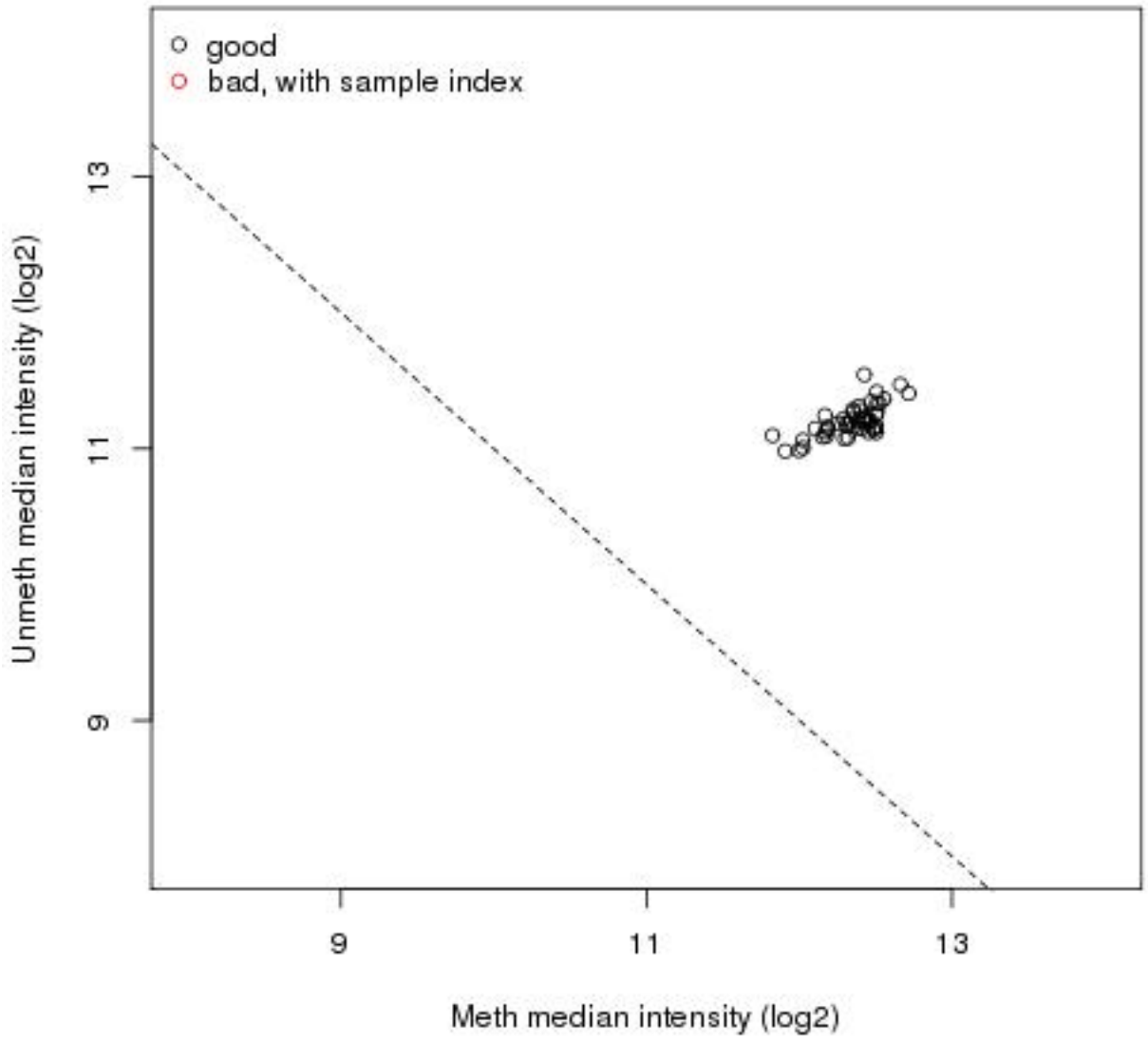
Vitamin K methylation preliminaries: QC and unsupervised analysis

