ID: MC13-0032

Validation of ^{1}H-NMR spectroscopy based metabolomics as a tool to detect lung cancer via a simple blood sample

Evelyne Louis[1], Liesbet Mesotten[1,2], Michiel Thomeer[1,2],Karolien Vanhove[1,3], Kurt Vandeurzen[4], Anna Sadowska[5], Els Wauters[6], Johan Vansteenkiste[6], Christophe Dooms[6], Gunter Reekmans[7], Peter Adriaensens[7]

[1]Faculty of Medicine and Life Sciences, Hasselt University, Diepenbeek, Belgium, [2]Ziekenhuis Oost-Limburg, Genk, Belgium, [3]Algemeen Ziekenhuis Vesalius, Tongeren, Belgium, [4]Mariaziekenhuis Noord-Limburg, Overpelt, Belgium, [5]Ziekenhuis Maas en Kempen, Maaseik, Belgium, [6]Universitair ziekenhuis, Campus Gasthuisberg Leuven, Belgium, [7]Institute of Materials Research, Hasselt University, Diepenbeek, Belgium

INTRODUCTION

- Lung cancer is the leading cause of cancer death worldwide.
- Detection methods with an improved sensitivity and specificity are urgently needed because lung cancer is often only diagnosed owing to symptoms of advanced disease.
- Accumulating evidence shows that the metabolism of cancer cells differs from that of normal cells. More specifically, the entire metabolism of cancer cells is reorganized to increase anabolic reactions that induce cell growth and survival. Disturbances in biochemical pathways which occur during the development of cancer provoke changes in the metabolic phenotype.
- Nuclear magnetic resonance (NMR) spectroscopy enables the identification and the quantitative analysis of complex mixtures of metabolites, as in plasma and serum, without an extended sample preparation. Recent studies have indicated the huge potential of this technique to detect different cancer types in an early stage. For several diseases, changes in the metabolic phenotype have been shown to correlate with the presence of a certain pathology. Identifying which changes in the metabolic phenotype correlate with the presence of lung cancer could allow us on the long-term to detect the presence of lung cancer in a simple blood sample.
- Recently, our research group has build a statistical classifier or model using multivariate orthogonal partial least squares-discriminant analysis (OPLS-DA). This classifier allows to discriminate between 77 lung cancer patients and 78 controls (**training cohort**) with a sensitivity of 89% and a specificity of 95% and an area under the curve (AUC) of 0.928.

OBJECTIVE

We aim to investigate the predictive accuracy of the build OPLS-DA model for an independent population (**validation cohort 1 and 2)**

SUBJECTS AND METHODS

Study population

In validation cohort 1, subjects in whom lung cancer was detected by computed tomography (CT) and confirmed by positron emission tomography (PET)/CT in the Limburg PET center (Hasselt, Belgium) were included. In validation cohort 2, lung cancer patients from University Hospital (UZ) Leuven were included. The diagnosis of lung cancer was confirmed by means of an anatomopathological (APO) biopsy or by a medical doctor with expertise in radiological or clinical data. The control group consists of subjects, who were referred to the department of Nuclear Medicine in Ziekenhuis Oost-Limburg (ZOL) for a stress examination of the heart, and of personnel from Hasselt University. This control group did not undergo a PET/CT. The study protocol was approved by the medical-ethical committees of ZOL, UZ Leuven and Hasselt University.

