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:

fit1 <- clogit(LUNG\_CANCER\_CASE ~ CpG1 + CpG2 + ... + CpGn + strata(CASESET))  
  
fit2 <- clogit(LUNG\_CANCER\_CASE ~ CpG\_score + strata(CASESET))

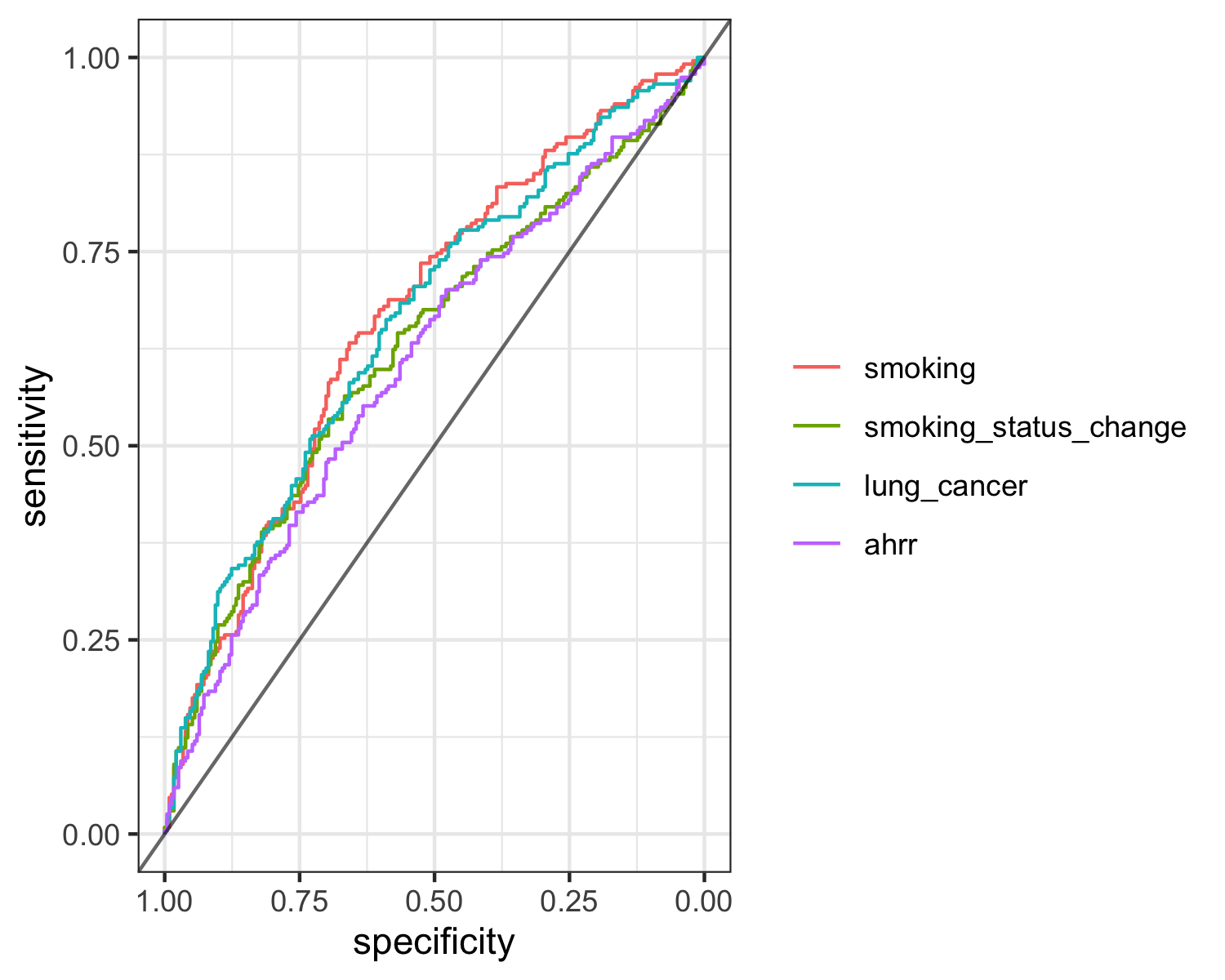
This code was run using 4 CpG sets: 1. EWAS of smoking 2. EWAS of change in smoking status 3. EWAS of lung cancer 4. cg05575921 *AHRR* only

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).

