GWAS: Name _____ IB 5/12 → 2.5/6

1. Formatting:

0.75/0.75

all margins 2.5cm	informative title
12 pt size	name on all pages
no raw R code or output	all pages numbered
max **10** pages	no blurry plots (**NOT png**)

0.75 2. Introduction/Background:

brief statement of scientific question

all variables defined – EDA?

0.5/2 3. PCA:

explain relation between PCs and population stratification

plot pc2 (y-axis) vs pc1 (x-axis)

4. Pre-processing / QC steps:

0.5/2

SNP QC: criteria and reasons

sample QC: criteria and reasons

Hardy-Weinberg equilibrium: what it means and how it relates to quality

Overall QC explanation

5. Association / post-association analysis:

gender/snp not defined

1/2

Describe association analysis in words and mathematically

Manhattan plot -log10

lambda analysis (including **SQUARE** QQ-normal plots) + explain

LD heatmap (optional – does NOT count); measure of LD

- Results?

- Most prevalent or Minor allele? (xboost)

3.5/7.75

0.5/1 ^everywhere or nowhere / logit p not logit $\hat{p}$

6. Write out final estimated model **mathematically** (for a given SNP)

hat on response variable MUST RELATE TO SNP ↳ define

0.75/1.25 7. Plots:

label size (not too small) captions

placement **NOT BLURRY**

0.25/1 8. Conclusions

recap analysis incomplete + vague state main findings

1/1 9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

1.5/4.25

Comments **Name:** JB

00 – informative title

A – eda

B – PCA + explain

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure

G – explain association test

H – write out final model *mathematically* Define gender/ snp

I – Manhattan plot (and explanation)

J – identify significant markers

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text)

Q – plot layout (see text)

R – overall organization and explanation of procedure

S – other:

- also cite P Breheny tutorial
- Define MAF, etc

GWAS: Name _____ GBjj 4.25/12 → 2.125/6

1. Formatting:

0.75/1

- all margins 2.5cm
- 12 pt size
- **no raw R code or output**
- max **10** pages
- informative title
- name on all pages
- all pages numbered
- no blurry plots (**NOT png**)

0.75/1

2. Introduction/Background:

- brief statement of scientific question
- all variables defined – minor allele?

0.25/2

3. PCA: not well-explained

- explain relation between PCs and population stratification
- plot pc2 (y-axis) vs pc1 (x-axis) dont need Figure 2b

0.5/2

4. Pre-processing / QC steps:

- SNP QC: criteria and reasons
- sample QC: criteria and reasons
- Hardy-Weinberg equilibrium: what it means and how it relates to quality
- Overall QC explanation

0.25/2

5. Association / post-association analysis:

- Describe association analysis in words and mathematically
- Manhattan plot + explain
- lambda analysis (including **SQUARE** QQ-normal plots) + explain
- LD heatmap (optional – does NOT count); measure of LD

2.5/7.75

model incorrectly specified

0/1

6. Write out final estimated model **mathematically** (for a given SNP)

hat on response variable MUST RELATE TO SNP

0.75/1.25 7. Plots: plot size small

label size (not too small) captions

placement **NOT BLURRY**

0.25/1 8. Conclusions

recap analysis state main findings

(*) interpretation

- much of what you say is generic and not specific to GWAS

0.75/ 9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

- cite refs in text, no general refs

- also cite P Breheny

- Don't cite me/Course

(*) cannot say how 'likely' any observed association is

- use your own words

1.75/4.25

Comments **Name:** GBii

00 – informative title

A – eda hists but box

B – PCA + explain

C – define and explain HWE – a/b or p/q?

D – define and explain HWE **test**

E – define λ

F – define LD measure

G – explain association test

H – write out final model *mathematically* incorrectly specified

I – Manhattan plot (and explanation)

J – identify significant markers

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text)

Q – plot layout (see text)

R – overall organization and explanation of procedure

S – other:

- GWAS doesn't require unrelated individuals
- you can have a GWAS for a family-based design - if case/control or cohort design then need independence
- define all terms (MAF, zygosity, etc)

GWAS: Name GBL 4.25/12 → 2.125/6

log

1. Formatting:

all margins 2.5cm — informative title

12 pt size — name on all pages

0.75/0.75 **no raw R code or output** — all pages numbered

max **10** pages — no blurry plots (**NOT png**)

← too many digits

2. Introduction/Background:

0.5/1 brief statement of scientific question — more on this

all variables defined

3. PCA:

0.5/2 explain relation between PCs and population stratification — unclear

plot pc2 (y-axis) vs pc1 (x-axis)

