Members of the jury

Prof. dr. I. Lambrichts, Hasselt University, Diepenbeek, Belgium, chairman

Prof. dr. M. Thomeer, Ziekenhuis Oost-Limburg, Genk, Belgium, promotor

Prof. dr. L. Mesotten, Ziekenhuis Oost-Limburg, Genk, Belgium, copromotor

Prof. dr. P. Adriaensens, Hasselt University, Diepenbeek, Belgium, copromotor

Prof. dr. V. Somers, Hasselt University, Diepenbeek, Belgium

Prof. dr. J-P. Noben, Hasselt University, Diepenbeek, Belgium

Prof. dr. Z. Shkedy, Hasselt University, Diepenbeek, Belgium

Prof. dr. K. Nackaerts, University Hospital Leuven, Leuven, Belgium

Prof. dr. R. Heeren, Maastricht University, Maastricht, The Netherlands

This study is part of the 'Limburg Clinical Research Program (LCRP) UHasselt-ZOL-Jessa', supported by the foundation Limburg Sterk Merk, province of Limburg, Flemish government, Hasselt University, Ziekenhuis Oost-Limburg and Jessa Hospital.

"If we knew what it was we were doing, it would not be called research, would it?"

Albert Einstein

Table of Contents

List of Figures

List of Tables

List of abbreviations

^{1}H	proton
^{18}F-FDG	^{18}F-fluorodeoxyglucose
α-KG	α-ketogluterate
ACC	acetyl CoA carboxylase
ACLY	ATP citrate lyase
ATP	adenosine triphosphate
AUC	area under the curve
bcl-2	b-cell lymphoma-2
COPD	chronic obstructive pulmonary disease
CPMG	Carr-Purcell-Meiboom-Gill
CXR	chest radiography
D_2O	deuterium oxide
DPCP	detectable preclinical phase
ER	estrogen receptor
FAS	fatty acid synthase
FC	fold change
FID	free induction decay
Gln	glutamine
GLS	glutaminase
Glu	glutamate
HER2	human epidermal growth factor receptor
HIF	hypoxia-inducible factor
HK-II	hexokinase-II
LDCT	low-dose computed tomography
LDH	lactate dehydrogenase
LiHe	lithium-heparin

LN_2	liquid nitrogen
MCT4	monocarboxylate transporter 4
MS	mass spectrometry
mSv	millisievert
NADH	nicotinamide adenine dinucleotide
NADPH	nicotinamide adenine dinucleotide phosphate
NLST	National Lung Screening Trial
NMR	nuclear magnetic resonance
NPV	negative predictive value
OAA	oxaloacetate
OPLS-DA	orthogonal partial least squares discriminant analysis
OXPHOS	oxidative phosphorylation
PC	principal component
PCA	principal component analysis
PDH	pyruvate dehydrogenase
PET	positron emission tomography
PFK-1	phosphofructokinase-1
PGM	phosphoglycerate mutase
PHDs	prolyl-4-hydroxylase domain proteins
PI3K	phosphatidylinositol 3-kinase
PIP2	phosphatidylinositol-4,5-diphosphate
PIP3	phophatidylinositol-3,4,5-triphosphate
PK	pyruvate kinase
PLCO	Prostate, Lung, Colorectal and Ovarian
PLS-DA	partial least squares discriminant analysis
ppm	parts per million
PPP	pentose phosphate pathway
PPV	positive predictive value

PR	progesterone receptor
PTEN	phosphate and tensin homologue
RCT	randomized controlled trial
RT	room temperature
S/N	signal-to-noise ratio
SCLC	small cell lung cancer
SPREC	Standard PREanalytical Code
TSP	trimethylsilyl-2,2,3,3-tetradeuteropropionic acid
VDAC	voltage-dependent anion channel
VHL	von Hippel Lindau
VIP	variable importance for the projection

CHAPTER 1

General introduction and objectives

Lung cancer: a global scourge

Lung cancer constitutes a major public health problem. Globally, it is the most common cancer in males and the fourth most common cancer in females, with approximately 1.8 million new cases diagnosed in 2012. Moreover, lung cancer represents the leading cause of cancer death worldwide, accounting for about 20% of all cancer-related deaths (1). The worldwide relative five-year survival rate of lung cancer is extremely poor, ranging from merely 5 to 10%, which is due to the fact that it is usually diagnosed at an advanced disease stage when curative treatment is limited. This, in turn can be attributed to the lack of symptoms during the early phases of lung cancer development (2).

In Belgium, lung cancer is the leading cause of cancer death in males and the second most common cause of cancer death in females. In 2008, the relative five-year survival rates for lung cancer ranged from 52% (stage I) to 2% (stage IV) and from 66% (stage I) to 5% (stage IV) for Belgian men and women, respectively (**Figure 1.1**) (3). Clearly, the prognosis of the patient is inversely related to the stage of the disease at time of diagnosis (4). If a closer look is taken at premature cancer death, i.e. cancer death before the age of 75 years, lung cancer is the most common and second most common cause in Belgian men and women, respectively (5). More specifically, 3,718 new cases of lung cancer were diagnosed among Belgian men in 2008 and 2,911 of these cases (78%) died as a direct consequence of their disease. Similarly, in the same time period, there were 1,470 new cases detected among Belgian women, with 1,016 of these patients (69%) dying from lung cancer (6).

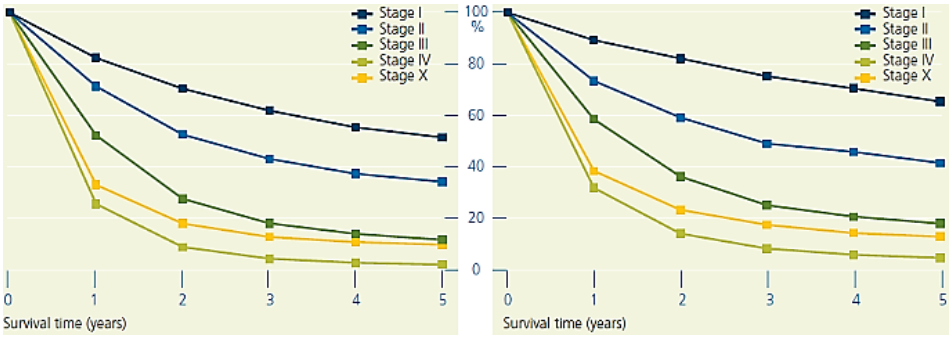


Figure 1.1. Relative lung cancer survival rates by stage in Belgian men (left) and women (right) between 2004 and 2008. Reprinted with permission from Francart et al. (3).

Chronic tobacco use is the principal cause of the development of lung cancer, with 80-90% of all lung cancer cases emerging as a consequence of cigarette smoking (7). Consequently, it seems evident that smoking cessation reduces the risk of developing lung cancer (8). However, even decades after smoking cessation, ex-smokers remain at elevated risk for lung cancer compared with people who have never smoked. Furthermore, smoking cessation does not have a beneficial impact on lung cancer mortality in the short-term, because lung cancer takes about 20 years to develop (9, 10). Hence, although prevention by means of tobacco control should stay the primary focus of public health campaigns, there is an imperative need for additional strategies beyond smoking cessation in the fight against lung cancer (10, 11). Early detection of lung cancer is required to allow effective treatment and to achieve a substantial decline in (premature) lung cancer mortality (9, 12).

Screening

Screening aims to detect lung cancer in an early stage, i.e. before patients experience clinical symptoms and when treatment is the most effective. This early detection should benefit the patient by increasing life expectancy and quality of life (13-15).

Criteria for effective screening and their application to lung cancer

In 2001, Obuchowski et al. (16) formulated ten criteria to evaluate whether screening tests are effective or not: two of these criteria apply to the disease screened for, five to the offered screening test and three to the available early stage treatment options (**Table 1.1**).

Table 1.1. Ten criteria for effective screening.

Characteristics of:	Specific criterion:	
Disease	Serious consequences	
	High prevalence of detectable preclinical phase	
Screening test	Acceptable levels of sensitivity and specificity	cost-effective
	Detection of disease before critical point	
	Little morbidity	
	Little pseudodisease	
	Affordability and accessibility	
Treatment	Existence	
	Effectiveness before symptoms begin	
	Little risks or toxicity	

Reprinted with permission from Obuchowski et al. (16).

Disease

The poor prognosis of lung cancer due to the lack of symptoms during early phases of cancer development makes lung cancer a perfect candidate for screening: even a minor advantage from screening could save many lives (17-19).

Furthermore, to warrant the costs of lung cancer screening, the detectable preclinical phase (DPCP) should have a high prevalence in the high-risk target population (**Figure 1.2**). The target population for lung cancer screening consists of asymptomatic men and women who are between 55 and 74 years old and who have a smoking history of at least 30 pack years (15). In this population, the prevalence of DPCP is 2-4% (20).

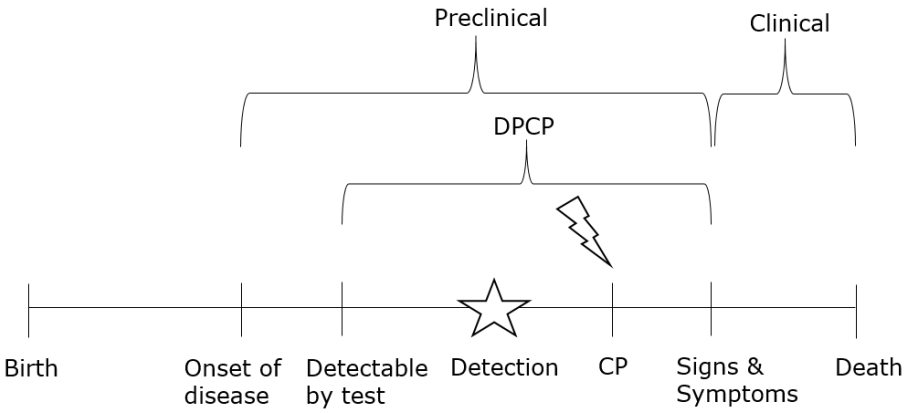


Figure 1.2. Diagram which displays a time line from birth to death from an individual. The natural history of a disease can be separated into a preclinical and a clinical phase. The preclinical phase is the time from the onset of disease to the appearance of clinical symptoms, whereas the clinical phase starts when clinical symptoms appear. The detectable preclinical phase is part of the preclinical phase, i.e. it is the interval of time when the disease is detectable by the screening test. A screening test is only effective when the critical point (i.e. when the primary tumor metastasizes) occurs during the detectable preclinical phase. CP: critical point; DPCP: detectable preclinical phase. Adapted with permission from Obuchowski et al. (16).

Screening tests

A screening test should be cost-effective, which means that it should not place a massive burden on the health care system and it should not cause more harm than good. The latter implies that the part of the population without the disease should not be harmed by the screening test and that the number of false positive results should be low to prevent unnecessary surgical interventions (15, 21, 22).

A cost-effective screening test 1) has acceptable levels of sensitivity and specificity, 2) is low-risk, accessible and affordable and 3) reduces disease-related mortality.

It is crucial to select a high-risk target population for screening in order to maximize the benefit-risk balance. This population consists of individuals who have the largest risk to develop the disease and who will experience the most benefits from screening (17, 21, 23). Limiting screening to high-risk individuals will lower the number of participants eligible for screening, thereby reducing the costs. Additionally, selecting high-risk individuals will also lead to a higher prevalence of the disease in the target population and a concomitant decrease in the incremental cost-effectiveness ratio (24).

In order to accurately select high-risk individuals for lung cancer screening, robust methods for risk prediction are required (17, 25). Current models to estimate lung cancer risk have tended to concentrate on epidemiological and clinical risk factors, such as age, gender, smoking behavior, occupational exposure to asbestos, prior diagnosis of pneumonia, previous history of cancer and family history of lung cancer (26-30).

Treatment

Effective treatment (i.e. surgery) is possible when lung cancer is detected in an early stage (31).

Screening biases

It has to be kept in mind that survival as a determinant of screening effectiveness is prone to well-known screening biases, including lead-time bias and overdiagnosis bias. Accordingly, an improvement in survival is not necessarily equivalent to a decrease in mortality. Hence, randomized controlled trials (RCTs), which compare mortality rates between screened individuals and a group of controls who do not undergo any form of screening are essential to determine whether screening allows to reduce disease-specific mortality, an endpoint for the evaluation of screening effectiveness that circumvents screening biases (14, 18, 23, 32).

Lead-time bias

The phenomenon of lead-time bias is depicted in **Figure 1.3**. Lead-time is the time between disease detection by screening (before patients experience any clinical symptoms) and the moment when the disease would have been diagnosed without screening (generally after the patient experiences clinical symptoms). Therefore, even when treatment in an early stage has no effect, the survival of patients identified by screening seems to be prolonged by the addition of the lead-time, while in reality the time of death has not been delayed (2, 14, 18).

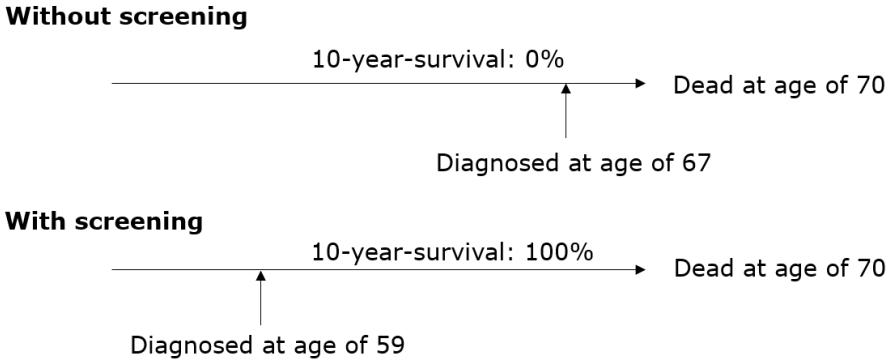


Figure 1.3. The phenomenon of lead-time bias. In case of an ineffective treatment, early detection of a disease by means of screening appears to increase survival, while in reality it has no effect on mortality. Adapted with permission from Welch et al. (18).

Overdiagnosis bias

Overdiagnosis bias represents the detection of pseudodisease, i.e. slow-growing indolent lesions that meet the pathological definition of a certain disease but which do not produce clinical symptoms before the patient dies of competing age-related causes (18, 33). Again, the detection of many people with pseudodisease will increase the survival rate, while mortality remains unaffected (**Figure 1.4**).

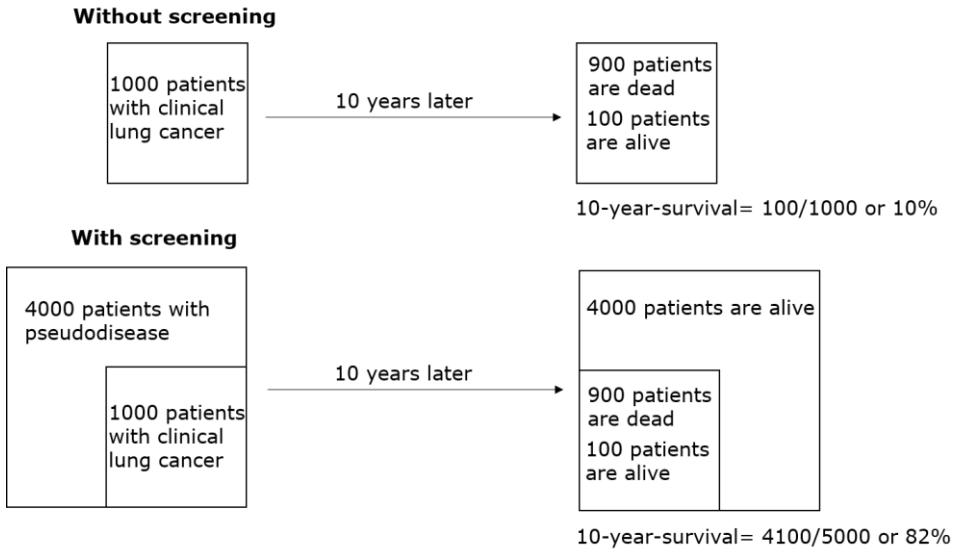


Figure 1.4. The phenomenon of overdiagnosis bias. The detection of pseudodisease appears to increase survival, while in reality the number of deaths remains unaffected. Adapted with permission from Welch et al. (18).

Lung cancer screening

Eighty-five percent of all lung cancers represent as non-small cell lung cancer and fifteen percent of the patients suffer from small cell lung cancer (SCLC) (15, 34). Due to its rapid doubling time and early metastases, it is generally believed that screening is not beneficial in patients with SCLC (35, 36). There are various methods available that can be used to screen for lung cancer, including chest radiography (CXR), sputum cytology and low-dose computed tomography (LDCT).

Chest radiography and sputum cytology

Although CXR has been widely used for lung cancer screening in the past, early large-scale RCTs failed to demonstrate a beneficial effect of CXR screening alone or in combination with sputum cytology on lung cancer mortality (37-41). However, many of these studies had methodological limitations (9, 21). Therefore, recently, the large Prostate, Lung, Colorectal and Ovarian (PLCO) trial was conducted to examine the effectiveness of annual CXR screening compared to usual care. The PLCO trial did not show a significant difference in lung cancer mortality between CXR screening and usual care, a finding which is consistent with the results of earlier RCTs (42). In addition, CXR appears to be a poor

diagnostic tool for detecting early stage lung cancer, as it does not allow to detect lung tumors smaller than two cm in diameter (43).

Low-dose computed tomography

LDCT, a major advance in imaging technology introduced in the mid-1990s, renewed interest in lung cancer screening following the disappointing results of RCTs using CXR (44, 45). LDCT offers fast image acquisition at a radiation dose (± 1.5 millisievert (mSv)) significantly lower than that of a standard CT scan (± 7 mSv) (13, 46). Nevertheless, the radiation dose necessary to perform LDCT is still about ten times larger than that needed for CXR (0.02 mSv) (15, 47).

Various single-arm cohort studies, including the Early Lung Cancer Action Project, have demonstrated that LDCT is far more sensitive in detecting lung cancer than CXR (20, 48-57). LDCT screening detects lung cancers at a smaller size (as small as 1 mm in diameter) and at an earlier stage which probably offers a higher likelihood of surgical cure compared with the lung tumors seen in clinical practice or detected by conventional CXR screening (21, 32, 58-60). However, because the aforementioned studies had no internal control group, they were subjected to screening biases, leaving them unable to answer the question whether LDCT screening allows to reduce lung cancer mortality (14, 45, 61). Accordingly, once the Lung Screening Study (a feasibility trial) had confirmed that LDCT allows to detect more lung cancers than CXR (60, 62), the results of the two largest prospective RCTs with lung cancer mortality as a primary endpoint – the National Lung Screening Trial (NLST) and the Dutch-Belgian lung cancer screening trial (Dutch acronym - NELSON) – were eagerly awaited to ascertain the true benefit of LDCT screening.

