictions
    - they have an empirical-Bayesian interpretation.

References

Benjamini, Y., and Hochberg, Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B* **57**:289-300.

Bickel, D.R. (2020) *Genomic Data Analysis. False Discovery Rates and Empirical Bayes Methods.* Boca Raton, Florida: CRC Press. 121 pp.

- Section 3.3. Nonlocal and local false discovery rates.

Bonferroni, C. E. (1936) Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze.*

Cule et al. (2011) Significance testing in ridge regression for genetic data. *BMC Bioinformatics* **12**:372. doi:10.1186/1471-2105-12-372

Deng, J.-T., Wang, X.-L., Chen, Y.-X., O'Brien, E.R., Gui, Y. and Walsh, M.P. (2015) The effects of knockdown of rho-associated kinase 1 and zipper-interacting protein kinase on gene expression and function in cultured human arterial smooth muscle cells. *PLoS ONE* **10**:e0116969. doi:10.1371/journal.pone.0116969

cont'd.../

References/…cont'd.

Efron, B. (2010). *Large-Scale Inference. Empirical Bayes Methods for Estimation, Testing, and Prediction.* (Institute of Mathematical Statistics Monographs, Series Number 1). Cambridge, UK: Cambridge University Press. 276pp. ISBN-10: 0521192498; ISBN-13: 978-0521192491.

Galwey, N.W. (2014) *Introduction to Mixed Modelling. Beyond Regression and Analysis of Variance.* 2nd edition. Chichester, UK: Wiley. 487 pp.

- Chapter 5. Estimation of random effects in mixed models: Best Linear Unbiased Predictions (BLUPs).

Galwey, N.W. (2023) A Q-Q plot aids interpretation of the false discovery rate. *Biometrical Journal* **65**. https://doi.org/10.1002/bimj.202100309

cont'd…/

References/…cont'd.

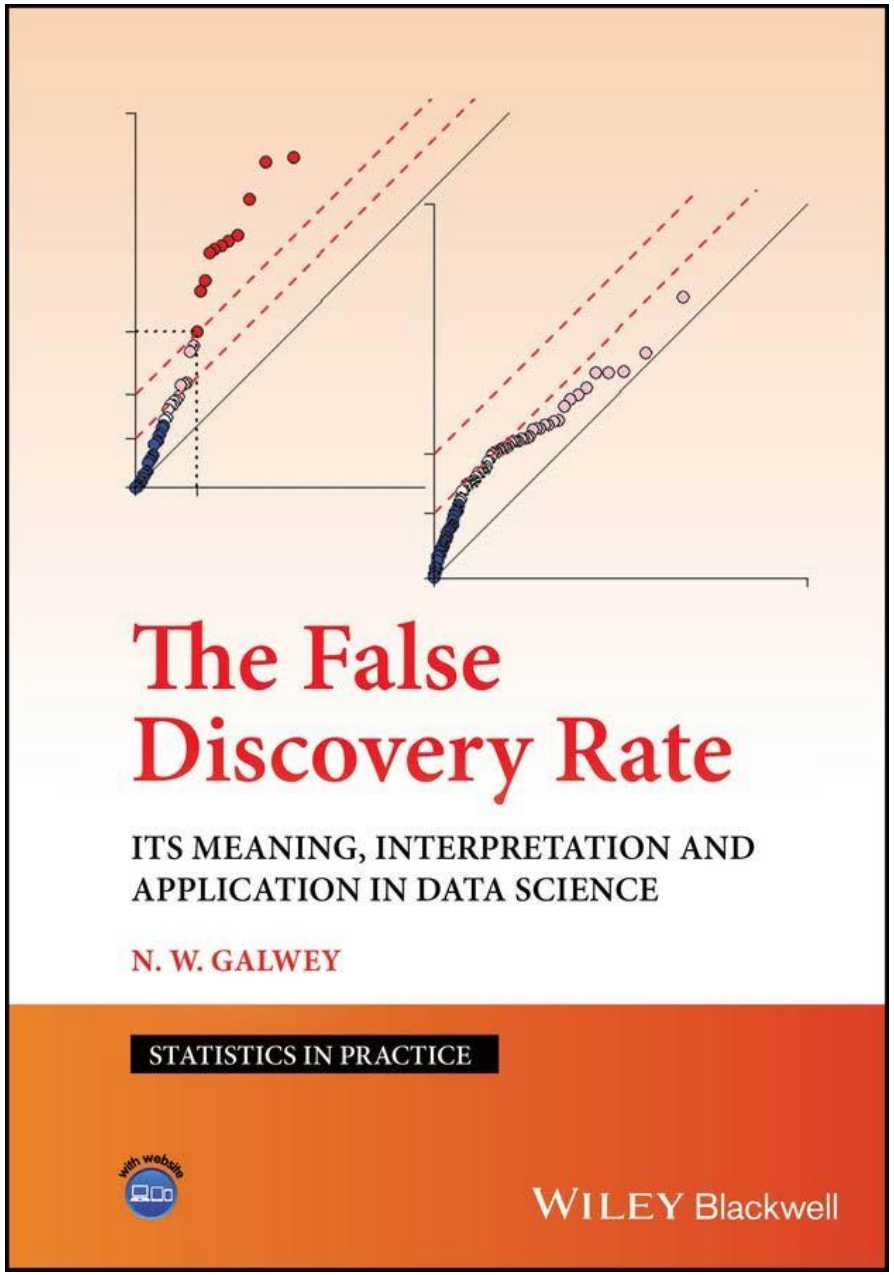
Galwey, N.W. (2025) *The False Discovery Rate: Its Meaning, Interpretation and Application in Data Science*. Chichester, UK: Wiley. 266pp. ISBN 9781119889779.

- Chapter 2. The Meaning of the False Discovery Rate (FDR).
- Chapter 6. The FDR in the context of Bayesian statistics.

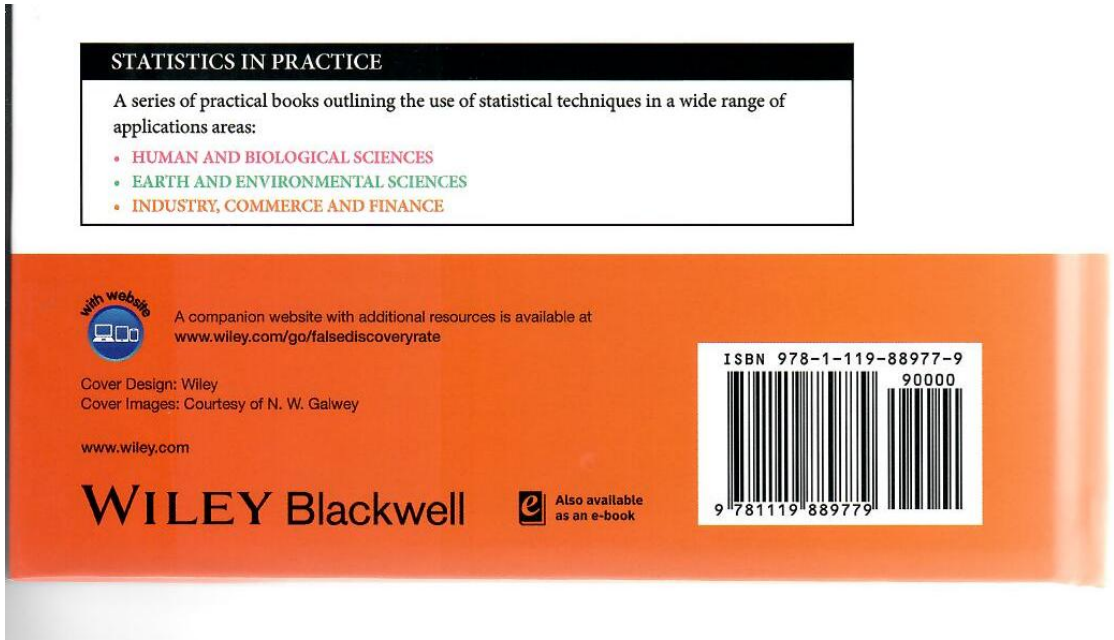
Ioannidis, J.P.A. (2005) Why most published research findings are false. *PLoS Medicine* **19**:e1004085. https://doi.org/10.1371/journal.pmed.0020124

Law, C.W., Chen, Y., Shi, W. and Smyth, G.K. (2014) Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology* **15**:R29. http://genomebiology.com/2014/15/2/R29

Storey, J.D. (2003) The positive false discovery rate: a Bayesian interpretation and the *q*-value. *The Annals of Statistics* **31**:2013-2035.



Galwey, N.W. (2025) *The False Discovery Rate: Its Meaning, Interpretation and Application in Data Science*. Chichester, UK: Wiley. 266pp. ISBN 9781119889779.

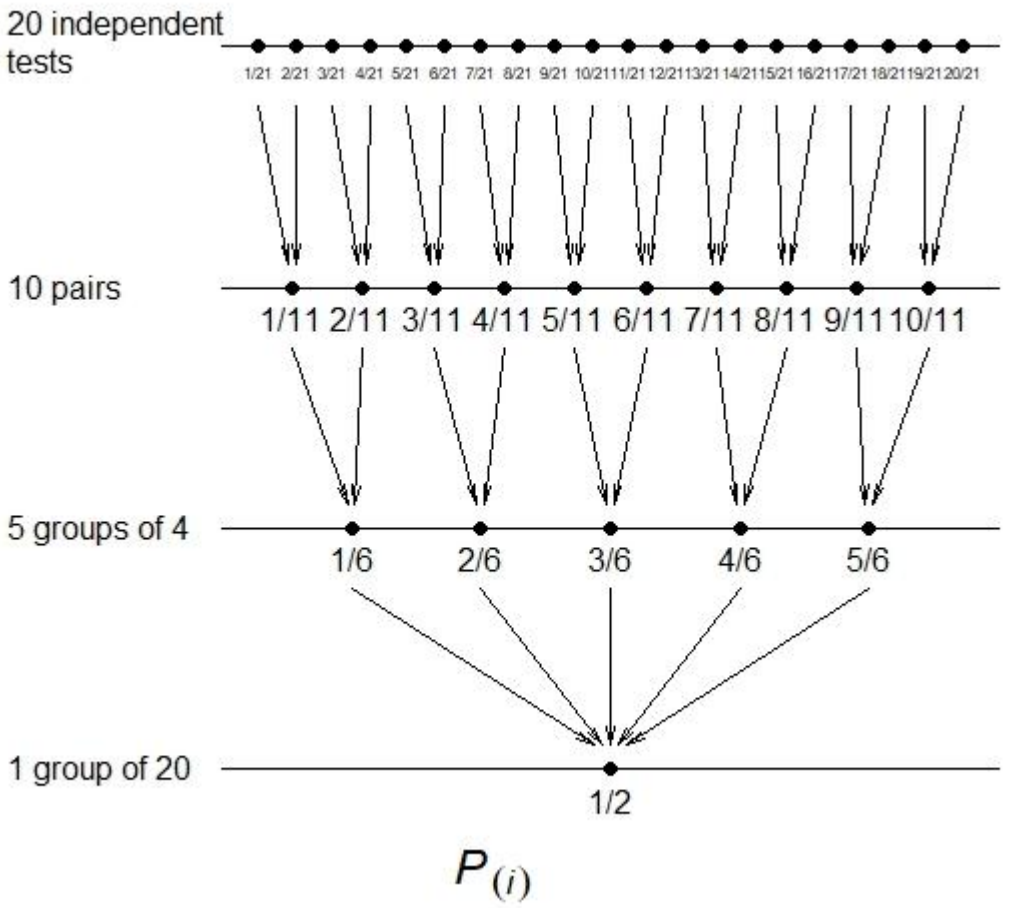


Backup slides

Correlation between tests; $-\log_{10}$ transformation and colour-coding of the Q-Q plot

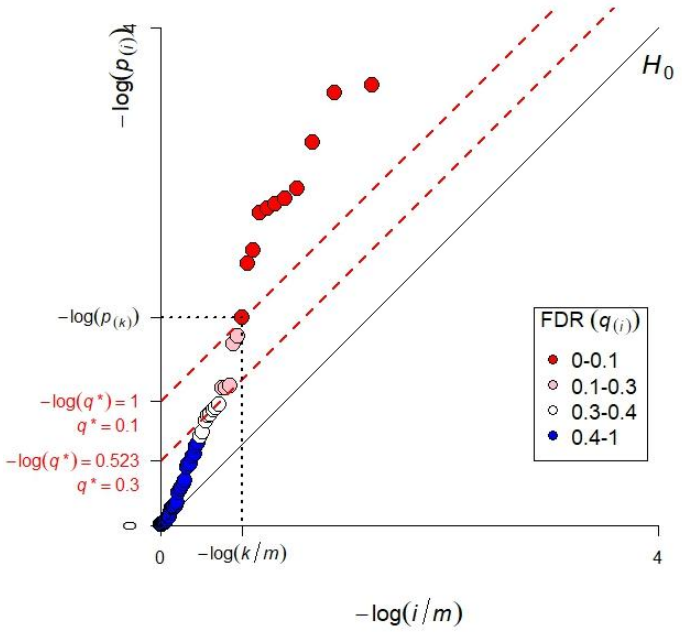
An example of correlation between tests

m tests comprise a smaller number of independent groups. Tests within each group are equivalent.

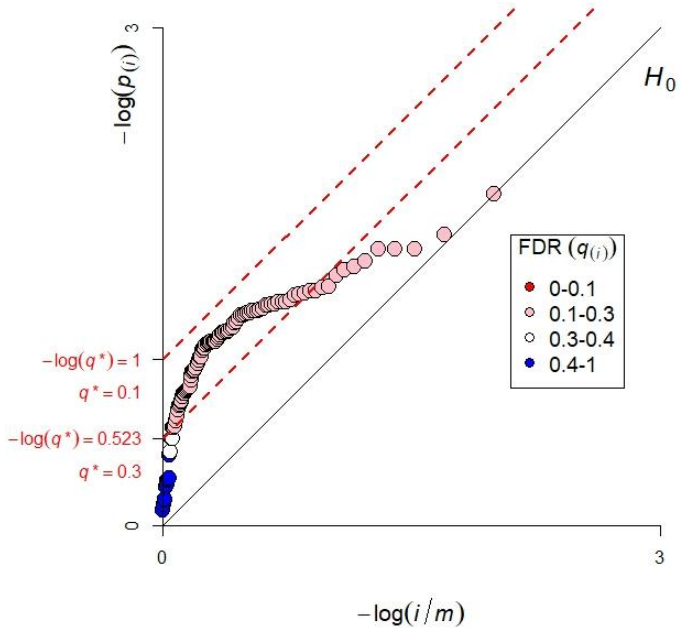


$\Rightarrow$ Few very small and very large values of $P_{(i)}$

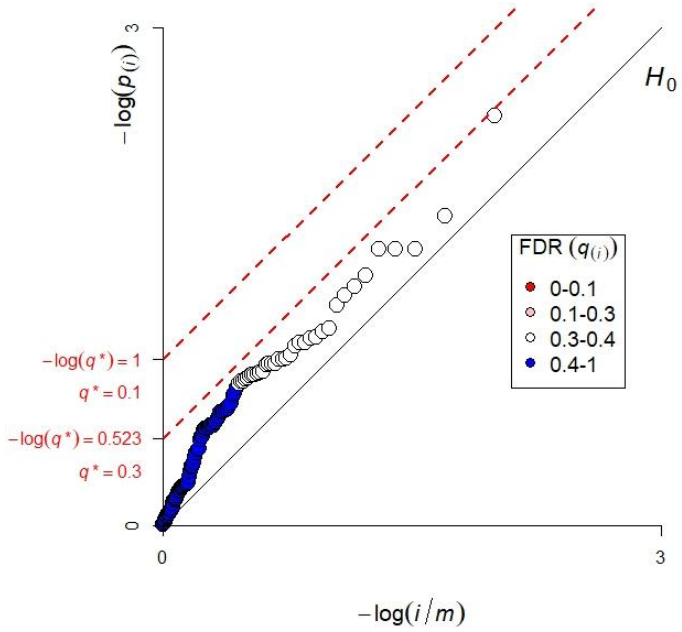
$-\log_{10}$ transformation and colour-coding of the Q-Q plot



H_1 true for some tests, tests uncorrelated



H_1 true for some tests, tests positively correlated



H_1 true for some tests, but weak effects and slight positive correlation

Additional set of conditions

$q_{(i)}$ = smallest value of q^* that causes $p_{(i)}$ to be significant.

Relationship between $p_{(i)}$ and $q_{(i)}$ is not strictly monotonic.

i.e. FDR is a feature of a ***set*** of tests.

Consequence of correlations between tests: Bayesian interpretation

Consequence of correlations between tests: Bayesian interpretation

BH criterion for control of FDR at level q^* :

$$\frac{p_{(k)}}{k/m} \leq q^* \qquad (1)$$

Unpackage $p_{(k)}$:

$$\mathrm{P}\left(P \leq p_{(k)} \middle| H_0\right) = p_{(k)}$$

This is true for all m_0 tests for which H_0 is true **only if** these tests are mutually independent.

If, for these tests, variables P are positively correlated, then (when $p_{(k)} < 0.5$)

$$\mathrm{P}\left(P \leq p_{(k)} \middle| H_0\right) \leq p_{(k)}$$

and q^* is conservative: Inequality (1) controls the FDR at a more stringent level.

Alternatively, set significance threshold at a level α independent of the data…

cont'd.../

Consequence of correlations between tests: Bayesian interpretation/…cont'd.