^{1}H-NMR spectroscopy

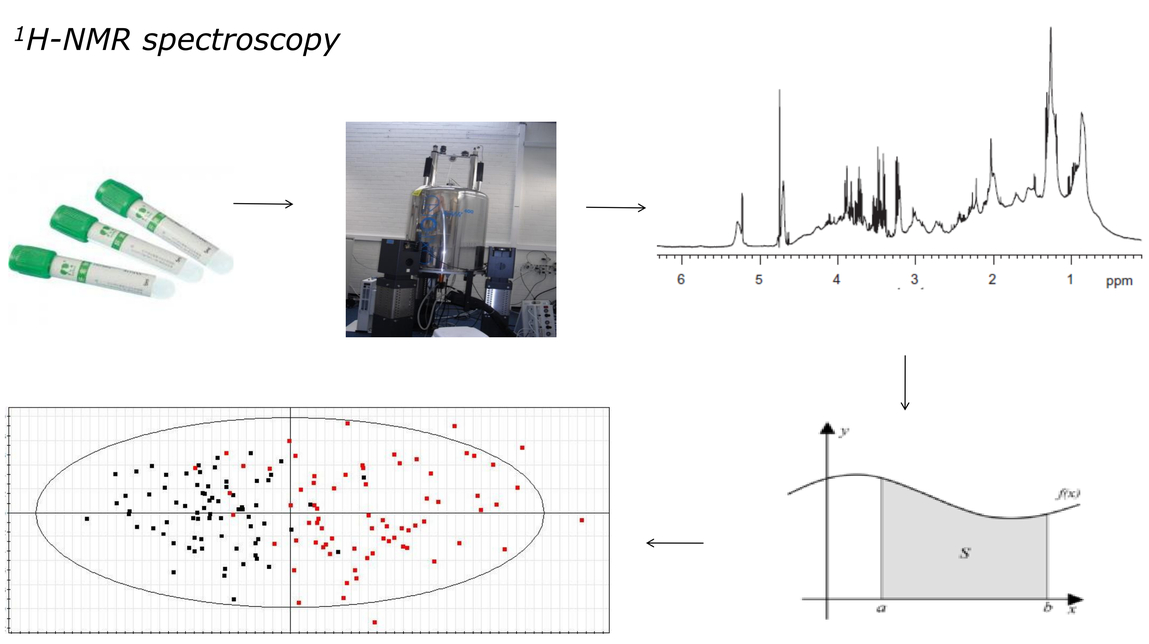

Figure 1. Determination of the metabolic phenotype of a blood sample by means of ^{1}H-NMR spectroscopy.

- Collection of fasting venous blood samples.
- Analysis of the metabolic composition of blood plasma by means of a 400 MHz NMR spectrometer.
- ^{1}H-NMR spectra were divided into 112 regions, of which the integration values were normalized to the total integrated area of all signals except these of remaining water, TSP, fructose and glucose. The end result corresponds to a set of 110 normalized integration values characteristic for 110 integration regions (all integration regions except those of water and TSP).

Statistical analysis

- Validation cohort 1 and 2 are imported as a secondary dataset in SIMCA (Version 12.0, Umetrics). Next, the classification of these cohorts is predicted by the build OPLS-DA model.
- The predictive accuracy of the model is also evaluated by means of a receiver operating characteristic (ROC) curve. The validation cohorts are both used as hold-out datasets.

RESULTS

Table 1. Subject characteristics of the training cohort and the validation cohorts.

	Training cohort		Validation cohorts		
	Lung cancer patients	Controls	Lung cancer patients (1)	Lung cancer patients (2)	Controls (1 and 2)
Number of subjects	77	78	80	78	80
Gender	M: 51 (66%)	M: 45 (58%)	M: 53 (66%)	M: 55 (71%)	M: 46 (58%)
	F: 26 (34%)	F: 33 (42%)	F: 27 (34%)	F: 23 (29%)	F: 34 (42%)
Age	67 ± 10	64 ± 13	68 ± 8	64 ± 9	61 ± 11
Smoking habits	Active: 49 (64%)	Active: 19 (24%)	Active: 51 (64%)	Active: 40 (51%)	Active: 13 (16%)
	Stopped > 6m: 25 (32%)	Stopped > 6m: 27 (35%)	Stopped > 6m: 26 (32%)	Stopped > 6m: 35 (45%)	Stopped > 6m: 34 (43%)
	Never: 3 (4%)	Never: 32 (41%)	Never: 3 (4%)	Never: 3 (4%)	Never: 33 (41%)

Table 2. Anatomopathological (APO) diagnosis of the lung tumors in the training cohort and the validation cohorts.