National Lung Screening Trial

The North-American NLST was a multicenter RCT which compared LDCT with CXR in the screening of a high-risk population for lung cancer. Remark that the NLST was already started before the current risk models were properly validated. Since lung cancer is principally due to chronic tobacco use and predominantly occurs in elderly people, the NLST selected eligible participants based on age and smoking behavior. More specifically, eligible participants for the NLST were aged between 55 and 74 years and had a smoking history of at least 30 pack years. Former

smokers were only included in the study if they had quit smoking within the past 15 years. From the 53,454 enrolled individuals, 26,722 were randomly assigned to three annual rounds of LDCT screening, while the remaining 26,732 individuals underwent three annual rounds of CXR screening. In all rounds, there was a higher rate of positive screening results in the LDCT group (T0: 27%, T1: 28%, T2: 17%) than in the CXR group (T0: 9%, T1: 6%, T2: 5%). In addition, LDCT screening introduced a stage shift, i.e. the number of early stage lung cancers detected by LDCT was increased whereas the number of advanced stage lung cancers was decreased relative to CXR screening (63). However, most of the positive results in both groups were false positive results. In the LDCT group, the chance that a positive result was a true positive result (positive predictive value or PPV) was only 3.8% (64).

Moreover, the NLST researchers showed a 20% relative reduction in lung cancer mortality and a 7% relative reduction in death from any cause in the LDCT group compared to the CXR group. To avoid one lung cancer death, 320 high-risk individuals needed to be screened by LDCT (63). A cost-effectiveness analysis of LDCT screening as performed by the NLST researchers showed a cost of $81,000 per quality-adjusted life-year gained (65). Although these results are already very promising, (until today) it is not clear yet whether they also apply to the general population.

Dutch-Belgian lung cancer screening (NELSON) trial

In Europe, a number of smaller RCTs are ongoing to investigate the mortality benefit of LDCT screening (66-70). The European RCTs differ from the NLST in that the control group does not undergo any form of screening as opposed to CXR screening. So far, three of these studies have demonstrated no mortality benefit of annual LDCT screening in their interim reports (66, 67, 69). Nonetheless, it can be argued that none of these studies had adequate statistical power at the time to identify a significant mortality reduction (12, 17). Currently, the mortality results of the second largest trial after the NLST, i.e. the Dutch-Belgian lung cancer screening (NELSON) trial, are eagerly awaited in order to confirm the findings of the NLST in a separate population (71).

The Dutch-Belgian NELSON trial evaluated whether three rounds of LDCT with an increasing length of screening interval (1, 2 and 2.5 years) relative to no screening

at all allows to reduce lung cancer mortality in a population at increased risk for lung cancer (71, 72). Note that also the NELSON trial was already begun before the existing risk models were accurately validated and that eligible participants were selected on the basis of age and smoking behavior. In particular, eligible participants for the NELSON trial were aged between 50 and 75 years, had smoked 15 or more cigarettes a day for more than 25 years (≥ 18.75 pack years) or 10 or more cigarettes a day for more than 30 years (≥ 15 pack years). Former smokers were only included in the study if they had quit smoking for less than ten years (73). From the 15,822 participants, 7,915 were randomly assigned to three rounds of LDCT screening (after 1 year, after 3 years and after 5.5 years) and 7, 907 individuals were randomly assigned to a control group without screening (74, 75). For all three screening rounds combined with two-year follow-up, sensitivity was 84.6% and specificity was 98.6%. In addition, the chance that a positive result was a true positive result (positive predictive value or PPV) was 40.4% and the chance that a negative result was a true negative result (negative predictive value or NPV) was 99.8%. This implies that more than half of the study participants assigned to the screening group were referred for false positive results (71). Nevertheless, the PPV observed for two-year follow-up in the NELSON trial was already much higher compared to the one found for approximately two-year follow-up in the NLST, namely 3.8% (64). In conclusion, using LDCT screening with an increasing length of screening interval has a high sensitivity and specificity for detecting lung cancer. Nevertheless, the results of the mortality analyses are still pending and are expected at the end of 2015 (21). The NELSON trial has 80% statistical power to show a relative reduction in lung cancer mortality of at least 25% during ten-year follow-up (72, 73).

Drawbacks of LDCT screening

The cost-effectiveness of LDCT screening must be rigorously examined. Therefore, the benefits of screening (i.e. high sensitivity and reduction in lung cancer mortality) need to be weighed against its potential harms and the costs (19). The financial costs of LDCT screening that need to be taken into account are the costs of the screening examination, the costs of the follow-up diagnostic tests in patients who have a positive screening result and the costs of treatment in patients who are diagnosed with lung cancer (61). The most important harms of LDCT screening are 1) the high rate of false positive results accompanied by

follow-up diagnostic tests and possible complications from these procedures, 2) the potential of overdiagnosis, and 3) possible radiation-induced carcinogenesis (12, 45). These harms entail a lot of additional costs (15).

False positive results

The main challenge of LDCT screening is the high rate of false positive results. Besides malignant pulmonary lesions, LDCT also identifies benign pulmonary lesions, such as inflammation and scars from previous infections (9, 13, 44). In the NLST, 96% of all positive screening results in the LDCT group were false positive (64). This results in emotional stress, needless financial costs and increased risks for healthy people (58). These risks encompass unnecessary radiation, biopsies and surgical procedures and are associated with higher morbidity and mortality rates (13, 14, 21).

Overdiagnosis

A second harm of LDCT screening is overdiagnosis. This refers to histopathologically confirmed lung cancers detected by screening that would not cause clinical symptoms or lead to death from the patient when left untreated. Overdiagnosis results from lesions that are so indolent that the patient will not die from them, but rather from competing causes of death (12, 21, 76). Consequently, these patients will not benefit from treatment and can only be harmed by it, since lung cancer resection is associated with significant morbidity and mortality (15, 58). Furthermore, these individuals have to deal with the anxiety associated with the diagnosis of lung cancer (18, 33). A recent study estimated that about 25% of all lung cancer cases detected by LDCT screening represent overdiagnosis (77).

Radiation exposure

A consequence of false positive results and overdiagnosis is additional exposure to ionizing radiation (1.5 mSv per LDCT examination) during unnecessary follow-up diagnostic tests (46). Although the dose from a single examination is small, the associated risks are not negligible. In particular, the excess radiation can interact synergistically with the damaging effects of cigarette smoking and might therefore induce the development of cancer, which can emerge even decades after exposure (14, 78). While the risks from a single examination are small, Brenner

et al. (78) have shown that annual LDCT screening in 50-year-old female smokers adds about 0.85% to the population-based expected risk of 16.9%, accounting for a 5% increase in risk. In 50-year-old male smokers, annual LDCT screening adds about 0.23% to the expected risk of 15.8%, a 1.5% increase in risk. Consequently, a mortality benefit of at least 5% is required to outweigh these radiation-related risks. Furthermore, de Gonzalez et al. (79) have revealed that lung cancer mortality needs to be reduced by 10% to offset the radiation-related risks of three annual LDCT screening rounds in 50-year-old smokers. Hence, the relative reduction of 20% in lung cancer mortality obtained in the NLST is sufficient to outweigh the radiation-related risks (63).

The need for complementary diagnostic cancer biomarkers

Due to the high false positive rate of LDCT, there is an increasing interest in improving the accuracy of current risk models by incorporating lung cancer risk-related biomarkers in order to better select high-risk individuals eligible for LDCT screening, thereby lowering the false positive rate and the corresponding financial burden (80, 81). Biomarkers are biological characteristics that are objectively measured and evaluated as indicators of normal biological processes, pathological processes or pharmacologic responses to a therapeutic intervention (82). A promising and clinically applicable biomarker for risk prediction should ideally be inexpensive, obtainable in a non-invasive way and exhibit a high sensitivity and specificity (83, 84). A blood-based diagnostic biomarker represents an appealing option to complement LDCT screening since blood samples can be obtained non-invasively and with minimal risk for the patient (85-87). Accordingly, recent studies have investigated the role of cell-free DNA and RNA, proteins and metabolites as diagnostic biomarkers for cancer (85, 88-92). Moreover, in the last couple of decades the -omics of systems biology, genomics, transcriptomics, proteomics and metabolomics have become increasingly popular **(Figure 1.5)** (93, 94). These sciences permit the global analysis of DNA, mRNA, proteins and metabolites in biological systems, respectively (95).

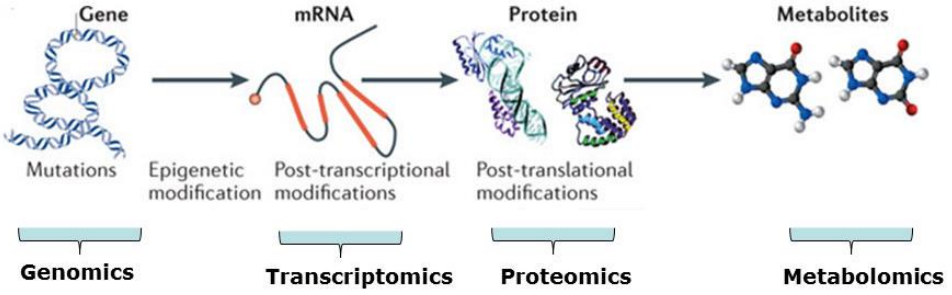


Figure 1.5. The –omics cascade of systems biology. Adapted with permission from Patti et al. (96).

Metabolomics

Metabolomics, also sometimes referred to as metabonomics, is defined as the analysis of the complete set of metabolites, i.e. the metabolome, within a biological system (97, 98). Metabolomics is an emerging field that combines analytical methods with multivariate projection methods (99). Although it is complementary to the other –omics fields, metabolomics offers a number of benefits (100, 101). Firstly, in contrast to the other –omics fields, it is high throughput and relatively cheap on a per sample basis (102-104). In addition, the metabolome is further down the line from gene to function and therefore reflects changes in the observed phenotype (105, 106). The metabolic phenotype is the end result of genetic and environmental (diet, physical activity) influences and provides a readout of the metabolic state of an individual (107). It is crucial to note that the metabolic phenotype is not only affected by disease processes, but also by confounding factors, such as age, gender, ethnicity, diet, drug administration and lifestyle (107-113).

Metabolomics studies are broadly divided into targeted and untargeted analyses. The untargeted approach, also referred to as metabolic profiling, focuses on the profiling of the total complement of metabolites in a biological sample. The targeted approach concentrates on the identification and quantification of a predetermined set of metabolites, such as those involved in a specific biochemical pathway (114). The Human Metabolome Project estimated that the number of metabolites is currently more than 40,000, an increase of 600% relative to the estimated number of 6,500 metabolites in 2007 (115, 116).

The main analytical platforms used in metabolomics studies are based on nuclear magnetic resonance (NMR) spectroscopy or mass spectrometry (MS) (103, 114, 117-119). The latter is combined with chromatographic separation of the metabolites using either gas- or liquid chromatography (103, 120). NMR spectroscopy and MS have their own specific advantages and limitations and are therefore considered to be ideally complementary. The advantages and limitations of both techniques are outlined in **Table 1.2**. Although MS is significantly more sensitive than NMR spectroscopy, the latter is highly reproducible (>98%), quantitative and non-destructive, provides structural information and requires minimal sample preparation (82, 103, 114, 121-123). Currently, no single analytical platform enables to identify and quantify all metabolites within a biological sample owing to their significant chemical diversity (114, 124).

Table 1.2. Advantages and limitations of NMR spectroscopy and MS, the main analytical platforms used in metabolomics studies.

	NMR spectroscopy	**MS**
Sample volume	Large sample (200-400 µl)	Small sample (1-10 µl)
Sample handling	All metabolites with concentrations above the detection limit can be detected in one measurement	Requires chromatographic separation techniques for different classes of metabolites
Sample recovery	Non-destructive	Destructive
Sample preparation	Minimal (addition of D_2O and chemical shift reference)	Substantial (separation techniques)
Sensitivity	Low (µM range)	High (pM range)
Structural information	High	Low
Analytical reproducibility	Very high	Moderate

D_2O: deuterium oxide; MS: mass spectrometry; NMR: nuclear magnetic resonance. Adapted from Lindon et al. (103), Van et al. (98), Bictash et al. (108), Aboud et al. (125) and Emwas et al. (114, 117).

NMR spectroscopy

NMR-based metabolomics investigations of biofluids have been in use since the 1980s (126, 127). NMR spectroscopy takes advantage of the fact that nuclei which possess a nuclear spin exist at different energy levels when placed in a strong magnetic field (94, 100). NMR active nuclei used to examine biological samples

encompass ^{1}H, ^{13}C, ^{15}N, ^{19}F and ^{31}P. Of these nuclei, the proton (^{1}H) is the most commonly used in the field of NMR-based metabolomics as it has the highest relative sensitivity, a natural abundance of 99.9%, and as it is ubiquitously present in metabolites (95, 106). A more thorough explanation of the basics of ^{1}H-NMR spectroscopy is added at the end of this chapter.

Workflow of a ^{1}H-NMR-based metabolomics study

Figure 1.6 shows a flow chart of the steps required for performing ^{1}H-NMR-based metabolomics studies. These steps include sample collection, storage and preparation, data acquisition, data processing and data analysis and biological interpretation. To minimize the impact of confounding factors, all experimental steps should be optimized carefully and performed according to standard operating procedures (128, 129).

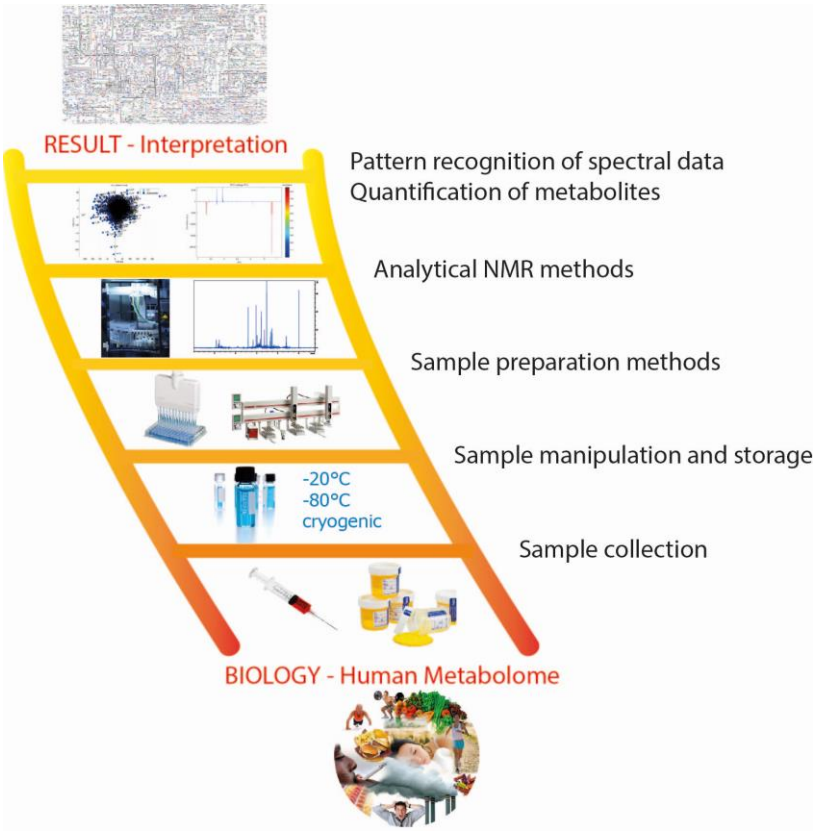


Figure 1.6. Overview of the steps required for performing a ^{1}H-NMR-based metabolomics study. Reprinted with permission from Dona et al. (130).

Sample collection, storage and preparation

^{1}H-NMR-based metabolomics studies mostly focus on biofluids, such as serum, plasma and urine (130-132). These biofluids can be obtained easily and in a relatively non-invasive way. Moreover, these body fluids can be analyzed by NMR with minimal sample preparation (82). In this thesis, the focus will be put on blood plasma.

Because of diurnal variation and to minimize variation associated with diet, it is desirable to collect blood samples in the morning and after an overnight fast (8-12 h) (110, 113, 133, 134). In order to enable collection of plasma, blood should be collected into tubes that contain an anti-coagulant agent in order to prevent clotting. For ^{1}H-NMR analysis of plasma, lithium-heparin is the preferred anti-coagulant because the signals of other anti-coagulants such as ethylenediamine tetraacetic acid and sodium citrate interfere with metabolite signals in the ^{1}H-NMR spectrum (95). Blood samples should be cooled until centrifugation in order to quench the metabolism of blood cells (95). After centrifugation, plasma should be collected in multiple aliquots in order to avoid erroneous results due to several freeze/thaw cycles. In order to ensure sample stability, long-term storage of plasma samples should be at -80°C (95, 128, 134). Furthermore, prior to ^{1}H-NMR analysis the thawed plasma sample is centrifuged to remove any cell debris. Furthermore, a deuterated solvent (deuterium oxide) is added as a magnetic field lock signal and a chemical shift reference is added as a calibration standard. For water-soluble metabolites, trimethylsilyl-2,2,3,3-tetradeuteropropionic acid (TSP) is often employed (95, 114, 135, 136).

Data acquisition

Most ^{1}H-NMR-based metabolomics studies of biofluids rely on 1D-NMR spectra. Since it takes only a few minutes to acquire these spectra, 1D-^{1}H-NMR is ideally suited for high-throughput screening (103). ^{1}H-NMR spectra of blood plasma (and serum) display broad signals from macromolecules (proteins and lipoproteins) with sharp peaks from metabolites superimposed on them (106). The broad signals of proteins and lipoproteins are generally attenuated by applying the Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence in order to improve the visibility of sharp signals from metabolites (82, 137). Furthermore, ^{1}H-NMR spectra of biofluids show a large signal arising from the water protons which obscures a large

part of the ^{1}H-NMR spectrum if unsuppressed. In general, this dominant signal is eliminated by implementing water suppression in the pulse sequence (103, 106).

Due to the narrow chemical shift range (10 parts per million (ppm)), 1D-^{1}H-NMR spectra of biofluids are very crowded and suffer from severe spectral overlap (138). As a result, assign NMR resonances to a specific metabolite is a challenging task, leading to uncertainties in signal assignments (114, 136). Assigning signals in ^{1}H-NMR spectra is generally based on chemical shifts, relative intensities, signal multiplicities and coupling constants. Furthermore, the presence of a suspected metabolite in the ^{1}H-NMR spectrum can be confirmed by adding a known compound to the studied biofluid (i.e. spiking) (106).