Alternatively, set significance threshold at a level α **independent of the data**

$$\frac{\alpha}{k/m} = q \qquad (1)$$

Unpackage α :

$$\mathrm{P}\left(P \leq \alpha \middle| H_0\right) = \alpha$$

This is true for all m_0 tests for which H_0 is true **even if** these tests are **not** mutually independent.

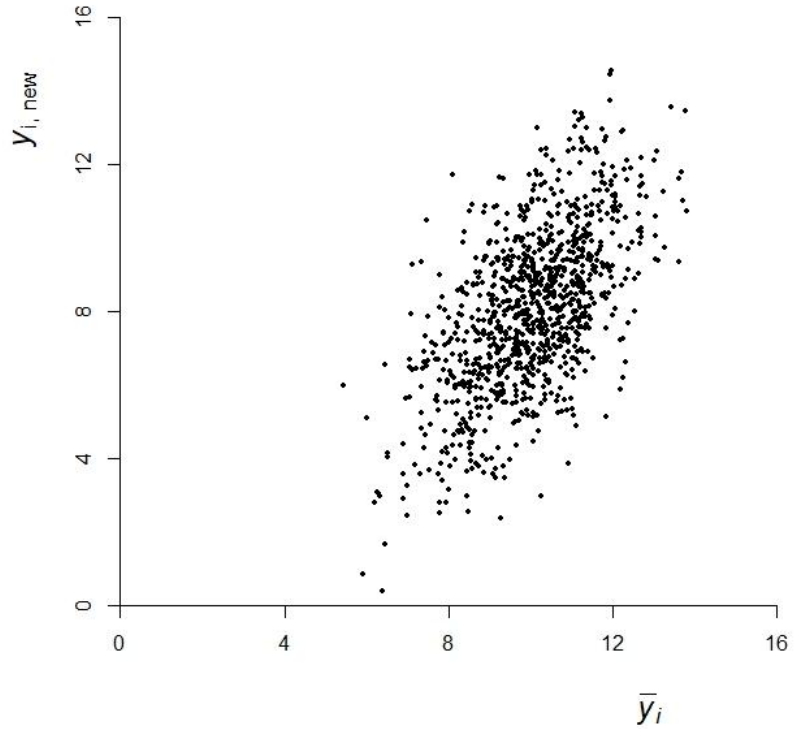
$\Rightarrow$ q is neither conservative nor anti-conservative: Equation (1) controls the FDR at this level…

…but statistical power is not maximised.

Shrinkage and regression towards the mean: the relationship

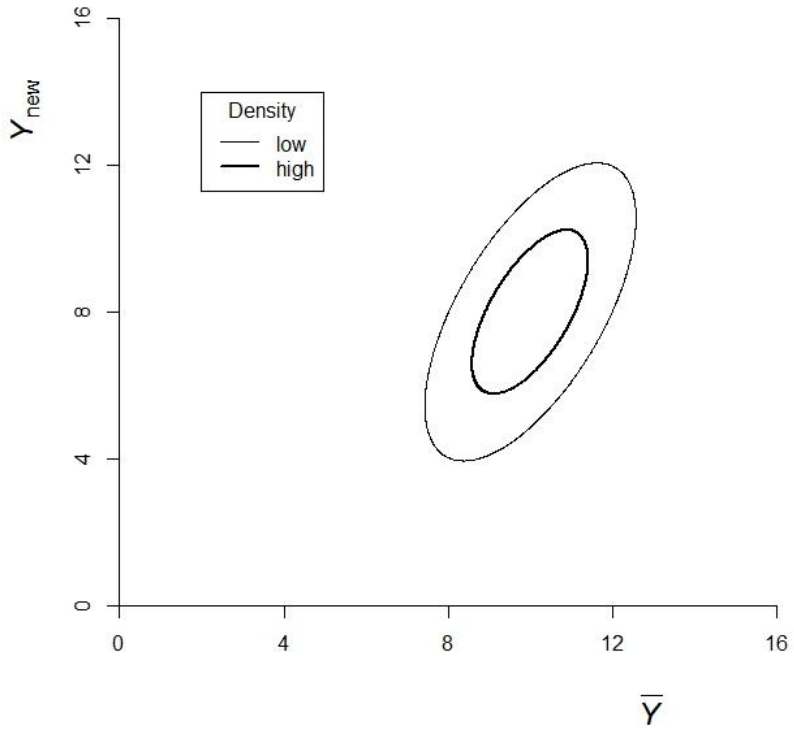
$\bar{y}_i$ is a predictor of $y_{i,\mathbf{new}}$

Random pairs of observations



summarise

Probability density contours

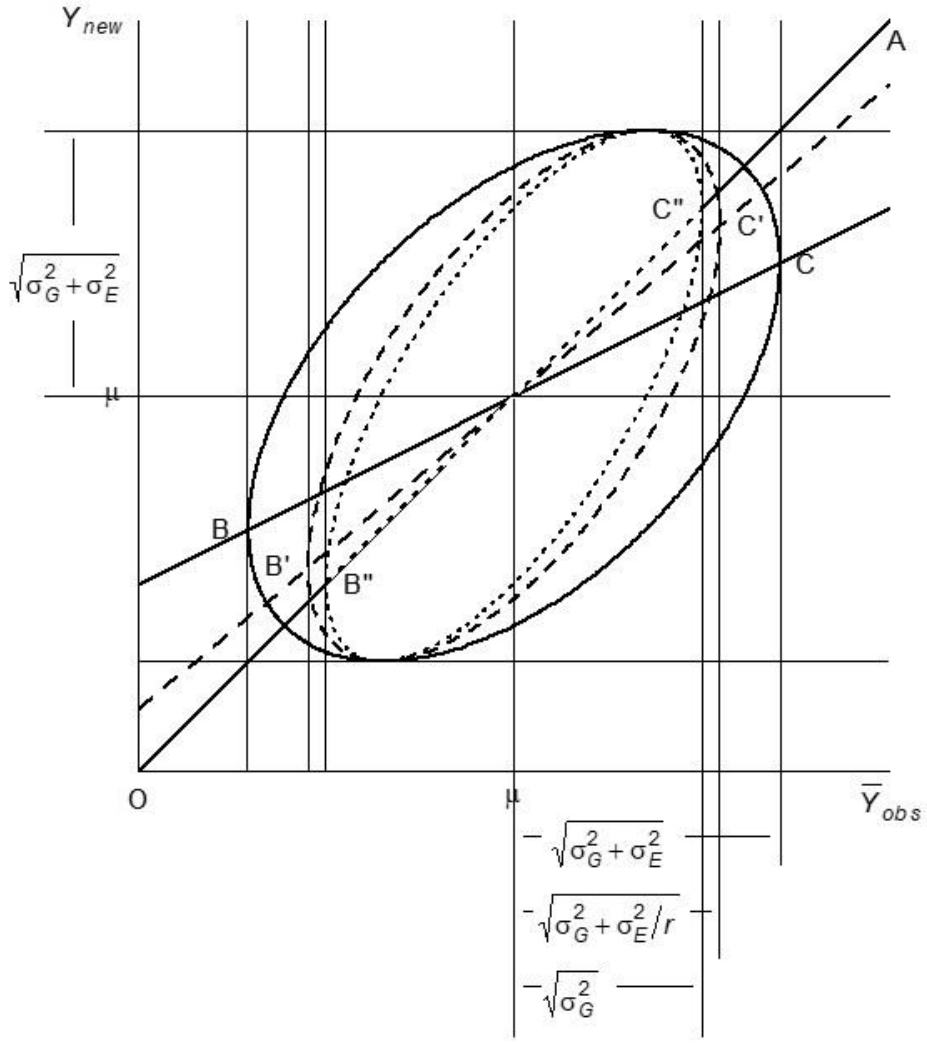


$$\bar{Y} \sim \mathrm{N}\left(\mu, \sqrt{\sigma_G^2 + \frac{\sigma_E^2}{r}}\right)$$

$$Y_{\mathrm{new}} \sim \mathrm{N}\left(\mu, \sqrt{\sigma_G^2 + \sigma_E^2}\right)$$

$$\mathrm{cov}(\bar{Y}, Y_{\mathrm{new}}) = \sigma_G^2$$

Means from larger samples need less shrinkage



Shrinkage factor is monotonically positively related to *r*.

As $r \to \infty$, $\dfrac{\sigma_G^2}{\sigma_G^2+\frac{\sigma_E^2}{r}} \to 1$

and there is no shrinkage.

FDR and shrunk estimates: unpackaging the common features

FDR and shrunk estimates: unpackaging the common features

Both address the Replication Crisis

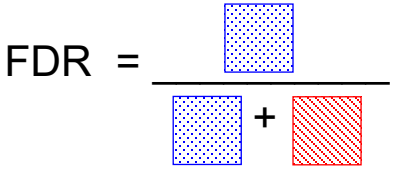
- FDR: defends against false positives while conserving statistical power
- Shrunk estimate: defends against over-optimistic effect-size estimation. (Concept of ‘power’ is irrelevant.)

cont’d…/

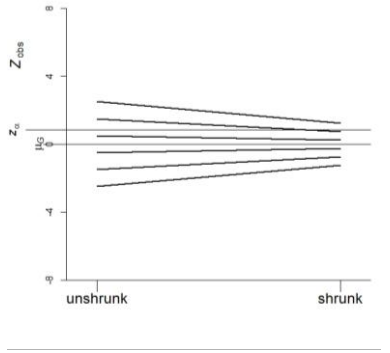
FDR and shrunk estimates: unpackaging the common features/…cont'd.

Both ask, 'What proportion could have happened by chance?'

- Both have an empirical-Bayes interpretation. (See Efron, 2010.)
  For the distribution to which 'What proportion…?' relates,
  - FDR uses the *k* significant test results
  - shrinkage uses the *m* group means.



- For 'What could have happened by chance?':
  - FDR uses a boundary value
  - shrinkage uses an unbiased estimate.



- Dependence on a boundary value is the price the FDR pays for dichotomising

FDR and shrunk estimates: unpackaging the common features/…cont'd.

Both provide predictions/expectations

- Prediction of E(future observation|current results) …
- FDR:

  Prediction of future results among the set of *k* hypotheses announced as discoveries.

  $$\text{FDR} \leq q^*$$
- Shrunk estimate:

  Prediction of mean of new observation from Group *i* .

  $$\mathrm{E}\left(Y_{i,\text{new}} | \bar{y}_{i,\text{ shrunk}}\right) = \bar{y}_{i,\text{ shrunk}}$$

cont'd…/

FDR and shrunk estimates: formal connection

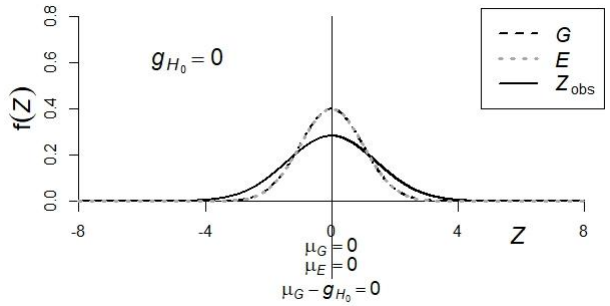
FDR and shrunk estimates: contexts for formal connection

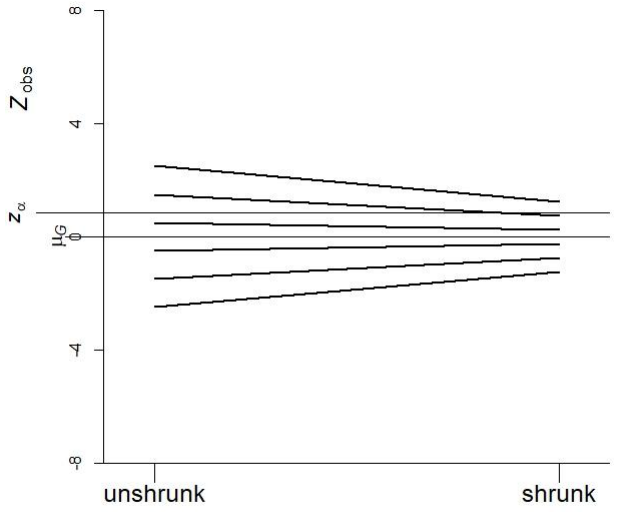
- Connection is illustrated in the context of group means
- Asymptotic normal-distribution model is used
- Unshrunk estimates are best linear unbiased estimates (BLUEs)
- Shrunk estimates are best linear unbiased predictors (BLUPs)

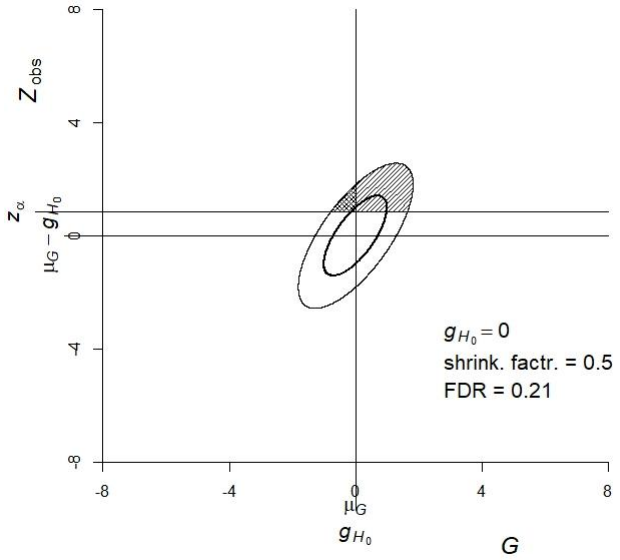
- Connection can probably be extended to shrunk estimates in other contexts, including regression models with
  - many explanatory variables (model parameters), one response (e.g. ridge regression)
  - simple predictive model (few parameters), many response variables
- Connection can probably ***not*** be extended to contexts where some estimates are shrunk to zero for parsimony, e.g. LASSO regression