4. Pre-processing / QC steps: ← superficial explanations

0.5/2 SNP QC: criteria and reasons

sample QC: criteria and reasons — mathematically

Hardy-Weinberg equilibrium: what it means and how it relates to quality

Overall QC explanation

↳ specifically

5. Association / post-association analysis:

0.25/2 Describe association analysis in words and mathematically

Manhattan plot + explain

lambda analysis (including **SQUARE** QQ-normal plots) + explain

LD heatmap (optional – does NOT count); measure of LD

vague - be specific and clear

2.5/7.75

0/1 model incorrectly specified

6. Write out final estimated model **mathematically** (for a given SNP)

hat on response variable MUST RELATE TO SNP (not defined)

0.75/1.25 7. Plots: plot sizes small

label size (not too small) captions

placement **NOT BLURRY**

0.25/1 8. Conclusions vague and out of order

recap analysis state main findings very generic, be specific

0.75/1 9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

- cite refs specifically in text - no general refs
- use your own words

1.75/4.25

Comments **Name:** GBl

00 – informative title

A – eda – hists not box

B – PCA + explain

C – define and explain HWE – mathematically

D – define and explain HWE **test**

E – define λ

F – define LD measure

G – explain association test

H – write out final model *mathematically* incorrect

I – Manhattan plot (and explanation)

J – identify significant markers

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M – interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text)

Q – plot layout (see text)

R – overall organization and explanation of procedure

S – other:

_____

_____

_____

_____

_____

GWAS: Name _____ AC 9.25/12 → 4.625/6

1. Formatting:

0.75/0.75

- all margins 2.5cm
- 12 pt size
- **no raw R code or output**
- max **10** pages
- informative title
- name on all pages
- all pages numbered
- no blurry plots (**NOT png**)

0.75/1

2. Introduction/Background:

- brief statement of scientific question
- all variables defined (no EDA)

1.75/2

3. PCA: mostly ok

- explain relation between PCs and population stratification
- plot pc2 (y-axis) vs pc1 (x-axis)

0.75/2

4. Pre-processing / QC steps: vague in parts

- SNP QC: criteria and reasons unclear / non-specific in parts
- sample QC: criteria and reasons very general
- Hardy-Weinberg equilibrium: what it means and how it relates to quality
- (Overall QC explanation) Lspecifically

1.25/2

5. Association / post-association analysis:

- Describe association analysis in words and mathematically ↳ not testing correlation / association
- Manhattan plot
- lambda analysis (including **SQUARE** QQ-normal plots)
- LD heatmap (optional – ~~does NOT~~ count); measure of LD

Be specific

5.25/7.75

1/1

6. Write out final estimated model **mathematically** (for a given SNP)

hat on response variable MUST RELATE TO SNP

1.25/1.25

7. Plots:

label size (not too small) captions

placement **NOT BLURRY**

0.75/1

8. Conclusions

recap analysis own words - state main findings

1/1

9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory (good) excellent

10. Other comments:

- use primary sources

- use your own words

4/4.25

Comments **Name:** AC

00 – informative title

A – eda not done?

B – PCA + explain

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure

G – explain association test

H – write out final model *mathematically*

I – Manhattan plot (and explanation)

J – identify significant markers

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text)

Q – plot layout (see text)

R – overall organization and explanation of procedure

S – other:

- Table of results for sig snps

GWAS: Name KD 7.75/12 → 3.875/6

1. Formatting: 0.75/1

- all margins 2.5cm
- 12 pt size
- **no raw R code or output**
- max **10** pages
- informative title
- name on all pages
- all pages numbered
- no blurry plots (**NOT png**)

2. Introduction/Background: 1/1

- brief statement of scientific question
- all variables defined

3. PCA: 2/2

- explain relation between PCs and population stratification
- plot pc2 (y-axis) vs pc1 (x-axis)

4. Pre-processing / QC steps: 0.5/2 mostly superficial

- SNP QC: criteria and reasons
- sample QC: criteria and reasons
- Hardy-Weinberg equilibrium: what it means and how it relates to quality
- Overall QC explanation

5. Association / post-association analysis: 0.5/2

- Describe association analysis in words and mathematically
- Manhattan plot — explain
- lambda analysis (including **SQUARE** QQ-normal plots) + explain
- LD heatmap (optional – does NOT count): measure of LD

4.75/7.75

0.5/1 write in terms of logit(P(y=1)) instead of $\hat{p}$

6. Write out final estimated model **mathematically** (for a given SNP)

hat on response variable MUST RELATE TO SNP

1/1.25 7. Plots (size) not defined

label size (not too small) captions

placement **NOT BLURRY**

0.5/1 8. Conclusions use paragraphing

recap analysis state main findings

1/1 9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

- Don't need table of Contents

- other refs?

