Aitoumellout (56)

Statistics for Genomic Data Analysis, Final Data Analysis Reports

2 **Background/Introduction 5%** incomplete - also clustering which chip? # probe sets?

9 **Quality assessment 15%** use more paragraphing

NUSE not defined you need to introduce model before weights

2/6 Description of quality assessment

3/3 Excluded poor quality slides (if relevant)

4/4 Graphical support for decisions horizontal line at 1.05 NUSE Boxplots

- square weights plots

3 **Normalization 10%**

incomplete

2/6 Description of normalization/expression measure used – what is the bg corr?

0/2 Graphical support for decisions - quant norm on bg corr

1/2 Appropriateness of normalization - PM* also quant norm PM

14 **Statistical analyses: Differential Expression 25%**

emp Bayes + hyps mathematically

3/5 Description of model (including design matrix, contrasts if necessary) – write parameterization mathematically

3/5 Multiple testing/adjusted p-values given OK, but use BH in general

2/5 Appropriate basis for gene ranking (and explanation) (**)

1/5 Estimated number of differentially expressed genes (and explanation) (***)

5/5 Volcano plot, other appropriate plots

14 **Statistical analyses: Cluster Analysis 20%** measures + silhouette

incomplete + define - use 1-cor, not Euclidean

4/10 Description of clustering algorithm used

5/5 Relevant graphics

5/5 Interpretation of resulting clusters (ok)

5 (*) **Conclusions 10%** not 'confirm effectiveness'

Other

2 **Gene list 5%**
- make 'pretty' column headings
- too many digits
- include probe set ID
- all on 1 page

5 **Reproducible R code 5%**

2 **Overall Presentation 5%**
- name chips P0, P1, ... (no '.cel')
- plot labels too small/illegible
- Don't need R details

(*) + make recap
+ DE & cluster results in separate paragraphs

- cite refs
- if you use MDS you must explain it

(**) Explain the procedure(s) mathematically, not as a list of R fns

(***) only based on adj p, NOT with FC constraint

Bachar 64

Statistics for Genomic Data Analysis, Final Data Analysis Reports

5 **Background/Introduction 5%**

Quality assessment 15% log 2 PM intensity

11
5/8 ⊛ Description of quality assessment - Define weighting / fitting procedure
3/3 Excluded poor quality slides (if relevant)
3/4 Graphical support for decisions - interpret figure!

Normalization 10%

7
3/6 Description of normalization/expression measure used - incomplete: what is the bg-corr?
2/2 Graphical support for decisions
2/2 Appropriateness of normalization - what is quant norm specifically

Statistical analyses: Differential Expression 25%

14
4/5 Description of model (including design matrix, contrasts if necessary) + emp Bayes (mathematically)
3/5 Multiple testing/adjusted p-values given + BH & FDR mathematically
2/5 Appropriate basis for gene ranking (and explanation) (**)
0/5 Estimated number of differentially expressed genes (and explanation)
5/5 Volcano plot, other appropriate plots

Statistical analyses: Cluster Analysis 20%

8
4/10 Description of clustering algorithm used incomplete - carefully describe algorithm + measures;
2/5 Relevant graphics + heatmap
2/5 Interpretation of resulting clusters which genes?

7 **Conclusions 10%** - more explicit recap

Other

4 **Gene list 5%** - make 'pretty' column names

5 **Reproducible R code 5%**

3 **Overall Presentation 5%** - name chips P0, P1, ... (no '.cel')
- label all chips in plots

⊛ M-est is Maximum Likelihood-like estimation (it is like MLE, but is NOT MLE - no distributional form IS assumed)

(**) only based on adj p, NOT with FC constaint
- also cite BH + Bolstad + Brettscheider et al. (+ only cite PRIMARY refs)

Lien 49

Statistics for Genomic Data Analysis, Final Data Analysis Reports

3 **Background/Introduction 5%** which chip? # probe sets?

Quality assessment 15% what does 'outliers' mean in this context?