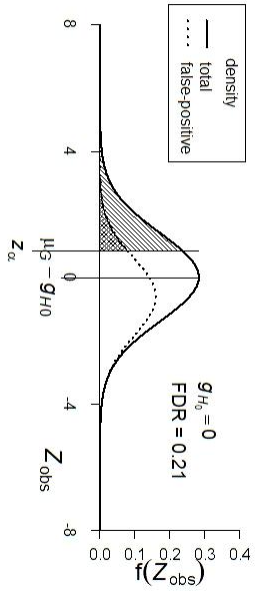
Explore case where $\mu_G \neq 0$

$\mu_G = 0$

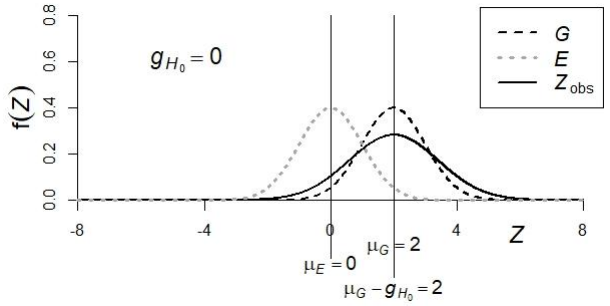


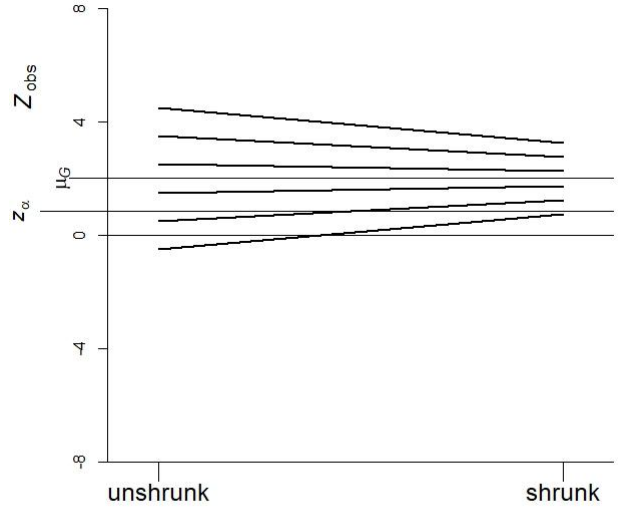


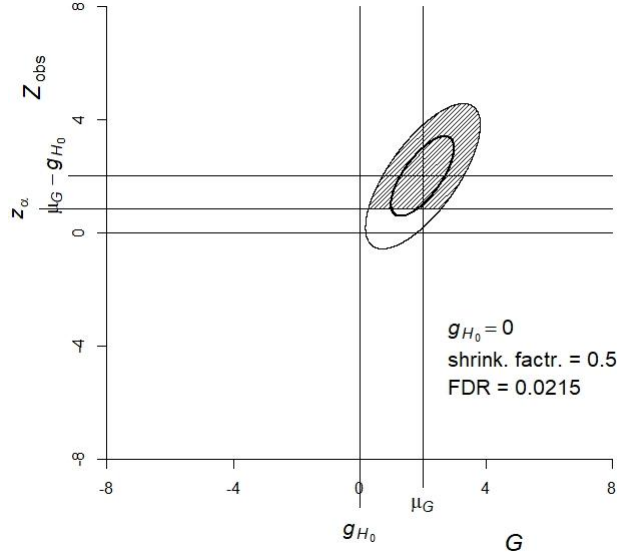


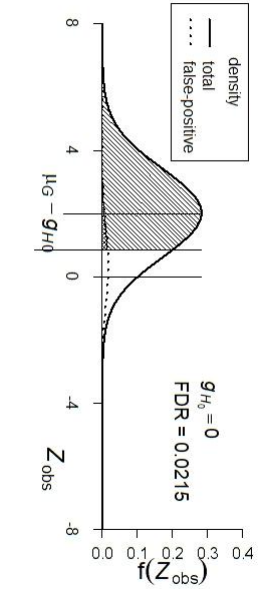


$\mu_G = 2$



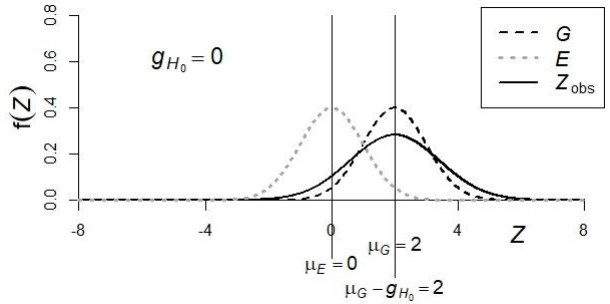


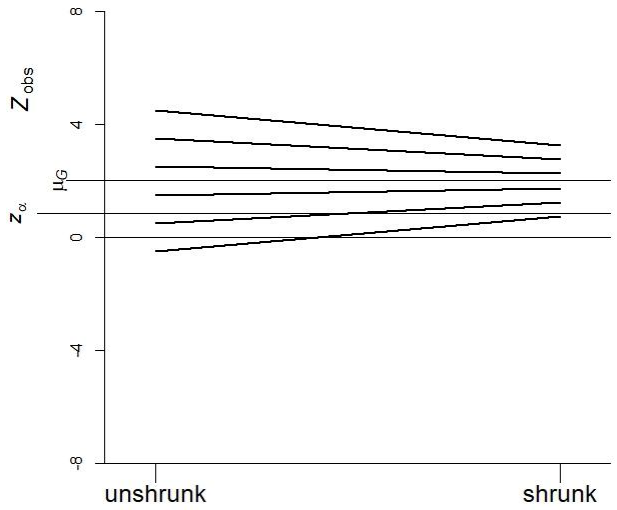


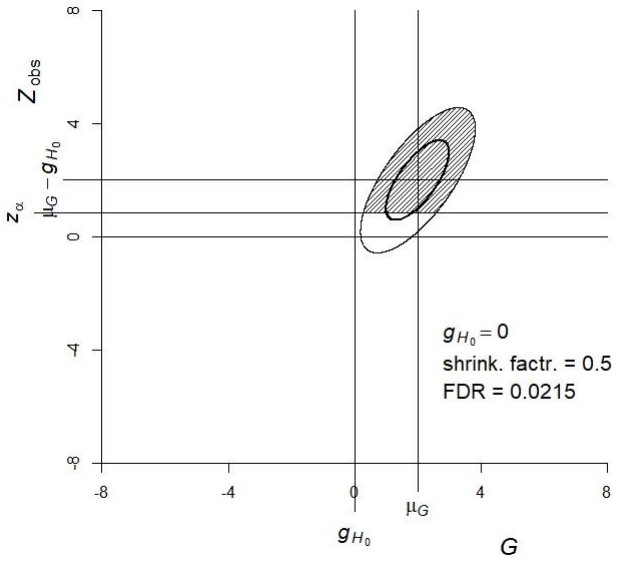


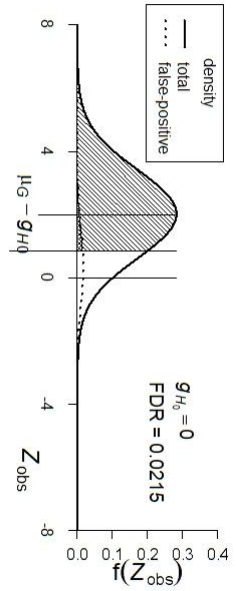
Explore case where $g_{H_0} \neq 0$

$\mu_G = 2, g_{H_0} = 0$

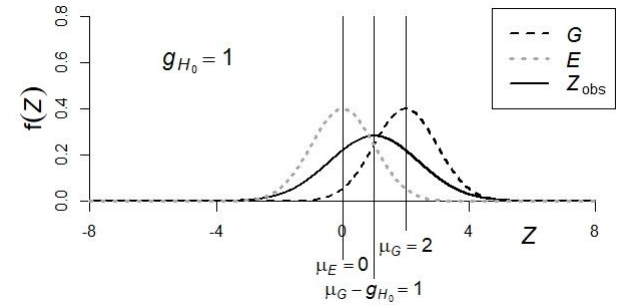


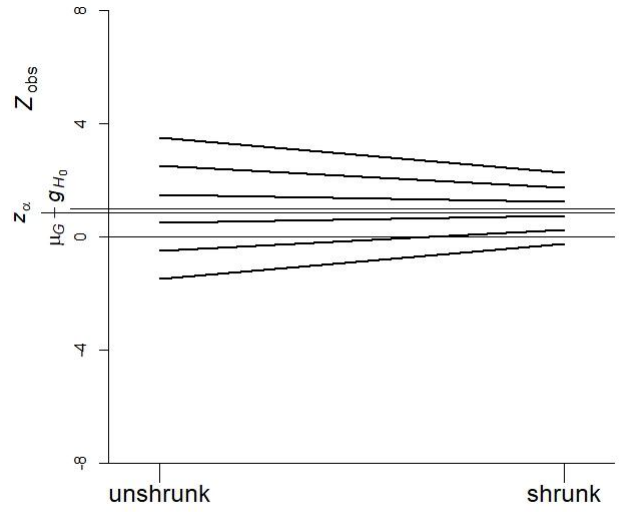


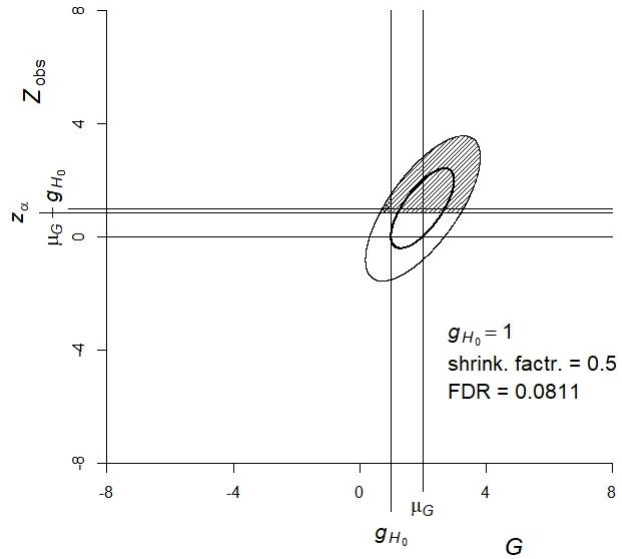


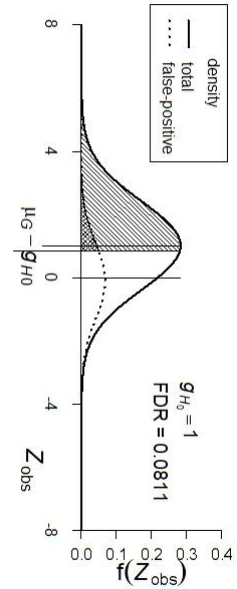


$\mu_G = 2, g_{H_0} = 1$





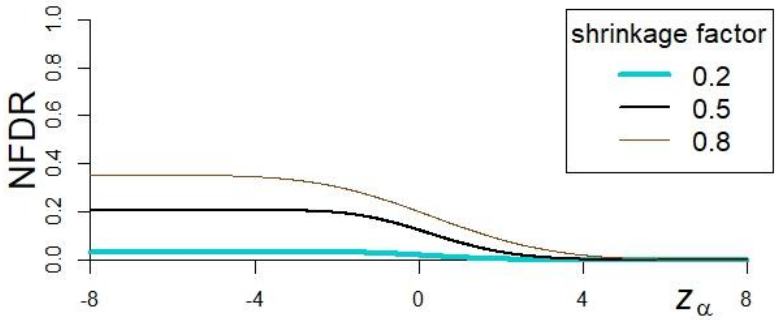




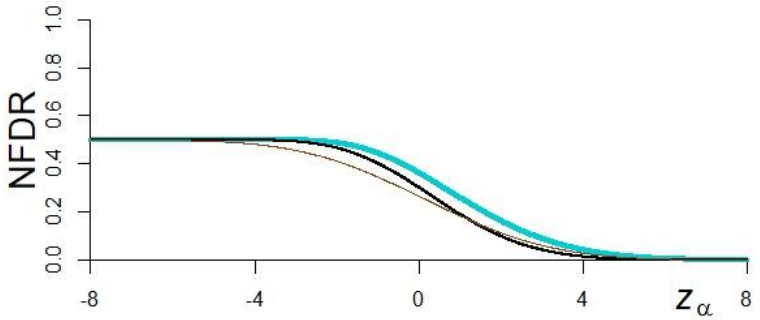
Relationship between FDR and significance threshold, extended over range of g_{H_0}