The spectral overlap issue is especially encountered on low- to medium-field NMR spectrometers (400-600 MHz), the instruments which are easy accessible in most research institutes and therefore mostly used in metabolomics studies (82, 139). Higher-field NMR instruments (800-950 MHz) provide an improved spectral resolution and signal-to-noise ratio (S/N) (106, 140, 141). Next to using higher-field NMR instruments, 2D-NMR experiments can be performed to improve signal dispersion and to elucidate the connectivities between signals, thereby increasing the confidence in a correct metabolite assignment (106, 142). However, the significantly longer acquisition times (hours instead of minutes for a 1D-NMR experiment) and the complex data analysis restrict common use of 2D-NMR experiments in metabolomics research (114, 136).

Data processing

Data processing is an essential step in the ^{1}H-NMR workflow in order to convert the obtained raw data in such a way that subsequent data analysis is easier and more accurate (95). First, the acquired ^{1}H-NMR spectra are subjected to manual phase correction, automated baseline correction and spectral referencing using NMR software (82). Subsequently, the spectra are subdivided into fixed (e.g. 0.01 or 0.04 ppm) regions (so-called bins or buckets) or in variably sized bins (i.e. intelligent bucketing) in order to reduce the number of data points and the impact of pH-induced changes on the chemical shift (95, 143, 144). Fixed spectral binning has the disadvantage of the possibility of splitting peaks between adjacent bins (114). The NMR spectra needs to be normalized in order to account for concentration differences between plasma samples (95, 136). The most common

method of normalization in the field of ^{1}H-NMR-based metabolomics is to divide each integral value by the total sum of integration values (i.e. integral or constant sum normalization) (145-147). However, this method is prone to the presence of extreme concentrations of single metabolites (e.g. glucose) in plasma samples, hampering subsequent accurate data analysis (147, 148). Furthermore, the normalized data needs to be scaled in order to avoid that the most abundant metabolites will dominate the constructed statistical models, while low-intensity metabolites, although biologically relevant, will be disregarded. A number of scaling methods are commonly used, i.e. mean centering, unit variance scaling, Pareto scaling and variable stability scaling (82, 147, 149, 150). The most used combination is mean centering and Pareto scaling (95). The end result of data processing is a table consisting of rows corresponding to the observations and columns corresponding to the metabolic variables (i.e. integration values), which will be subjected to data analysis (136).

Data analysis and biological interpretation

The objective of pattern recognition methods such as multivariate statistical analyses is to reduce the complexity of the generated NMR data and to present the information in a simple and interpretable format in order to extract useful information from the flood of data, i.e. the integration values which contribute most to the discrimination between the patients and the controls and the underlying disturbed biochemical pathways (95, 120, 136, 151). The first step is unsupervised principal component analysis (PCA), an exploratory tool which summarizes the majority of the variation in the original dataset into a small number of principal components (PCs) without a priori knowledge of sample class, thereby reducing the dimensionality of the data (120, 152). Each PC is a linear combination of the original variables whereby each successive PC describes the maximum amount of variance which was not accounted for by the preceding PCs (82, 95). Subsequently, supervised orthogonal partial least squares discriminant analysis (OPLS-DA) is performed in order to construct a classification model that optimally differentiates between healthy and diseased subjects (120, 153). In order to avoid model overfitting, the predictive ability of the constructed classification model needs to be evaluated in an independent (i.e. data not used to construct the model) validation set (103, 108). More information on

multivariate projection methods which can be used in ^{1}H-NMR-based metabolomics studies is added at the end of this chapter.

^{1}H-NMR-based metabolomics in cancer research

^{1}H-NMR-based metabolomics holds great potential in cancer diagnosis, prediction of therapy response and development of new therapies since cancer cells are characterized by profound metabolic alterations (154, 155). More specifically, the complete metabolism of cancer cells is reorganized to meet their aberrant demands for nutrients to support growth, proliferation and survival under suboptimal conditions (156-158).

Recently, the hallmarks of cancer, i.e. self-sufficiency in growth signals, insensitivity to anti-growth signals, evasion of apoptosis, limitless replicative potential, sustained angiogenesis and tumor invasion and metastasis, described by Hanahan et al. were revisited (159). Deregulated cellular metabolism is now included amongst the hallmarks of cancer (**Figure 1.7**) (160). In this thesis, two major metabolic pathways which are altered in cancer cells are discussed, i.e. enhanced "aerobic" glycolysis and increased glutaminolysis.

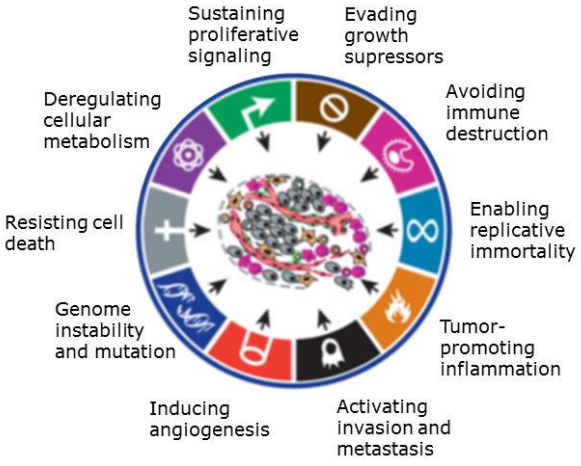


Figure 1.7. The revisited hallmarks of cancer. Deregulated cellular metabolism is now included among the hallmarks of cancer. Adapted with permission from Hanahan et al. (160).

Metabolic reprogramming in cancer cells

In the presence of O_2 (aerobic conditions), normal cells predominantly metabolize glucose to pyruvate through glycolysis and fully oxidize the produced pyruvate to CO_2 in the mitochondria through the Krebs cycle and oxidative phosphorylation (OXPHOS) (**Figure 1.8**). In the absence of O_2 (anaerobic conditions), normal cells reduce pyruvate and NAD^+ to lactate and nicotinamide adenine dinucleotide (NADH) by means of lactate dehydrogenase (LDH) (161). Subsequently, lactate is released into the extracellular space by monocarboxylate transporter 4 (MCT4), regenerating NAD^+ in order to continue glycolysis (**Figure 1.9**). The process that converts glucose into pyruvate generates 2 molecules of adenosine triphosphate (ATP) per molecule of glucose, whereas the complete oxidation of glucose through the Krebs cycle and OXPHOS produces 30 molecules of ATP (155).

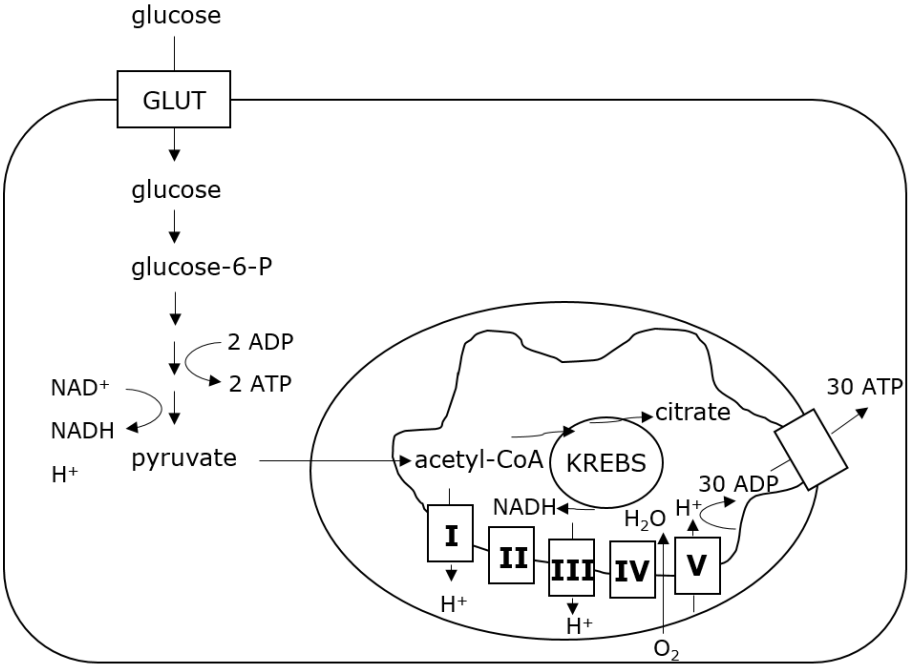


Figure 1.8. The metabolism of glucose in normal cells under aerobic conditions. Acetyl-CoA: acetyl coenzyme A; ADP: adenosine diphosphate; ATP: adenosine triphosphate; glucose-6-P: glucose-6-phosphate; GLUT: glucose transporter; NADH: nicotinamide adenine dinucleotide. Adapted with permission from Bracke et al. (162).

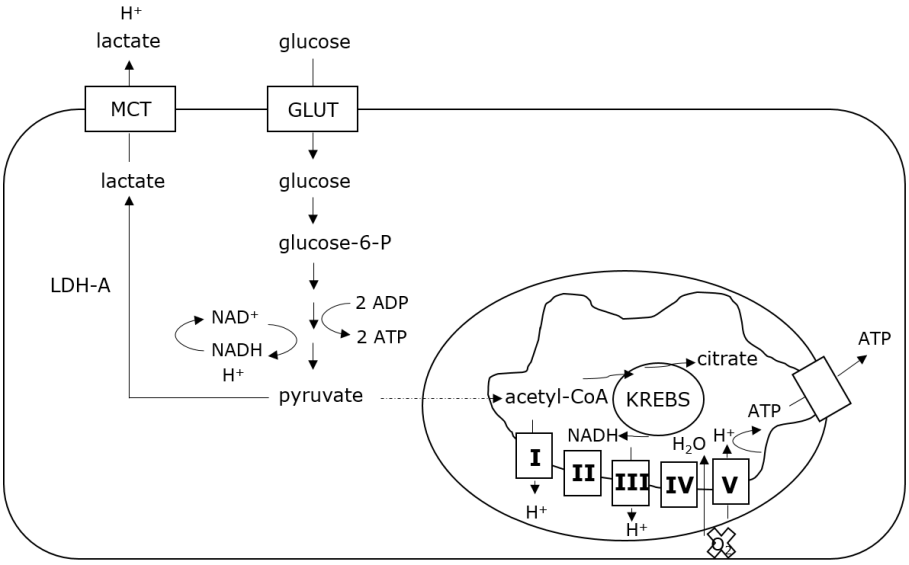


Figure 1.9. The metabolism of glucose in normal cells under anaerobic conditions. Acetyl-CoA: acetyl coenzyme A; ADP: adenosine diphosphate; ATP: adenosine triphosphate; glucose-6-P: glucose-6-phosphate; GLUT: glucose transporter; LDH-A: lactate dehydrogenase A; MCT: monocarboxylate transporter; NADH: nicotinamide adenine dinucleotide. Adapted with permission from Bracke et al. (162).

Otto Warburg observed decades ago that cancer cells consume glucose at a surprisingly high rate compared to normal cells. Furthermore, in contrast to normal cells cancer cells preferentially metabolize glucose via glycolysis and secrete the glucose-derived carbon as lactate even in the presence of sufficient O_2 to support mitochondrial OXPHOS, a phenomenon referred to as "aerobic glycolysis" or the Warburg effect **(Figure 1.10)** (163, 164). The widely used diagnostic technique positron emission tomography takes advantage of the Warburg effect to image cancer cells using the radioactive tracer ^{18}F-fluorodeoxyglucose (165).

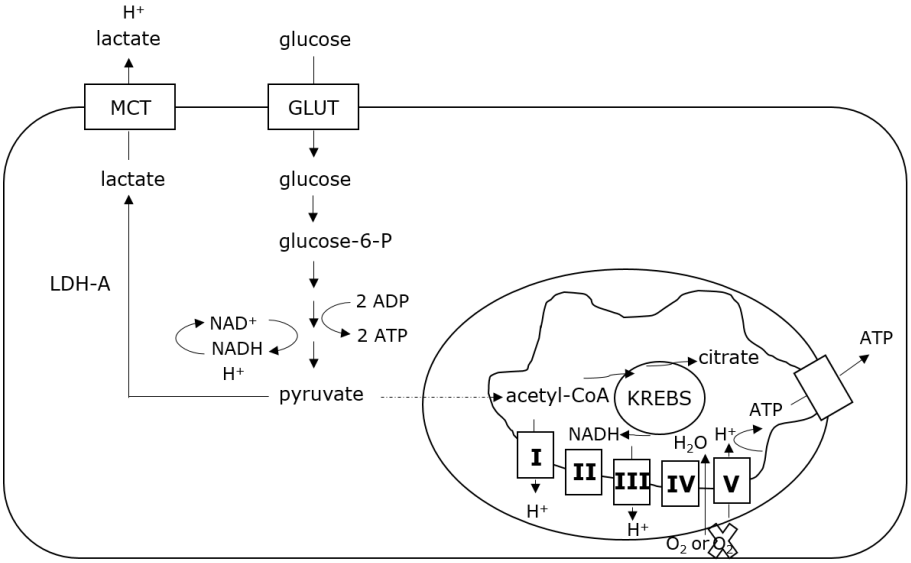


Figure 1.10. The metabolism of glucose in cancer cells. Even in the presence of sufficient O_2 to maintain mitochondrial OXPHOS, cancer cells prefer to convert glucose into lactate. This phenomenon is known as the Warburg effect or "aerobic glycolysis". Acetyl-CoA: acetyl coenzyme A; ADP: adenosine diphosphate; ATP: adenosine triphosphate; glucose-6-P: glucose-6-phosphate; GLUT: glucose transporter; LDH-A: lactate dehydrogenase A; MCT: monocarboxylate transporter; NADH: nicotinamide adenine dinucleotide. Adapted with permission from Bracke et al. (162).

Besides "aerobic glycolysis", cancer cells exhibit increased uptake and metabolism of glutamine (Gln) (166). Gln is deaminated to glutamate (Glu) by the mitochondrial enzyme glutaminase (GLS). Subsequently, Glu is transformed into α-ketogluterate (α-KG) (glutaminolysis), an intermediate of the Krebs cycle, by glutamate dehydrogenase (158, 167, 168). Gln-derived α-KG can then be oxidized to malate or oxaloacetate (OAA) (anaplerosis), thereby sustaining Krebs cycle function despite constant efflux of intermediates to support biosynthetic pathways (169, 170). Furthermore, Gln-derived α-KG can be converted into citrate via reductive carboxylation (171).

Causes of the metabolic transformation

Otto Warburg hypothesized that irreversible mitochondrial dysfunction causes the switch from OXPHOS to glycolysis in cancer cells in order to compensate for the loss of ATP (163). However, it is now clear that the oxidative capacity of mitochondria is not impaired in most cancer cells and that the Warburg effect has to be perceived as an uncoupling of glycolysis and OXPHOS and not as a switch from OXPHOS to glycolysis (172-175). Instead, the metabolic transformation of cancer cells is the result of the activation of oncogenes and/or the inactivation of tumor suppressor genes, indicating that it is intrinsically linked to oncogenic transformation (176-178). Moreover, the adaptation of cancer cells to unfavorable conditions (i.e. the hypoxic microenvironment) plays a central role in the metabolic transformation of cancer cells (158, 179). In the following sections, the most important factors involved in the metabolic transformation of cancer cells are discussed.

Oncogenes and tumor suppressor genes

PI3K-AKT pathway

The phosphatidylinositol 3-kinase (PI3K)-AKT pathway is one of the most commonly altered signaling pathways in cancer cells. This pathway is activated when a growth factor binds to its growth factor receptor, thereby activating PI3K, the kinase which converts phosphatidylinositol-4,5-diphosphate (PIP2) to phophatidylinositol-3,4,5-triphosphate (PIP3). The phosphatase phosphate and tensin homologue (PTEN) catalyzes the opposite reaction, dephosphorylating PIP3 to PIP2, thereby inhibiting the activation of the PI3K-AKT pathway. PIP3 attracts the serine/threonine kinase AKT (or protein kinase B) and phosphoinositide-dependent kinases, which phosphorylate and activate AKT (162). Constitutive activation of the PI3K-AKT pathway is due to overexpression of growth factors, overexpression of growth factor receptors, activating mutations in PI3K or AKT and/or inactivating mutations in PTEN (177).

Activation of AKT by phosphorylation results in enhanced cellular glucose uptake by inducing the expression and membrane translocation of glucose transporters (156, 180). Furthermore, activated AKT stimulates "aerobic glycolysis" by activation of the glycolytic enzymes hexokinase-II (HK-II) and

phosphofructokinase-1 (PFK-1) (177). Activation of AKT also induces the translocation of HK-II to the outer mitochondrial membrane where it binds to the voltage-dependent anion channel (VDAC). Translocation of HK-II makes it less sensitive to product feedback inhibition and brings it closer to OXPHOS-derived ATP, thereby increasing the rate at which phosphorylation of glucose occurs. Additionally, the association of HK-II with VDAC prevents binding of pro-apoptotic factors (e.g. cytochrome c and b-cell lymphoma-2 (bcl-2) proteins bax and bcl-2) to VDAC, thereby linking metabolic reprogramming to apoptosis inhibition (174, 181, 182). The impact of constitutive activation of the PI3K-AKT pathway in cancer cells is depicted in **Figure 1.11**.

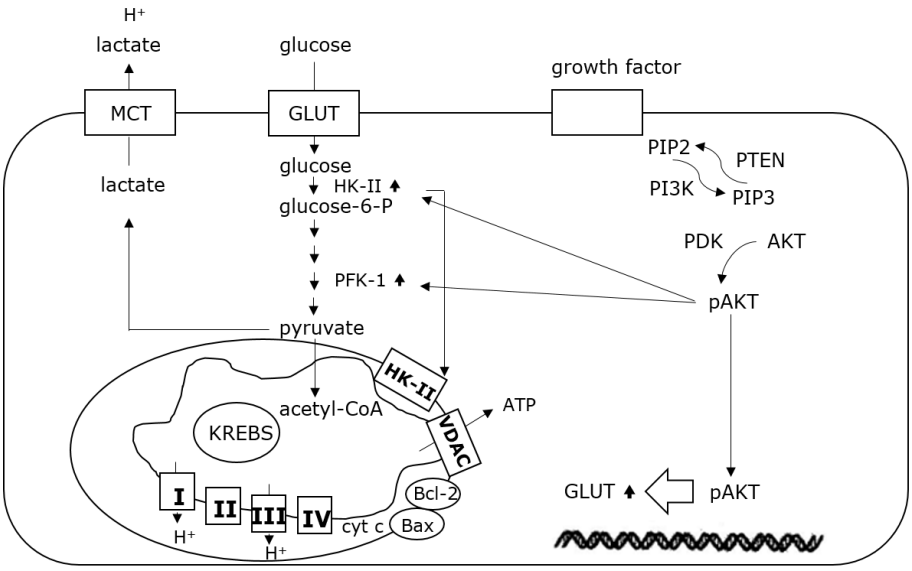


Figure 1.11. AKT as a regulator of "aerobic glycolysis". Constitutive activation of the PI3K-AKT pathway in cancer cells results in enhanced glucose uptake, increased "aerobic glycolysis" and inhibition of apoptosis. Acetyl-CoA: acetyl coenzyme A; ATP: adenosine triphosphate; Bcl-2: b-cell lymphoma-2; cyt c: cytochrome c; glucose-6-P: glucose-6-phosphate; GLUT: glucose transporter; HK-II: hexokinase-II; MCT: monocarboxylate transporter; pAKT: phosphorylated AKT; PDK: phosphoinositide-dependent kinases; PFK-1: phosphofructokinase-1; PI3K: phosphatidylinositol-3-kinase; PIP2: phosphatidylinositol-4,5-diphosphate; PIP3: phosphatidylinositol-3,4,5-triphosphate; PTEN: phosphate and tensin homologue; VDAC: voltage-dependent anion channel. Adapted with permission from Bracke et al. (162).

c-myc

The transcription factor c-myc is often overexpressed in cancer cells. Once activated, c-myc forms a complex with its co-factor max. Subsequently, this

complex binds to E-boxes and induces transcription of several target genes involved in the regulation of glucose and glutamine metabolism (177). Overexpression of c-myc induces the expression of glucose transporters and stimulates "aerobic glycolysis" by activating nearly all glycolytic enzymes, e.g. HK-II, PFK-1, glyceraldehyde-3-phosphate dehydrogenase, enolase, pyruvate kinase (PK), and LDH-A (183, 184). C-myc also regulates Gln uptake and metabolism. More specifically, it stimulates the expression of Gln transporters and increases the activity of GLS, the enzyme that converts Gln into Glu (177, 185). The influence of overexpressed c-myc in cancer cells is illustrated in **Figure 1.12**.

