The heterosis problem: a comparison of Eric's method with edgeR, baySeq, and ShrinkBayes

Will Landau, Eric Mittman

Iowa State University

November 15, 2014

The heterosis problem: a comparison of Eric's method with edgeR, baySeq, and ShrinkBayes

Will Landau, Eric Mittman

The workflow

Simulated

edgeR baySeq ShrinkBayes

The contest

ROC (receiver operating characteristic) curves The results

The workflow

Simulated data

The contenders

baySeq ShrinkRay

ShrinkBayes

The contest

ROC (receiver operating characteristic) curve

The heterosis problem: a comparison of Eric's method with edgeR, baySeq, and ShrinkBayes

Will Landau, Eric Mittman

The workflow

Simulated

The contende edgeR baySeq

The contest

ROC (receiver operating characteristic) curves

Simulation workflow

- Simulate 30 datasets as above:
 - 10 datasets with 4 samples (libraries, columns, etc.) per group
 - ▶ 10 with 8 per group
 - ▶ 10 with 16 per group
- For each simulated dataset, test for heterosis with
 - empirical Bayes with STAN (Eric's method)
 - ▶ edgeR
 - baySeq
 - ▶ ShrinkBayes
- Compare methods with ROC curves

The heterosis problem: a comparison of Eric's method with edgeR, baySeq, and ShrinkBayes

Will Landau, Eric Mittman

The workflow

Simulated

The contende edgeR baySeq

The contest

ROC (receiver operating characteristic) curves The results

The workflow

Simulated data

The contenders edgeR baySeq

The contest

ROC (receiver operating characteristic) curves

The heterosis problem: a comparison of Eric's method with edgeR, baySeq, and ShrinkBayes

Will Landau, Eric Mittman

The workflow

Simulated data

The contender edgeR baySeq

ShrinkBayes

ROC (receiver operating

The results

Simulate heterosis data with known heterosis genes

Parent (1) Parent (2) Hybrid (3) Truth **HPH** Feature 1 700190018251860 **HPH** Feature 2 18 50 501 400 90 100 225 15 300 106 200 400 70 279 100 123 0 Feature 3 LPH Feature 4 893 400 760 901 100d 513 760 580 5 Feature 902 912 999 825 819 761 800 465 0 10 25000

The heterosis problem: a comparison of Eric's method with edgeR, baySeq, and ShrinkBayes

Will Landau, Eric Mittman

Simulated data

edgeR

Apply edgeR to real data to get simulation parameters

Normalization factors



Main effects and dispersions

Parent (1)	Parent (2)	Hybrid (3)	Dispersion	
$\mu_{1,1}$	$\mu_{1,2}$	$\mu_{1,3}$	ϕ_1	
$\mu_{2,1}$	$\mu_{2,2}$	$\mu_{2,3}$	ϕ_2	
$\mu_{27888,1}$	$\mu_{27888,2}$	$\mu_{27888,3}$	φ ₂₇₈₈₈	

The heterosis problem: a comparison of Eric's method with edgeR, baySeq, and ShrinkBayes

Will Landau, Eric Mittman

The workflow

Simulated data

The contenders

edgeR baySeq ShrinkBayes

The conte

OC (receiver perating haracteristic) curves

he results

Truth: which genes have heterosis?

Feature 1	$I(\mu_{1,3} > \max(\mu_{1,1}, \mu_{1,2}) \text{ or } < \min(\mu_{1,1}, \mu_{1,2}))$
Feature 2	$I(\mu_{2,3} > \max(\mu_{2,1}, \mu_{2,2}) \text{ or } < \min(\mu_{2,1}, \mu_{2,2}))$
Feature 27888	$I(\mu_{27888,3} > \max(\mu_{27888,1}, \mu_{27888,2}) \text{ or } < \min(\mu_{2,1}, \mu_{27888,2}))$

The heterosis problem: a comparison of Eric's method with edgeR, baySeq, and ShrinkBayes

Will Landau, Eric Mittman

The workflow

Simulated data

The contenders
edgeR
baySeq
ShrinkBayes

The contest

ROC (receiver
operating
characteristic) curves

iid negative binomial counts (parent 1)

$NB(e^{c_1+\mu_{1,1}},\phi_1)$	$NB(e^{c_2+\mu_{1,1}},\phi_1)$	$NB(e^{c_3+\mu_{1,1}},\phi_1)$	$NB(e^{c_4+\mu_{1,1}},\phi_1)$
$NB(e^{c_1+\mu_{2,1}},\phi_2)$	$NB(e^{c_2+\mu_{2,1}},\phi_2)$	$NB(e^{c_3+\mu_{2,1}},\phi_2)$	$NB(e^{c_4+\mu_{2,1}},\phi_2)$
$NB(e^{c_1 + \mu_{27888,1}}, \phi_{27888})$	${\rm NB}(e^{c_2+\mu_{27888,1}},\phi_{27888})$	$\text{NB}(e^{c_3 + \mu_{27888,1}}, \phi_{27888})$	$NB(e^{c_4 + \mu_{27888,1}}, \phi_{27888})$