Threshold expressed as z_α

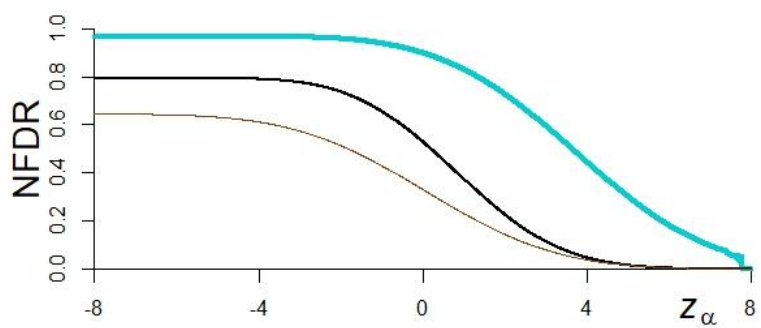
$g_{H_0} = -1$



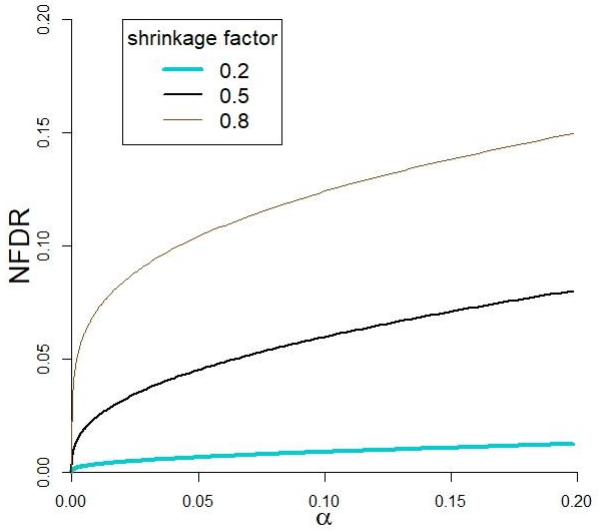
$g_{H_0} = 0$

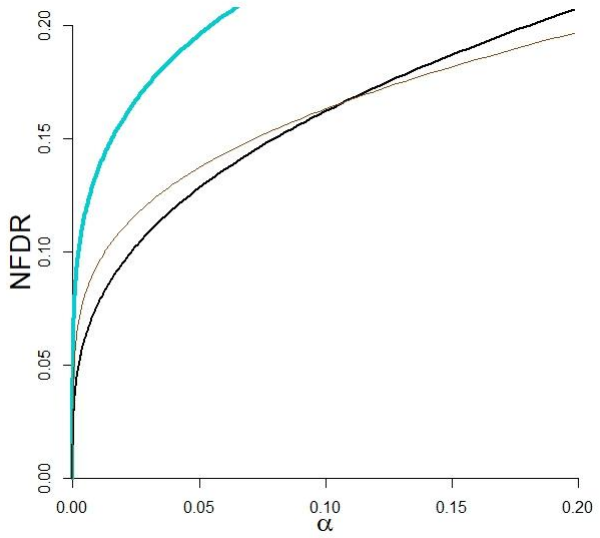


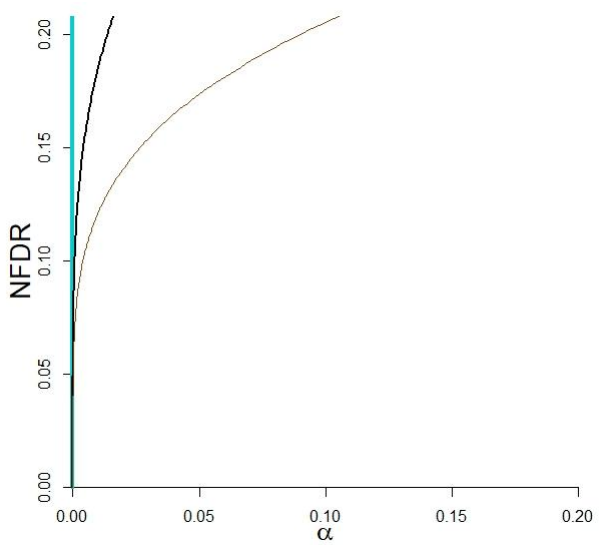
$g_{H_0} = +1$



Threshold expressed as α

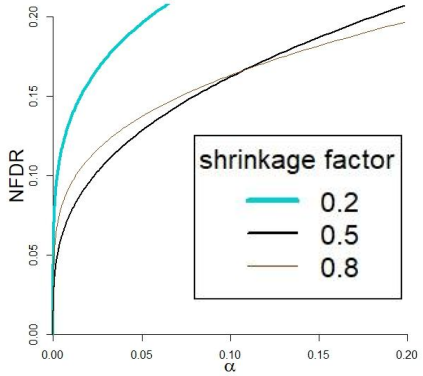


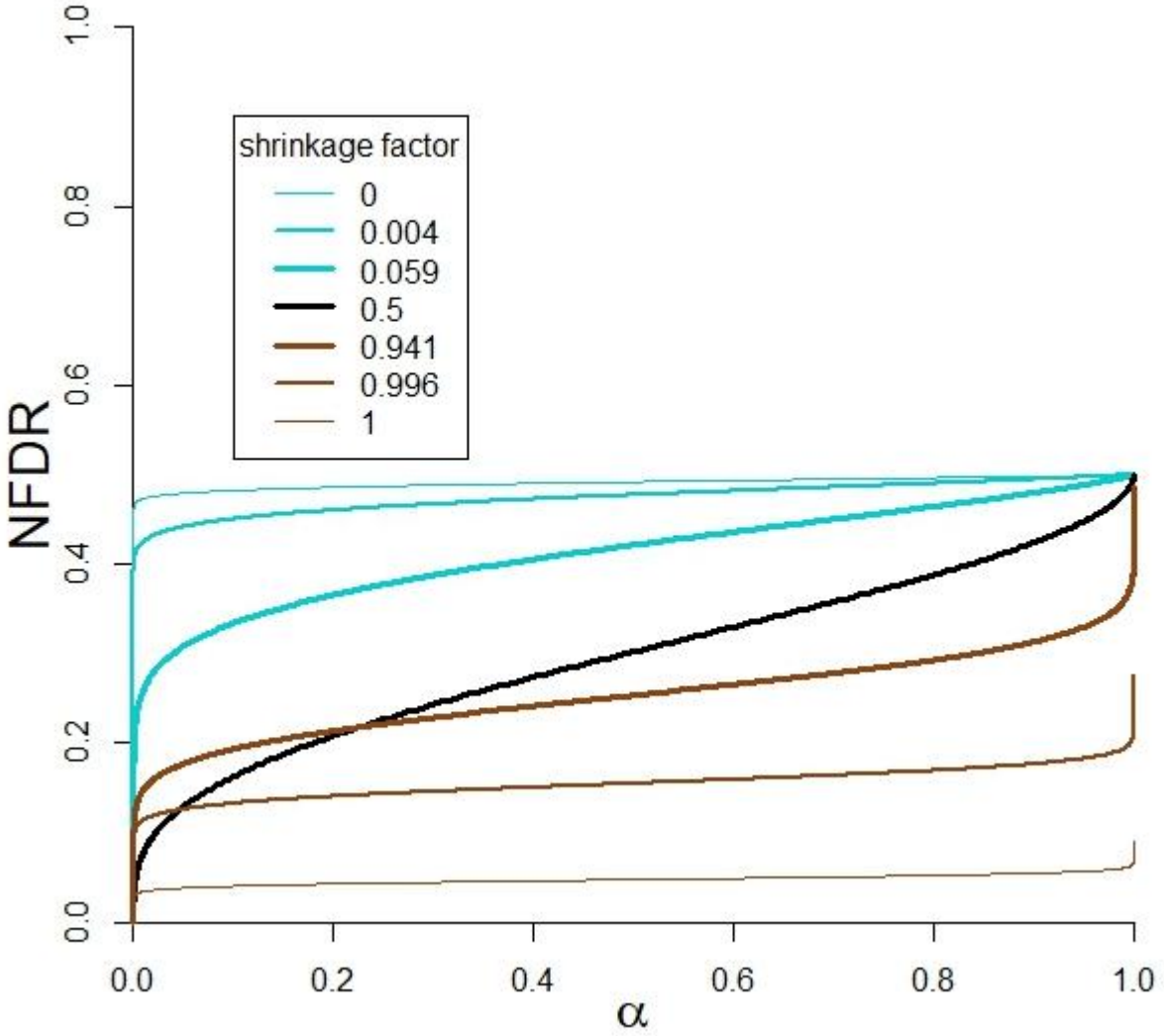




N.B. max(α) = 0.20

Relationship between FDR and α, $g_{H_0} = 0, 0 \leq \alpha \leq 1$

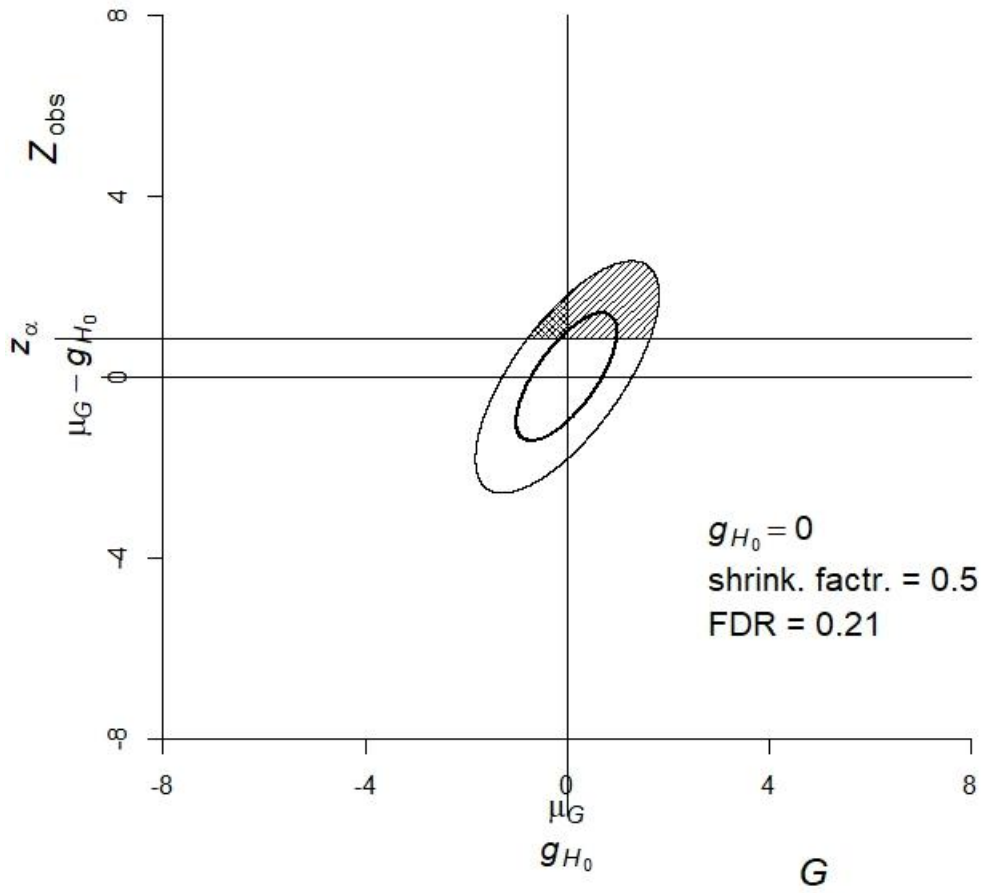




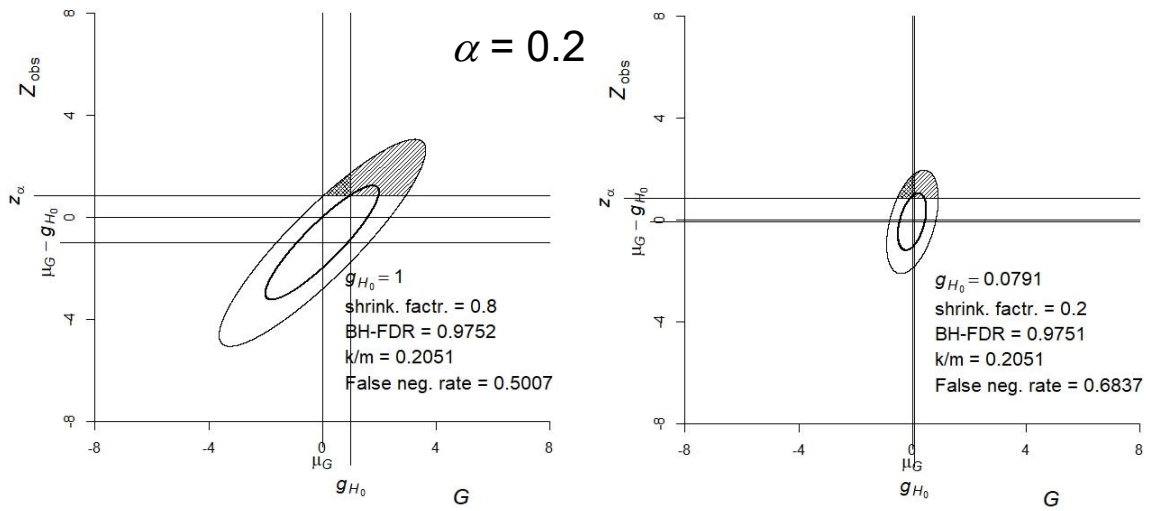
- S.F. small: stringent α required to achieve low NFDR
- SF large: lenient α is sufficient
- Crossovers at low α: counterintuitive?

Practical considerations

The dichotomisation of the significance test exacts a price

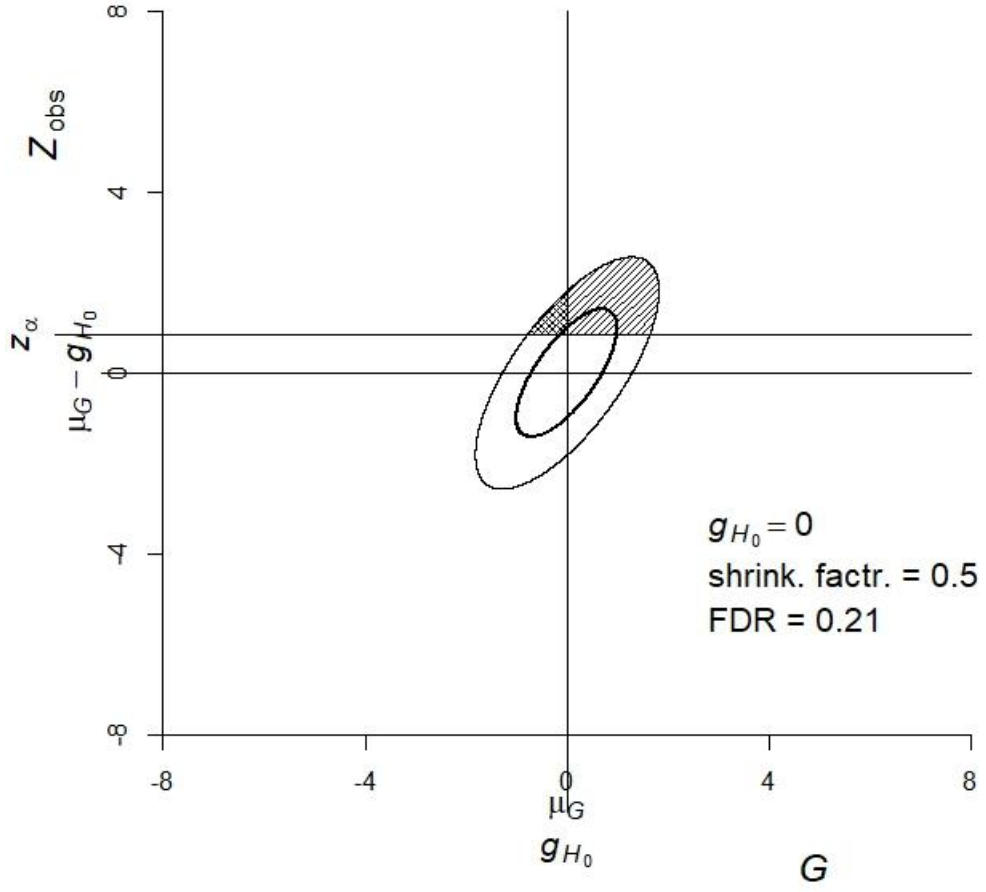


- Not all the information in the bivariate distribution Z_{obs} vs. G is used.
- The excess of 'discoveries' over those expected by chance could be due to:
  - few large effects, of which we detect nearly all, or
  - many small effects, of which we detect only a few.



cont'd…/

The dichotomisation of the significance test exacts a price/…cont'd.



Conservative NFDR = $\dfrac{\alpha}{k/m}$

- Not all the information in the bivariate distribution Z_{obs} vs. G is used.
- The probability mass ▨ is less than $P(Z \geq z_\alpha | G = g_{H0}) = \alpha$, because
  - H_0 is not true for all m tests
  - for the m_0 tests for which H_0 is true, the one-sided test specifies $G \leq g_{H0}$, not $G = g_{H0}$. Hence $P(Z \geq z_\alpha | H_0) \leq \alpha$.

FDR or shrunk estimates: which approach should be preferred?

BH-FDR model:

E(FDR) = $\frac{m_0\alpha}{m_0\alpha+m_1\cdot \mathrm{P}(Z\geq z_\alpha|H_1)} = \frac{m_0\alpha}{k} = \frac{{}^{m_0}/_{m}\alpha}{{}^{k}/_{m}} \leq \frac{\alpha}{{}^{k}/_{m}}$

As ${}^{m_1}/_{m} \to 0$, ${}^{m_0}/_{m} \to 1$ and E(FDR) $\to$ $\frac{\alpha}{{}^{k}/_{m}}$

- Hence $\frac{\alpha}{{}^{k}/_{m}}$ is a nearly-unbiased estimate of FDR when m_1 is small

Shrunk estimates model:

$Y = \mu + G + E$, $G \sim \mathrm{N}(\mu, \sigma_G)$, $E \sim \mathrm{N}(0, \sigma_E)$

One-sided test. $H_0: G \leq g_{H_0}$ may be arbitrary and unhelpful.
Two-sided test. $H_0: G = 0$ is (almost) never true.
With this H_0 , ${}^{m_1}/_{m} \to 1$.

- Hence presentation of shrunk estimates may be more appropriate

FDR or shrunk estimates: what happens in practice?

Discipline	Distribution of real effects	m_1/m	Method	
			Current	**Appropriate**
genetics	rare	~0	Bonferroni-type corrected *p*-values	BH-FDR
gene expression	small effects ubiquitous	~1	BH-FDR	shrunk estimates

Cule et al. (2011) Significance testing in ridge regression for genetic data.
BMC Bioinformatics **12**:372. doi:10.1186/1471-2105-12-372

Q. *(Lead author)* But why are we performing significance tests on ridge-regression parameter shrunk estimates?

A. *(Supervisor)* Because geneticists like *p*-values.

But why not present FDR ***and*** shrunk estimates?

Steps for extraction and collation of gene-expression data

An official website of the United States government Here's how you know

National Library of Medicine
National Center for Biotechnology Information

Log in

GEO Profiles | GEO Profiles | Search | Advanced | Help

GEO Profiles

This database stores individual gene expression profiles from curated DataSets in the Gene Expression Omnibus (GEO) repository. Search for specific profiles of interest based on gene annotation or pre-computed profile characteristics.

Getting Started

- GEO Documentation
- GEO FAQ
- About GEO Profiles
- Construct a Query
- Download Options

GEO Tools

- Submit to GEO
- Advanced Search
- DataSet Browser
- Programmatic Access

More Resources

- GEO Home
- GEO DataSets
- SRA

Example Searches

Gene symbol	CYP1A1[Gene Symbol]
Gene symbols in DataSets that contain specific keywords	(CYP1A1[Gene Symbol] OR ME1[Gene Symbol]) AND (smok* OR diet)
Partial gene name in a specific DataSet	kinase[Gene Description] AND GDS182
Gene Ontology(GO) term in a specific DataSet	apoptosis[Gene Ontology] AND GDS182
Chromosome region and species	(8[Chromosome] AND 10000:3000000[Base Position]) AND mouse[organism]
Genes that show subset effects in DataSets that examine the effect of an agent	agent[Flag Information] AND "value subset effect"[Flag Type]

Inbox - nwgalwey@gmail.com | Google Calendar - Week of 23 | CM22 - BBC Weather | Empirical Bayes method - Wikip | (20+) Facebook | GEO DataSet Browser

www.ncbi.nlm.nih.gov/sites/GDSbrowser/

NCBI

DATASET BROWSER (CURATED)

Gene Expression Omnibus

Search for [ ] Search | Clear | Show All | Advanced Search

Page size 20

4348 DataSet records

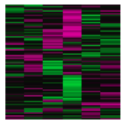
Page 1 of 218 > >>

DataSet	Title	Organism(s)	Platform	Series	Samples
GDS5826	Multiple myeloma cell lines with acquired resistance to chemotherapeutic agent carfilzomib	Homo sapiens	GPL570	GSE69078	12
GDS5825	Interleukin-1α deficiency effect on injured spinal cord	Mus musculus	GPL6246	GSE70302	12
GDS5881	Nebulin deficiency effect on the soleus	Mus musculus	GPL6246	GSE70213	12
GDS5880	Nebulin deficiency effect on the quadriceps	Mus musculus	GPL6246	GSE70213	12
GDS5913	SRPIN803 small molecule inhibitor of SRPK1 effect on retinal pigment epithelial cell line	Homo sapiens	GPL570	GSE62947	6
GDS5665	Pathogen-associated molecular-pattern curdlan effect on interleukin-2 deficient GM-CSF myeloid dendritic cells	Mus musculus	GPL6246	GSE58120	12
GDS5662	Histone demethylase KDM3A-deficiency effect on estrogen-stimulated breast cancer cells in vitro	Homo sapiens	GPL10558	GSE68918	11
GDS5948	Zipper-interacting protein kinase deficiency effect on coronary artery smooth muscle cells in vitro	Homo sapiens	GPL6244	GSE56819	6
GDS5659	Rho-associated kinase deficiency effect on coronary artery smooth muscle cells in vitro	Homo sapiens	GPL6244	GSE56819	6
GDS5658	Retinoid X receptor activation effect on stimulated RAW264.7 cells	Mus musculus	GPL1261	GSE62107	6

DataSet Record GDS5948: Expression Profiles | Data Analysis Tools | Sample Subsets

Title:	Zipper-interacting protein kinase deficiency effect on coronary artery smooth muscle cells in vitro
Summary:	Analysis of cultured vascular smooth muscle cells following knockdown of zipper-interacting protein kinase (ZIPK). ZIPK is phosphorylated and activated by Rho-associated kinase 1 (ROCK1). These results, together with those from GDS5659, provide further insight into ROCK1 and ZIPK functions.
Organism:	*Homo sapiens*
Platform:	GPL6244: [HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]
Citation:	Deng JT, Wang XL, Chen YX, O'Brien ER et al. The effects of knockdown of rho-associated kinase 1 and zipper-interacting protein kinase on gene expression and function in cultured human arterial smooth muscle cells. *PLoS One* 2015;10(2):e0116969. PMID: 25723491
Reference Series:	GSE56819
Sample count:	6
Value type:	transformed count
Series	

Cluster Analysis



Download

- DataSet full SOFT file
- DataSet SOFT file
- Series family SOFT file
- Series family MINiML file
- Annotation SOFT file

HHS Vulnera... ibility

Scroll down through DataSet rows, to Title that corresponds to that of Deng et al. (2015).

Select this row to display DataSet Record GDS5948, including the citation of Deng et al.

Then click on Series 56819.

NCBI

Gene Expression Omnibus

HOME | SEARCH | SITE MAP | GEO Publications | FAQ | MIAME | Email GEO

NCBI > GEO > **Accession Display** Reviewer access | Sign Out

GEO help: Mouse over screen elements for information.

Scope: Self Format: HTML Amount: Quick GEO accession: GSE56819 GO

Series GSE56819 Query DataSets for GSE56819

Status	Public on May 15, 2015
Title	Effects of knockdown of ROCK1 and ZIPK on gene expression in cultured human vascular smooth muscle cells
Organism	Homo sapiens
Experiment type	Expression profiling by array
Summary	Rho-associated kinase (ROCK) and zipper-interacting protein kinase (ZIPK) have been implicated in diverse physiological functions, including smooth muscle contraction, cell proliferation, cell adhesion, apoptosis, cell migration and inflammation. Many aspects of regulation via ROCK and ZIPK, however, remain unclear. In this study, we utilized an siRNA approach to knock down ROCK1 and ZIPK in cultured human arterial smooth muscle cells. Microarray analysis was performed, using a whole-transcript expression chip, to identify changes in gene expression profiles induced by ROCK1 and ZIPK knockdown. ROCK1 knockdown affected the expression of 553 genes (355 down-regulated and 198 up-regulated), while ZIPK knockdown affected the expression of 390 genes (219 down-regulated and 171 up-regulated). A high incidence of up- and down-regulation of transcription regulator genes was observed in both ROCK1 and ZIPK knockdowns. Other markedly affected groups included transporters, kinases, peptidases, transmembrane and G protein-coupled receptors, growth factors, phosphatases and ion channels. Three microRNAs (mir-145, mir-199 and mir-622) were up-regulated by ROCK1 knockdown, whereas ZIPK knockdown had no effect on microRNA expression. 76 differentially expressed genes were common to ROCK1 and ZIPK knockdown, of which 41 were down-regulated and 26 up-regulated by both treatments, while the other 9 genes were differentially up/down-regulated. Ingenuity Pathway Analysis identified five pathways shared between the two knockdowns, which are mainly involved in cell cycle regulation. Marked differences in the effects of ROCK1 and ZIPK knockdown on the genes involved in cell cycle regulation suggested that ROCK1 and ZIPK regulate the cell cycle by different mechanisms. ROCK1, but not ZIPK knockdown significantly reduced the viability of vascular SMC. ROCK1 knockdown also affected several cytokine signaling pathways with up-regulation of 5 and down-regulation of 4 cytokine genes, in contrast to ZIPK knockdown, which affected the expression of only two cytokine genes (both down-regulated). IL-6 gene expression and secretion of IL-6 protein were up-regulated by ROCK1 knockdown, whereas ZIPK knockdown reduced IL-6 mRNA expression and IL-6 protein secretion and ROCK1 protein expression, suggesting that ROCK1 may inhibit IL-6 secretion. IL-1β mRNA and protein levels were increased in response to ROCK1 knockdown. Finally, ROCK1 but not ZIPK knockdown inhibited proliferation of vascular smooth muscle cells. We conclude that ROCK1 and ZIPK have diverse, but predominantly distinct regulatory functions in vascular smooth muscle

Copy and paste ‘Samples’ table to a text file, then convert it to an Excel spreadsheet.