		Training cohort	Validation cohort (1)	Validation cohort (2)
Total number of tumors		82 (5 patients with 2 tumors)	82 (2 patients with 2 tumors)	79 (1 patient with 2 tumors)
Histology	Adenocarcinoma	28 (34%)	36 (44%)	37 (47%)
	Adenosquamous carcinoma	3 (4%)	2 (2%)	0 (0%)
	Spinocellular carcinoma	22 (27%)	18 (22%)	23 (29%)
	Large cell carcinoma	4 (5%)	2 (2%)	5 (6%)
	Carcinoid	1 (1%)	1 (1%)	0 (0%)
	Small cell lung cancer	14 (17%)	10 (12%)	10 (13%)
	No APO	10 (12%)	13 (16%)	4 (5%)
Stage	I	25 (30%)	24 (29%)	9 (12%)
	II	7 (9%)	11 (14%)	7 (9%)
	III	29 (35%)	23 (28%)	20 (25%)
	IV	21 (26%)	24 (29%)	42 (53%)
	?	0 (0%)	0 (0%)	1 (1%)

Investigate the predictive accuracy of the build model

- By using the build model, 62 out of 80 (78%) lung cancer patients and 54 out of 80 (68%) controls (**validation cohort 1**) and 61 out of 78 (78%) lung cancer patients and 54 out of 80 controls (80%) (**validation cohort 2**) are correctly classified.
- When validation cohort 1 is used as a hold-out dataset, the AUC of the build model is is 0.852. When validation cohort 2 is used as a hold-out dataset, the AUC of the build model is 0.872.

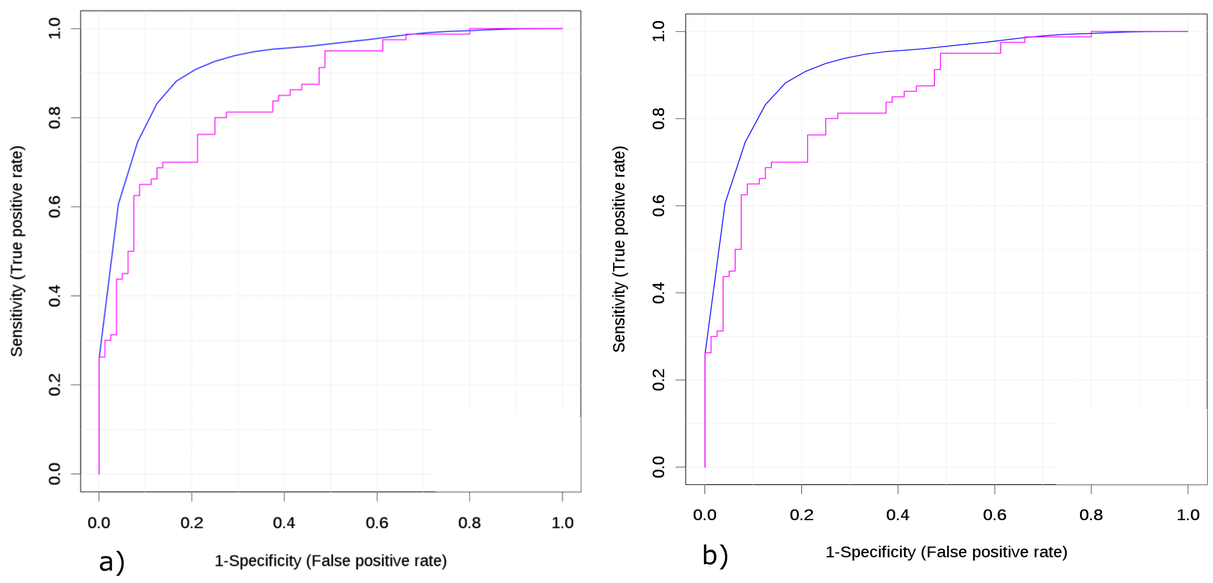

Figure 2. a) ROC curve of the build model, using the data of the training cohort (blue, AUC of 0.928) and using the data of validation cohort 1 (pink, AUC of 0.852), b) ROC curve of the build model, using the data of the training cohort (blue, AUC of 0.928) and using the data of validation cohort 2 (pink, AUC of 0.872).

CONCLUSION

The constructed model allows to classify the majority of the lung cancer patients and controls of the validation cohorts correctly. Once we have collected sufficient samples to validate this method, we want to investigate at random whether it allows to detect lung cancer in a population with a low prevalence, actually whether it can be used as a valid screening tool.

ACKNOWLEDGEMENTS

This study is part of the Limburg Clinical Research Program (LCRP) UHasselt-ZOL-Jessa, supported by the Foundation Limburg Sterk Merk, Hasselt University, Ziekenhuis Oost-Limburg and Jessa Hospital. Samples are stored at the University Biobank Limburg (UBiLim).

evelyne.louis@uhasselt.be