The heterosis problem: a comparison of Eric's method with edgeR, baySeq, and ShrinkBayes

Will Landau, Eric Mittman

The workflow

Simulated data

The contenders edgeR baySeq ShrinkBayes

The contest

ROC (receiver operating characteristic) curves The results

Remove low-count rows to get 25000 features

Parent (1) Parent (2) Hybrid (3) Truth **HPH** Feature 1 700|900|825|860 **HPH** Feature 2 501 400 Feature 3 100 225 15 300 106 200 400 70 279 100 123 0 LPH 893 400 760 901 100d 513 760 580 Feature 4 Feature 902 912 999 825 819 761 800 465 0 25000

The heterosis problem: a comparison of Eric's method with edgeR, baySeq, and ShrinkBayes

Will Landau, Eric Mittman

The workflow

Simulated data

l he contend: edgeR baySeq ShrinkBayes

The cont

ROC (receiver operating characteristic) curves
The results

The contenders

The heterosis problem: a comparison of Eric's method with edgeR, baySeq, and ShrinkBayes

Will Landau, Eric

The contenders

- Fit a loglinear model to estimate main effects $\mu_{f,t}$
 - Feature f = 1, ..., 25000
 - Treatment group 1 (parent), 2 (parent), 3 (hybrid)
- Likelihood ratio tests to get p-values $p_{f,1}$, $p_{f,2}$

$$H_{0,1}: \mu_{f,3} = \mu_{f,1}$$
 $H_{a,1}: \mu_{f,3} \neq \mu_{f,1}$
 $H_{0,2}: \mu_{f,3} = \mu_{f,2}$ $H_{a,2}: \mu_{f,3} \neq \mu_{f,2}$

Final p-value	if
$p_{f,2}/2$	$\begin{split} \widehat{\mu}_{f,3} < \widehat{\mu}_{f,1} &\leq \widehat{\mu}_{f,2} \text{ or } \widehat{\mu}_{f,3} > \widehat{\mu}_{f,1} \geq \widehat{\mu}_{f,2} \\ \widehat{\mu}_{f,3} < \widehat{\mu}_{f,2} &\leq \widehat{\mu}_{f,1} \text{ or } \widehat{\mu}_{f,3} > \widehat{\mu}_{f,2} \geq \widehat{\mu}_{f,1} \\ \widehat{\mu}_{f,1} &\leq \widehat{\mu}_{f,3} \leq \widehat{\mu}_{f,2} \text{ or } \widehat{\mu}_{f,2} \leq \widehat{\mu}_{f,3} \leq \widehat{\mu}_{f,1} \end{split}$
1	$\widehat{\mu}_{f,1} \leq \widehat{\mu}_{f,3} \leq \widehat{\mu}_{f,2} \text{ or } \widehat{\mu}_{f,2} \leq \widehat{\mu}_{f,3} \leq \widehat{\mu}_{f,1}$

The heterosis problem: a comparison of Eric's method with edgeR, baySeq, and ShrinkBayes

Will Landau, Eric Mittman

edgeR

- ▶ Estimate main effects $\mu_{f,t}$ using edgeR.
- Calculate the posterior probability that each feature satisfies:

Model	Constraint
M_1	All $\mu_{f,t}$'s equal
M_2	$\mu_{f,1} = \mu_{f,2}$
M_3	$\mu_{f,1} = \mu_{f,3}$
M_4	$\mu_{f,2} = \mu_{f,3}$
M_5	All $\mu_{f,t}$'s distinct

▶ Final posterior probabilities of heterosis:

Posterior probability	if
Posterior probability 1 $P(M_3 \mid {\sf data}) + P(M_5 \mid {\sf data})$	$\widehat{\mu}_{f,1} \leq \widehat{\mu}_{f,3} \leq \widehat{\mu}_{f,2}$ or
	$\widehat{\mu}_{f,2} \leq \widehat{\mu}_{f,3} \leq \widehat{\mu}_{f,1}$
$P(M_3 \mid data) + P(M_5 \mid data)$	otherwise

The heterosis problem: a comparison of Eric's method with edgeR, baySeq, and ShrinkBayes

Will Landau, Eric Mittman

The workflow

Simulated

edgeR baySeq ShrinkBayes

- ▶ Built on inla (integrated nested Laplace approximation).
- empirical Bayes with a zero-inflated NB likelihood and normal priors.
- ▶ I reparameterize

$$\begin{split} \phi_f &= \frac{\mu_{f,1} + \mu_{f,2}}{2} \qquad \text{(parental mean)} \\ \alpha_f &= \frac{\mu_{f,2} - \mu_{f,1}}{2} \qquad \text{(half parental difference)} \\ \delta_f &= \mu_{f,3} - \frac{\mu_{f,1} + \mu_{f,2}}{2} \qquad \text{(hybrid effect)} \end{split}$$

