Lung cancer is the leading cause of death in the U.S. Non-small cell lung cancer (NSCLC), the vast majority of lung cancers, has three main subtypes: adenocarcinoma, large cell carcinoma and squamous cell carcinoma. These subtypes are difficult to distinguish histologically but demand distinct courses of treatment, and applying the wrong treatment can lead to serious adverse consequences. We have developed a gene expression-based classifier to distinguish squamous and non-squamous NSCLC by using quantitative Nuclease Protection Assay (qNPA). The chemistry does not require RNA extraction, reverse transcription, or amplification and is fully automated on the HTG Edge System.

### Methods

We first identified a set of candidate biomarker genes through microarray analysis of both fresh frozen and formalin fixed paraffin-embedded (FFPE) samples from 134 NSCLC patients, plus in silico analysis of six microarray datasets from GEO. We then refined the set of marker genes and developed a Support Vector Machine (SVM) classifier on an independent cohort of 161 FFPE samples. These 161 samples are represented to adenocarcinomas. The classifier, combining with an ALK screening assay can provide increased accuracy for NSCLC subtyping.

The classifier's robustness was further demonstrated by diluting the tumor content with normal adjacent tissues, with as little as 20% of the original tumor content in the final sample. All diluted samples were predicted correctly (Fig 3). In addition, the estimated class probabilities did not vary significantly by dilution ratio, indicating that the classifications are robust to low tumor content.

When adenocarcinoma and squamous lysate were titrated together, the prediction scores varied roughly linearly with the adenocarcinoma or squamous cell concentration, reflecting the biological changes in the sample mixture (Fig 3).

### Results

The performance of the classifier was measured by its call concordance with three pathologists' IHC panel consensus reads. The classifier distinguished squamous and non-squamous NSCLC of the 97 FFPE independent samples with AUC of 0.98 (Fig 1) and accuracy of 94% (Fig 4). Two of the six discordant samples were confirmed as positive in our ALK screening assay. ALK fusions are generally limited in adenocarcinomas. The classifier, combining with an ALK screening assay can provide increased accuracy for NSCLC subtyping.

### Conclusions

We have developed a robust classifier that can accurately distinguish Squamous from non-squamous NSCLC. Its performance has been demonstrated by independent clinical validation samples.

### Data Sets

<table>
<thead>
<tr>
<th>Platform</th>
<th>Sample Type</th>
<th>Sample size</th>
<th>Adenocarcina</th>
<th>Large cell</th>
<th>Squamous</th>
</tr>
</thead>
<tbody>
<tr>
<td>qNPA</td>
<td>Fresh frozen*</td>
<td>134</td>
<td>134</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>FFPE</td>
<td></td>
<td>161</td>
<td>136</td>
<td>25</td>
<td></td>
</tr>
</tbody>
</table>

* Samples from the two discovery sets are Fresh frozen and FFPE matched tissues.

### References