QC and preprocessing

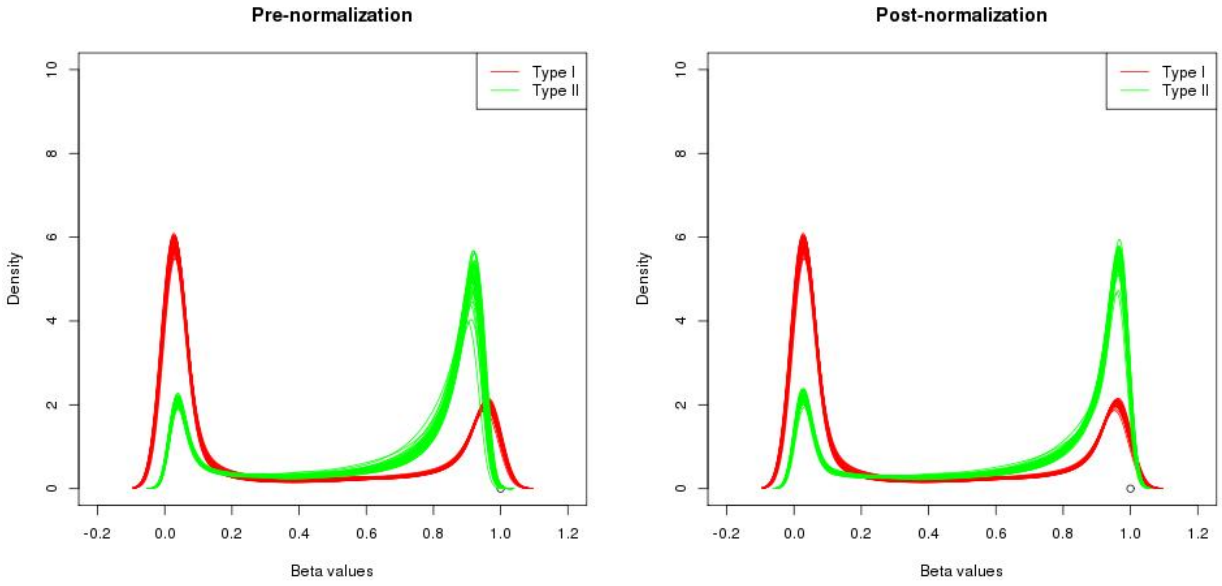
First, we wanted to gain a basic confidence in the validity of the methylation data as returned from YCGA. A plot of microarray intensity values for methylated and unmethylated channels shows that no sample has grossly aberrant signal detection levels on the whole.



QC steps involved removal of samples for which a notable proportion of probes were undetectable (none