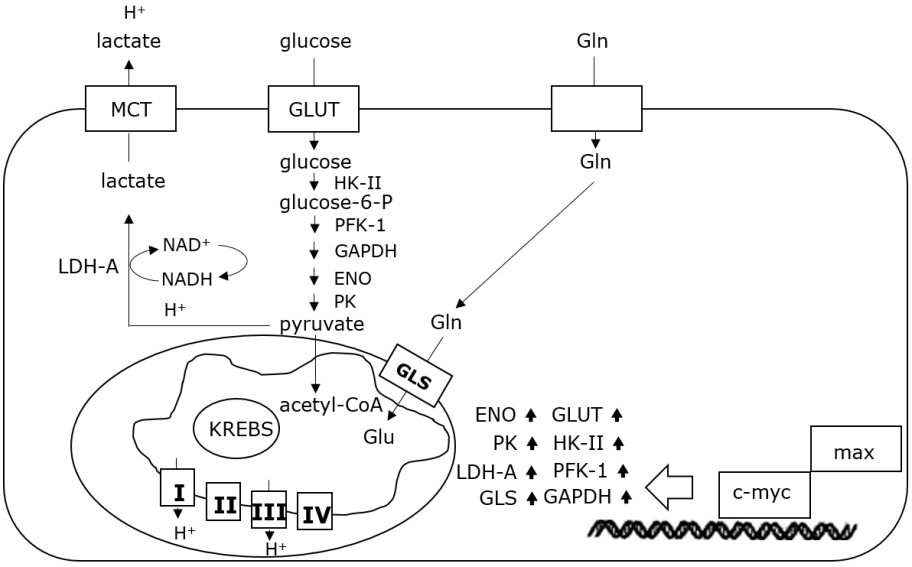


Figure 1.12. c-myc as an activator of "aerobic glycolysis". Overexpression of c-myc in cancer cells induces the expression of glucose transporters and promotes "aerobic glycolysis" as well as glutamine uptake and metabolism. acetyl-CoA: acetyl coenzyme A; ENO: enolase; GAPDH: glyceraldehyde-3-phosphate dehydrogenase; Gln: glutamine; GLS: glutaminase; Glu: glutamate; glucose-6-P: glucose-6-phosphate; GLUT: glucose transporter; HK-II: hexokinase-II; LDH-A: lactate dehydrogenase A; MCT: monocarboxylate transporter; NADH: nicotinamide adenine dinucleotide; PFK-1: phosphofructokinase-1; PK: pyruvate kinase. Adapted with permission from Bracke et al. (162).

p53

In normal cells, the transcription factor p53 is activated by DNA damage. Subsequently, it inhibits the cell cycle in order to permit DNA repair. If DNA repair is not possible anymore, it induces apoptosis (177). As part of its tumor suppressor function, it promotes mitochondrial OXPHOS and inhibits glycolysis

(177). However, in most cancers p53 is mutated, inducing the opposite effect: stimulation of "aerobic glycolysis" and inhibition of mitochondrial OXPHOS. Loss of function of p53: 1) stimulates glucose uptake by promoting the expression of glucose transporters, 2) activates the glycolytic enzyme phosphoglycerate mutase (PGM), 3) prevents expression of the gene encoding cytochrome c oxidase 2, whose product is required for the assembly of the cytochrome c oxidase complex (complex IV) of the mitochondrial respiratory chain and 4) inhibits activation of TP53-induced glycolysis and apoptosis regulator, an enzyme which decreases the levels of fructose-2,6-biphosphate, a potent activator of the glycolytic enzyme PFK-1 (177, 186, 187). The impact of loss of function of p53 in cancer cells is depicted in **Figure 1.13**.

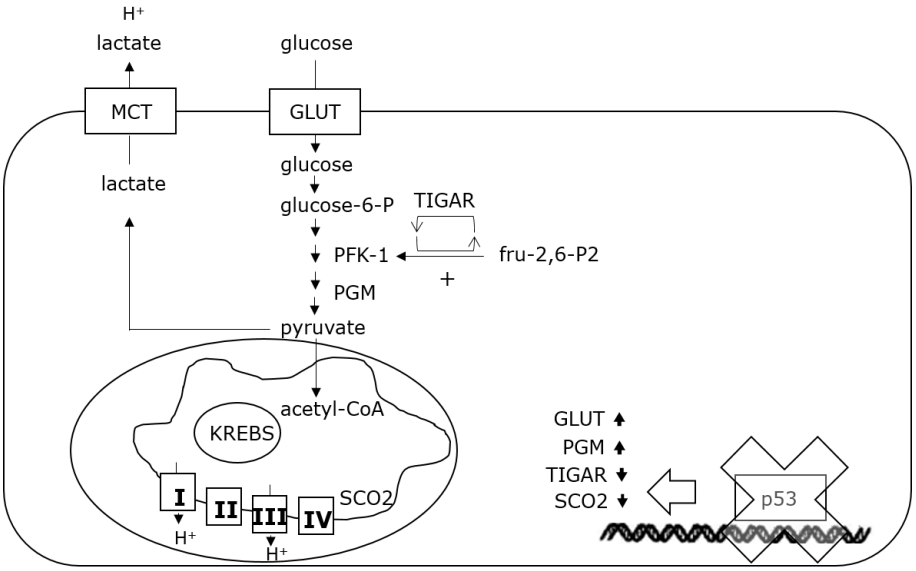


Figure 1.13. p53 as a regulator of "aerobic glycolysis" and mitochondrial OXPHOS. Loss of function of p53 in cancer cells induces the expression of glucose transporters, promotes "aerobic glycolysis" and suppresses mitochondrial OXPHOS. Acetyl-CoA: acetyl coenzyme A; fru-2,6-P2: fructose-2,6-biphosphate; glucose-6-P: glucose-6-phosphate; GLUT: glucose transporter; MCT: monocarboxylate transporter; OXPHOS: oxidative phosphorylation; PFK-1: phosphofructokinase-1; PGM: phosphoglycerate mutase; SCO2: cytochrome c oxidase 2; TIGAR: TP53-induced glycolysis and apoptosis regulator. Adapted with permission from Bracke et al. (162).

Adaptation to the hypoxic microenvironment

Hypoxia-inducible factor-1 (HIF-1) is a heterodimeric transcription factor consisting of a labile α- and a stabile β-subunit. Under normal oxygen (normoxic)

conditions, prolyl residues of HIF-1α are hydroxylated by prolyl-4-hydroxylase domain proteins (PHDs). Subsequently, hydroxylated HIF-1α is ubiquitinated by the von Hippel Lindau (VHL) protein and degraded by the proteasome (188).

However, as solid tumors proliferate rapidly, they outgrow the local blood supply, leading to hypoxia and concomitant stabilization of HIF-1α. Consequently, it enters the nucleus, heterodimerizes with HIF-1β and induces the expression of several genes involved in glucose metabolism, angiogenesis, invasion and survival (177, 189, 190). Stabilization of HIF-1α induces "aerobic glycolysis" by means of induction of the expression of glucose transporters to enhance glucose uptake, upregulation of the glycolytic enzymes HK-II, PGM and LDH-A as well as the plasma membrane MCT4 which transports lactate into the extracellular space (191, 192). In addition, it attenuates mitochondrial OXPHOS by activating the gene encoding pyruvate dehydrogenase kinase-1. This enzyme phosphorylates and inactivates the pyruvate dehydrogenase (PDH) complex, preventing the conversion of pyruvate into acetyl coenzyme A (acetyl-CoA) by PDH and thereby promoting "aerobic glycolysis" (193, 194).

It is now clear that besides hypoxia, also oncogenic events contribute to the stabilization of HIF-1α in cancer cells. These events give rise to pseudohypoxia, a condition in which hypoxic signaling is preserved in normoxic conditions (195). For instance, the activation of the PI3K-AKT pathway and the inactivation of the VHL protein lead to the stabilization of HIF-1α under pseudohypoxia (188, 196-198). Moreover, loss-of-function mutations of succinate dehydrogenase and/or fumarate hydratase lead to the accumulation of the Krebs cycle intermediates succinate and/or fumarate, which interfere with the α-KG-dependent prolyl hydroxylation of HIF-1α by PHDs, thereby preventing degradation of HIF-1α under normoxic conditions (199-201). The influence of HIF-1α stabilization in cancer cells is illustrated in **Figure 1.14**.

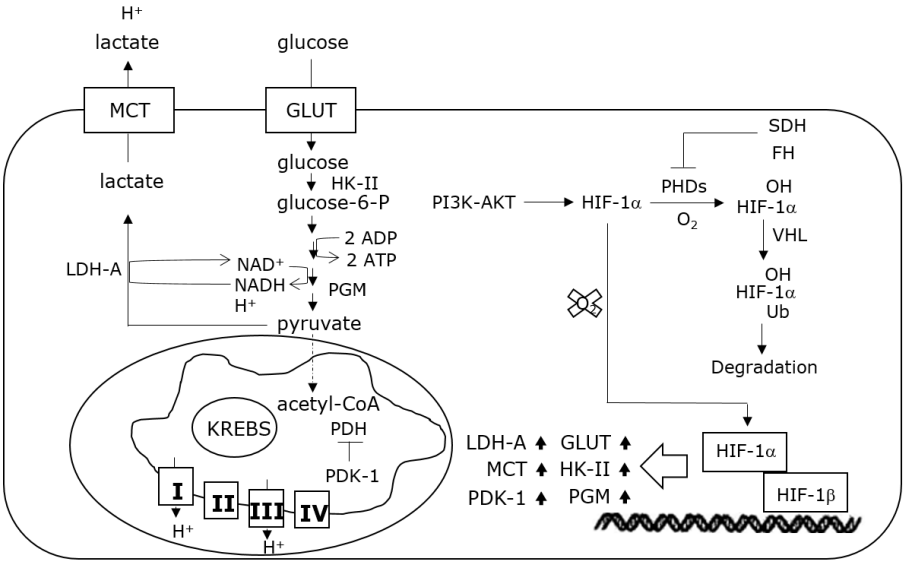


Figure 1.14. HIF-1α stabilization as a regulator of glycolysis and mitochondrial OXPHOS. Stabilization of HIF-1α promotes glucose uptake and "aerobic glycolysis" and suppresses mitochondrial OXPHOS. acetyl-CoA: acetyl coenzyme A; ADP: adenosine disphosphate; ATP: adenosine triphosphate; FH: fumarate hydratase; glucose-6-P: glucose-6-phosphate; GLUT: glucose transporter: HK-II: hexokinase-II; LDH-A: lactate dehydrogenase A; MCT: monocarboxylate transporter; NADH: nicotinamide adenine dinucleotide; OH: hydroxylation; PDH: pyruvate dehydrogenase; PDK-1: pyruvate dehydrogenase kinase-1; PGM: phosphoglycerate mutase; PHDs: prolyl-4-hydroxylase domain proteins; PI3K: phosphatidylinositol-3-kinase; SDH: succinate dehydrogenase; Ub: ubiquitination; VHL: von Hippel Lindau. Adapted with permission from Bracke et al. (162).

Advantages of the metabolic transformation for cancer cells

Although the preferential use of "aerobic glycolysis" appears to be unfavorable for cancer cells, it offers several advantages for them (190). In the following sections, the diverse benefits of the glycolytic switch and increased uptake of glutamine for cancer cells are described.

Survival under conditions of fluctuating oxygen availability

Cancer cells which rely on "aerobic glycolysis" for ATP production are able to survive under unfavorable conditions encountered within the microenvironment (e.g. fluctuating oxygen and nutrient availability), while cancer cells which depend on OXPHOS cannot produce enough ATP under hypoxia and die (202).

Promotion of invasion and metastasis

"Aerobic glycolysis" in cancer cells leads to acidification of the microenvironment due to the action of MCTs which co-transport H^+ with monocarboxylate anions such as lactate (203). Extracellular acidification promotes invasion and metastasis by pH-dependent activation of cathepsins and metalloproteinases that disrupt extracellular matrix and basement membranes (204). Furthermore, acidification of the microenvironment restrains anticancer T-cell immune responses (205).

Defense against oxidative stress

Increased "aerobic glycolysis" leads to the production of reducing equivalents in the form of nicotinamide adenine dinucleotide phosphate (NADPH), resulting in a good redox status and therefore a decrease in reactive oxygen production and oxidative stress in cancer cells (166, 175, 206).

Evasion from apoptosis

The glycolytic switch protects cancer cells from apoptosis. The link between metabolic reprogramming and resistance to apoptosis is caused by the association of HK-II to the outer mitochondrial membrane. HK-II competes with pro-apoptotic factors for binding to VDAC and impacts the balance between pro- and anti-apoptotic factors which regulate membrane permeabilization (207).

Production of ATP at a faster rate

Although "aerobic glycolysis" is less efficient than the Krebs cycle and OXPHOS in yielding ATP (2 molecules of ATP per molecule of glucose relative to 30 molecules of ATP), it can generate more ATP than OXPHOS by producing it at a faster rate (208).

Supply of building blocks for biosynthetic processes

Last but not least, enhanced "aerobic glycolysis" and the metabolism of glutamine provide building blocks required for the biosynthesis of macromolecules (nucleic acids, proteins and lipids) essential for cell division, thereby supporting the creation of new daughter cells (190, 209, 210). In the following sections, the biosynthesis of nucleic acids and lipids in cancer cells are described in more detail.

Biosynthesis of nucleic acids

A key biosynthetic pathway fed by the first steps of the glycolytic pathway is the pentose phosphate pathway (PPP), which consists of a non-reversible oxidative and a reversible non-oxidative branch. The oxidative branch of the PPP is initiated by the conversion of the glycolytic intermediate glucose-6-phosphate to 6-phosphogluconolactone by glucose-6-phosphate dehydrogenase. This branch generates ribose-5-phosphate, a critical precursor for nucleic acids, and NADPH which is required to facilitate biosynthetic reactions and to combat oxidative stress. In contrast, fructose-6-phosphate and glyceraldehyde-3-phosphate are shunted in the non-oxidative branch of the PPP which produces ribose-5-phosphate.

Cancer cells prefer to synthesize nucleic acids via the non-oxidative branch of the PPP due to the action of PI3K-AKT, c-myc, p53 and HIF-1 (211). Moreover, cancer cells promote the biosynthesis of nucleic acids by the selective expression of the M2 embryonic isoform of PK (PKM2) instead of the usual somatic M1 isoform. The M2 isoform can switch between a highly active tetrameric form and a less active dimeric form. Phosphorylation by tyrosine kinases renders PKM2 inactive, causes a bottleneck, slowing down the rate of glycolysis at the end and increasing the availability of upstream glycolytic intermediates for the biosynthetic PPP (212-214).

Biosynthesis of lipids

The breakdown of glucose via the glycolytic pathway results in the formation of mitochondrial acetyl-CoA which condenses with OAA to form citrate. Subsequently, citrate is exported from mitochondria to the cytosol where it is converted to acetyl-CoA and OAA by the enzyme ATP citrate lyase (ACLY). Next, acetyl CoA carboxylase (ACC) converts acetyl-CoA to malonyl-CoA. Finally, fatty acid synthase (FAS) condenses acetyl-CoA and malonyl-CoA in fatty acids (161, 190). Cancer cells utilize these fatty acids to synthesize lipids, which are required to assemble membranes and to modify membranes-targeted proteins (176, 215).

The export of citrate needs to be compensated in order to maintain the function of the Krebs cycle. Therefore, Gln-derived α-KG is converted to OAA via an anaplerotic reaction (see **section 7.2**: increased uptake and metabolism of

glutamine) (161, 190). Moreover, Gln-derived α-KG can be converted to citrate via reductive carboxylation by isocitrate dehydrogenase (171, 216, 217).

The increased lipid synthesis in cancer cells is predominantly regulated by the PI3K-AKT pathway, which activates the lipogenic transcription factor sterol regulatory element binding protein-1 and thereby results in the overexpression of the lipogenic enzymes ACLY, ACC and FAS (218, 219).

Application of metabolomics in lung cancer research

In the past decade, several studies have investigated lung cancer metabolism. Most of these studies employed MS to study lung cancer-induced metabolic alterations in tumor tissue (220), plasma (221-225), serum (226-228) or urine (229), while only a few studies were based on ^{1}H-NMR spectroscopy. Although metabolic phenotyping of plasma/serum has the benefit to evaluate more directly the complex interaction between tumor and host, the majority of these ^{1}H-NMR-based metabolomics studies focused on the metabolic composition of lung cancer tissue (230, 231) rather than on body fluids (232, 233). To the best of our knowledge, only the Portuguese research group of Rocha et al. (232) examined lung cancer-induced metabolic changes in plasma by means of ^{1}H-NMR spectroscopy. They have shown that the plasma metabolic phenotype enables to discriminate between 85 lung cancer patients and 78 controls with a sensitivity of 92% and a specificity of 89%. Nevertheless, a huge drawback of this and many other metabolomics studies is that their findings were not validated in an independent cohort.

