Answers

Q Which cohort did not upload their data in the requested format

A Sample B

Q Write code to fix the format for Sample B

A

awk '{ print $2":" $3, $0}' sampleB.txt > sampleB.txt.corrected

Q Write code to QC the data files A

echo "MarkerName CHR BP A1 A2 Freq1 R2 Beta SE P N" >sampleA.qced

awk '{ if ($6>=.01 && $6<=.99 && $7>=.6) print $0}' sampleA.txt >> sampleA.qced echo "MarkerName RS CHR BP A1 A2 Freq1 R2 Beta SE P N" >sampleB.qced

awk '{ if ($7>=.01 && $7<=.99 && $8>=.6) print $0}' sampleB.txt.corrected >> sampleB.qced

echo "MarkerName CHR BP A1 A2 Freq1 R2 Beta SE P N" >sampleC.qced

awk '{ if ($6>=.01 && $6<=.99 && $7>=.6) print $0}' sampleC.txt >> sampleC.qced

Q How many variants were in the file before QC? How many survived the QC filters? A

N variants before QC after QC

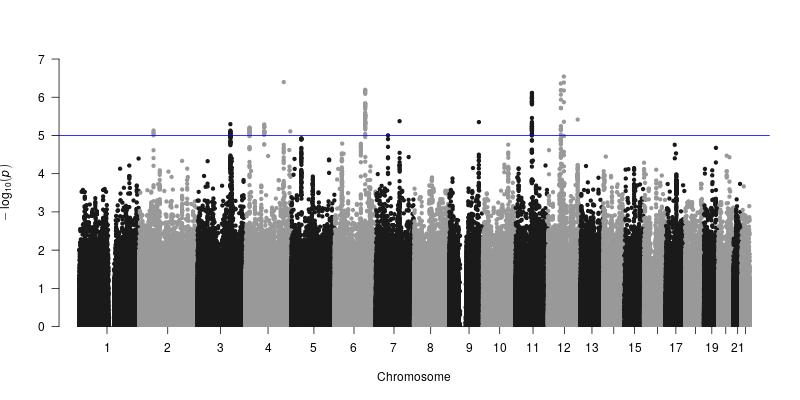
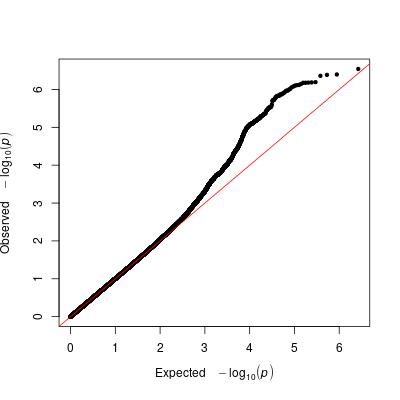
Cohort A 1527493 1329420

Cohort B 1527493 1329342

Cohort C 1527493 1329553

Q Take a look at the SE.metal.out file and find the top hit. Paste the marker name and the P value into the box below.

A Smallest p-value is 2.872e-07 at marker '12:66343810'

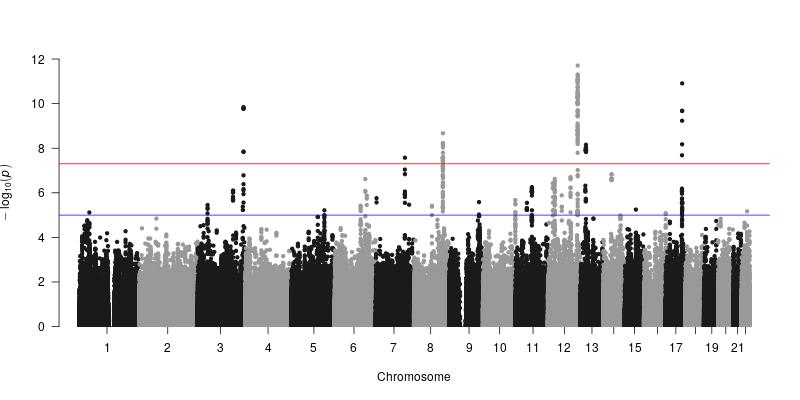
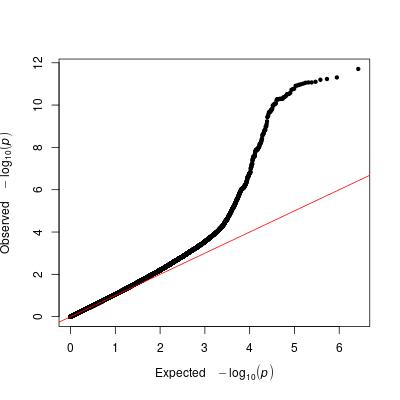


Q How many genome-wide significant regions do you see in the Manhattan plot? A 0

Q Does the QQ plot show any signs of technical artefacts? A No

Q Take a look at the N.metal.out file and find the top hit. Paste the marker name and the P value into the box below. A Smallest p-value is 1.961e-12 at marker '12:123758235'

Q Is this the same top hit as you found with the inverse variance weighted analysis? A No - different variant and different locus



Q Do you like these results better or worse? A I'm much happier with the results

Q For each regression record the Nagelkerke's pseudo r-square. Provide these values in the table below. A

N variants AD\_meta Morningness

PRS A 0.174748 0.00165849

PRS B 0.198440 0.01047640

PRS C 0.236648 0.01777100

PRS D 0.269725 0.00280292

Q What does the plot tell us about the genetic architecture of Alzheimer's disease in this sample?

A Some large and significant genetic variants but substantial contributions from less significant variants.