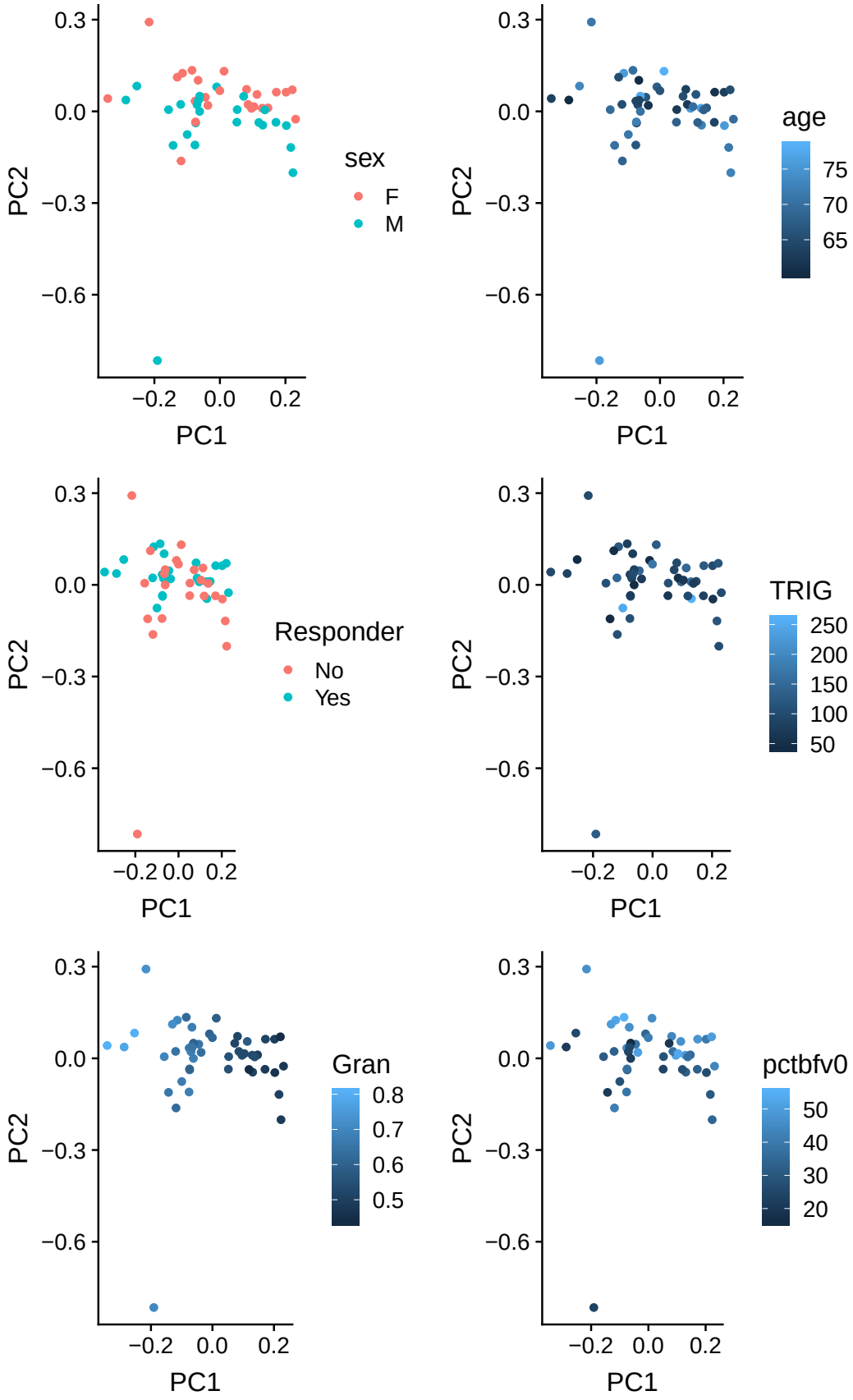
were removed), and removal of probes based on detection consistency across samples, associated SNPs, cross-hybridizing probes, non-CpG probes, and others.

The BMIQ normalization procedure was performed in order to approximately match methylation profiles between Type I (separate bead types for methylated vs. unmethylated) and Type II (single bead type for both) probes. The pre- and post-BMIQ plots confirm that it accomplished its intended effect.