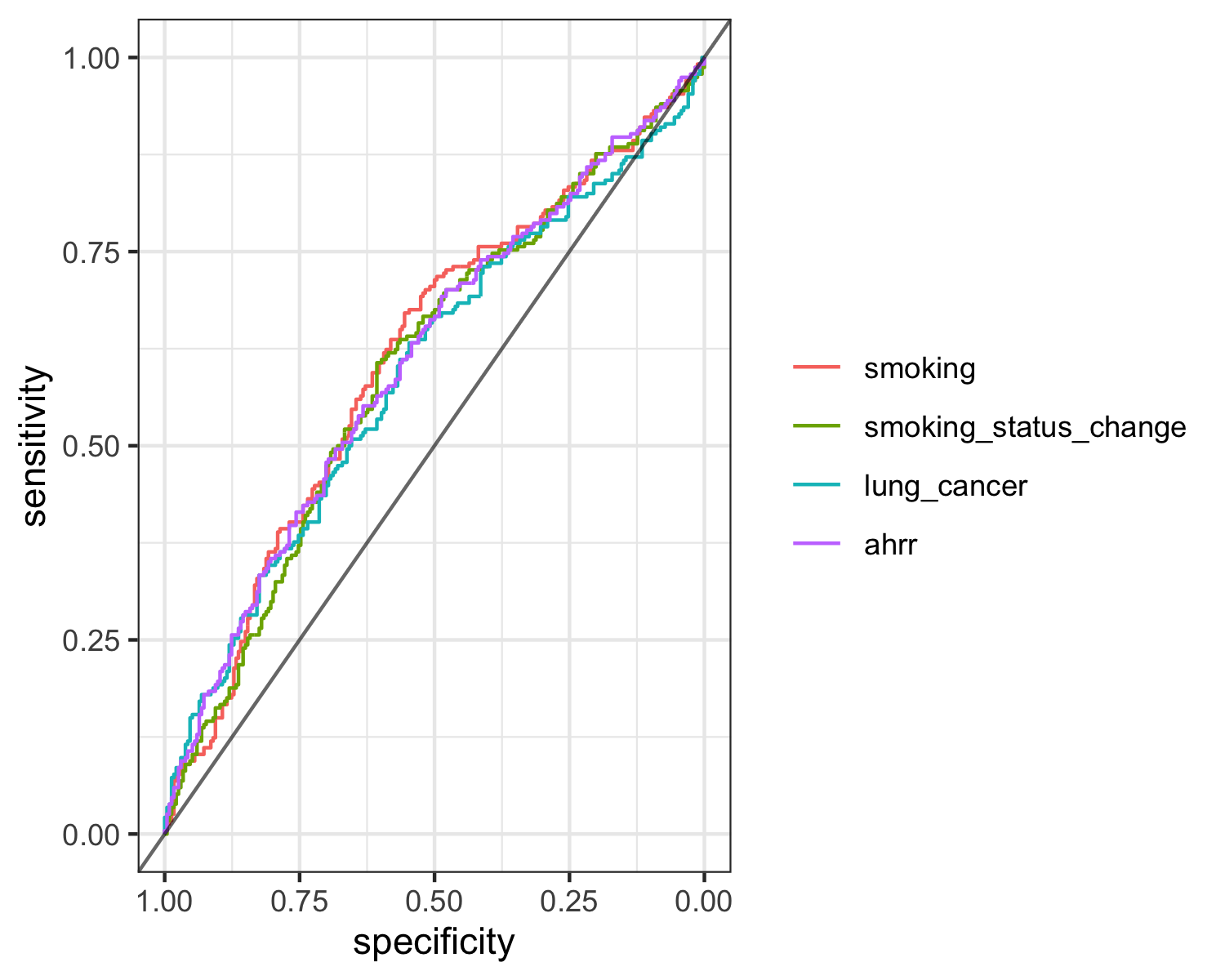
## Results

|  |  |  |  |
| --- | --- | --- | --- |
| 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



ROC curve using fit1 (CpG sites treated separately in the model)



ROC curve using fit2 (CpG sites combined into a single score)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| cpg\_set | model | auc | ci\_lower | ci\_upper |
| smoking | separate\_sites | 0.667 | 0.6181 | 0.7158 |
| smoking | score | 0.6134 | 0.5622 | 0.6646 |
| smoking\_status\_change | separate\_sites | 0.6275 | 0.5769 | 0.6782 |
| smoking\_status\_change | score | 0.6005 | 0.5491 | 0.6519 |
| lung\_cancer | separate\_sites | 0.6587 | 0.6096 | 0.7079 |
| lung\_cancer | score | 0.594 | 0.5425 | 0.6456 |
| ahrr | separate\_sites | 0.608 | 0.5569 | 0.659 |
| ahrr | score | 0.608 | 0.5569 | 0.659 |

AUCs from the ROC curves above. separate\_sites = fit1, score = fit2.