Lung cancer prediction using DNA methylation

Thomas Battram

Goal:

To test the predictive performance of the top 76 CpGs identified in a lung cancer EWAS performed in HUNT, in an independent dataset (NSHDS).

Quality control of data:

478 samples present with DNA methylation data

41 of the 76 probes were present in the 450k dataset.

Probes that have detection p values of >0.01 on 5% or more samples and samples that have detection p values of >0.01 for 5% or more probes were removed. None of the 41 probes were removed. 477 samples were left. Finally, after removing incomplete case-control pairs 468 were used in the analysis.

Models used

Conditional logistic regression code:

```
fit <- clogit(LUNG_CANCER_CASE ~ CpG1 + CpG2 + ... + CpGn + strata(CASESET))
```

Results

AUC = 0.67 (95% CI: 0.62 − 0.72)