Then click on the code for the first sample.

Overall design	Human coronary artery smooth muscle cells were transfected with siRNA targeting ROCK1 or ZIPK or with negative control siRNA that does not target any gene product. 48 h later, total RNA was isolated, reverse transcribed, amplified, labeled with the Ambion WT Express kit and hybridized to Human Gene 1.0 ST arrays (Affymetrix) at 45 oC for 16 h. The probe arrays were washed and stained on an Affymetrix GeneChip Fluidics-450 and scanned on an Affymetrix GeneChip Scanner 3000 7G System. Triplicates were prepared under all three conditions for microarray analysis.
Contributor(s)	Walsh MP, Deng JT
Citation(s)	Deng JT, Wang XL, Chen YX, O'Brien ER et al. The effects of knockdown of rho-associated kinase 1 and zipper-interacting protein kinase on gene expression and function in cultured human arterial smooth muscle cells. *PLoS One* 2015;10(2):e0116969. PMID: 25723491

Analyze with GEO2R

Submission date	Apr 15, 2014
Last update date	Jul 26, 2018
Contact name	Michael Patrick Walsh
E-mail(s)	walsh@ucalgary.ca
Phone	403-220-3021
Organization name	University of Calgary
Department	Biochemistry & Molecular Biology
Street address	3330 Hospital Drive NW
City	Calgary
State/province	Alberta
ZIP/Postal code	T3A 0M9
Country	Canada

Platforms (1)	GPL6244 [HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]
Samples (9)	GSM1369856 Control siRNA 1
Less...	GSM1369857 Control siRNA 2
	GSM1369858 Control siRNA 3
	GSM1369859 ROCK1 siRNA 1
	GSM1369860 ROCK1 siRNA 2
	GSM1369861 ROCK1 siRNA 3
	GSM1369862 ZIPK siRNA 1
	GSM1369863 ZIPK siRNA 2
	GSM1369864 ZIPK siRNA 3

Relations

BioProject PRJNA244678

Download family	Format
SOFT formatted family file(s)	SOFT
MINiML formatted family file(s)	MINiML
Series Matrix File(s)	TXT

ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1369856

E-mail(s)	walsh@ucalgary.ca
Phone	403-220-3021
Organization name	University of Calgary
Department	Biochemistry & Molecular Biology
Street address	3330 Hospital Drive NW
City	Calgary
State/province	Alberta
ZIP/Postal code	T3A 0M9
Country	Canada

Platform ID	GPL6244	
Series (1)	GSE56819	Effects of knockdown of ROCK1 and ZIPK on gene expression in cultured human vascular smooth muscle cells

Data table header descriptions

ID_REF	
VALUE	RMA signal

Data table

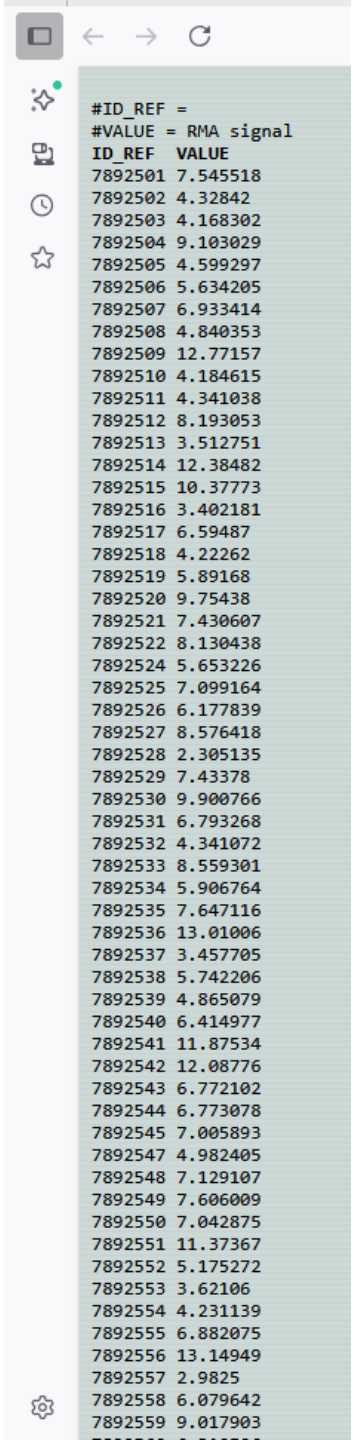
ID_REF	VALUE
7892501	7.545518
7892502	4.32842
7892503	4.168302
7892504	9.103029
7892505	4.599297
7892506	5.634205
7892507	6.933414
7892508	4.840353
7892509	12.77157
7892510	4.184615
7892511	4.341038
7892512	8.193053
7892513	3.512751
7892514	12.38482
7892515	10.37773
7892516	3.402181
7892517	6.59487
7892518	4.22262
7892519	5.89168

Total number of rows: **33297**

Table truncated, full table size **549 Kbytes**.

View full table...

Click on ‘View full table…’



```
#ID_REF =
#VALUE = RMA signal
ID_REF  VALUE
7892501 7.545518
7892502 4.32842
7892503 4.168302
7892504 9.103029
7892505 4.599297
7892506 5.634205
7892507 6.933414
7892508 4.840353
7892509 12.77157
7892510 4.184615
7892511 4.341038
7892512 8.193053
7892513 3.512751
7892514 12.38482
7892515 10.37773
7892516 3.402181
7892517 6.59487
7892518 4.22262
7892519 5.89168
7892520 9.75438
7892521 7.430607
7892522 8.130438
7892524 5.653226
7892525 7.099164
7892526 6.177839
7892527 8.576418
7892528 2.305135
7892529 7.43378
7892530 9.900766
7892531 6.793268
7892532 4.341072
7892533 8.559301
7892534 5.906764
7892535 7.647116
7892536 13.01006
7892537 3.457705
7892538 5.742206
7892539 4.865079
7892540 6.414977
7892541 11.87534
7892542 12.08776
7892543 6.772102
7892544 6.773078
7892545 7.005893
7892547 4.982405
7892548 7.129107
7892549 7.606009
7892550 7.042875
7892551 11.37367
7892552 5.175272
7892553 3.62106
7892554 4.231139
7892555 6.882075
7892556 13.14949
7892557 2.9825
7892558 6.079642
7892559 9.017903
```

Copy and paste full table for Sample GSM1369856 to a text file.

Repeat previous steps for other samples.

NCBI

Gene Expression Omnibus

HOME | SEARCH | SITE MAP | GEO Publications | FAQ | MIAME | Email GEO

NCBI > GEO > **Accession Display** Reviewer access | Sign Out

GEO help: Mouse over screen elements for information.

Scope: Self Format: HTML Amount: Quick GEO accession: GSE56819 GO

Series GSE56819 Query DataSets for GSE56819

Status	Public on May 15, 2015
Title	Effects of knockdown of ROCK1 and ZIPK on gene expression in cultured human vascular smooth muscle cells
Organism	Homo sapiens
Experiment type	Expression profiling by array
Summary	Rho-associated kinase (ROCK) and zipper-interacting protein kinase (ZIPK) have been implicated in diverse physiological functions, including smooth muscle contraction, cell proliferation, cell adhesion, apoptosis, cell migration and inflammation. Many aspects of regulation via ROCK and ZIPK, however, remain unclear. In this study, we utilized an siRNA approach to knock down ROCK1 and ZIPK in cultured human arterial smooth muscle cells. Microarray analysis was performed, using a whole-transcript expression chip, to identify changes in gene expression profiles induced by ROCK1 and ZIPK knockdown. ROCK1 knockdown affected the expression of 553 genes (355 down-regulated and 198 up-regulated), while ZIPK knockdown affected the expression of 390 genes (219 down-regulated and 171 up-regulated). A high incidence of up- and down-regulation of transcription regulator genes was observed in both ROCK1 and ZIPK knockdowns. Other markedly affected groups included transporters, kinases, peptidases, transmembrane and G protein-coupled receptors, growth factors, phosphatases and ion channels. Three microRNAs (mir-145, mir-199 and mir-622) were up-regulated by ROCK1 knockdown, whereas ZIPK knockdown had no effect on microRNA expression. 76 differentially expressed genes were common to ROCK1 and ZIPK knockdown, of which 41 were down-regulated and 26 up-regulated by both treatments, while the other 9 genes were differentially up/down-regulated. Ingenuity Pathway Analysis identified five pathways shared between the two knockdowns, which are mainly involved in cell cycle regulation. Marked differences in the effects of ROCK1 and ZIPK knockdown on the genes involved in cell cycle regulation suggested that ROCK1 and ZIPK regulate the cell cycle by different mechanisms. ROCK1, but not ZIPK knockdown significantly reduced the viability of vascular SMC. ROCK1 knockdown also affected several cytokine signaling pathways with up-regulation of 5 and down-regulation of 4 cytokine genes, in contrast to ZIPK knockdown, which affected the expression of only two cytokine genes (both down-regulated). IL-6 gene expression and secretion of IL-6 protein were up-regulated by ROCK1 knockdown, whereas ZIPK knockdown reduced IL-6 mRNA expression and IL-6 protein secretion and ROCK1 protein expression, suggesting that ROCK1 may inhibit IL-6 secretion

Click on ‘Analyse with GEO2R’.

N.B. We do not really want the results of this analysis. It is just a technique to obtain a table mapping ID_REF values to Gene.symbol and Gene.title values.

Overall design	Human coronary artery smooth muscle cells were transfected with siRNA targeting ROCK1 or ZIPK or with negative control siRNA that does not target any gene product. 48 h later, total RNA was isolated, reverse transcribed, amplified, labeled with the Ambion WT Express kit and hybridized to Human Gene 1.0 ST arrays (Affymetrix) at 45 oC for 16 h. The probe arrays were washed and stained on an Affymetrix GeneChip Fluidics-450 and scanned on an Affymetrix GeneChip Scanner 3000 7G System. Triplicates were prepared under all three conditions for microarray analysis.
Contributor(s)	Walsh MP, Deng JT
Citation(s)	Deng JT, Wang XL, Chen YX, O'Brien ER et al. The effects of knockdown of rho-associated kinase 1 and zipper-interacting protein kinase on gene expression and function in cultured human arterial smooth muscle cells. *PLoS One* 2015;10(2):e0116969. PMID: 25723491

Analyze with GEO2R

Submission date	Apr 15, 2014
Last update date	Jul 26, 2018
Contact name	Michael Patrick Walsh
E-mail(s)	walsh@ucalgary.ca
Phone	403-220-3021
Organization name	University of Calgary
Department	Biochemistry & Molecular Biology
Street address	3330 Hospital Drive NW
City	Calgary
State/province	Alberta
ZIP/Postal code	T3A 0M9
Country	Canada

Platforms (1)	GPL6244 [HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcript (gene) version]
Samples (9) Less...	GSM1369856 Control siRNA 1 GSM1369857 Control siRNA 2 GSM1369858 Control siRNA 3 GSM1369859 ROCK1 siRNA 1 GSM1369860 ROCK1 siRNA 2 GSM1369861 ROCK1 siRNA 3 GSM1369862 ZIPK siRNA 1 GSM1369863 ZIPK siRNA 2 GSM1369864 ZIPK siRNA 3

Relations

BioProject PRJNA244678

Download family	Format
SOFT formatted family file(s)	SOFT
MINiML formatted family file(s)	MINiML
Series Matrix File(s)	TXT

Use GEO2R to compare two or more groups of Samples in order to identify genes that are differentially expressed across experimental conditions. Results are presented as a table of genes ordered by significance. Full instructions

GEO accession GSE56819 Set Effects of knockdown of ROCK1 and ZIPK on gene expression in cultured human vascular smooth muscle cells

Samples Define groups Selected **0** out of **9** samples

Columns Set

Group	Accession	Title	Source name	Tissue	Cell type	Cell line
-	GSM1369856	Control siRNA 1	Human CASMCs, control siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
-	GSM1369857	Control siRNA 2	Human CASMCs, control siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
-	GSM1369858	Control siRNA 3	Human CASMCs, control siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
-	GSM1369859	ROCK1 siRNA 1	Human CASMCs, ROCK1 siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
-	GSM1369860	ROCK1 siRNA 2	Human CASMCs, ROCK1 siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
-	GSM1369861	ROCK1 siRNA 3	Human CASMCs, ROCK1 siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
-	GSM1369862	ZIPK siRNA 1	Human CASMCs, ZIPK siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
-	GSM1369863	ZIPK siRNA 2	Human CASMCs, ZIPK siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
-	GSM1369864	ZIPK siRNA 3	Human CASMCs, ZIPK siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583

GEO2R | Options | Profile graph | R script

Quick start

- Specify a GEO Series accession and a Platform if prompted.
- Click 'Define groups' and enter names for the groups of Samples you plan to compare, e.g., test and control.
- Assign Samples to each group. Highlight Sample rows then click the group name to assign those Samples to the group. Use the Sample metadata (title, source and characteristics) columns to help determine which Samples belong to which group.
- Click 'Analyze' to perform the calculation with default settings.
- You may change settings in the Options tab.

How to use

Analyze

www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE56819

NCBI

Gene Expression Omnibus

GEO Publications | FAQ | MIAME | Email GEO

NCBI » GEO » GEO2R » GSE56819 Login

Use GEO2R to compare two or more groups of Samples in order to identify genes that are differentially expressed across experimental conditions. Results are presented as a table of genes ordered by significance. Full instructions

GEO accession GSE56819 Set Effects of knockdown of ROCK1 and ZIPK on gene expression in cultured human vascular smooth muscle cells

Samples Define groups Selected **0** out of **9** samples

Enter a group name: List

ZIPK

Cancel selection

Columns Set

Group	Accession		Source name	Tissue	Cell type	Cell line
-	GSM1369856		Human CASMCs, control siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
-	GSM1369857	Control siRNA 2	Human CASMCs, control siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
-	GSM1369858	Control siRNA 3	Human CASMCs, control siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
-	GSM1369859	ROCK1 siRNA 1	Human CASMCs, ROCK1 siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
-	GSM1369860	ROCK1 siRNA 2	Human CASMCs, ROCK1 siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
-	GSM1369861	ROCK1 siRNA 3	Human CASMCs, ROCK1 siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
-	GSM1369862	ZIPK siRNA 1	Human CASMCs, ZIPK siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
-	GSM1369863	ZIPK siRNA 2	Human CASMCs, ZIPK siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
-	GSM1369864	ZIPK siRNA 3	Human CASMCs, ZIPK siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583

GEO2R | Options | Profile graph | R script

Quick start

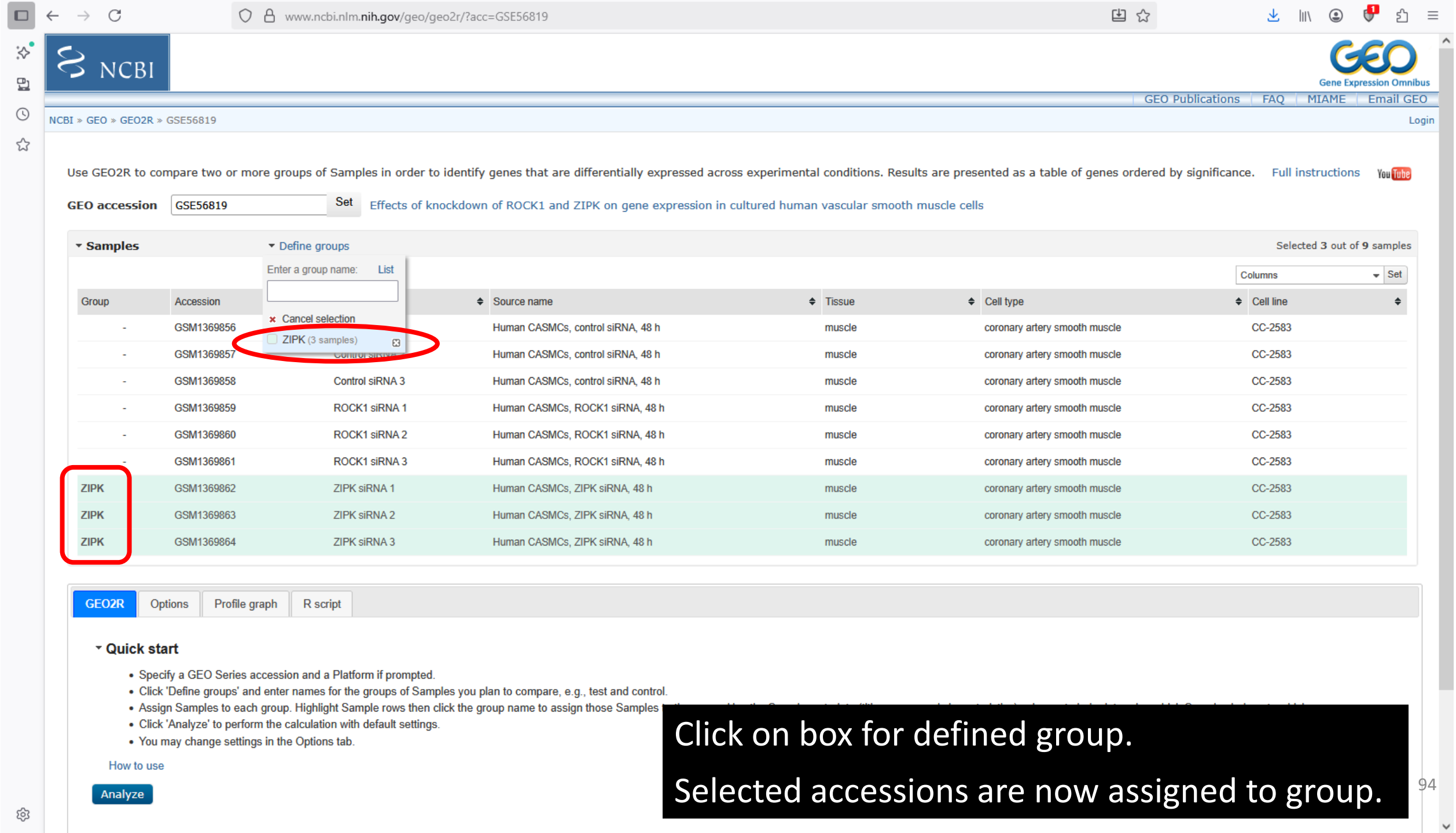
- Specify a GEO Series accession and a Platform if prom
- Click 'Define groups' and enter names for the groups of
- Assign Samples to each group. Highlight Sample rows t
- Click 'Analyze' to perform the calculation with default se
- You may change settings in the Options tab.