PCA

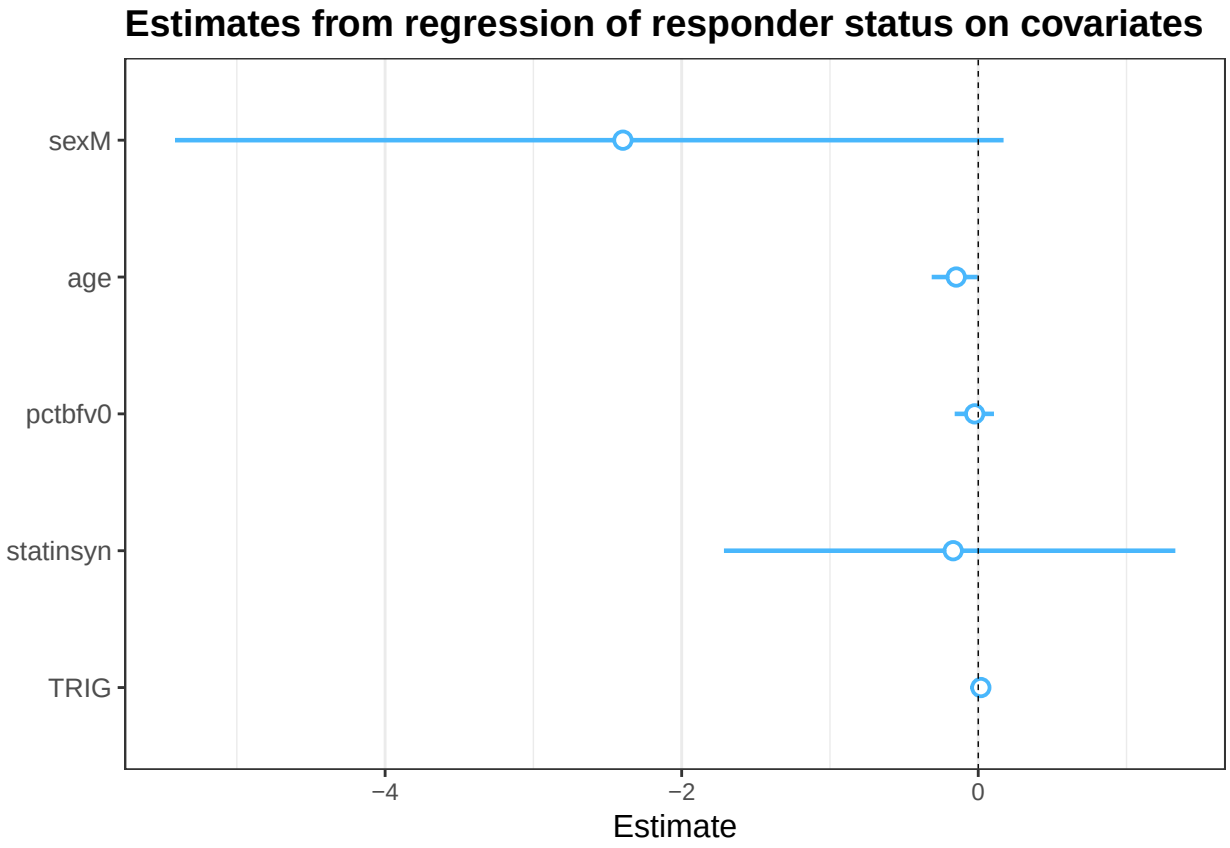
PCA was used to investigate sample grouping with respect to large-scale patterns in the methylation data.