7 You need to introduce MODEL before weights

2/8 Description of quality assessment

3/3 Excluded poor quality slides (if relevant) – median NUSE > 1.05

2/4 Graphical support for decisions – + weights pseudo-images

– All NUSE boxplots in same graph

Normalization 10%

1. bg 2. quant norm 3. probe set summarization

7 3/6 Description of normalization/expression measure used – what is quantile norm specifically

2/2 Graphical support for decisions

2/2 Appropriateness of normalization

12 **Statistical analyses: Differential Expression 25%**

+ emp Bayes

2/5 Description of model (including design matrix, contrasts if necessary) – mathematically

0/5 Multiple testing/adjusted p-values given – explain procedure + hyp tests mathematically

5/5 Appropriate basis for gene ranking (and explanation)

0/5 Estimated number of differentially expressed genes (and explanation)

5/5 Volcano plot, other appropriate plots

Statistical analyses: Cluster Analysis 20%

5 0/10 Description of clustering algorithm used

5/5 Relevant graphics

0/5 Interpretation of resulting clusters

5 **Conclusions 10%** – more recap + DE & cluster results in separate paragraphs

Other – point size too small

2 **Gene list 5%** – make 'pretty' column labels – too many digits

5 **Reproducible R code 5%**

3 **Overall Presentation 5%** – name chips P0, P1, ... (no '.cel')

– plot labels too small

– cite refs

Scialanga (63)

Statistics for Genomic Data Analysis, Final Data Analysis Reports

3 **Background/Introduction 5%** which chip? # probe sets?

5 **Quality assessment 15%**

2/8 NUSE not defined need model before weights
Description of quality assessment
(*) 0/3 Excluded poor quality slides (if relevant)
3/4 Graphical support for decisions plot weights pseudo images (not resids)
add horizontal line at 1.05 NUSE boxplots

8 **Normalization 10%**

4/6 Description of normalization/expression measure used $\log_2 PM^*$ ← bg corr/quant norm PM
2/2 Graphical support for decisions — what is the bg corr?
2/2 Appropriateness of normalization

17 **Statistical analyses: Differential Expression 25%**

4/5 Description of model (including design matrix, contrasts if necessary) + emp Bayes (mathematically)
3/5 Multiple testing/adjusted p-values given — hyps + procedure mathematically
2/5 Appropriate basis for gene ranking (and explanation)
3/5 Estimated number of differentially expressed genes (and explanation) (**)
5/5 Volcano plot, other appropriate plots

12 **Statistical analyses: Cluster Analysis 20%**

4/10 incomplete description of algorithms + measures
Description of clustering algorithm used use 1-cor (not Euclidean)
5/5 Relevant graphics
3/5 Interpretation of resulting clusters — say a little more

8 **Conclusions 10%** 'successfully'? 'robust'?

Other

3 **Gene list 5%** - make 'pretty' column labels
- too many digits

5 **Reproducible R code 5%**

2 **Overall Presentation 5%** - name chips P0, P1, etc (no '.cel')
- no raw R output (make 'pretty')
- plot labels too small

(*) exclude if median NUSE > 1.05

(**) only based on adj p, NOT with FC constraint

- cite refs

Sikkong ⑤⑥

Statistics for Genomic Data Analysis, Final Data Analysis Reports

5 **Background/Introduction 5%**

(*)

Quality assessment 15%

2/4 - need to introduce model before weights - NUSE not defined

7

2/4 Description of quality assessment
3/3 Excluded poor quality slides (if relevant)
2/4 Graphical support for decisions - weights pseudo-images (not resids) - square shape - add horizontal line at 1.05 NUSE

7 **Normalization 10%**

3/6 Description of normalization/expression measure used - what is the boxplots
2/2 Graphical support for decisions
2/2 Appropriateness of normalization - what is bg corr, quant norm specifically

Statistical analyses: Differential Expression 25%

16

4/5 Description of model (including design matrix, contrasts if necessary) + emp Bayes (mathematically)
2/5 Multiple testing/adjusted p-values given - hyp tests + procedure + FDR mathematically
2/5 Appropriate basis for gene ranking (and explanation) (**)
3/5 Estimated number of differentially expressed genes (and explanation)
5/5 Volcano plot, other appropriate plots

4 **Statistical analyses: Cluster Analysis 20%**

2/10 Description of clustering algorithm used
0/5 Relevant graphics
2/5 Interpretation of resulting clusters

Very incomplete + use 1-cor (not Euclidean); carefully describe algorithm + measures; which genes?