Aims and outline of the study

From the general introduction, is has become clear that lung cancer constitutes a major public health problem, which accounts for approximately one out of five of all cancer-related deaths worldwide. Furthermore, the current method used for lung cancer screening, namely low-dose computed tomography (LDCT), has several limitations. An important disadvantage of LDCT screening is the high rate of false positive results, which leads to unnecessary and possibly harmful follow-up diagnostic tests and puts a huge financial burden on the health care system. Therefore, there is an increasing interest in strengthening current risk models by incorporating lung cancer-risk related biomarkers in order to improve the selection of high-risk individuals eligible for LDCT screening, thereby decreasing the false positive rate and the corresponding financial burden.

Metabolomics holds great potential for cancer diagnosis since the metabolism of cancer cells differs from that of normal cells. Metabolites are the end products of cellular metabolism and therefore reflect changes in the observed metabolic phenotype. ^{1}H-NMR spectroscopy, one of the main analytical platforms used in metabolomics studies, is a highly reproducible tool that enables a fast and non-invasive identification and quantification of complex mixtures of metabolites, as in blood plasma, with minimal sample preparation and relatively low costs on a per sample basis. Hence, ^{1}H-NMR-based metabolomics of plasma represents an attractive tool in the search for blood-based diagnostic biomarkers in order to complement current risk models for LDCT screening.

The first aim was to optimize a NMR analysis protocol which enables a proper evaluation of ^{1}H-NMR-based metabolomics of plasma as a tool to detect lung cancer. Hereto, human blood plasma was spiked with 37 different metabolites in relevant concentrations and analyzed on a medium-field 400 MHz NMR spectrometer in order to accurately assign the resonance signals present in the plasma ^{1}H-NMR spectrum. The resulting information was used to rationally divide the 400 MHz plasma spectra into 110 well-defined integration regions, representing the relative metabolite concentrations (i.e. plasma metabolic phenotype). Subsequently, the plasma metabolic phenotype was used to classify a small group of 53 breast cancer patients and 52 controls in order to show the

proof of principle. Furthermore, the impact of noisy integration regions (or noisy variables) on multivariate group classification was investigated (**Chapter 2**).

The second aim was a detailed signal assignment of human plasma metabolites in order to rationally divide the ^{1}H-NMR spectrum in integration regions, which represent the plasma metabolic phenotype. To this end, human blood plasma was spiked with the same 37 metabolites and analyzed on a high-field 900 MHz spectrometer at the "Institut de Reserche Interdisciplinaire" in Lille, France. On the basis of the metabolite spiking on the 900 MHz spectrometer, the i) peak assignment in the 400 MHz spectra (used in **Chapter 4-7**) could be somewhat further fine-tuned and ii) the 900 MHz spectra could be segmented into 105 variable-sized integration regions, representing the plasma metabolic phenotype. Additionally, the plasma metabolic phenotype (determined on a 400 MHz and 900 MHz spectrometer) of a case-control dataset of 69 lung cancer patients and 74 controls was used to train a classification model in discriminating between both groups in order to find out advantages and disadvantages of a high-field 900 MHz spectrometer versus a medium-field 400 MHz spectrometer (**Chapter 3**).

The third aim was to develop a standardized protocol concerning sample handling (collection, processing, freezing and storage) which enables the implementation of ^{1}H-NMR-based metabolomics in clinical practice. Hereto, the impact of relevant preanalytical conditions (a half-filled blood tube, exposure of blood to oxygen, hemolysis, prolonged processing delays at 4°C, centrifugation at room temperature, freezing of plasma samples on dry ice and in liquid nitrogen before storage at -80°C and long-term storage of plasma at -80°C) on the plasma metabolic phenotype was examined. In addition, we investigated the value of the Standard PREanalytical Code, a method which has recently been developed by the biobank community to encode preanalytical conditions and to exclude plasma samples that were subjected to undesirable, interfering preanalytical conditions, within the field of clinical ^{1}H-NMR-based metabolomics (**Chapter 4**).

The following aim was to evaluate whether ^{1}H-NMR-based metabolomics of blood plasma can be used as a tool to detect lung cancer. Hereto, a classification model was trained in discriminating between 233 lung cancer patients and 226 controls based on data input from their plasma metabolic phenotype. Next, the predictive accuracy of the trained classification model was investigated in an independent

validation cohort consisting of 98 lung cancer patients and 89 controls. Moreover, we examined which plasma metabolites are responsible for group discrimination in order to get more insight into the disturbed metabolism of lung cancer. Notwithstanding the fact that the number of early stage lung cancer patients is still limited, in a next step, we made an effort to investigate whether the plasma metabolic phenotype permits to differentiate between early stage lung cancer patients and controls (**Chapter 5**).

The next aim was to find out whether the plasma metabolic phenotype allows to discriminate between lung and breast cancer. To this end, a classifier was trained in differentiating between 80 female breast cancer patients and 54 female lung cancer patients, all with an adenocarcinoma. Subsequently, the validity of the trained classifier was assessed in an independent validation cohort consisting of 60 female breast cancer patients and 81 male lung cancer patients. Furthermore, we examined which plasma metabolites are responsible for group discrimination in order to improve the understanding of the disturbed metabolism in both cancer types (**Chapter 6**).

The final aim was to further examine whether the plasma metabolic phenotype represents a specific diagnostic tool or rather a common cancer biomarker. Hereto, a classification model was constructed to discriminate between 37 colorectal patients, 37 breast cancer patients and 37 lung cancer patients, all with an adenocarcinoma (**Chapter 7**).

ANNEX

Basics of ^{1}H-NMR spectroscopy and multivariate projection methods

This technical note describes the basics of ^{1}H-NMR spectroscopy which rely on quantum mechanical laws, and are therefore quite complex and need to be explained in order to gain a better understanding of this technique. Furthermore, multivariate projection methods used for statistical processing of multidimensional ^{1}H-NMR datasets are described.

Basics of ^{1}H-NMR spectroscopy

^{1}H-NMR spectroscopy is a technique that measures the resonance frequencies of hydrogen nuclei that are subjected to a low-frequency radiation and a strong magnetic field. The resonance of nuclei at a specific frequency occurs in the radiofrequency range of the electromagnetic spectrum (**Figure S1**).

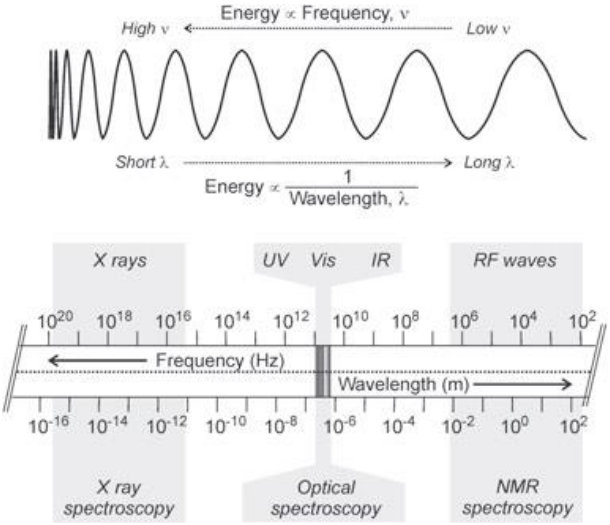


Figure S1. The electromagnetic spectrum. IR: infrared; NMR: nuclear magnetic resonance; RF: radiofrequency; UV: ultraviolet; Vis: visible. Reprinted with permission from Bothwell et al. (234).

Nuclear spin in a static magnetic field

NMR spectroscopy relies on the spin (or angular momentum) of protons and neutrons. Protons and neutrons are spin ½ particles, i.e. they have a spin with a spin quantum number s=1/2. The nuclear spin of an atom is expressed as the vector sum of unpaired proton and neutron spins and its nuclear spin quantum number I = n/2 with n the sum of unpaired protons and neutrons. Nuclei with an even number of <u>both</u> protons and neutrons have a spin quantum number I=0, for example, ^{12}C and ^{16}O. These types of nuclei do not have a net nuclear spin and

will not give rise to a NMR signal, and are therefore called NMR inactive nuclei. Nuclei with an odd number of protons as well as neutrons have a spin quantum number I=1, for example, ^{2}H or ^{14}N. However, if the number of protons or neutrons is odd, then I=half-integer. For example, the hydrogen nucleus consists of only one proton (^{1}H) and therefore has a nuclear spin quantum number I=1/2, making it NMR-active. The positively charged hydrogen nucleus creates a small magnetic field around the hydrogen atom which is called the magnetic dipole moment (μ-vector) (**Figure S2**).

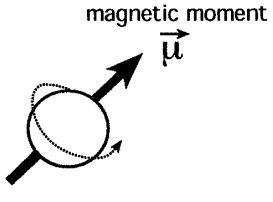


Figure S2. The magnetic dipole moment μ of the positively charged hydrogen nucleus. The orientations of the magnetic moment are defined by magnetic quantum number (m) values. The magnetic moment is directly proportional to the angular momentum ($\mu = \gamma \times P$, where μ: magnetic moment of the nucleus; γ: gyromagnetic constant; P: spin angular momentum).

If the ^{1}H atom is placed in a homogenous, strong and static external magnetic field (B_0), the μ-vector can appear in two states: parallel (spin-up or m=+1/2; α-state or low energy state) or anti-parallel (spin-down or m=-1/2; β-state or higher energy state) with respect to B_0 (**Figure S3**). Due to the interaction between the magnetic moment and the magnetic field, the magnetic moment will precess around the magnetic field B_0.

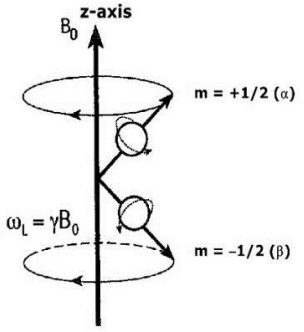


Figure S3. Two energy states (spin-up or spin-down) of the proton magnetic moment which precesses around an externally applied magnetic field B_0 oriented along the z-axis.

Energy states and population

For each spin state, there is energy associated with it, which is characterized by the frequency of the precession of nuclei in the presence of a magnetic field. The frequency with which the μ-vector spins around B_0 is called the Larmor frequency (ω_0) which depends on the strength of B_0 and the proton gyromagnetic constant (γ):

$$\omega = \gamma B_0 \quad \text{or} \qquad \nu = \frac{\gamma B_0}{2\pi}$$

where γ is the nuclear gyromagnetic constant which is a characteristic constant for a specific nucleus; B_0 is the magnetic field strength in the units of Tesla; ω is expressed in rad/s and ν in Hz. Therefore, the energy difference of the allowed transition is given by:

$$\Delta E = \gamma \hbar B_0 = \hbar \omega$$

where $\hbar$ is the Planck constant h divided by 2π (**Figure S4**).

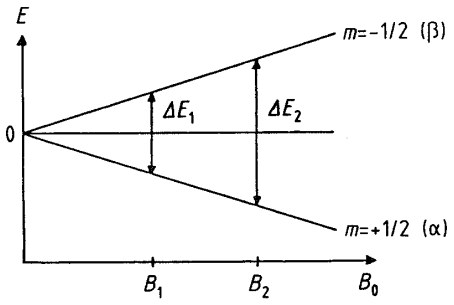


Figure S4. Energy difference between two Zeeman energy states. The intensity of the NMR signal relates to the population difference between two Zeeman energy states of the transition.

Consequently, spin transitions can be induced between the two different energy states by irradiation with electromagnetic radiation of the appropriate frequency. To this end, the following resonance conditions must be met:

$$E = h\nu = h\nu_L = h\frac{\gamma B_0}{2\pi} = \Delta E$$

The Boltzmann equation quantitatively describes the ratio of the spin populations in the two energy states. Both the energy difference between the transition states and the population difference between the states increases with the magnetic field strength. At room temperature, the amount of spins (μ-vectors) in the low energy state is slightly higher than in the high energy state. Thus, the sum of all these μ-vectors along the z-axis results in a net z-magnetization along the z-axis, and is presented as the magnetization vector M_0 (**Figure S5**). This indicates that only a small fraction of the spins will contribute to the NMR signal intensity due to the small energy difference between the two states and hence NMR spectroscopy intrinsically is a spectroscopic technique with rather low sensitivity. Therefore, stronger magnetic fields lead to improvement in sensitivity, in addition to a higher resolution.

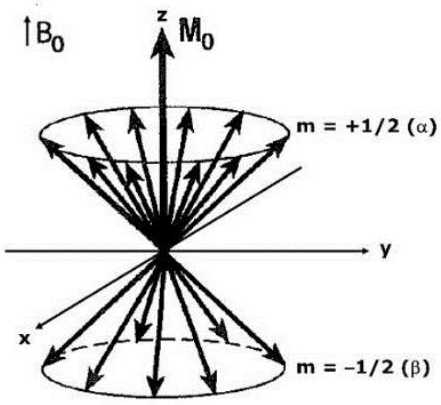


Figure S5. Statistical distribution of μ-vectors over two cone halves which results in net or longitudinal magnetization, M_0, which is oriented in the same direction as vector B_0. The vector sum of the components of the nuclear magnetic moments in the x-y plane is zero. At equilibrium, M_0 is generated by the small population difference between α and β states, and is parallel to the direction of the static magnetic field B_0.

Rotating frame

In order to create a NMR signal, the proton spins have to be flipped over from a low energy level to a high energy level. This is achieved by imposing a short but strong radio frequency pulse or B_1-pulse into the right angles on the B_0 field. This causes a simultaneous excitation of all protons by energy transfer from the B_1-pulse to the protons, called resonance. Because the Larmor frequency is not observed in an NMR experiment, a new coordinate frame is introduced to remove

the effect of the Larmor frequency, called the rotating frame (rotating with frequency ν). If a 90° pulse is applied on the spins, the M_0 vector will end up in the transversal x'-y' plane (**Figure S6**). Nuclear moments are no longer spinning along the z-axis but are stationary in this rotating frame. It should be noted that M_0 shifts away from the z-axis in a clockwise angle α, depending on the duration and strength of the B_1 field.

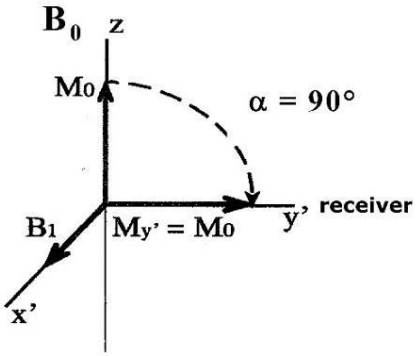


Figure S6. Vector representation of transversal magnetization ($M_{y'}=M_0$) upon applying a 90° RF-pulse along the x'-axis by the B_1 field in the rotating frame which results in a flip over of M_0 towards the y'-axis (receiver).

In practice, the B_1-pulse consists of a linear oscillating B_1 field along the x'-axis with carrier frequency ν (same as the frequency of the rotating frame and in the order of the Larmor frequency) and magnitude $2B_1$ and which is easily produced by applying an electric current through the probe coil (**Figure S7**). The linear oscillating B_1 field can be decomposed in two circular oscillating magnetic fields with Larmor frequency and magnitude B_1. Resonance can now be achieved by interaction of the μ-vectors with the circular magnetic B_1-field that is oscillating in the same direction and with frequency ν. This situation is called resonance condition.

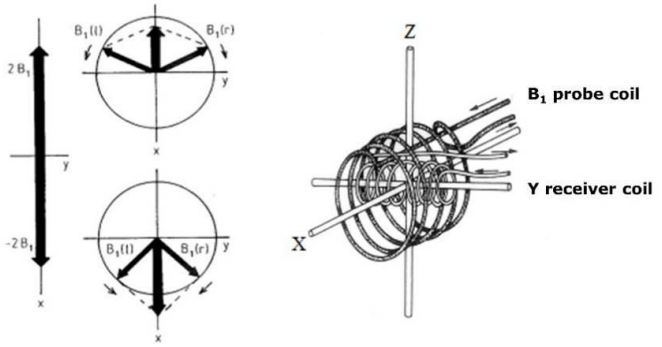


Figure S7. Oscillating B_1-fields in the x'-y'-plane are achieved by applying a linear oscillating electromagnetic field with magnitude of $2B_1$ along the x'-axis. The B_1 probe coil is oriented along the x-axis and the receiver coil along the y-axis.

A 90° pulse arises when the B_1 field is turned on and then turned off while M_0 moves from the z-axis to the x'-y' plane. The time during which B_1 is applied is called the 90° pulse width (in the order of µs). During a 90° pulse, the maximal NMR signal is obtained by the occurrence of two phenomena: 1) phase coherence, i.e. clustering of a part of the µ-vectors at the surface of the precession coin resulting in a shift of the equilibrium magnetization M_0 from the z-axis (longitudinal axis) towards the y'-axis of the transversal x'-y' plane ($M_{y'}$); 2) a net energy transfer from the B_1-field (transmitter) to the nuclear spins, leading to a flip over of some of the spins of the low energy state to the high energy state (**Figure S8**).

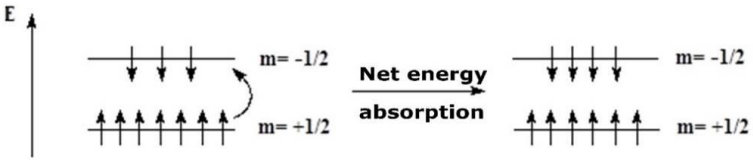


Figure S8. Net energy transfer from the B_1-field to the nuclei leading to a flip over of some nuclear spins from the low energy state to the high energy state.

After a 90° pulse, the magnetization along the z-axis becomes zero resulting in an equal distribution of nuclear spins over the two energy states, i.e. an equal amount of spins (µ-vectors) that are oriented parallel or anti-parallel to B_0 (**Figure S9**). The population difference between the two energy states than also becomes zero. In addition, the y-component achieves a maximal M_0 value ($M_{y'}$) and a

magnetization vector is created in the transversal plane ($M_{y'}=M_0$), i.e. transversal (detectable) magnetization.

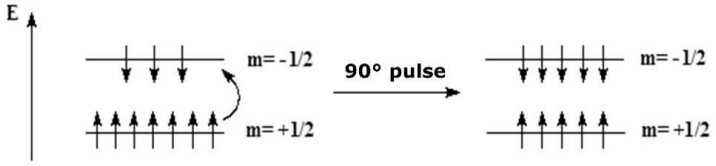


Figure S9. Orientation of nuclear spins after applying a 90° pulse.

