Using paired tissue and serum samples to characterize human lung cancer metabolomics with ex vivo 1H HRMAS MRS

Hailiang Huang[1], Emily Decelle[1], Yannick Berker[1], Andreas Schuler[1,2], Isabel Dittman[1,2], Li Su[3], Eugene J Mark[1], Mark J Daly[1], David C Christiani[1,3], and Leo L Cheng[1]

[1]*Massachusetts General Hospital, Boston, Massachusetts, United States,* [2]*Charité Universitätsmedizin, Berlin, Germany,* [3]*Harvard School of Public Health, Boston, Massachusetts, United States*

Target Audience: Lung cancer is one of the primary causes of cancer related death in both men and women globally. Although tests such as CT and PET can identify cancerous legions, their radiation hazards and high costs exclude them from feasible screening tools for the general population. New screening methods allowing for the diagnosis of lung cancer at an early, asymptomatic stage would improve the prognosis for many patients, for whom the 5-year survival rate remains less than 16 percent.

Purpose: We use high resolution magic angle spinning (HRMAS) proton MRS (1HMRS) to identify metabolomic profiles of lung cancer tissue and matched serum samples. Previous studies have reported that profiles for lung cancer tissue may be predictive of the profiles of matched serum samples. In this study, we further search correlations of metabolomic data with traditional histopathology from the same tissue samples, and identify serum lung cancer metabolomic markers based on matched tissue analysis.

Methods: Paired tissue and serum samples from 107 patients of adenocarcinoma (AC) and squamous cell carcinoma (SCC), and 29 serum samples from control subjects without lung disease, were analyzed. Clinical information is listed in Table 1. *MR Spectroscopy*. MR experiments were carried out on a Bruker AVANCE spectrometer operating at 600 MHz (14.1T) and pre-cooled to 4°C. A 4mm zirconia rotor was used with Kel-F inserts to create a 10µl sample space, and D_2O was added for ^2H field locking. Rotor spinning rate was regulated by a MAS controller, and verified by measuring the inter-SSB distances from spectra with an accuracy of 1.0Hz. A repetition time of 5s and 128 transients were used to acquire each spectrum. Spectra were collected with a spinning rate of 3600Hz, with a rotor synchronized CPMG filter to reduce broad resonances; 360 CPMG cycles were applied with one pi-pulse between two rotor cycles in each CPMG cycle to result in a filter time of 200ms. Spectra were analyzed by an in-house MatLab based program. *Histopathology*. After spectroscopy, tissue samples were fixed in formalin, embedded in paraffin, cut into sets of 5µm sections at 100µm intervals, and stained with hematoxylin and eosin. Volume percentages of histological features (cancer, stroma, necrosis, lymphatic structures, and cartilage) were analyzed and quantified by a pathologist.

Table 1.

Patient Group	Age (years)	Gender	Tumor Stage	Smoker Status	Packs per year
AC 56 patients	M=63.69 SD=8.49	female=30 male=26	I=31 II=10 III=9 IV=6	Current=34 Former=21 Never=1	M=50.19 SD=29.52
SCC 51 patients	M=68.07 SD=8.62	female=16 male=35	I=31 II=10 III=8 IV=2	Current=31 Former=20 Never=0	M=71.99 SD=35.57
Control 29 patients	M=66.87 SD=9.52	female=11 male=18	N/A	N/A	M=41.34 SD=37.00

Results: We use Lasso, a feature selection method using L1 regularized linear regression, to select a subset of peaks from the spectral results to build a linear model to predict AC/SCC readings. The number of peaks selected by Lasso can be tuned by a shrinkage parameter, which is estimated using 5-fold cross-validation. We randomly split the samples into 5 datasets of about equal sizes. Each dataset in turn is used as the validation data, while the other 4 datasets are used as the training data. The validation error is averaged across all 5 validation datasets. The shrinkage parameter that has the smallest validation error is then used in Lasso to perform the model selection. We repeat this procedure 100 times to access its stability. A 2-peak model (peaks at 2.14-2.10 and 0.89-0.89 ppm) has been identified as the most optimal model (Figure 1A). The predicted probability from this model agrees well with the AC/SCC readings (Figure 1B). We use the same procedure to build predictive models for quantitative histopathology readings from serum and tissue samples. Out of the four histopathology readings (%Cancer, %Fibrosis/ Inflammation, %Necrosis and %Cartilage), we were able to build a model for %Fibrosis/Inflammation (p-value=1E-5) using 3 peaks from serum samples (peaks at 2.09-2.03, 1.33-1.32 and 3.27-3.24), and a model for %cancer cells (p-value=0.002) using one peak from tissue samples (peak at 3.73-3.71). Further data analyses are still underway in our laboratory.

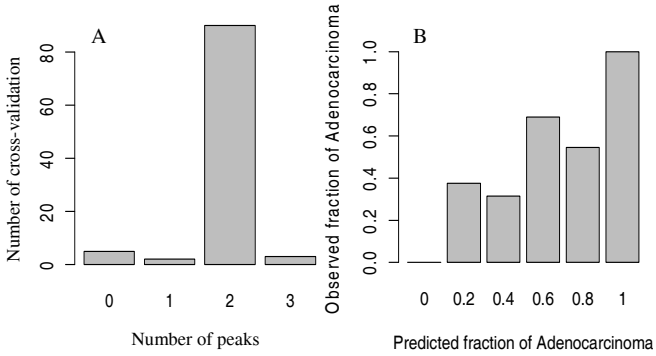

Figure 1. A) Number of peaks selected from 100 cross-validations. The 2-peak model is significantly more often selected than models of other sizes. B) Samples are binned based on the predicted probability of AC using the 2-peak model. The fraction of AC samples in each bin agrees well with the predicted probability.

Discussion and Conclusion: The advantage of using the HRMAS technique to analyze intact tissue samples non-destructively has been demonstrated by numerous studies. When applying this method on analyses of serum samples, we can routinely obtain high resolution spectra with a small amount of sample. These preliminary results, showing the potential to predict AC/SCC status with metabolomic profiles from serum, could have important implications in the screening for lung cancer. If patients could be identified by suspicious serum metabolomic profiles, and subsequently undergo more advanced radiological testing, lung cancer could be diagnosed at an earlier stage where effective treatment and a good prognosis is more likely.