8 **Conclusions 10%** - more recap

Other

2 **Gene list 5%** - incomplete - looks like wrong parameter/contrast - make 'pretty' column labels

5 **Reproducible R code 5%**

2 **Overall Presentation 5%** - Don't need R fn details - name chips P0, P1 (no '.cel') - plot labels too small

(*) exclude chips with median NUSE > 1.05

(**) only based on adj p, NOT FC constraint

- more refs

Good (59)

Statistics for Genomic Data Analysis, Final Data Analysis Reports

2 **Background/Introduction 5%** Use paragraphing, don't need B-stat which chip? # probe sets?

12 **Quality assessment 15%**
where do the weights come from/what are they?
5/8 Description of quality assessment ⊛
3/3 Excluded poor quality slides (if relevant)
4/4 Graphical support for decisions square weights plots

7 **Normalization 10%**
probe-level PM measurements
3/6 Description of normalization/expression measure used – incomplete: need probe set summarization – what is → quant norm
2/2 Graphical support for decisions
2/2 Appropriateness of normalization
(run-on sentences – a little hard to follow)

12 **Statistical analyses: Differential Expression 25%**
5/5 Description of model (including design matrix, contrasts if necessary)
2/5 Multiple testing/adjusted p-values given – hyps + BH procedure specifically + FDR mathematically
(***) 2/5 Appropriate basis for gene ranking (and explanation)
0/5 Estimated number of differentially expressed genes (and explanation)
3/5 Volcano plot, other appropriate plots – show y-axis = $-\log_{10}$ adj p (not B)

11 **Statistical analyses: Cluster Analysis 20%**
4/10 Description of clustering algorithm used – incomplete – which genes? explain algorithm(s) (**)
5/5 Relevant graphics
2/5 Interpretation of resulting clusters (not 'confirms')

5 **Conclusions 10%** ← + recap

Other
– too many digits
– make 'pretty' column labels
2 **Gene list 5%** – all on 1 page
– put on last page
5 **Reproducible R code 5%**
3 **Overall Presentation 5%** – name chips P0, P1, ... (no '.cel')
– plot labels too small

⊛ QC: y = bg-corr, quant norm, $\log_2$

(**) Use Ward's (not Euclidean + single linkage) with 1-cor

– cite references

(***) only based on adj p, NOT FC constraint

59

give informative title

Xiang

Statistics for Genomic Data Analysis, Final Data Analysis Reports

3 **Background/Introduction 5%** which chip? # probe sets?

NUSE/RLE not defined

7 **Quality assessment 15%**

you need to introduce MODEL before measures

2/8 Description of quality assessment

3/3 Excluded poor quality slides (if relevant) - median NUSE > 1.05

2/4 Graphical support for decisions - + weights pseudo-images

- add horizontal line at 1.05 NUSE boxplots

6 **Normalization 10%**

- what is the model?

2/6 Description of normalization/expression measure used - what is the log corr?

2/2 Graphical support for decisions - What is quant norm specifically

2/2 Appropriateness of normalization

15 **Statistical analyses: Differential Expression 25%**

- write model in terms of 2 pars + emp. Bayes (mathematically)

3/ ~~2~~/5 Description of model (including design matrix, contrasts if necessary)

2/5 Multiple testing/adjusted p-values given hyp st procedure mathematically + FDR

2/5 Appropriate basis for gene ranking (and explanation) (**)

3/5 Estimated number of differentially expressed genes (and explanation)

5/5 Volcano plot, other appropriate plots

13 **Statistical analyses: Cluster Analysis 20%**

- incomplete description, use 1-cor + Wards (not Euclidean + complete)

3/10 Description of clustering algorithm used

5/5 Relevant graphics

5/5 Interpretation of resulting clusters

5 **Conclusions 10%** - give re-cap + DE & cluster results in separate paragraphs

Other

- use scientific notation for p-values
- make 'pretty' column headings
- need probe set ID
- use gene symbol (not long description)

2 (*) **Gene list 5%**

5 **Reproducible R code 5%**

3 **Overall Presentation 5%** - no png plots

- plot labels too small

(Don't need RLE, just NUSE)

(*) Probe set Gene Symbol

(**) only based on adj p, NOT with FC constraint

- Incomplete refs