Free induction decay

After the B_1-pulse is switched off, the NMR signal is detected while the nuclei return back to their equilibrium (relaxation, see next paragraph). The detection of the NMR signal is achieved when μ-vectors – with their unique frequency ν_i (or in the rotating frame $\Delta\nu_i = \nu_i - \nu$) – will spin back around the z-axis under the influence of B_0. This induces a fluctuating electric current in the detection coil (receiver) along the y'-axis. In the rotating frame, spins with $\nu_i = \nu$ ($\Delta\nu_i = 0$) will not spin around the z-axis, and as a function of time they induce a constant current in the receiver coil. Nuclei with $\nu_i > \nu$ will oscillate with respect to the y'-axis and as a function of time they induce a fluctuating current with frequency $\Delta\nu_i$, allowing to retrieve $\nu_i = \Delta\nu_i + \nu$. Nuclei with $\nu_i >> \nu$ will oscillate faster with respect to the y'-axis and as a function of time they induce a fluctuating current with a higher frequency (**Figure S10**). Hence, a free induction decay (FID; Free of the influence in the B_1 field; Induced in the receiver coil; Decaying back to the equilibrium) is induced in the detector. This is a complex and time dependent signal which is induced in the detector as a result of the different resonance frequencies of protons having a different electron (chemical) environment. By means of a Fourier transformation, the complex time signal can be unraveled in its constituent frequencies and amplitudes, establishing the NMR spectrum.

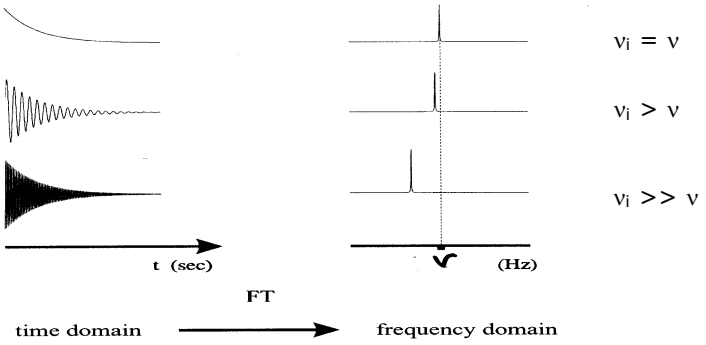


Figure S10. The free induction decay (FID) is the decay of the transversal magnetization which forms the observable NMR signals detected by the receiver in the x'y'-plane of the rotating frame. The FID is the sum of many time domain signals and is transformed via Fourier Transformation into the frequency domain signals or NMR spectrum. FT: Fourier transformation.

Relaxation

After the B_1 pulse, the μ-vectors will also be subjected to relaxation. Two relaxation processes, i.e. T_1 and T_2 relaxation, can be described by an exponential function with a characteristic relaxation decay time constant. The inverse of the relaxation decay time constant is a measure of the speed of relaxation: the shorter the relaxation time is, the more efficient the relaxation will be.

The longitudinal relaxation or T_1 relaxation is the decay time by which the z-magnetization (M_z) aims to reach its original value M_0 by energy exchange between the nuclei and the environment (back to the Boltzmann equilibrium: excess of spins in the low energy state). Longitudinal relaxation corresponds with a change in energy because the energy absorbed by the spins under influence of the B_1-pulse has to be returned to the environment. Hence, the original equilibrium distribution of the two spin states will be restored (**Figure S11**).

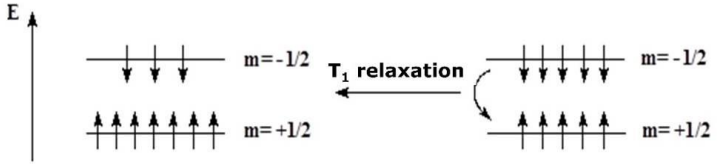


Figure S11. Longitudinal or T_1 relaxation.

Transversal relaxation or T_2 relaxation is the decay time by which the x'-y' magnetization disappears. The individual μ-vectors of the hydrogen protons will

lose their phase coherence by shared spin-spin interactions (**Figure S12**). Consequently, the transversal magnetization $M_{x'y'}$ completely disappears by strong spin-spin interactions, while the equilibrium magnetization might still not reached ($M_z \ll M_0$). Generally, T_2 is equal to or shorter than T_1.

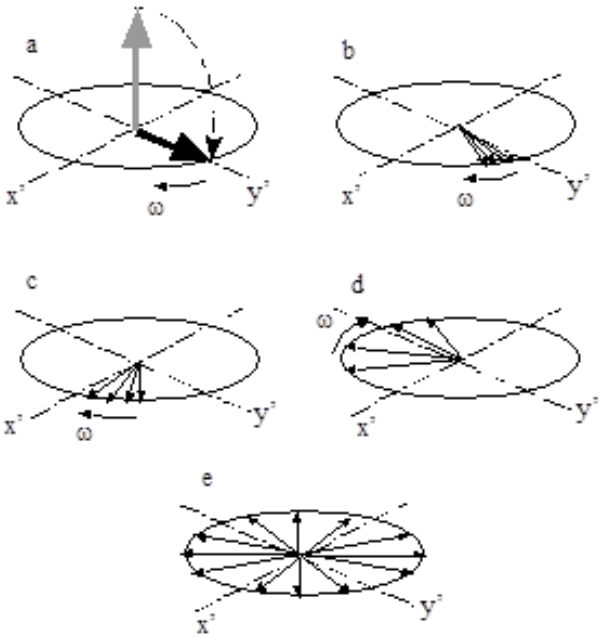


Figure S12. Transversal or T_2 relaxation.

The CPMG pulse sequence

For this thesis, slightly T_2-weighted spectra were acquired using the Carr-Purcell-Meiboom-Gill (CPMG)-presat pulse sequence which is called after its inventors (235, 236). The CPMG pulse sequence uses the faster T_2 relaxation of protons in macromolecules (such as proteins and polysaccharides), i.e. protons with shorter T_2 relaxation times, to suppress these particular signals and to generate spectra in which only the signals of small molecule metabolites are observed. In addition, water suppression (CPMG-presat, i.e. presaturation) was performed during 3 s in order to allow optimal detection and quantification of the signals close to the water resonance. The CPMG-presat pulse sequence has the form $[RD\text{-}90°\text{-}(\tau\text{-}180°\text{-}\tau)_m\text{-}ACQ]_n$ where RD is the relaxation delay, 90° is the 90° RF-pulse, 180° is the 180° RF-pulse, $mx2\tau$ is the spin-echo delay, m represent the number of loops, and ACQ is the acquisition time. During RD (0.5 s), the water signal is irradiated. Another

advantage of this spin-echo pulse-sequence is the elimination of the influence of irregularities in B_0 over the sample volume (by dephasing-rephasing). By this, the detected echo signal is only attenuated by real T_2 relaxation. An increase in the number of 'spin-echo' cycles (m) corresponds with an extension of the total spin-echo delay.

In this thesis, 400 MHz NMR spectra were acquired at 21.2°C with 96 scans and a spectral width of 6,000 Hz on a Varian Inova spectrometer (9.4 Tesla, Agilent Technologies Inc.) equipped with an Agilent OneNMR 5mm probe. Furthermore, 900 MHz NMR spectra were acquired at 21.2°C with 64 scans and a spectral width of 14,423 Hz on a Bruker Avance spectrometer (21.2 Tesla, Bruker Biospin) equipped with a triple resonance cryoprobe. Each FID was zero-filled to 65 K points and multiplied by a 0.7 Hz exponential line-broadening function prior to Fourier transformation.

The ^{1}H-NMR spectrum

The ^{1}H-NMR spectrum is determined by three parameters that can be related to the structure of the molecule: (i) chemical shift or the location of ^{1}H signals in the spectrum; (ii) integration of signal area or signal intensity; (iii) J-coupling patterns or the shape of the signals in the spectrum (**Figure S13**).

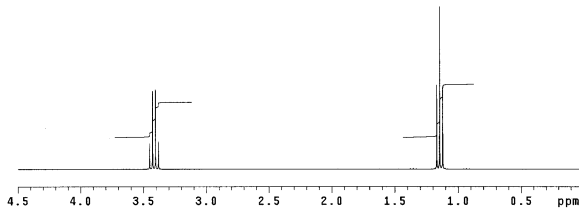


Figure S13. ^{1}H-NMR spectrum of diethyl ether ($C_4H_{10}O$). The place in the spectrum is determined by the chemical shift, the lines above the signals indicate the integration values and regions, and the quadruplet and triplet pattern is determined by the J-coupling. ppm: parts per million.

Chemical shift

When an atom is placed in a magnetic field, it will not sense the true value of the external B_0 field since local neighboring electrons shield the nucleus from the B_0 field. Hence, the effective field (B_{0i}) that a hydrogen nucleus feels depends on its chemical environment determining the neighboring electron density. The higher

the electron density around the hydrogen nucleus, the more it is shielded. For a nucleus i with a dimensionless magnetic shielding constant σi the following applies:

$$B_{0i} = B_0 - \sigma_i B_0 = (1-\sigma_i)B_0$$

so that $\omega_i = \gamma(1-\sigma_i)B_0 = \gamma B_{0i}$ or $\nu_i = \frac{\gamma}{2\pi}(1-\sigma_i)B_0 = \frac{\gamma}{2\pi}B_{0i}$

Hence, B_{0i} will be different from B_0 explaining why the resonance frequency ω_i (or ν_i in Hz) is different for different chemical environments. Absolute frequencies are rarely used. The location of the resonance signal in the NMR spectrum is represented by the chemical shift (δ):

$$\delta = \frac{\nu_i - \nu_{TMS}}{\nu_{TMS}} \cdot 10^6$$

in which ν_i is the resonance frequency of the proton considered and νTMS is the resonance frequency of the generally accepted reference compound tetramethylsilane. In most cases $\nu_i > \nu$TMS.

For example, for protons of a –O-CH_3 group on a 200 MHz (4.7 Tesla) NMR spectrometer applies:

ν_i = 200,000,646 Hz if νTMS = 200,000,000 Hz

$$\delta_{OCH_3} = \frac{200{,}000{,}646 - 200{,}000{,}000}{200\ x\ 10^6}\ x\ 10^6 = 3.23 ppm$$

On a 400 MHz (9.4 Tesla) NMR device the following applies:

ν_i = 400,001,292 Hz if νTMS = 400,000,000 Hz

$$\delta_{OCH_3} = \frac{400{,}001{,}292 - 400{,}000{,}000}{400\ x\ 10^6}\ x\ 10^6 = 3.23 ppm$$

Hence, on a 200 MHz NMR spectrometer, 200 Hz equals 1 ppm and on a 400 MHz NMR spectrometer – as used in this thesis – 400 Hz equals also 1 ppm. This is

explained by the fact that the chemical shift is independent of B_0. Therefore, the resolution of a 400 MHz NMR spectrometer is greater than that of a 200 MHz NMR spectrometer.

The chemical shift is dimensionless and expressed in parts per million (ppm). Most of the proton resonance signals are located between 0 and 12 ppm (**Figure S14**). The reference used in this thesis is TSP which is composed of three equivalent CH_3 methyl groups single bonded to a silicon atom. All CH_3 group protons have the same electronic environment, and therefore result in only a single ^{1}H-NMR signal at 0.015 ppm.

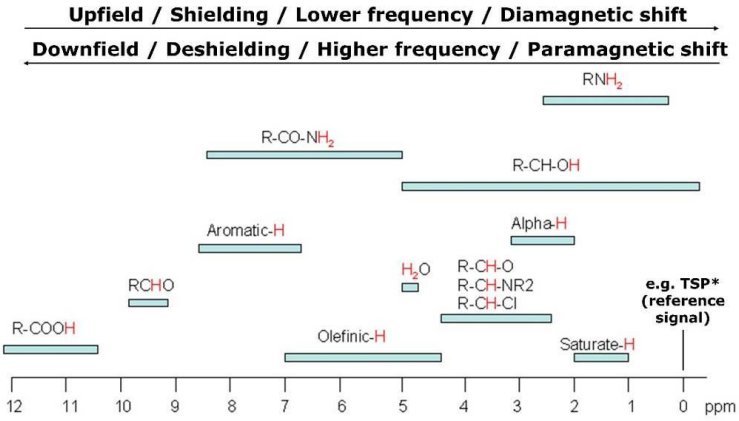


Figure S14. ^{1}H-NMR chemical shifts (δ in ppm) for common functional chemical groups (237). *^{1}H-NMR spectra of this thesis were referenced to the methyl resonance of trimethylsilyl-2,2,3,3-tetradeuteropropionic acid (TSP) at 0.015 ppm. Ppm: parts per million.

Integration region

The integration value of an integration (or spectral) region is a measure for the number of contributing protons. Integration values are obtained by putting integration lines over well-defined integration regions (**Figure S13**). For this thesis, 400 MHz spectra, in the range between 8 and 0.8 ppm, were segmented into 110 variable-sized spectral regions, excluding the water region between 5.2 and 4.7 ppm and the TSP signal between 0.3 and -0.3 ppm. The 110 spectral regions were defined by means of spiking of plasma with known metabolites on a 400 MHz spectrometer (see **Chapter 2**). Likewise, 900 MHz spectra, in the range between 8.5 and 0.8 ppm, were divided into 105 variable-sized spectral regions, excluding the water region between 5.1 and 4.7 and the TSP signal between 0.3

and -0.3 ppm. The 105 spectral regions were defined by means of spiking of plasma with known metabolites on a 900 MHz spectrometer (see **Chapter 3**). The resulting spectral regions in the 400 MHz and 900 MHz spectra were integrated and normalized relative to the total integrated area of all spectral regions.

J-coupling

Two nuclei with a different chemical environment (non-equivalent nuclei with different chemical shift values) can induce the phenomena of spin-spin coupling or J-coupling via binding electrons. The J-coupling pattern is determined by the rule of multiplicity: $(2n_1*I_1 +1) * (2n_2*I_2 +1)$, where n is the number of nuclei equivalent to each other but not equivalent to the nucleus under investigation, and I is the spin quantum number.

For example, chloroethane (CH_3CH_2Cl) consists of CH_3 and CH_2 protons. In this case, the CH_3 signal splits into a triplet: 2*2*1/2 +1 = 3, and the CH_2 signal into a quadruplet: 2*3*1/2 +1 = 4. In other words, the CH_3 protons sense the orientation of the CH_2 proton spins, resulting in a triplet with a 1-2-1 intensity ratio: (A) ↑↑; (B) ↑↓ or ↓↑; (C) ↓↓. And the other way around for the CH_2 protons sensing the CH_3 proton spins and producing a quadruplet with a 1-3-3-1 intensity ratio: (A) ↑↑↑; (B) ↑↑↓ or ↑↓↑ or ↓↑↑; (C) ↓↓↑ or ↓↑↓ or ↑↓↓; (D) ↓↓↓ (**Figure S15**).

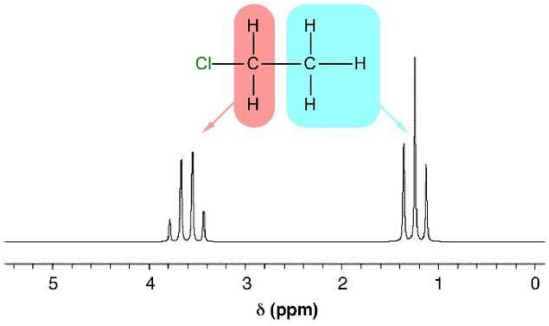


Figure S15 J-coupling patterns of chloroethane, i.e. a triplet and quadruplet (238). ppm: parts per million.

In general, J-couplings can result in singlets, doublets, triplets, etc. in the ^{1}H-NMR spectrum. It should be noted that the J-coupling between chemically equivalent protons is not observed in the ^{1}H-NMR spectrum.

Multivariate projection methods

Unsupervised PCA analysis

Principal component analysis (PCA) constitutes the basis for multivariate data analysis and it is performed on a data matrix X with N rows (observations) and K columns (variables) (**Figure S16**). Subsequently, the data matrix X is converted to a variable space with as many dimensions as there are variables. Each variable represents one co-ordinate axis.

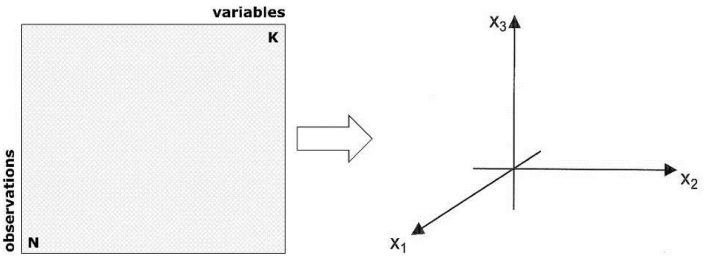


Figure S16. Structure of the data matrix X which is used for PCA analysis. The N observations (rows) can be biological individuals, analytical samples, etc. The K variables (columns) can be NMR variables derived from spectral data.

Next, each observation of the X-matrix is placed in the K-dimensional variable space (**Figure S17A**). After mean-centering and scaling of the data, the co-ordinate system is re-positioned such that the average point now is the origin.

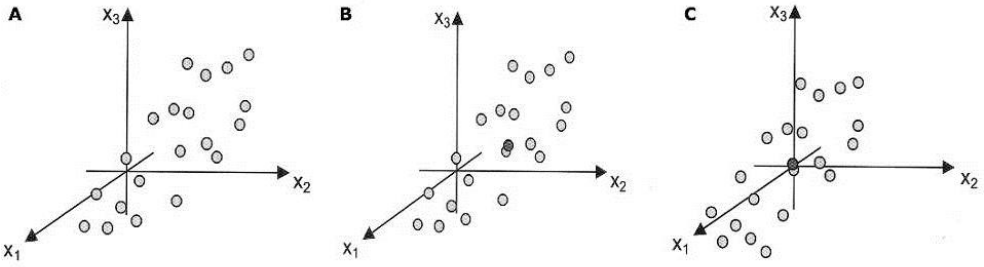


Figure S17. Plotting the observations in a K-dimensional space. (A) Observations (N, rows) in data matrix X are a group of points in the variable space K; (B) The vector of variable averages is computed and is a point (black dot) situated in the middle of the group of points; (C) After mean-centering, the origin of the co-ordinate system is moved to coincide with the average point (black dot).

The data matrix X can then be converted to principal components (PCs). The first principal component (PC1) is the line in the K-dimensional space that best

approximates the shape of the group of points (observations), i.e. it describes the largest variation in the data (**Figure S18A**). Each observation is projected onto this line in order to get a value which is known as a score. However, one component is insufficient to model the systemic variation of a multidimensional dataset. Therefore, the second principal component (PC2) which is orthogonal to PC1 and improves the approximation of the X-data to a feasible extent, is also calculated (**Figure S18B**). This results in a plane into the K-dimensional variable space (**Figure S19**).

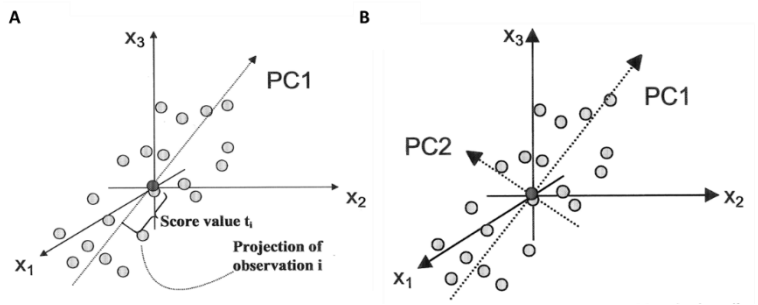


Figure S18. Principal components for PCA. (A) The first principal component, PC1; (B) The second principal component, PC2.