The heterosis problem: a comparison of Eric's method with edgeR, baySeq, and ShrinkBayes

Will Landau, Eric Mittman

The workflow

Simulated

The contenders
edgeR
baySeq
ShrinkBayes

ShrinkBayes

ϕ_{f}	α_{f}	δ_f
parental mean	half parental difference	hybrid effect

Use contrasts to calculate final posterior probabilities of heterosis:

Posterior probability	if
0	$ \delta_f < \alpha_f $, otherwise:
$P(\delta_f + lpha_f > 0 \mid ext{data}) \ P(\delta_f - lpha_f > 0 \mid ext{data}) \ P(\delta_f - lpha_f < 0 \mid ext{data})$	$\delta_f > -\alpha_f$
$P(\delta_f - lpha_f > 0 \mid data)$	$\delta_f > \alpha_f$
$P(\delta_f - lpha_f < 0 \mid data)$	$\delta_f < \alpha_f$
$P(\delta_f + \alpha_f < 0 \mid data)$	$\delta_f < -\alpha_f$

The heterosis problem: a comparison of Eric's method with edgeR, baySeq, and ShrinkBayes

Will Landau, Eric Mittman

The workflow

Jiiilulatec

The contender edgeR baySeq ShrinkBayes

The workflow

Simulated data

The contenders

edgeR

baySeq

ShrinkBayes

The contest

ROC (receiver operating characteristic) curves

The heterosis problem: a comparison of Eric's method with edgeR, baySeq, and ShrinkBayes

Will Landau, Eric Mittman

The workflow

Simulated

edgeR
baySeq

The contest

ROC (receiver operating characteristic) curves

- \triangleright N_{true} heterosis features, N_{false} null features.
- ► Results of testing each feature for heterosis (25000 columns here):

pval	0.802	0.935	0.539	0.001		0.500	0.603
truth	0	0	1	1		1	0

Sort table by p-value (or other binary classifier)

pval	0.000	0.001	0.005	0.006		0.901	1.000
truth	1	1	0	1		0	0

The heterosis problem: a comparison of Eric's method with edgeR, baySeq, and ShrinkBayes

Will Landau, Eric Mittman

The workflow

Simulated

edgeR baySeq ShrinkBaves

Calculating false positive rate (FPR) and true positive rate (TPR)

▶ In practice, we would declare the lowest-p-value features to have heterosis.

pval	0.000	0.001	0.005	0.006	 0.901	1.000
truth	1	1	0	1	 0	0

▶ With 2 heterosis genes and 1 null gene,

$$FPR = \frac{1}{N_{false}}$$
 $TPR = \frac{2}{N_{true}}$

Repeat for multiple cutoffs to get multiple (FPR, TPR) pairs.

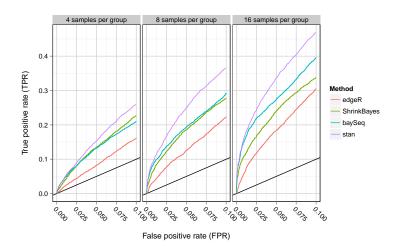
The heterosis problem: a comparison of Eric's method with edgeR, baySeq, and ShrinkBayes

Will Landau, Eric

ROC (receiver operating characteristic) curves

/Will Landau, Eric Mittman (Iowa State UniveThe heterosis problem: a comparison of Eric's

Example ROC curves



The heterosis problem: a comparison of Eric's method with edgeR, baySeq, and ShrinkBayes

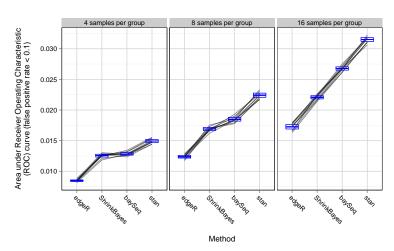
Will Landau, Eric Mittman

The workflow

Simulate

The contende
edgeR
baySeq
ShrinkBayes

Areas under ROC curves



The heterosis problem: a comparison of Eric's method with edgeR, baySeq, and ShrinkBayes

Will Landau, Eric Mittman

The workflow

Simulate

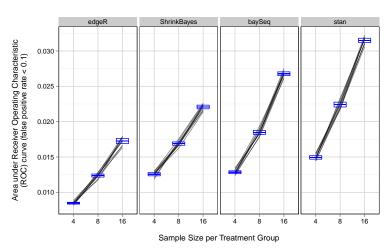
edgeR
baySeq
ShrinkBayes

The contest

ROC (receiver operating characteristic) curves

The results

Areas under ROC curves



The heterosis problem: a comparison of Eric's method with edgeR, baySeq, and ShrinkBayes

Will Landau, Eric Mittman

The workflow

Simulate

The contend edgeR baySeq ShrinkBayes