How to use

Analyze

Select accessions in control group. To select, point-and-click for first accession. Control + point-and-click for subsequent accessions

Enter group name in 'Define groups' field. Press Enter.


www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE56819
NCBI
Gene Expression Omnibus
GEO Publications
FAQ
MIAME
Email GEO
NCBI » GEO » GEO2R » GSE56819
Login
Use GEO2R to compare two or more groups of Samples in order to identify genes that are differentially expressed across experimental conditions. Results are presented as a table of genes ordered by significance.
Full instructions
GEO accession
GSE56819
Set
Effects of knockdown of ROCK1 and ZIPK on gene expression in cultured human vascular smooth muscle cells
Samples
Define groups
Selected 3 out of 9 samples
Enter a group name:
List
Cancel selection
ZIPK (3 samples)
Columns
Set
Group
Accession
Source name
Tissue
Cell type
Cell line
- GSM1369856 Human CASMCs, control siRNA, 48 h muscle coronary artery smooth muscle CC-2583
- GSM1369857 Human CASMCs, control siRNA, 48 h muscle coronary artery smooth muscle CC-2583
- GSM1369858 Control siRNA 3 Human CASMCs, control siRNA, 48 h muscle coronary artery smooth muscle CC-2583
- GSM1369859 ROCK1 siRNA 1 Human CASMCs, ROCK1 siRNA, 48 h muscle coronary artery smooth muscle CC-2583
- GSM1369860 ROCK1 siRNA 2 Human CASMCs, ROCK1 siRNA, 48 h muscle coronary artery smooth muscle CC-2583
- GSM1369861 ROCK1 siRNA 3 Human CASMCs, ROCK1 siRNA, 48 h muscle coronary artery smooth muscle CC-2583
ZIPK GSM1369862 ZIPK siRNA 1 Human CASMCs, ZIPK siRNA, 48 h muscle coronary artery smooth muscle CC-2583
ZIPK GSM1369863 ZIPK siRNA 2 Human CASMCs, ZIPK siRNA, 48 h muscle coronary artery smooth muscle CC-2583
ZIPK GSM1369864 ZIPK siRNA 3 Human CASMCs, ZIPK siRNA, 48 h muscle coronary artery smooth muscle CC-2583
GEO2R
Options
Profile graph
R script
Quick start
Specify a GEO Series accession and a Platform if prompted.
Click 'Define groups' and enter names for the groups of Samples you plan to compare, e.g., test and control.
Assign Samples to each group. Highlight Sample rows then click the group name to assign those Samples
Click 'Analyze' to perform the calculation with default settings.
You may change settings in the Options tab.
How to use
Analyze
Click on box for defined group.
Selected accessions are now assigned to group.
94

www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE56819

NCBI

Gene Expression Omnibus

GEO Publications | FAQ | MIAME | Email GEO

NCBI » GEO » GEO2R » GSE56819 Login

Use GEO2R to compare two or more groups of Samples in order to identify genes that are differentially expressed across experimental conditions. Results are presented as a table of genes ordered by significance. Full instructions

GEO accession GSE56819 Set Effects of knockdown of ROCK1 and ZIPK on gene expression in cultured human vascular smooth muscle cells

Samples Define groups Selected 3 out of 9 samples

Enter a group name: List

ROCK1

Cancel selection

ZIPK (3 samples)

Columns Set

Group	Accession		Source name	Tissue	Cell type	Cell line
-	GSM1369856		Human CASMCs, control siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
-	GSM1369857		Human CASMCs, control siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
-	GSM1369858	Control siRNA 3	Human CASMCs, control siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
-	GSM1369859	ROCK1 siRNA 1	Human CASMCs, ROCK1 siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
-	GSM1369860	ROCK1 siRNA 2	Human CASMCs, ROCK1 siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
-	GSM1369861	ROCK1 siRNA 3	Human CASMCs, ROCK1 siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
ZIPK	GSM1369862	ZIPK siRNA 1	Human CASMCs, ZIPK siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
ZIPK	GSM1369863	ZIPK siRNA 2	Human CASMCs, ZIPK siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583
ZIPK	GSM1369864	ZIPK siRNA 3	Human CASMCs, ZIPK siRNA, 48 h	muscle	coronary artery smooth muscle	CC-2583

GEO2R | Options | Profile graph | R script

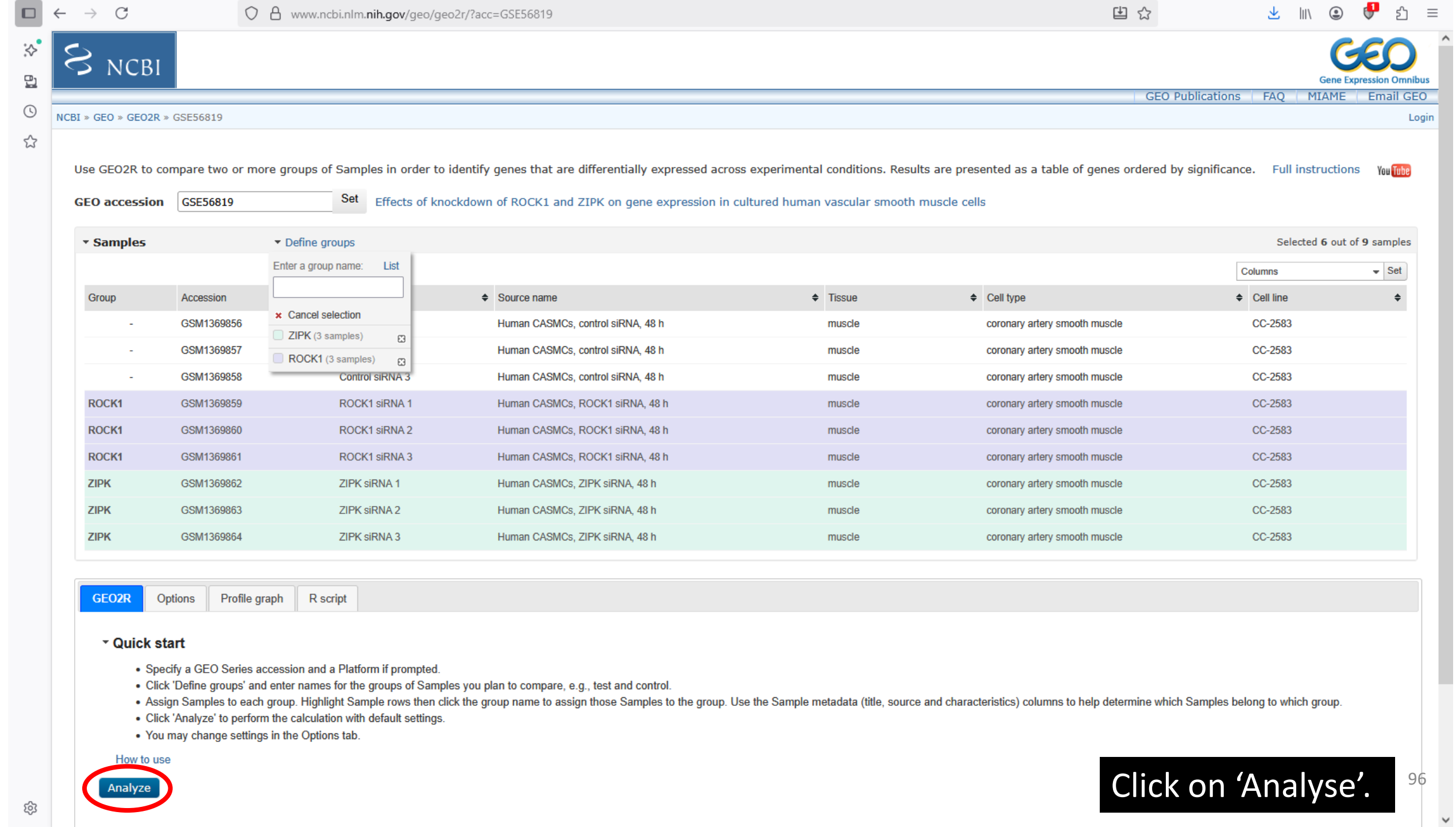
Quick start

- Specify a GEO Series accession and a Platform if prompted.
- Click 'Define groups' and enter names for the groups of Samples you plan to compare, e.g., test and control.
- Assign Samples to each group. Highlight Sample rows then click the group name to assign those Samples to the group. Use the Sample metadata (title, source and characteristics) columns to help determine which Samples belong to which group.
- Click 'Analyze' to perform the calculation with default settings.
- You may change settings in the Options tab.

How to use

Analyze

Repeat preceding steps to create a second group.


www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE56819
NCBI
Gene Expression Omnibus
GEO Publications | FAQ | MIAME | Email GEO
NCBI » GEO » GEO2R » GSE56819
Login
Use GEO2R to compare two or more groups of Samples in order to identify genes that are differentially expressed across experimental conditions. Results are presented as a table of genes ordered by significance.
Full instructions
GEO accession
GSE56819
Set
Effects of knockdown of ROCK1 and ZIPK on gene expression in cultured human vascular smooth muscle cells
Samples
Define groups
Selected 6 out of 9 samples
Enter a group name:
List
Cancel selection
ZIPK (3 samples)
ROCK1 (3 samples)
Columns
Set
Group
Accession
Source name
Tissue
Cell type
Cell line
- GSM1369856 Human CASMCs, control siRNA, 48 h muscle coronary artery smooth muscle CC-2583
- GSM1369857 Human CASMCs, control siRNA, 48 h muscle coronary artery smooth muscle CC-2583
- GSM1369858 Control siRNA 3 Human CASMCs, control siRNA, 48 h muscle coronary artery smooth muscle CC-2583
ROCK1 GSM1369859 ROCK1 siRNA 1 Human CASMCs, ROCK1 siRNA, 48 h muscle coronary artery smooth muscle CC-2583
ROCK1 GSM1369860 ROCK1 siRNA 2 Human CASMCs, ROCK1 siRNA, 48 h muscle coronary artery smooth muscle CC-2583
ROCK1 GSM1369861 ROCK1 siRNA 3 Human CASMCs, ROCK1 siRNA, 48 h muscle coronary artery smooth muscle CC-2583
ZIPK GSM1369862 ZIPK siRNA 1 Human CASMCs, ZIPK siRNA, 48 h muscle coronary artery smooth muscle CC-2583
ZIPK GSM1369863 ZIPK siRNA 2 Human CASMCs, ZIPK siRNA, 48 h muscle coronary artery smooth muscle CC-2583
ZIPK GSM1369864 ZIPK siRNA 3 Human CASMCs, ZIPK siRNA, 48 h muscle coronary artery smooth muscle CC-2583
GEO2R
Options
Profile graph
R script
Quick start
Specify a GEO Series accession and a Platform if prompted.
Click 'Define groups' and enter names for the groups of Samples you plan to compare, e.g., test and control.
Assign Samples to each group. Highlight Sample rows then click the group name to assign those Samples to the group. Use the Sample metadata (title, source and characteristics) columns to help determine which Samples belong to which group.
Click 'Analyze' to perform the calculation with default settings.
You may change settings in the Options tab.
How to use
Analyze
Click on ‘Analyse’.

www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE56819

NCBI

Gene Expression Omnibus

GEO Publications | FAQ | MIAME | Email GEO

NCBI » GEO » GEO2R » GSE56819 Login

Use GEO2R to compare two or more groups of Samples in order to identify genes that are differentially expressed across experimental conditions. Results are presented as a table of genes ordered by significance. Full instructions

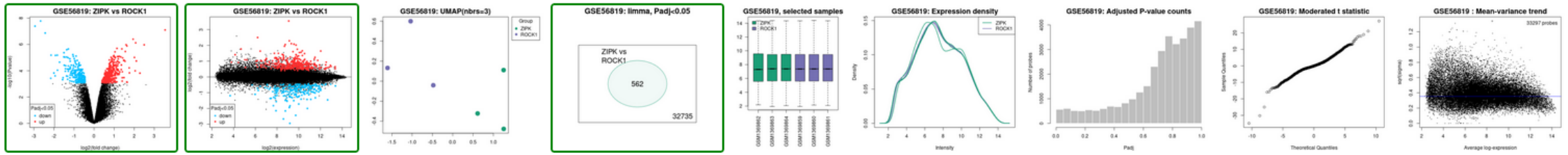
GEO accession GSE56819 Set Effects of knockdown of ROCK1 and ZIPK on gene expression in cultured human vascular smooth muscle cells

▸ **Samples** ▸ Define groups Selected **6** out of **9** samples

GEO2R | Options | Profile graph | R script

Reanalyze if you changed any options.