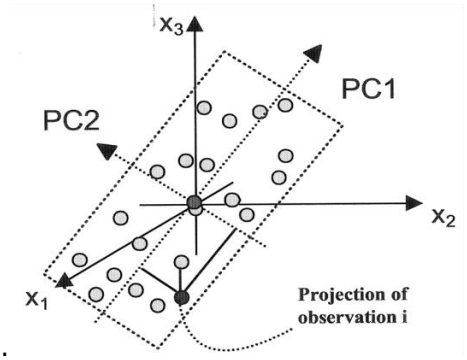


Figure S19. Two principal components define a model plane.

The co-ordinate values of the observations on this plane are called scores, and hence the plotting of such a projected configuration is known as a score plot. A score plot is always accompanied by a loading plot that reveals which variables are responsible for the patterns seen among the observations (239) (**Figure S20**).

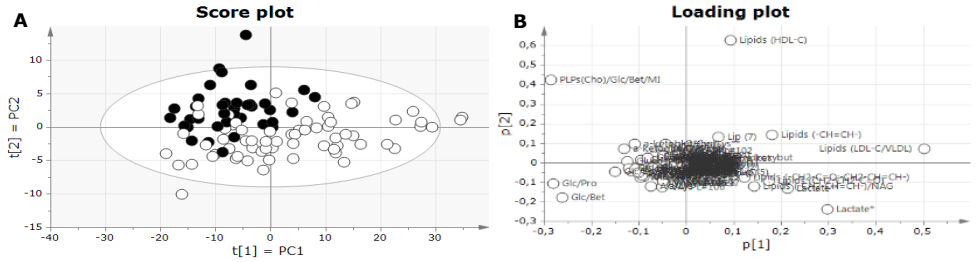


Figure S20. PCA score and loading plot. (A) PCA score plot of the two first PCs: PC1 and PC2 of the PCA score plot explain the largest and second largest variation within the data, respectively. Each point in the score plot represents an observation. Observations that lie outside the white circle (95% confidence interval) can be assumed as being outliers. (B) PCA loading plot of the two first PCs (p_2 vs p_1). Each point in the loading plot represents a variable which is responsible for the pattern seen among the observations. PC: principal component.

Consequently, PCA score and loading plots can be used to identify clustering patterns and dominant variation in the dataset which may not be associated with the real biological effect but could also be due to a secondary effect such as diet, age, gender and instrumental error (e.g. batch effect). In addition, PCA score plots permit to identify outliers, i.e. observations that lie outside the 95% confidence interval (**Figure S20**). However, it is recommended to perform additional methods such as distance to model or a Hotelling's T^2 range plot in order to identify outliers in the orthogonal plane (**Figure S21**).

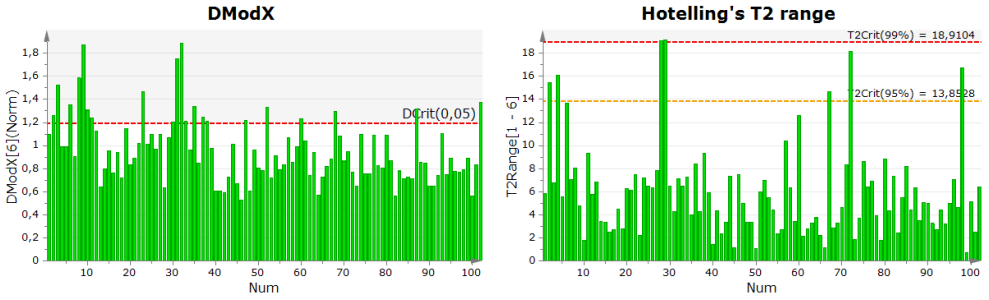


Figure S21. Outlier detection. (A) DModX is the distance of an observation in the dataset to the X model plane. DModX is displayed as the absolute DModX divided by the pooled residual standard deviation of the model. The critical value of DModX (Dcrit) is computed from the F-distribution. Observations with a DModX twice as large as Dcrit are moderate outliers. (B) Hotelling's T^2 range plot displays the distance from the origin in the model plane (score space) for each selected observation. This plot shows the T^2 calculated as the sum over the selected range of components of the scores, i.e. 1 to 6 in this case, in square divided by their standard deviations in square. Values larger than the 95% confidence interval are suspect, and values larger than 99% confidence interval can be considered as serious. DModX: distance to model.

Supervised PLS and OPLS-DA analysis

Unsupervised methods such as PCA are commonly used together with supervised methods such as partial least squares (PLS) and discriminant function analysis (e.g. PLS-DA and OPLS-DA) (240).

PLS is a method for relating two data matrices, X and Y, to each other by a linear multivariate model, i.e. it models the association between X and Y by regression. Data matrix X with N observations and K factors/predictors (independent variables) is related to data matrix Y with N observations and M responses (dependent variables) (**Figure S22**).

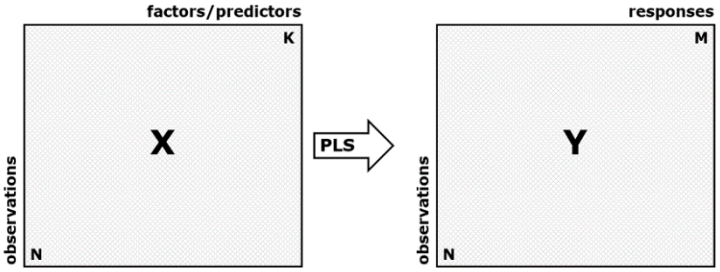


Figure S22. Structure of data matrix X and Y used for PLS analysis. The N observations (rows) can be biological individuals, analytical samples, etc. The X-variables (K, factors/predictors) can be NMR variables derived from spectral data. The Y-variables (M, responses) are often gathered to reflect properties of samples, clinical variables, etc.

In contrast to PCA, each row (observation) now represents two points instead of one, one in the X-space and one in the Y-space (**Figure S23**). Also here, data have to be mean-centered and scaled.

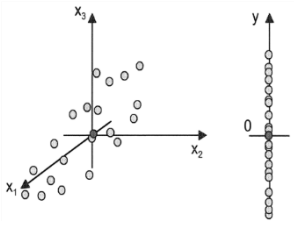


Figure S23. In a regression model, the observations are two groups of points, one in the predictor (X) space and one in the response (Y) space. Note that in this figure, a single Y-variable instead of a matrix Y of responses is considered.

In PLS, the first component is a line in the X-space that well approximates the groups of points and provides a good correlation with the Y-vector (**Figure S24**). The co-ordinate of an observation along this line is obtained by projecting the sample onto the line. This co-ordinate is termed the score, t_{i1}, of observation i. The scores of all the observations constitute the first X-score vector t_1 (a latent variable). The score vector t_1 can then be used to obtain an estimate of y, $\hat{y}_1$, after PC1 of PLS, which is t_1 multiplied by the weight of the y-vector, c_1. The differences between the measured and estimated response data are called residuals. The y-residuals represent the variation that is left unexplained by the first PLS component. The residual vector f_1 is calculated as $y - \hat{y}_1$.

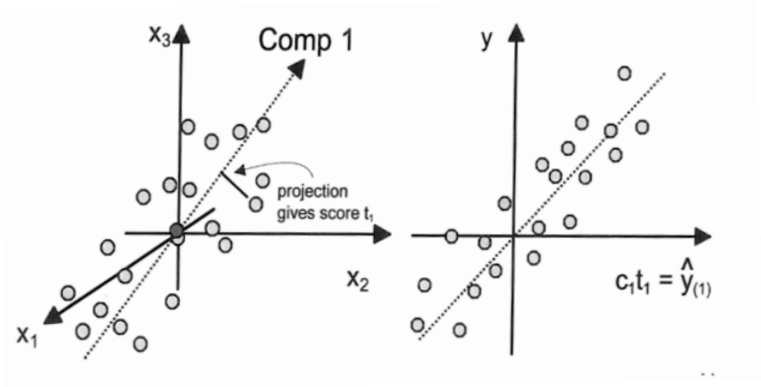


Figure S24. First principal component for PLS. With one single Y-variable, the Y-space reduces to a one dimensional vector. PC1 will orient itself so that it well describes the group of points in the X-space while at the same time giving a good correlation with the Y-vector. Score vector t_1 summarizes the information in the original X-variables.

Here too, one PLS component is insufficient to adequately model the variation in the Y-data. Therefore, a second component is calculated which is also a line in the X-space, which passes through the origin and is orthogonal to the first component (**Figure S25**). This component finds the direction in the X-space that improves the description of the X-data as much as possible, while providing a good correlation with the y-residuals, f_1, remaining after the first component. The second set of score values of the observations arises from the co-ordinates along the second projection direction in the X-space, i.e. second score vector t_2. In general, the tighter the scatter around the diagonal, the stronger the correlation between X and Y (239).

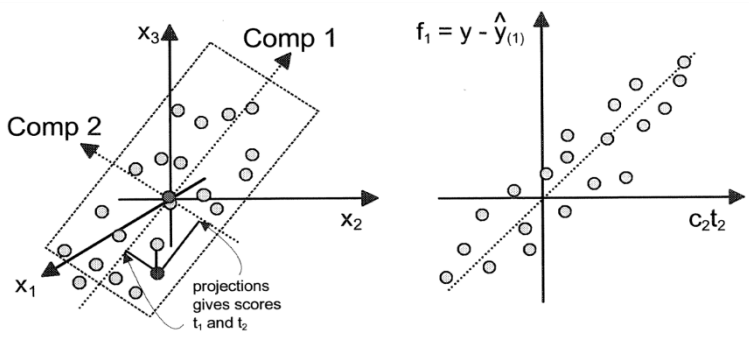


Figure S25. Second principal component for PLS in the X-space is orthogonal to the first one. By projecting the observations onto this line, one obtains the score vector t_2. The second score vector times the second weight of the y-data, c_2, correlates with the y-residual, f_1, after the first dimension. The two components together define a plane in the X-space.

In addition, the combined power of t_1 and t_2 in modeling and predicting y can also be examined with PLS. An estimate of y after two model components, $\hat{y}_2$, is obtained through computing $c_1t_1 + c_2t_2$. The y-variable is better modeled by two components than by one because the agreement between observed and estimated Y-data improves.

Supervised methods such as discriminant analysis (DA) can also be used for sample classification (e.g. disease vs. healthy). Orthogonal PLS (OPLS) is an extension of PLS which separates the systemic variation in X into two parts, one part that is correlated (predictive) to Y and one part that is uncorrelated (orthogonal) to Y. In other words, it models the variation orthogonal to the Y response, resulting in models that are equally predictive but easier to interpret than conventional PLS (153) (**Figure S26**). In OPLS-DA, knowledge about the class membership (e.g. disease vs. healthy) is used to discriminate groups of metabolites that are significant in combination (e.g. biomarker signature/classifier).

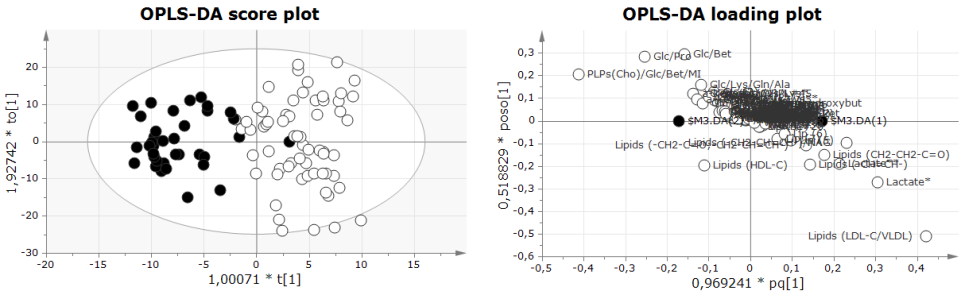


Figure S26. OPLS-DA score and loading plot. (A) OPLS-DA score plot: The first predictive component (t[1]) explains the variation between both groups and the first orthogonal component (to[1]) explains the variation within both groups. (B) OPLS-DA loading plot: The horizontal axis displays the X-loadings p and the Y-loadings q of the predictive component. The vertical axis displays the X-loadings p(o) and the Y-loadings s(o) for the Y-orthogonal component.

The model performance can be examined by evaluating the total amount of predictive and orthogonal variation in X ($R^2X(cum)$), the total amount of variation in Y ($R^2Y(cum)$), and the predictive ability of the model ($Q^2(cum)$) as determined by 7-fold cross-validation. The higher these values, the higher the association between X and Y variables, the better the model (**Figure S26**).

In OPLS-DA it is important to define both a training set (original data) and an independent validation set (a class of separate samples that can be predicted based on a series of mathematical models derived from the training set) in order to validate study results (239) (**Figure S27**).

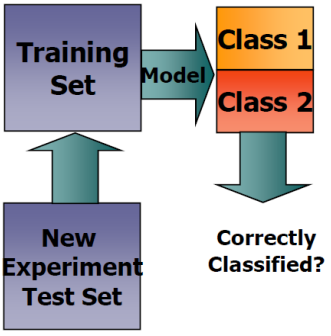


Figure S27. Classification models. First, the model must be trained on representative data (training set). Next, the model must be tested using new data (validation set).

Diagnostic tools

OPLS-DA provides many diagnostic tools which assist in the model interpretation, and in the assessment of model performance and relevance. The misclassification table summarizes how well the OPLS-DA models classify the observations into known classes (**Figure S28**). Sensitivity is defined as the percentage of patients that are actually classified as patients, while specificity is explained as the percentage of controls that are actually classified as controls.

	Members	Correct	0	1
0 = Specificity	36	92%	33	3
1 = Sensitivity	61	95%	3	58

Figure S28. Misclassification table. 33 out of 36 controls are correctly classified and 58 out of 61 patients are correctly classified, corresponding to a specificity of 92% and a sensitivity of 95%, respectively.

The variable influence on projection (VIP) parameter gives the importance of the X-variables, both for X- and Y-models (**Figure S29**). The VIP is a weighted sum of squares of the OPLS-DA weights, taking into account the amount of explained Y-variance in each dimension. Hence, one VIP-vector summarizes all components and Y-variables. Predictors with a large VIP are most influential for the model. The VIP value and/or visual thresholds on the VIP plot may be used for variable selection. Note, however, that variable selection should be carried out with caution as there are strong correlations among the X-variables.

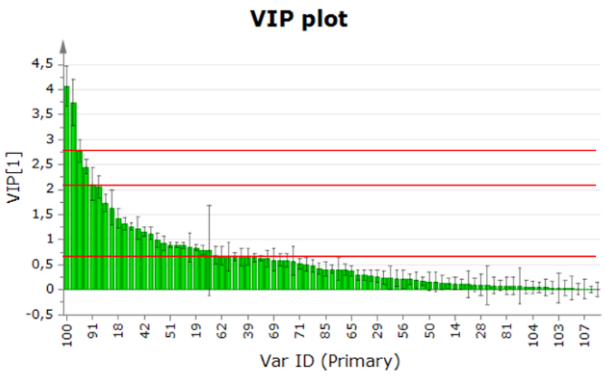


Figure S29. Variable influence on projection (VIP) plot. The red lines are indicating different thresholds which might be used and tested for variable selection.

Especially for NMR data, an S-plot of the variables responsible for the pattern seen among the observations, is very useful and can be illustrated as points or a line (**Figure S30**). The S-plot is a practical and reliable way to identify important discriminating variables (241).

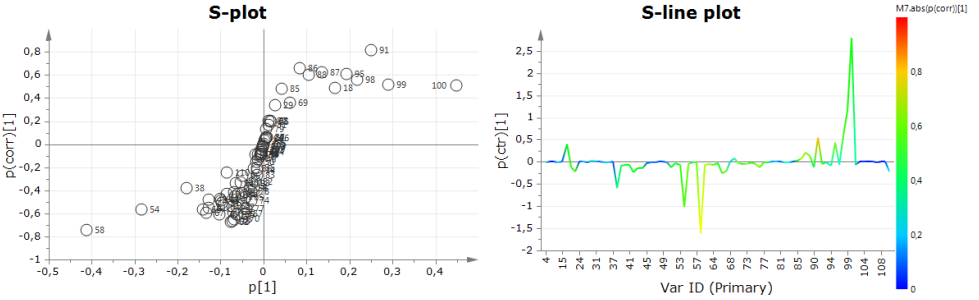


Figure S30. S-plots of an OPLS-DA model.

(A) S-plot with variables in the bottom left or upper right as being those with a strong contribution to the model and high statistical reliability. (B) S-line plot has the advantage of taking the spectral order of NMR data into account. The relevance of the model is indicated by the signal amplitude and the significance by color. Strongly discriminating variables combine a high numerical loading value and red to orange color.

A permutation plot can be developed for PLS-DA to assess the risk that the current model is spurious, i.e. the model fits the training set well but does not predict Y well for new observations (**Figure S31**).

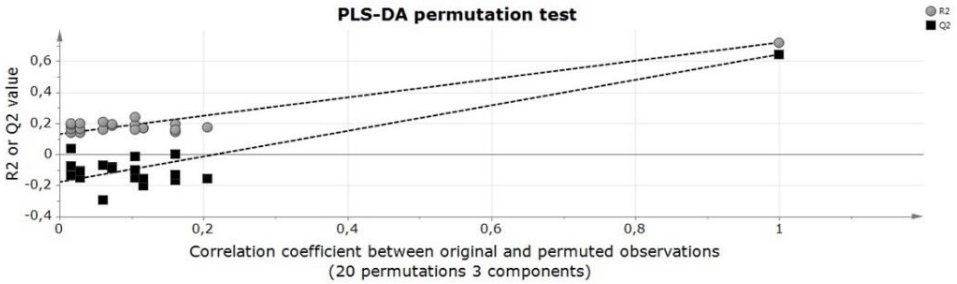


Figure S31. PLS-DA permutation test. The PLS-DA permutation test illustrates the validity of the model. For a model to be valid, R^2 (grey dots) and Q^2 (black boxes) values for each permuted observation have to be lower than the R2 and Q2 values of the original model on the right.

The permutation plot compares the goodness of fit (R^2 and Q^2) of the original model with the goodness of fit of several models based on data where the order of the Y-observations has been randomly permuted, while the X-matrix has been kept intact. The plot shows, for a selected Y-variable, on the vertical axis the values of R^2 and Q^2 for the original model (on the right side) and of the Y-permuted

models (on the left side). The horizontal axis shows the correlation between the permuted Y-vectors and the original Y-vector for the selected Y. Criteria for a valid model are: (i) all Q^2 values on the left are lower than the original point on the right; (ii) the regression line of the Q^2 points intersects the vertical axis (on the left) at or below zero; (iii) all R^2 values to the left are lower than the original point to the right.

