Lung cancer prediction using DNA methylation

Thomas Battram

Goal:

To test the performance of CpGs identified via smoking and lung cancer EWAS to predict lung cancer within NSHDS. These CpGs were identified via EWAS in HUNT

Quality control of data:

478 samples present with DNA methylation data

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 probes of interest 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))
```

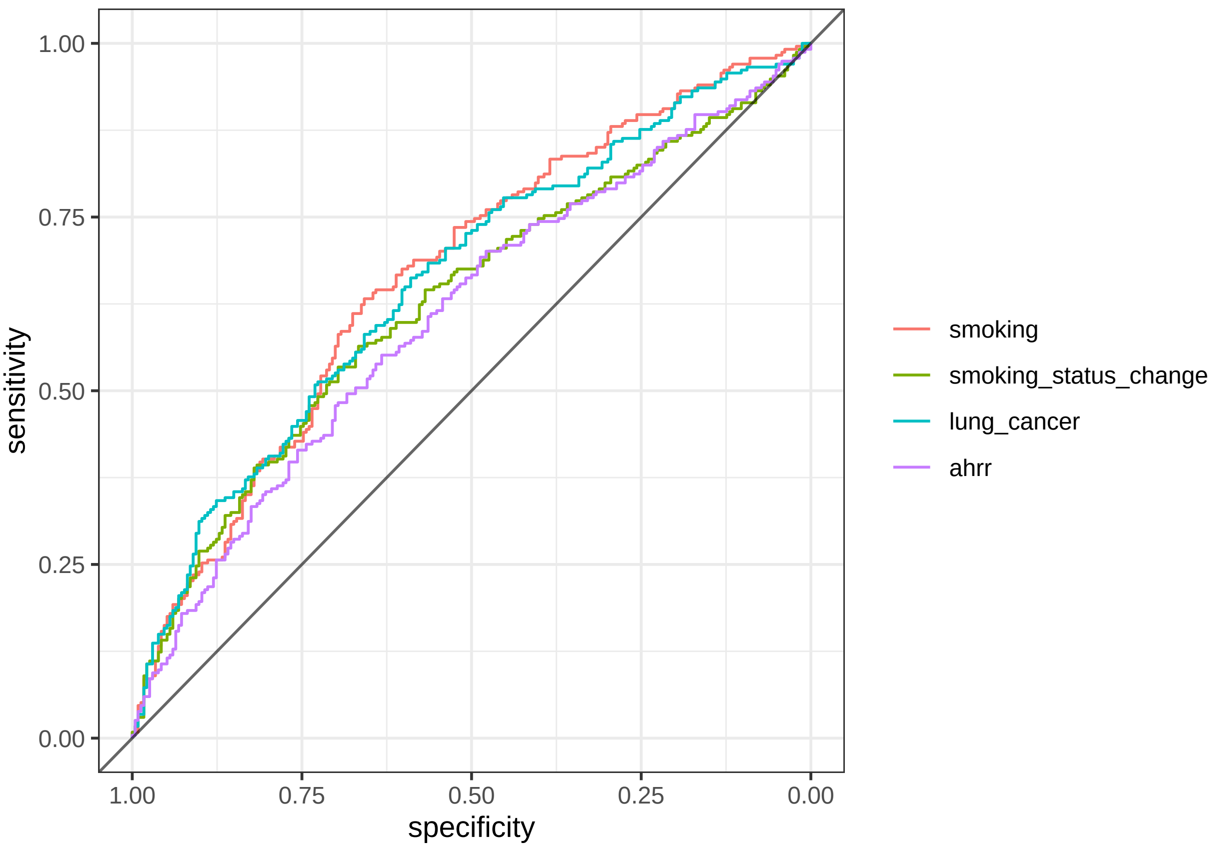
This code was run using 4 CpG sets: 1. EWAS of smoking 2. EWAS of change in smoking status 3. EWAS of lung cancer 4. cg05575921 *AHRR* only

CpGs were weighted according to beta coefficients from their EWAS. For the cg05575921 *AHRR* only model, the CpG beta value was weighted by the beta coefficient for that CpG from the smoking EWAS. For the lung cancer EWAS CpG set, the CpG beta values were weighted by their log(OR).

Results

ewas	n_cpgs	n_cpgs_in_450k	n_unique_cpgs_in_450k
smoking	76	41	24
smoking_status_change	9	6	0
lung_cancer	50	25	10
combined	135	51	51

Summary of the number of CpG sites in each CpG set



cpg_set	auc	ci_lower	ci_upper
smoking	0.667	0.6181	0.7158
smoking_status_change	0.6275	0.5769	0.6782
lung_cancer	0.6587	0.6096	0.7079
ahrr	0.608	0.5569	0.659

AUCs from the ROC curves above