Visualization

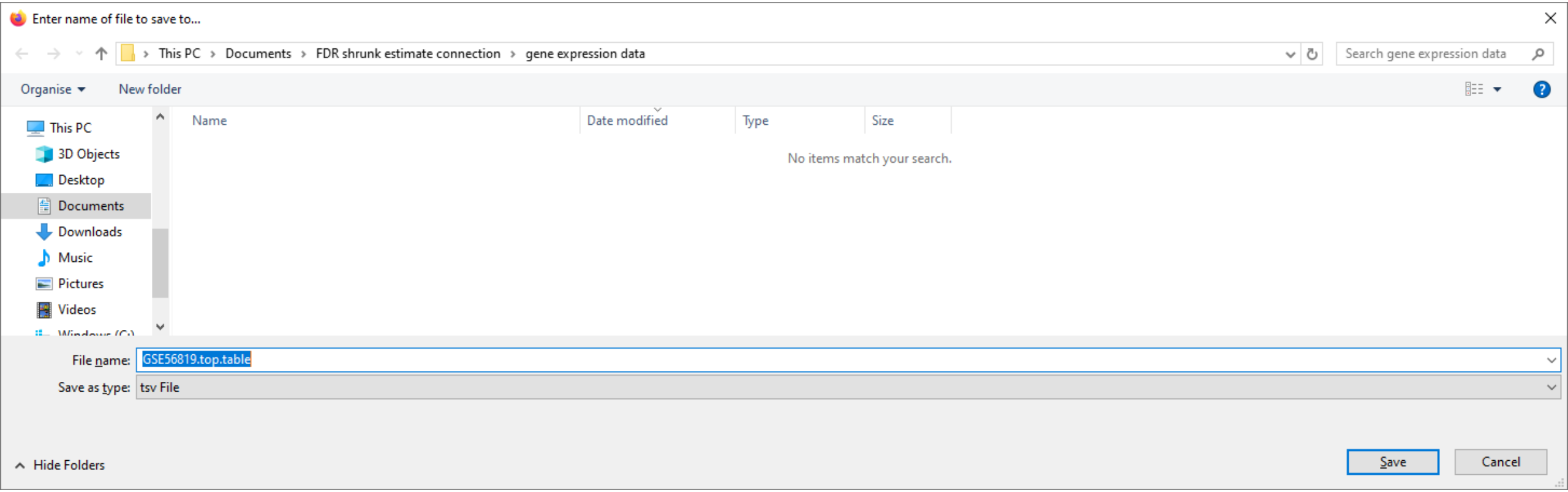


Top differentially expressed genes

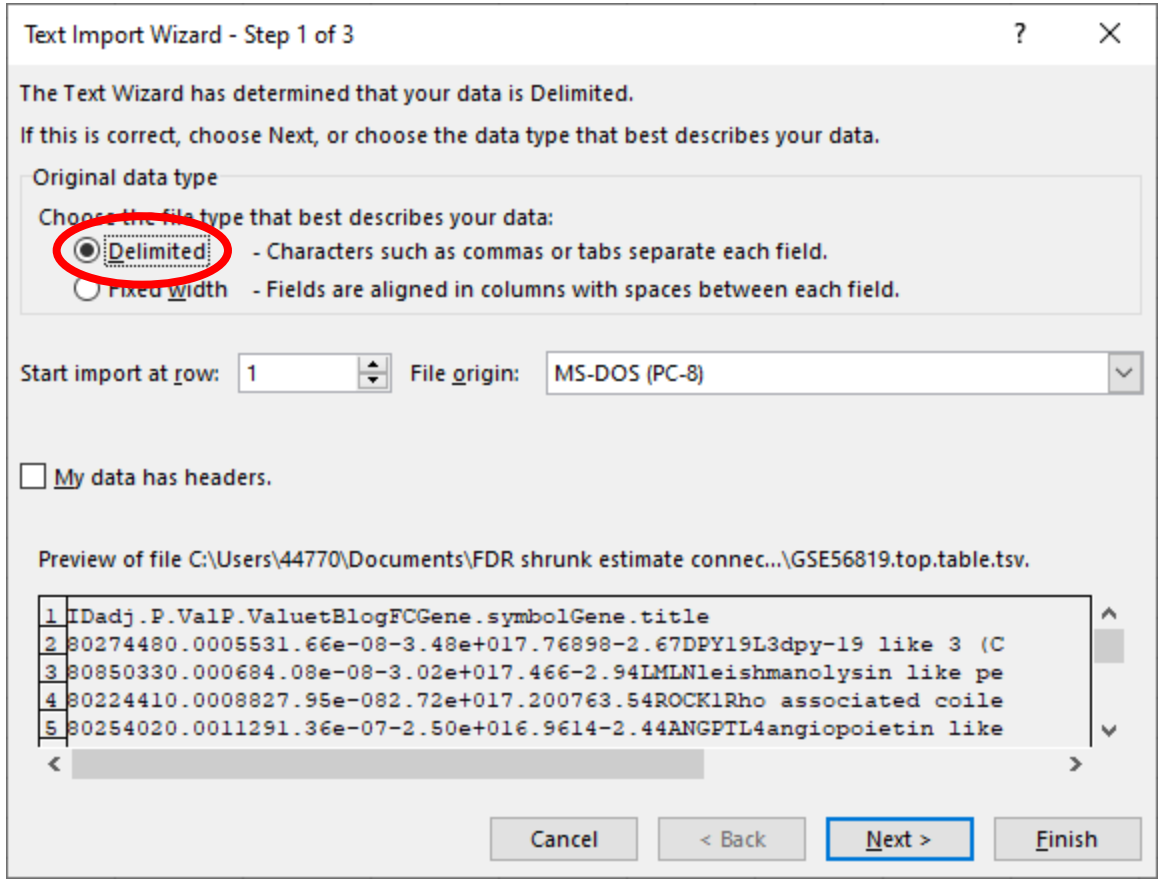
Download full table | Select columns

ID	adj.P.Val	P.Value	t	B	logFC	Gene.symbol	Gene.title
8027448	0.000553	1.66e-08	-34.8	7.77	-2.668	DPY19L3	dpy-19 like 3 (C. elegans)
8085033	0.00068	4.08e-08	-30.19	7.47	-2.945	LMLN	leishmanolysin like peptidase
8022441	0.000882	7.95e-08	27.17	7.2	3.54	ROCK1	Rho associated coiled-coil contai…
8025402	0.001129	1.36e-07	-24.97	6.96	-2.438	ANGPTL4	angiopoietin like 4
7999279	0.002657	3.99e-07	21.04	6.4	1.944	NAGPA	N-acetylglucosamine-1-phospho…
7902476	0.004219	7.60e-07	18.98	6.02	2.072	MIGA1	mitoguardin 1
8163807	0.004483	1.07e-06	17.99	5.8	1.599	PHF19	PHD finger protein 19
8020382	0.004483	1.08e-06	17.96	5.79	2.		
8038407	0.004816	1.30e-06	17.42	5.67	1.		
7962349	0.00623	2.01e-06	16.26	5.37	1.		
8114787	0.00623	2.20e-06	16.02	5.3	1.388	GNPDA1	glucosamine-6-phosphate deami

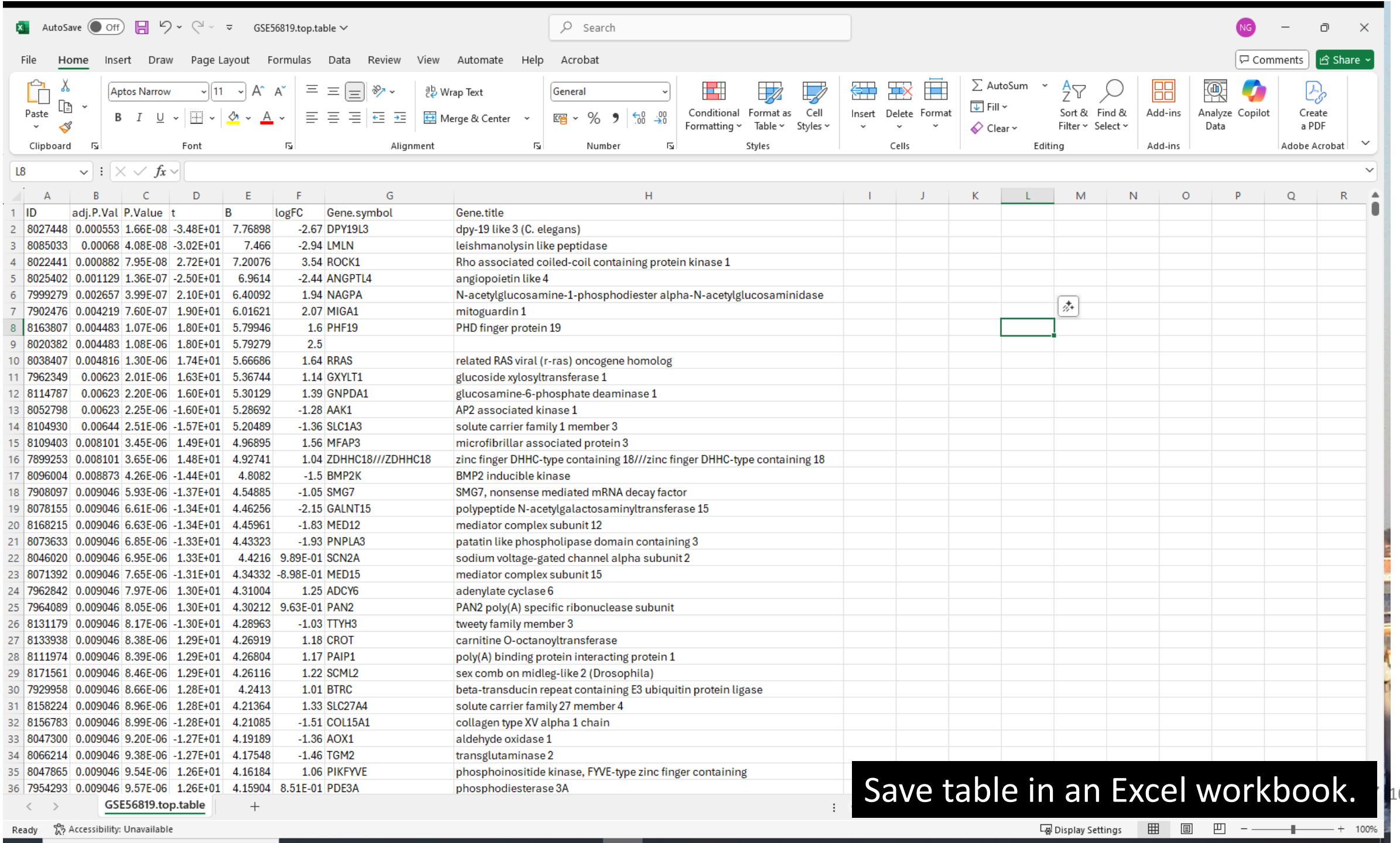
Click on ‘Download full table’.



Save downloaded full table, which maps ID_REF values to Gene.symbol and Gene.title values, to a .tsv file.



Open the .tsv file in Excel.
Accept default data conversions.



	A	B	C	D	E	F	G	H
1	ID	adj.P.Val	P.Value	t	B	logFC	Gene.symbol	Gene.title
2	8027448	0.000553	1.66E-08	-3.48E+01	7.76898	-2.67	DPY19L3	dpy-19 like 3 (C. elegans)
3	8085033	0.00068	4.08E-08	-3.02E+01	7.466	-2.94	LMLN	leishmanolysin like peptidase
4	8022441	0.000882	7.95E-08	2.72E+01	7.20076	3.54	ROCK1	Rho associated coiled-coil containing protein kinase 1
5	8025402	0.001129	1.36E-07	-2.50E+01	6.9614	-2.44	ANGPTL4	angiopoietin like 4
6	7999279	0.002657	3.99E-07	2.10E+01	6.40092	1.94	NAGPA	N-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase
7	7902476	0.004219	7.60E-07	1.90E+01	6.01621	2.07	MIGA1	mitoguardin 1
8	8163807	0.004483	1.07E-06	1.80E+01	5.79946	1.6	PHF19	PHD finger protein 19
9	8020382	0.004483	1.08E-06	1.80E+01	5.79279	2.5		
10	8038407	0.004816	1.30E-06	1.74E+01	5.66686	1.64	RRAS	related RAS viral (r-ras) oncogene homolog
11	7962349	0.00623	2.01E-06	1.63E+01	5.36744	1.14	GXYLT1	glucoside xylosyltransferase 1
12	8114787	0.00623	2.20E-06	1.60E+01	5.30129	1.39	GNPDA1	glucosamine-6-phosphate deaminase 1
13	8052798	0.00623	2.25E-06	-1.60E+01	5.28692	-1.28	AAK1	AP2 associated kinase 1
14	8104930	0.00644	2.51E-06	-1.57E+01	5.20489	-1.36	SLC1A3	solute carrier family 1 member 3
15	8109403	0.008101	3.45E-06	1.49E+01	4.96895	1.56	MFAP3	microfibrillar associated protein 3
16	7899253	0.008101	3.65E-06	1.48E+01	4.92741	1.04	ZDHHC18///ZDHHC18	zinc finger DHHC-type containing 18///zinc finger DHHC-type containing 18
17	8096004	0.008873	4.26E-06	-1.44E+01	4.8082	-1.5	BMP2K	BMP2 inducible kinase
18	7908097	0.009046	5.93E-06	-1.37E+01	4.54885	-1.05	SMG7	SMG7, nonsense mediated mRNA decay factor
19	8078155	0.009046	6.61E-06	-1.34E+01	4.46256	-2.15	GALNT15	polypeptide N-acetylgalactosaminyltransferase 15
20	8168215	0.009046	6.63E-06	-1.34E+01	4.45961	-1.83	MED12	mediator complex subunit 12
21	8073633	0.009046	6.85E-06	-1.33E+01	4.43323	-1.93	PNPLA3	patatin like phospholipase domain containing 3
22	8046020	0.009046	6.95E-06	1.33E+01	4.4216	9.89E-01	SCN2A	sodium voltage-gated channel alpha subunit 2
23	8071392	0.009046	7.65E-06	-1.31E+01	4.34332	-8.98E-01	MED15	mediator complex subunit 15
24	7962842	0.009046	7.97E-06	1.30E+01	4.31004	1.25	ADCY6	adenylate cyclase 6
25	7964089	0.009046	8.05E-06	1.30E+01	4.30212	9.63E-01	PAN2	PAN2 poly(A) specific ribonuclease subunit
26	8131179	0.009046	8.17E-06	-1.30E+01	4.28963	-1.03	TTYH3	tweety family member 3
27	8133938	0.009046	8.38E-06	1.29E+01	4.26919	1.18	CROT	carnitine O-octanoyltransferase
28	8111974	0.009046	8.39E-06	1.29E+01	4.26804	1.17	PAIP1	poly(A) binding protein interacting protein 1
29	8171561	0.009046	8.46E-06	1.29E+01	4.26116	1.22	SCML2	sex comb on midleg-like 2 (Drosophila)
30	7929958	0.009046	8.66E-06	1.28E+01	4.2413	1.01	BTRC	beta-transducin repeat containing E3 ubiquitin protein ligase
31	8158224	0.009046	8.96E-06	1.28E+01	4.21364	1.33	SLC27A4	solute carrier family 27 member 4
32	8156783	0.009046	8.99E-06	-1.28E+01	4.21085	-1.51	COL15A1	collagen type XV alpha 1 chain
33	8047300	0.009046	9.20E-06	-1.27E+01	4.19189	-1.36	AOX1	aldehyde oxidase 1
34	8066214	0.009046	9.38E-06	-1.27E+01	4.17548	-1.46	TGM2	transglutaminase 2
35	8047865	0.009046	9.54E-06	1.26E+01	4.16184	1.06	PIKFYVE	phosphoinositide kinase, FYVE-type zinc finger containing
36	7954293	0.009046	9.57E-06	1.26E+01	4.15904	8.51E-01	PDE3A	phosphodiesterase 3A

Application to simulated data

Simulation of gene expression data

Model:

Control-treatment exp'tal units: $Y = 0 + E$
Active-treatment exp'tal units: $Y = G + E$

$G \sim \mathrm{N}(0, \sigma_G^2)$ $\quad \sigma_G^2 = 0.25, 0.5, 1, 2, 4$
$E \sim \mathrm{N}(0, \sigma_E^2)$ $\quad \sigma_E^2 = 1$

G = treatment effect

Data:

$y_{0ij} = 0 + e_{0ij}$
$y_{1ij} = g_i + e_{1ij}$

$i = 1 \ldots m$
$j = 1 \ldots r$

$m = 100$
$r = 3$
Number of simulations = 100

Realisations g_i , e_{0ij} , e_{1ij} are mutually independent

Two treatments (Control, Active); *m* response variables, $i = 1 \ldots m$;
$2r$ observations (y_{0ij} , y_{1ij}) of each response variable

Analysis of simulated gene expression data

Two treatments (Control, Active); m response variables, $i = 1 \ldots m$; $m = 100$
$2r$ observations (y_{0ij} , y_{1ij}) of each response variable $r = 3$

Two-sample, one-sided t test, performed separately on each response variable

Hypotheses:

H_0: $G \leq 0$
H_1: $G > 0$

$DF_{Resid} = 2 \times (r - 1) = 4$

Results from Simulation 1

Histogram of *p*-values

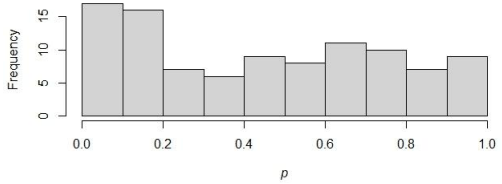
-$\log_{10}$-transformed Q-Q plot

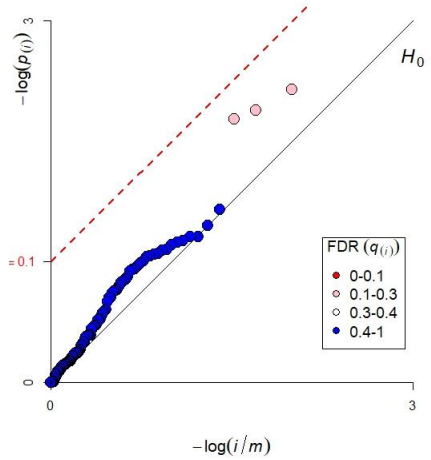
Shrunk v. unshrunk estimate

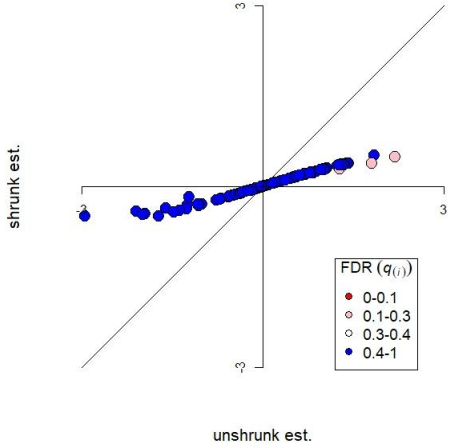
Unshrunk estimate v. true effect

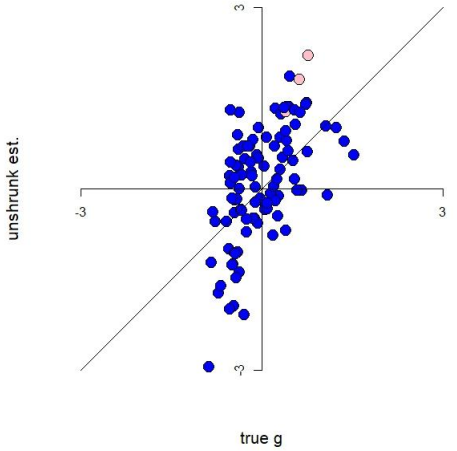
Shrunk estimate v. true effect

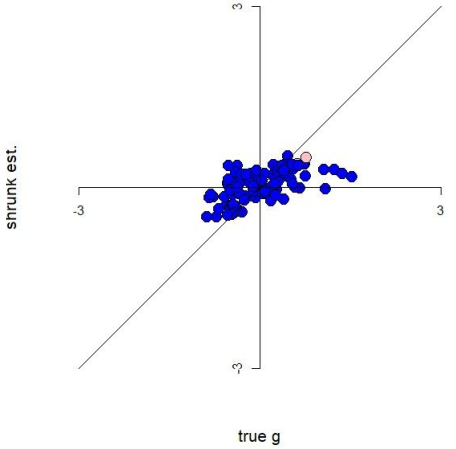
$\sigma_G^2 = 0.25$



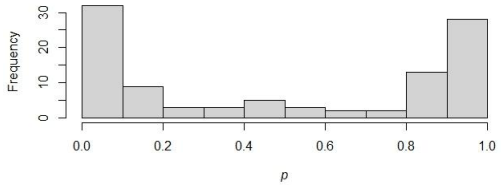


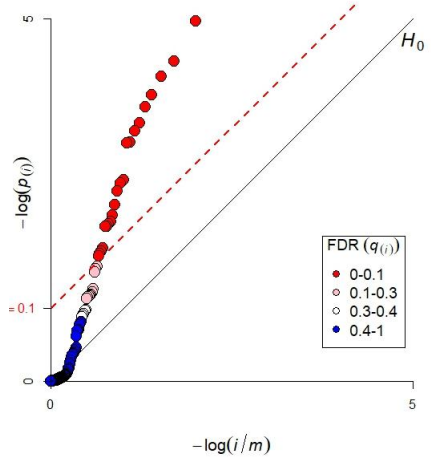


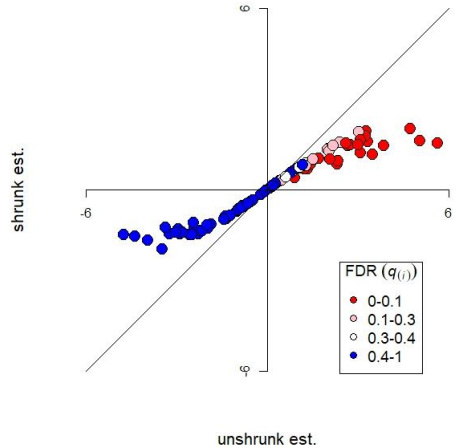


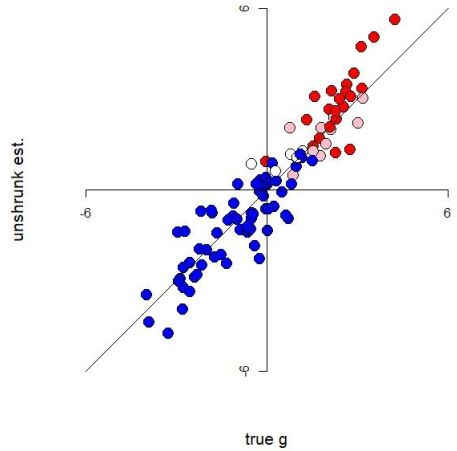


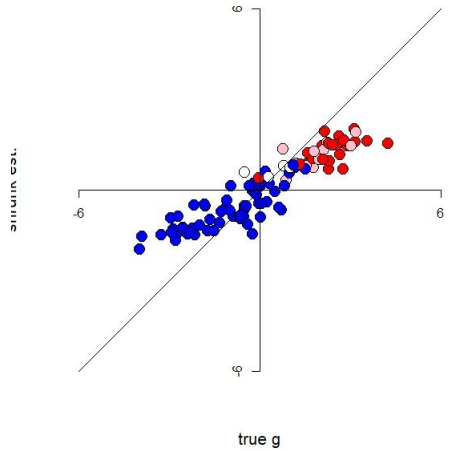
$\sigma_G^2 = 4$











Assessment of performance of shrunk estimates

σ_G^2	Quantile probability	$r_{unshrunk,shrunk}$	$r_{unshrunk,true}$	$r_{shrunk,true}$	$MSE_{unshrunk,true}$	$MSE_{shrunk,true}$
0.25	Min.	0.980	0.332	0.310	0.446	0.108
	0.25	0.989	0.483	0.476	0.610	0.169
	0.5	0.992	0.521	0.512	0.670	0.186
	0.75	0.994	0.565	0.562	0.715	0.209
	Max.	0.998	0.717	0.706	0.913	0.255
1	Min.	0.962	0.640	0.650	0.490	0.338
	0.25	0.982	0.748	0.736	0.600	0.434
	0.5	0.986	0.767	0.757	0.682	0.470
	0.75	0.990	0.795	0.782	0.742	0.519
	Max.	0.995	0.858	0.859	0.921	0.655
4	Min.	0.949	0.886	0.860	0.511	0.854
	0.25	0.966	0.917	0.884	0.623	1.171
	0.5	0.971	0.923	0.894	0.669	1.254
	0.75	0.976	0.931	0.908	0.725	1.384
	Max.	0.984	0.951	0.927	0.835	1.765

$$\text{MSE} = \sum_{i=1}^{m} \frac{(\hat{g}_i - g_i)^2}{m}$$

where

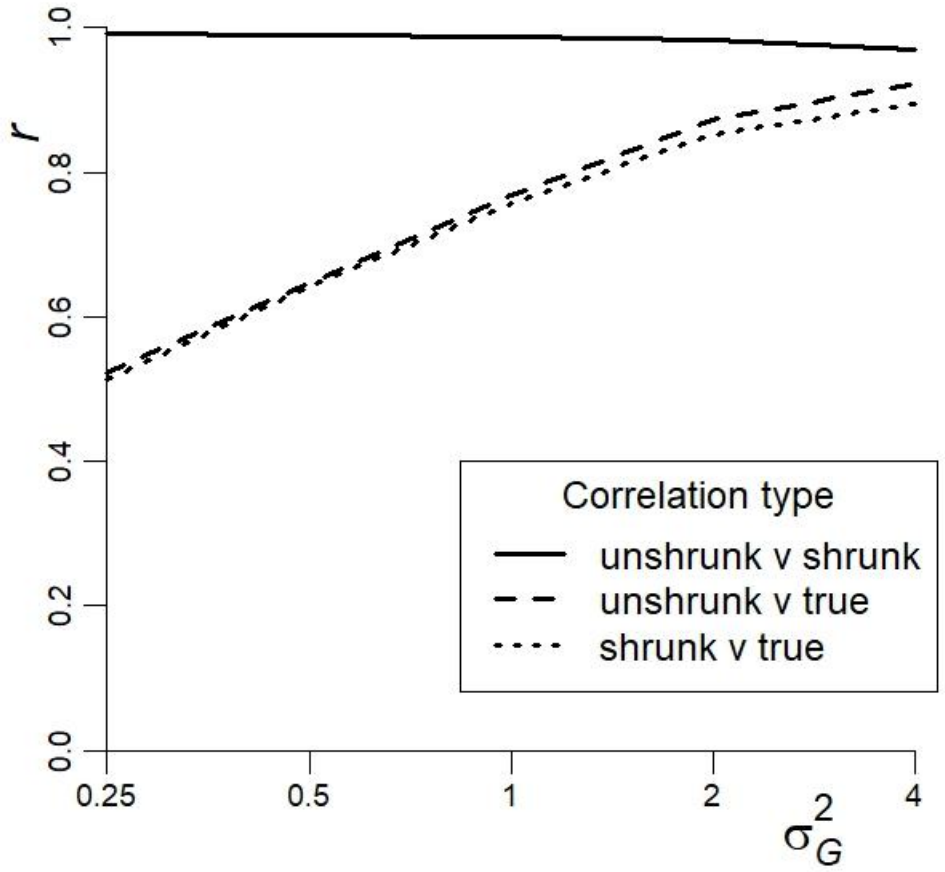
g_i = realisation of G in test i

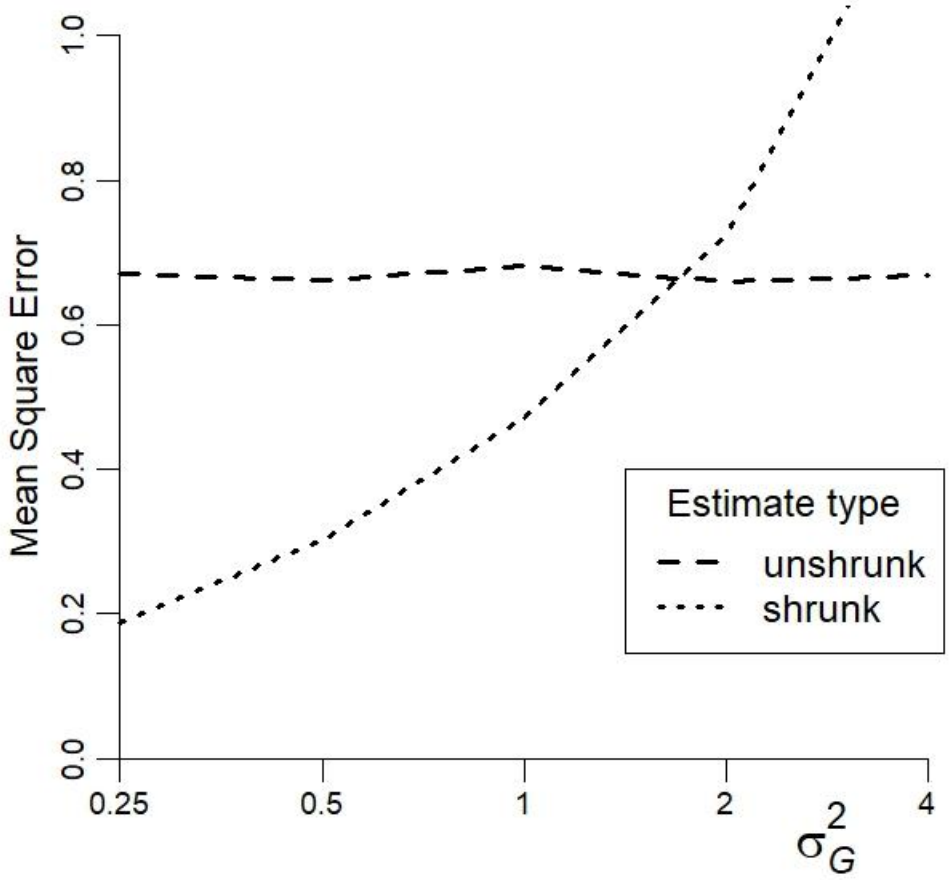
- When σ_G^2 is small, shrinkage reduces MSE
- When σ_G^2 is large, shrinkage increases MSE

- Bias-variance trade-off?

cont'd.../

Assessment of performance of shrunk estimates/…cont'd.





cont'd…/

Assessment of performance of shrunk estimates/…cont'd.

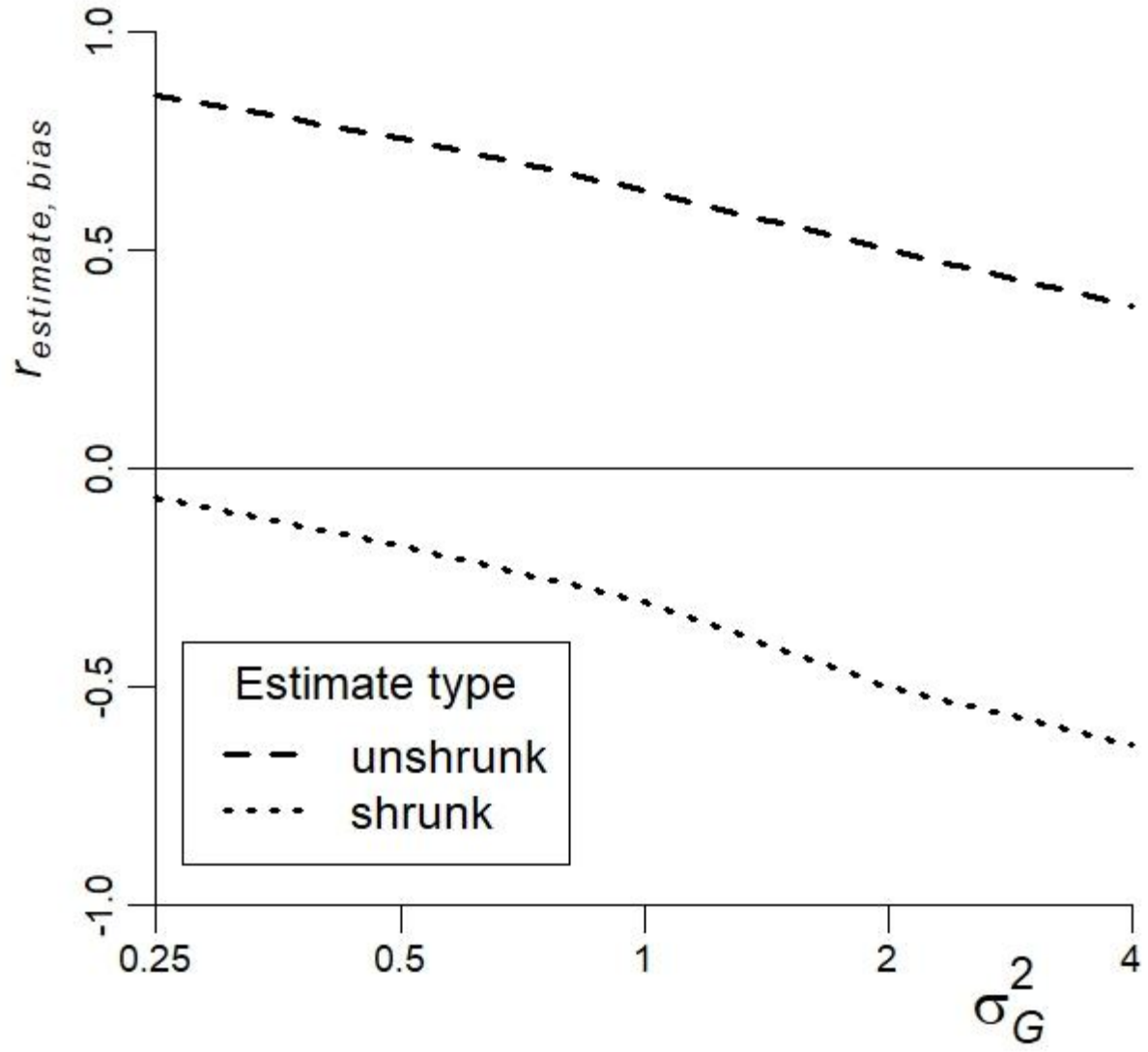
bias =
estimate – true effect

Estimate may be either unshrunk or shrunk

σ_G^2	Quantile probability	$r_{estimate,bias}$	
		Unshrunk est.	Shrunk est.
0.25	Min.	0.757	-0.508
	0.25	0.833	-0.181
	0.5	0.856	-0.067
	0.75	0.872	0.066
	Max.	0.916	0.417
1	Min.	0.448	-0.570
	0.25	0.594	-0.386
	0.5	0.638	-0.308
	0.75	0.664	-0.238
	Max.	0.734	-0.098
4	Min.	0.165	-0.731
	0.25	0.305	-0.661
	0.5	0.370	-0.633
	0.75	0.420	-0.593
	Max.	0.597	-0.510

cont'd…/

Assessment of performance of shrunk estimates/…cont'd.



- Shrinkage on the basis of

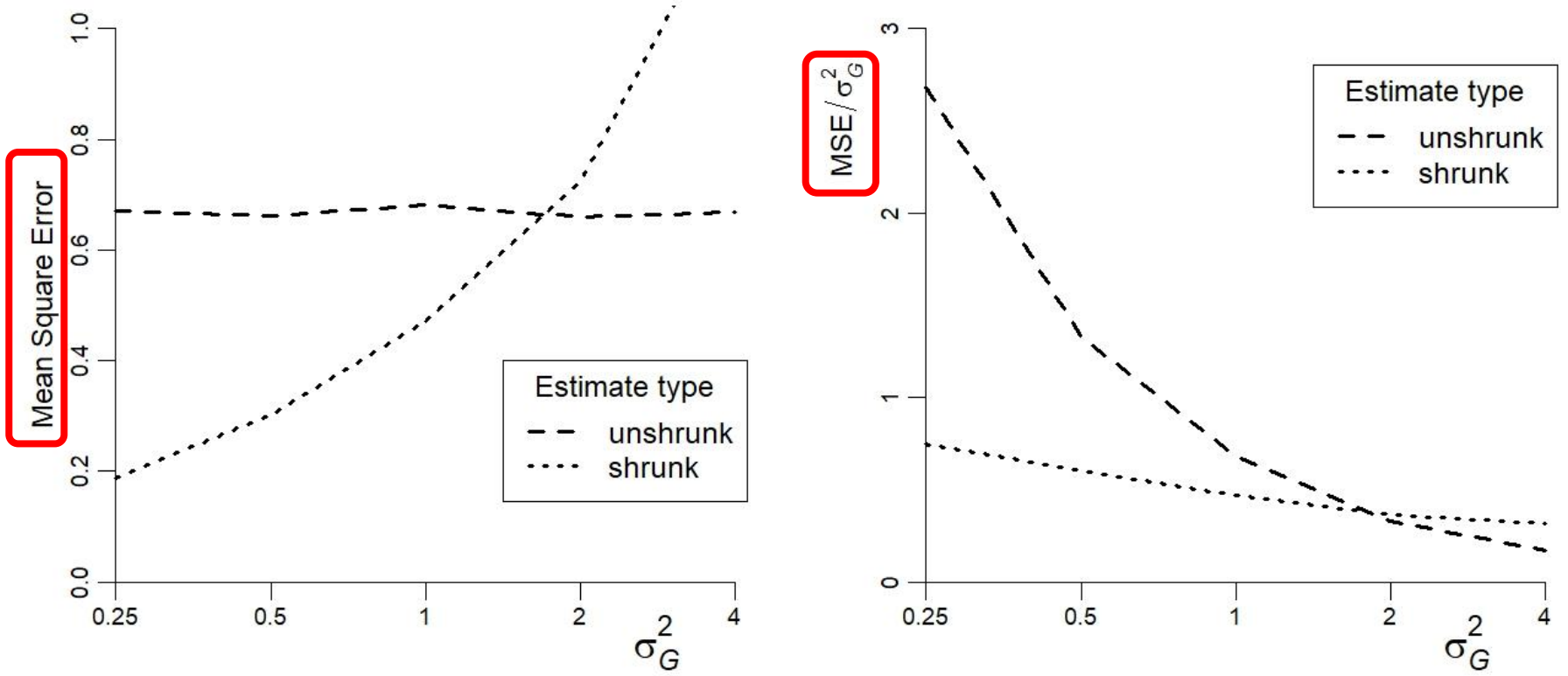
$$\text{S. F.} = \frac{\widehat{\text{var}}(Z_{\text{obs}}) - 1}{\widehat{\text{var}}(Z_{\text{obs}})}$$

over-compensates for bias…

- …especially when $\boldsymbol{\sigma_G^2}$ is large…

cont'd…/

Assessment of performance of shrunk estimates/…cont'd.



- …but overcompensation of shrinkage does not matter much when $\boldsymbol{\sigma_G^2}$ is large.