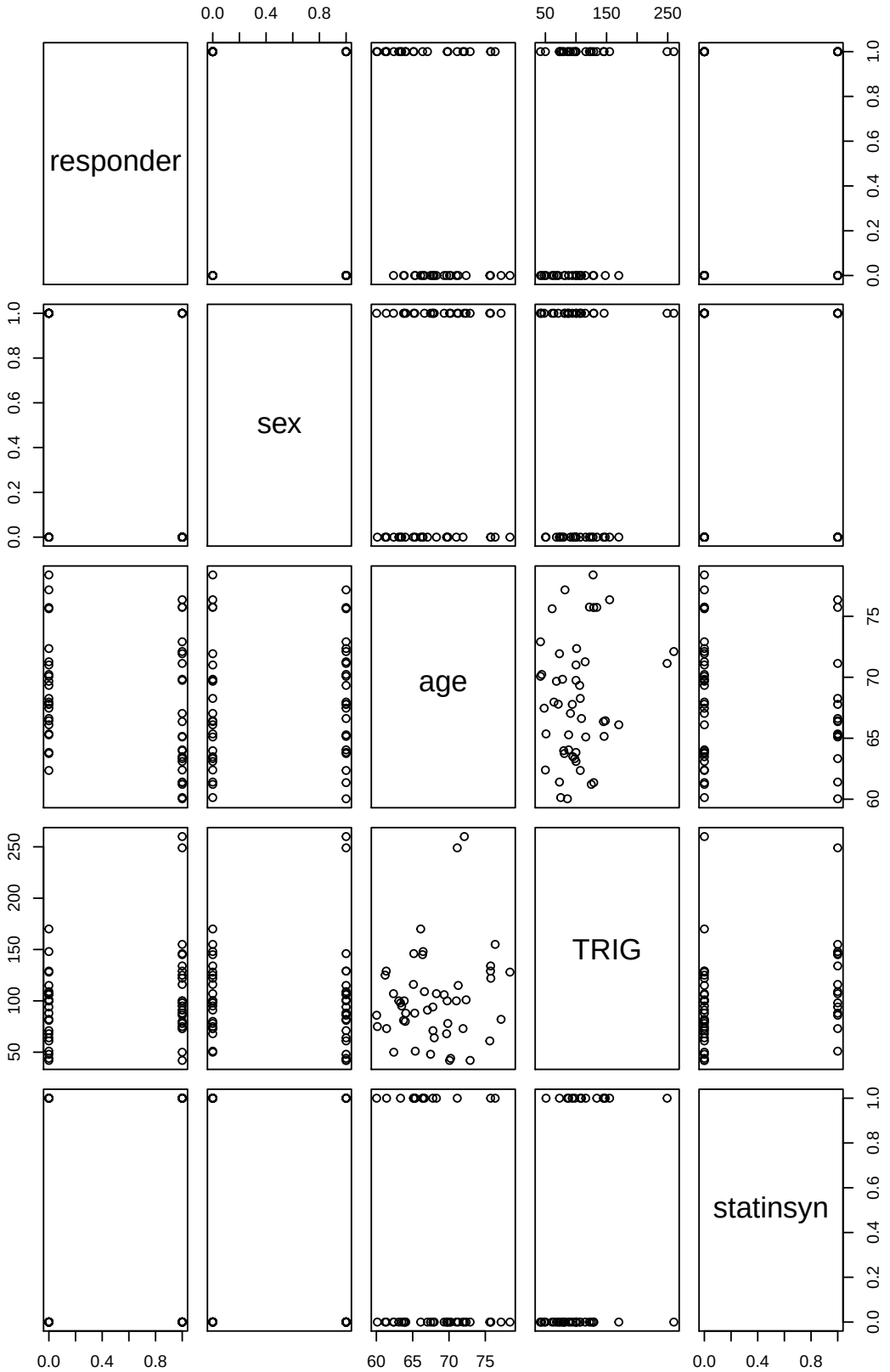

0.3
0.0
−0.3
−0.6
PC2
−0.2
0.0
0.2
PC1
sex
F
M
age
75
70
65
Responder
No
Yes
TRIG
250
200
150
100
50
Gran
0.8
0.7
0.6
0.5
pctbfv0
50
40
30
20

PC1 from the beta-value matrix (~16% of variance explained) seems to load heavily on granulocytes (mostly made up of neutrophils). Sex seems to be related to PC2 (~7% variance explained).