3/4.25

Comments **Name:** KD

log

00 – informative title

A – eda hists not box + Figure 4 unnecessary

B – PCA + explain why is substructure of 'of concern'

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure

G – explain association test – specifically

H – write out final model *mathematically* snp not defined

I – Manhattan plot (and explanation)

J – identify significant markers

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M – interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text) Manhattan too small

Q – plot layout (see text)

R – overall organization and explanation of procedure

S – other:

- Define all terms (Call rate, MAF, etc)
- assume unrelated in case/control or cohort designs (like this one) - you can also do guess on family designs

GWAS: Name WL 8.5/12 → 4.25/6

1. Formatting:

0.75/0.75

- all margins 2.5cm
- 12 pt size
- **no raw R code or output**
- max **10** pages
- informative title
- name on all pages
- all pages numbered
- no blurry plots (**NOT png**)

0.5/1

2. Introduction/Background:

- brief statement of scientific question — be specific
- all variables defined — no EDA

1.75/2

3. PCA: good (BUT: use your own words and give refs)

- explain relation between PCs and population stratification
- plot pc2 (y-axis) vs pc1 (x-axis)

0.75/2

4. Pre-processing / QC steps:

- SNP QC: criteria and reasons
- sample QC: criteria and reasons — be specific
- Hardy-Weinberg equilibrium: what it means and how it relates to quality
- Overall QC explanation — why are 2 individuals removed? (explain specifically /score's?)

0.75/2

5. Association / post-association analysis:

- Describe association analysis in words and mathematically — sex/snp not defined
- Manhattan plot — be more specific on y-axis
- lambda analysis (including **SQUARE** QQ-normal plots) + explain
- LD heatmap (optional – does NOT count); measure of LD

- Results

4.5/7.75

either ^everywhere or nowhere / logit $\hat{p}$

0.5/1 6. Write out final estimated model **mathematically** (for a given SNP)

not logit $\hat{p}$

hat on response variable MUST RELATE TO SNP

└ not defined

1.25/1.25 7. Plots:

label size (not too small) captions

placement **NOT BLURRY**

1/1 8. Conclusions good

recap analysis state main findings

1.25/1 9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory (good) excellent

10. Other comments:

- also ref P Breheny tutorial

4/4.25

Comments **Name:** WL

00 – informative title

A – eda – missing?

B – PCA + explain

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure

G – explain association test

H – write out final model *mathematically* – define sex/snp

I – Manhattan plot (and explanation)

J – identify significant markers

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text)

Q – plot layout (see text)

R – overall organization and explanation of procedure

S – other:

very good !!

GWAS: Name AM 5/12 → 2.5/6

log

1. Formatting:

0.75/0.75

all margins 2.5cm — informative title

12 pt size — name on all pages

no raw R code or output — all pages numbered

max **10** pages — no blurry plots (**NOT png**)

0.75/1

2. Introduction/Background:

brief statement of scientific question — say a little more

all variables defined

0.75/2

3. PCA: incomplete

explain relation between PCs and population stratification

plot pc2 (y-axis) vs pc1 (x-axis)

4. Pre-processing / QC steps:

0.5/2

SNP QC: criteria and reasons

sample QC: criteria and reasons

Hardy-Weinberg equilibrium: what it means and how it relates to quality

Overall QC explanation

explain clearly

0.5/2

5. Association / post-association analysis:

Describe association analysis in words and mathematically

Manhattan plot + explain

lambda analysis (including **SQUARE** QQ-normal plots) + explain

LD heatmap (optional – does NOT count); measure of LD

3.25/7.75

not clear

0/1

6. Write out final estimated model **mathematically** (for a given SNP)

hat on response variable MUST RELATE TO SNP

0.75/1.25

7. Plots: size

label size (not too small) captions

placement **NOT BLURRY**

0/1

8. Conclusions not done

recap analysis state main findings

1/1

9. Overall presentation (clarity of explanations, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

- no refs

1.75/4.75

Comments **Name:** AM

log

00 – informative title

A – eda hists not box + Don't need figure 6

B – PCA + explain your plot interpretations are unclear – explain specifically

C – define and explain HWE

D – define and explain HWE **test** specifically

E – define λ

F – define LD measure

G – explain association test

H – write out final model *mathematically* – what model are you fitting?