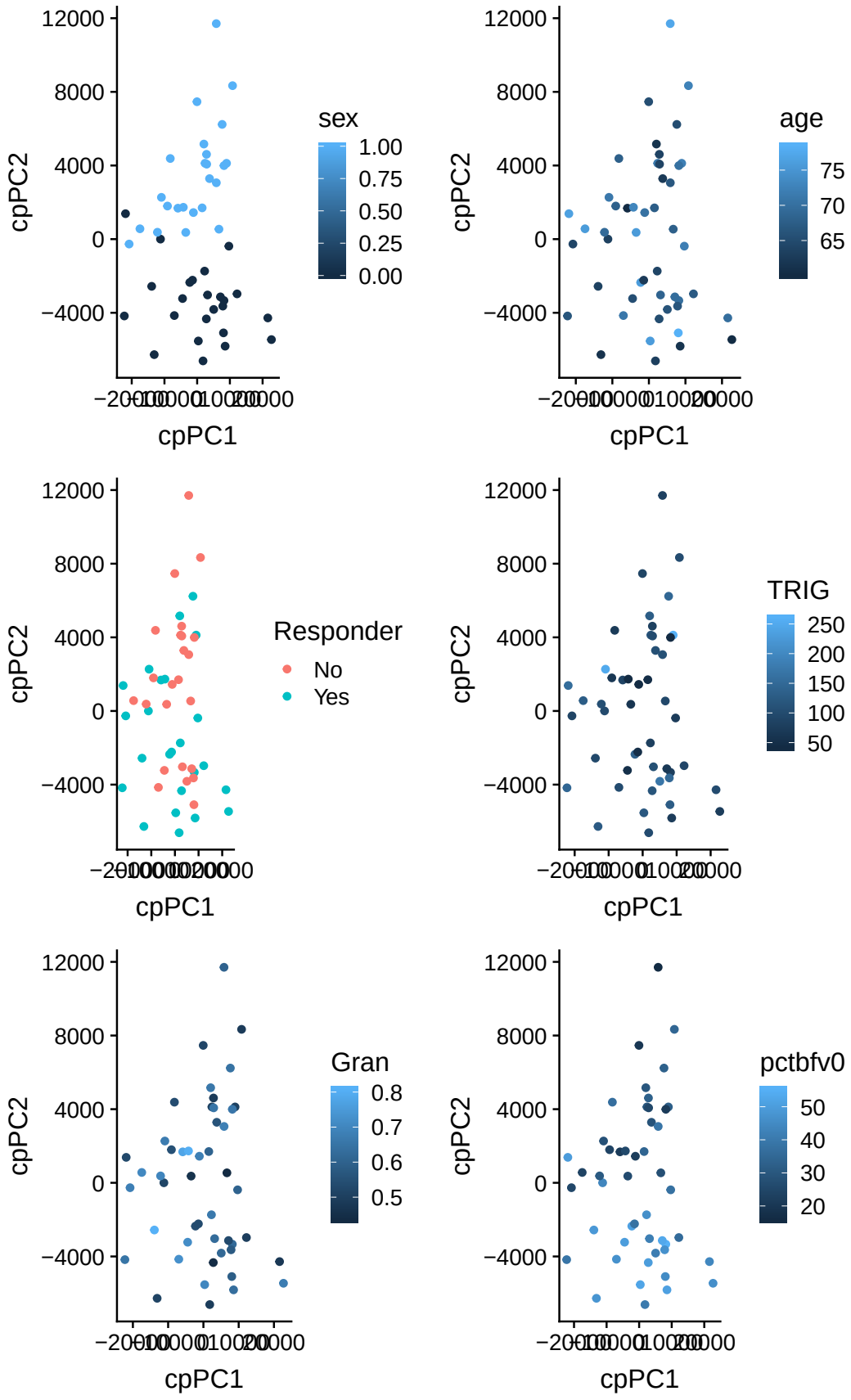
Outcome and covariate relationships

Additionally, before conducting any methylation-phenotype tests, we examined the correlations between various technical and biological variables and our outcome to understand covariate correlations and distributions.




responder
sex
age
TRIG
statinsyn
0.0
0.4
0.8
50
150
250
1.0
0.6
0.2
60
65
70
75

Supplementary: same PCA plots but with control-probe PC1&2


12000
8000
4000
0
−4000
cpPC2
cpPC1
sex
1.00
0.75
0.50
0.25
0.00
age
75
70
65
Responder
No
Yes
TRIG
250
200
150
100
50
Gran
0.8
0.7
0.6
0.5
pctbfv0
50
40
30
20