I – Manhattan plot (and explanation)

J – identify significant markers

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions

N – **no raw R**

O – plot labels too small

P – plot size (see text) APP 6 plots too small

Q – plot layout (see text)

R – overall organization and explanation of procedure

S – other:

- Define all terms (mAF, zygosity, etc)
- Figure 8?
- incomplete refs

GWAS: Name MM 4.75/12 → 2.375/6

1. Formatting: log

0.75/0.75

all margins 2.5cm — informative title

12 pt size — name on all pages

no raw R code or output — all pages numbered

max **10** pages — no blurry plots (**NOT png**)

0.25/1 2. Introduction/Background:

brief statement of scientific question

all variables defined

0.5/2 3. PCA: + explain scree plot

explain relation between PCs and population stratification

plot pc2 (y-axis) vs pc1 (x-axis)

sex = 1/2? which is M/F?

0.5/2 4. Pre-processing / QC steps: incomplete

SNP QC: criteria and reasons

sample QC: criteria and reasons

Hardy-Weinberg equilibrium: what it means and how it relates to quality

↳ specifically

Overall QC explanation

0.25/2 5. Association / post-association analysis: incomplete model

Describe association analysis in words and mathematically

Manhattan plot — not explained

lambda analysis (including **SQUARE** QQ-normal plots)

LD heatmap (optional – does NOT count); measure of LD

2.25/7.75

logit $P(Y=1)$ not logit $(P(\hat{Y}=1))$

0.25/1 6. Write out final estimated model **mathematically** (for a given SNP)

hat on response variable MUST RELATE TO SNP

1/1.25 7. Plots:

label size (not too small) captions

placement **NOT BLURRY**

0.5/1 8. Conclusions

recap analysis - more clearly state main findings vague and generic

0.75/1 9. Overall presentation (clarity of ~~explanations~~, appropriate citations / references) :

poor satisfactory good excellent

10. Other comments:

- no footnotes, ref cite in text
- explain relation between missing values to qualify
- Weinberg (not Weinberg)
- Don't ref course [1]
- your report is very superficial, you need to explain things clearly and specifically

2.5/4.25

Comments **Name:** MM

log

00 – informative title

A – eda hists not box/ figure 4 unnecessary

B – PCA + explain incomplete

C – define and explain HWE

D – define and explain HWE **test**

E – define λ + analysis incomplete

F – define LD measure

G – explain association test

H – write out final model *mathematically* incomplete + include SNP

I – Manhattan plot (and explanation)

J – identify significant markers

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M - interpret conclusions

↳ these are vague and generic

N – **no raw R**

O – plot labels too small

P – plot size (see text)

Q – plot layout (see text)

R – overall organization and explanation of procedure

S – other:

- Explain all Figures
- Define ALL terms (MAF, etc)
- use paragraphing
- Don't need Table 1

GWAS: Name _____ AP 7.25/12 → 3.675/6

1. Formatting:

0.75/0.75

- all margins 2.5cm
- 12 pt size
- **no raw R code or output**
- max **10** pages
- informative title
- name on all pages
- all pages numbered
- no blurry plots (**NOT png**)

2. Introduction/Background:

1/1

- brief statement of scientific question
- all variables defined

3. PCA:

1.5/2

- explain relation between PCs and population stratification
- plot pc2 (y-axis) vs pc1 (x-axis)

4. Pre-processing / QC steps: incomplete

0.75/2

- SNP QC: criteria and reasons
- sample QC: criteria and reasons
- Hardy-Weinberg equilibrium: what it means and how it relates to quality
- Overall QC explanation

5. Association / ~~post-association analysis~~:

0.75/2

- Describe association analysis in words and mathematically
- Manhattan plot + explain
- lambda analysis (including **SQUARE** QQ-normal plots) + explain
- LD heatmap (optional – does NOT count); measure of LD

4.75/7.75

0.5/1 — you left out age + sex

6. Write out final estimated model **mathematically** (for a given SNP)

hat on response variable MUST RELATE TO SNP

1/1.25 7. Plots: make 'pretty' labels → incorrect definition

label size (not too small) captions

placement **NOT BLURRY**

0/1 8. Conclusions not done

recap analysis state main findings

1/1 9. Overall presentation (clarity of explanations, appropriate citations / references) : → incomplete

poor satisfactory good excellent

10. Other comments:

2.5/4.25

Comments **Name:** AP

00 – informative title

A – eda lists not box

B – PCA + explain

C – define and explain HWE

D – define and explain HWE **test**

E – define λ

F – define LD measure

G – explain association test

H – write out final model *mathematically* – Snp. incorrectly defined

I – Manhattan plot (and explanation)

J – identify significant markers

K – square QQ plots

L - fix blurry plots (use jpeg or pdf, NOT png)

M – interpret conclusions (not done)

N – **no raw R**

O – plot labels too small

P – plot size (see text)

Q – plot layout (see text)

R – overall organization and explanation of procedure

(S) – other:

- also cite Breheny tutorial
- don't need LD heatmap
- incomplete - please let me know if you have questions! ☺