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. International journal of cancer. 2015;136(5):E359-86.
2. Boyle P, Levin B. World Health Organization. International agency for research on Cancer. World Cancer Report. 2008.
3. Francart J, De Gendt C, Vandeven J, Adam M. Cancer Survival in Belgium. Belgian Cancer Registry, 2012.
4. Goldstraw P, Crowley J, Chansky K, Giroux DJ, Groome PA, Rami-Porta R, et al. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours. Journal of thoracic oncology. 2007;2(8):706-14.
5. Renard F, Tafforeau J, Deboosere P. Premature mortality in Belgium in 1993-2009: leading causes, regional disparities and 15 years change. Archives of public health. 2014;72(1):34.
6. Renard F. Voortijdige kankersterfte bij volwassenen in België: belangrijkste oorzaken en evolutie van 1993-1994 tot 2008-2009. Onco Hemato. 2015;9(3):21-30.
7. Shopland DR, Eyre HJ, Pechacek TF. Smoking-attributable cancer mortality in 1991: is lung cancer now the leading cause of death among smokers in the United States? Journal of the National Cancer Institute. 1991;83(16):1142-8.
8. Peto R, Darby S, Deo H, Silcocks P, Whitley E, Doll R. Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. BMJ. 2000;321(7257):323-9.
9. Manser R, Lethaby A, Irving LB, Stone C, Byrnes G, Abramson MJ, et al. Screening for lung cancer. Cochrane Database of Systematic Reviews 2004. 2013(1):CD001991.
10. Halpern MT, Gillespie BW, Warner KE. Patterns of absolute risk of lung cancer mortality in former smokers. Journal of the National Cancer Institute. 1993;85(6):457-64.
11. Crispo A, Brennan P, Jockel KH, Schaffrath-Rosario A, Wichmann HE, Nyberg F, et al. The cumulative risk of lung cancer among current, ex- and never-smokers in European men. British journal of cancer. 2004;91(7):1280-6.
12. Woolf SH, Harris RP, Campos-Outcalt D. Low-dose computed tomography screening for lung cancer: how strong is the evidence? JAMA internal medicine. 2014;174(12):2019-22.
13. Bourzac K. Diagnosis: Early warning system. Nature. 2014;513(7517):S4-6.
14. Shlomi D, Ben-Avi R, Balmor GR, Onn A, Peled N. Screening for lung cancer: time for large-scale screening by chest computed tomography. The European respiratory journal. 2014;44(1):217-38.
15. Wood DE, Eapen GA, Ettinger DS, Hou L, Jackman D, Kazerooni E, et al. Lung cancer screening. Journal of the National Comprehensive Cancer Network. 2012;10(2):240-65.
16. Obuchowski NA, Graham RJ, Baker ME, Powell KA. Ten criteria for effective screening: their application to multislice CT screening for pulmonary and colorectal cancers. American journal of roentgenology. 2001;176(6):1357-62.
17. Field JK, Oudkerk M, Pedersen JH, Duffy SW. Prospects for population screening and diagnosis of lung cancer. Lancet. 2013;382(9893):732-41.
18. Welch HG, Woloshin S, Schwartz LM, Gordis L, Gotzsche PC, Harris R, et al. Overstating the evidence for lung cancer screening: the International Early Lung Cancer Action Program (I-ELCAP) study. Arch Intern Med. 2007;167(21):2289-95.
19. Humphrey LL, Deffebach M, Pappas M, Baumann C, Artis K, Mitchell JP, et al. Screening for lung cancer with low-dose computed tomography: a systematic review to update the US Preventive services task force recommendation. Annals of internal medicine. 2013;159(6):411-20.
20. Henschke CI, McCauley DI, Yankelevitz DF, Naidich DP, McGuinness G, Miettinen OS, et al. Early Lung Cancer Action Project: overall design and findings from baseline screening. Lancet. 1999;354(9173):99-105.
21. Tammemagi MC, Lam S. Screening for lung cancer using low dose computed tomography. BMJ. 2014;348:g2253.

22. Swensen SJ, Jett JR, Hartman TE, Midthun DE, Mandrekar SJ, Hillman SL, et al. CT screening for lung cancer: five-year prospective experience. Radiology. 2005;235(1):259-65. Epub 2005/02/08.
23. Field JK, Duffy SW. Lung cancer screening: the way forward. British journal of cancer. 2008;99(4):557-62.
24. Whynes DK. Could CT screening for lung cancer ever be cost effective in the United Kingdom? Cost effectiveness and resource allocation. 2008;6:5.
25. Field JK, Chen Y, Marcus MW, McRonald FE, Raji OY, Duffy SW. The contribution of risk prediction models to early detection of lung cancer. Journal of surgical oncology. 2013;108(5):304-11.
26. Cassidy A, Myles JP, van Tongeren M, Page RD, Liloglou T, Duffy SW, et al. The LLP risk model: an individual risk prediction model for lung cancer. British journal of cancer. 2008;98(2):270-6.
27. Spitz MR, Hong WK, Amos CI, Wu X, Schabath MB, Dong Q, et al. A risk model for prediction of lung cancer. Journal of the National Cancer Institute. 2007;99(9):715-26.
28. Bach PB, Kattan MW, Thornquist MD, Kris MG, Tate RC, Barnett MJ, et al. Variations in lung cancer risk among smokers. Journal of the National Cancer Institute. 2003;95(6):470-8.
29. Hoggart C, Brennan P, Tjonneland A, Vogel U, Overvad K, Ostergaard JN, et al. A risk model for lung cancer incidence. Cancer Prev Res 2012;5(6):834-46.
30. Tammemagi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, et al. Selection criteria for lung-cancer screening. The New England journal of medicine. 2013;368(8):728-36.
31. Ganti AK, Mulshine JL. Lung cancer screening: panacea or pipe dream? Annals of oncology. 2005;16 Suppl 2:ii215-9.
32. van Klaveren RJ, Habbema JDF, Pedersen JH, de Koning HJ, Oudkerk M, Hoogsteden HC. Lung cancer screening by low-dose spiral computed tomography. The European respiratory journal. 2001;18(5):857-66.
33. Black WC. Overdiagnosis: An underrecognized cause of confusion and harm in cancer screening. Journal of the National Cancer Institute. 2000;92(16):1280-2.
34. Devesa SS, Bray F, Vizcaino AP, Parkin DM. International lung cancer trends by histologic type: male:female differences diminishing and adenocarcinoma rates rising. International journal of cancer. 2005;117(2):294-9.
35. Cuffe S, Moua T, Summerfield R, Roberts H, Jett J, Shepherd FA. Characteristics and outcomes of small cell lung cancer patients diagnosed during two lung cancer computed tomographic screening programs in heavy smokers. Journal of thoracic oncology. 2011;6(4):818-22.
36. Jett JR. Limitations of screening for lung cancer with low-dose spiral computed tomography. Clinical cancer research. 2005;11(13 Pt 2):4988s-92s.
37. Brett GZ. The value of lung cancer detection by six-monthly chest radiographs. Thorax. 1968;23(4):414-20.
38. Fontana RS, Sanderson DR, Taylor WF, Woolner LB, Miller WE, Muhm JR, et al. Early lung cancer detection: results of the initial (prevalence) radiologic and cytologic screening in the Mayo Clinic study. The American review of respiratory disease. 1984;130(4):561-5.
39. Kubik AK, Parkin DM, Zatloukal P. Czech Study on Lung Cancer Screening: post-trial follow-up of lung cancer deaths up to year 15 since enrollment. Cancer. 2000;89(11 Suppl):2363-8.
40. Melamed MR, Flehinger BJ, Zaman MB, Heelan RT, Perchick WA, Martini N. Screening for early lung cancer. Results of the Memorial Sloan-Kettering study in New York. Chest. 1984;86(1):44-53.
41. Tockman MS. Survival and Mortality from Lung-Cancer in a Screened Population - the Johns-Hopkins Study. Chest. 1986;89(4):S324-S5.
42. Oken MM, Hocking WG, Kvale PA, Andriole GL, Buys SS, Church TR, et al. Screening by chest radiograph and lung cancer mortality: the Prostate, Lung, Colorectal, and Ovarian (PLCO) randomized trial. Jama. 2011;306(17):1865-73.
43. Sone S, Li F, Yang ZG, Takashima S, Maruyama Y, Hasegawa M, et al. Characteristics of small lung cancers invisible on conventional chest radiography and

detected by population based screening using spiral CT. The British journal of radiology. 2000;73(866):137-45.
44. Bach PB, Mirkin JN, Oliver TK, Azzoli CG, Berry DA, Brawley OW, et al. Benefits and harms of CT screening for lung cancer: a systematic review. Jama. 2012;307(22):2418-29.
45. Aberle DR, Abtin F, Brown K. Computed tomography screening for lung cancer: has it finally arrived? Implications of the national lung screening trial. Journal of clinical oncology. 2013;31(8):1002-8.
46. Larke FJ, Kruger RL, Cagnon CH, Flynn MJ, McNitt-Gray MM, Wu X, et al. Estimated radiation dose associated with low-dose chest CT of average-size participants in the National Lung Screening Trial. American journal of roentgenology. 2011;197(5):1165-9.
47. Aberle DR, Berg CD, Black WC, Church TR, Fagerstrom RM, Galen B, et al. The National Lung Screening Trial: overview and study design. Radiology. 2011;258(1):243-53.
48. Sone S, Li F, Yang ZG, Honda T, Maruyama Y, Takashima S, et al. Results of three-year mass screening programme for lung cancer using mobile low-dose spiral computed tomography scanner. British journal of cancer. 2001;84(1):25-32.
49. Sone S, Takashima S, Li F, Yang Z, Honda T, Maruyama Y, et al. Mass screening for lung cancer with mobile spiral computed tomography scanner. Lancet. 1998;351(9111):1242-5.
50. Diederich S, Wormanns D, Semik M, Thomas M, Lenzen H, Roos N, et al. Screening for early lung cancer with low-dose spiral CT: prevalence in 817 asymptomatic smokers. Radiology. 2002;222(3):773-81.
51. Henschke CI, McCauley DI, Yankelevitz DF, Naidich DP, McGuinness G, Miettinen OS, et al. Early lung cancer action project: a summary of the findings on baseline screening. The oncologist. 2001;6(2):147-52.
52. Nawa T, Nakagawa T, Kusano S, Kawasaki Y, Sugawara Y, Nakata H. Lung cancer screening using low-dose spiral CT: results of baseline and 1-year follow-up studies. Chest. 2002;122(1):15-20.
53. Sobue T, Moriyama N, Kaneko M, Kusumoto M, Kobayashi T, Tsuchiya R, et al. Screening for lung cancer with low-dose helical computed tomography: anti-lung cancer association project. Journal of clinical oncology. 2002;20(4):911-20.
54. Kaneko M, Eguchi K, Ohmatsu H, Kakinuma R, Naruke T, Suemasu K, et al. Peripheral lung cancer: screening and detection with low-dose spiral CT versus radiography. Radiology. 1996;201(3):798-802.
55. Swensen SJ, Jett JR, Sloan JA, Midthun DE, Hartman TE, Sykes AM, et al. Screening for lung cancer with low-dose spiral computed tomography. American journal of respiratory and critical care medicine. 2002;165(4):508-13.
56. Henschke CI, Yankelevitz DF, Libby DM, Pasmantier MW, Smith JP, Miettinen OS. Survival of patients with stage I lung cancer detected on CT screening. The New England journal of medicine. 2006;355(17):1763-71.
57. Tiitola M, Kivisaari L, Huuskonen MS, Mattson K, Koskinen H, Lehtola H, et al. Computed tomography screening for lung cancer in asbestos-exposed workers. Lung Cancer. 2002;35(1):17-22.
58. Humphrey LL, Teutsch S, Johnson M. Lung cancer screening with sputum cytologic examination, chest radiography, and computed tomography: an update for the U.S. Preventive Services Task Force. Annals of internal medicine. 2004;140(9):740-53.
59. Mulshine JL. New developments in lung cancer screening. Journal of clinical oncology. 2005;23(14):3198-202.
60. Gohagan JK, Marcus PM, Fagerstrom RM, Pinsky PF, Kramer BS, Prorok PC, et al. Final results of the Lung Screening Study, a randomized feasibility study of spiral CT versus chest X-ray screening for lung cancer. Lung Cancer. 2005;47(1):9-15.
61. Black WC. Computed tomography screening for lung cancer: review of screening principles and update on current status. Cancer. 2007;110(11):2370-84.
62. Gohagan J, Marcus P, Fagerstrom R, Pinsky P, Kramer B, Prorok P. Baseline findings of a randomized feasibility trial of lung cancer screening with spiral CT scan vs chest radiograph: the Lung Screening Study of the National Cancer Institute. Chest. 2004;126(1):114-21.

63. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. The New England journal of medicine. 2011;365(5):395-409.
64. Church TR, Black WC, Aberle DR, Berg CD, Clingan KL, Duan F, et al. Results of initial low-dose computed tomographic screening for lung cancer. The New England journal of medicine. 2013;368(21):1980-91.
65. Black WC, Gareen IF, Soneji SS, Sicks JD, Keeler EB, Aberle DR, et al. Cost-effectiveness of CT screening in the National Lung Screening Trial. The New England journal of medicine. 2014;371(19):1793-802.
66. Infante M, Cavuto S, Lutman FR, Brambilla G, Chiesa G, Ceresoli G, et al. A randomized study of lung cancer screening with spiral computed tomography: three-year results from the DANTE trial. American journal of respiratory and critical care medicine. 2009;180(5):445-53.
67. Saghir Z, Dirksen A, Ashraf H, Bach KS, Brodersen J, Clementsen PF, et al. CT screening for lung cancer brings forward early disease. The randomised Danish Lung Cancer Screening Trial: status after five annual screening rounds with low-dose CT. Thorax. 2012;67(4):296-301.
68. Lopes Pegna A, Picozzi G, Falaschi F, Carrozzi L, Falchini M, Carozzi FM, et al. Four-year results of low-dose CT screening and nodule management in the ITALUNG trial. Journal of thoracic oncology. 2013;8(7):866-75.
69. Pastorino U, Rossi M, Rosato V, Marchiano A, Sverzellati N, Morosi C, et al. Annual or biennial CT screening versus observation in heavy smokers: 5-year results of the MILD trial. Eur J Cancer Prev. 2012;21(3):308-15.
70. Becker N, Motsch E, Gross ML, Eigentopf A, Heussel CP, Dienemann H, et al. Randomised study on early detection of lung cancer with MSCT in Germany: results of the first 3 years of follow-up after randomisation. Journal of thoracic oncology. 2015.
71. Horeweg N, Scholten ET, de Jong PA, van der Aalst CM, Weenink C, Lammers JW, et al. Detection of lung cancer through low-dose CT screening (NELSON): a prespecified analysis of screening test performance and interval cancers. The Lancet Oncology. 2014;15(12):1342-50.
72. van Klaveren RJ, Oudkerk M, Prokop M, Scholten ET, Nackaerts K, Vernhout R, et al. Management of lung nodules detected by volume CT scanning. The New England journal of medicine. 2009;361(23):2221-9.
73. van Iersel CA, de Koning HJ, Draisma G, Mali WP, Scholten ET, Nackaerts K, et al. Risk-based selection from the general population in a screening trial: selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON). International journal of cancer Journal. 2007;120(4):868-74.
74. van den Bergh KA, Essink-Bot ML, Bunge EM, Scholten ET, Prokop M, van Iersel CA, et al. Impact of computed tomography screening for lung cancer on participants in a randomized controlled trial (NELSON trial). Cancer. 2008;113(2):396-404.
75. van den Bergh KA, Essink-Bot ML, Borsboom GJ, Th Scholten E, Prokop M, de Koning HJ, et al. Short-term health-related quality of life consequences in a lung cancer CT screening trial (NELSON). British journal of cancer. 2010;102(1):27-34.
76. Patz EF, Jr., Pinsky P, Gatsonis C, Sicks JD, Kramer BS, Tammemagi MC, et al. Overdiagnosis in low-dose computed tomography screening for lung cancer. JAMA internal medicine. 2014;174(2):269-74.
77. Veronesi G, Maisonneuve P, Bellomi M, Rampinelli C, Durli I, Bertolotti R, et al. Estimating overdiagnosis in low-dose computed tomography screening for lung cancer: a cohort study. Annals of internal medicine. 2012;157(11):776-84.
78. Brenner DJ. Radiation risks potentially associated with low-dose CT screening of adult smokers for lung cancer. Radiology. 2004;231(2):440-5.
79. Berrington de Gonzalez A, Kim KP, Berg CD. Low-dose lung computed tomography screening before age 55: estimates of the mortality reduction required to outweigh the radiation-induced cancer risk. Journal of medical screening. 2008;15(3):153-8.
80. Raji OY, Agbaje OF, Duffy SW, Cassidy A, Field JK. Incorporation of a genetic factor into an epidemiologic model for prediction of individual risk of lung cancer: the Liverpool Lung Project. Cancer Prev Res. 2010;3(5):664-9.

81. Spitz MR, Etzel CJ, Dong Q, Amos CI, Wei Q, Wu X, et al. An expanded risk prediction model for lung cancer. Cancer Prev Res. 2008;1(4):250-4.
82. Smolinska A, Blanchet L, Buydens LM, Wijmenga SS. NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review. Analytica chimica acta. 2012;750:82-97.
83. Broadhurst DI, Kell DB. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics. 2006;2(4):171-96.
84. Knottnerus JA, van Weel C, Muris JW. Evaluation of diagnostic procedures. BMJ. 2002;324(7335):477-80.
85. Tsay JC, DeCotiis C, Greenberg AK, Rom WN. Current readings: blood-based biomarkers for lung cancer. Seminars in thoracic and cardiovascular surgery. 2013;25(4):328-34.
86. Mamas M, Dunn WB, Neyses L, Goodacre R. The role of metabolites and metabolomics in clinically applicable biomarkers of disease. Archives of toxicology. 2011;85(1):5-17.
87. Vermeersch KA, Styczynski MP. Applications of metabolomics in cancer research. Journal of carcinogenesis. 2013;12:9.
88. Hasan N, Kumar R, Kavuru MS. Lung cancer screening beyond low-dose computed tomography: the role of novel biomarkers. Lung. 2014;192(5):639-48.
89. Hassanein M, Callison JC, Callaway-Lane C, Aldrich MC, Grogan EL, Massion PP. The state of molecular biomarkers for the early detection of lung cancer. Cancer Prev Res. 2012;5(8):992-1006.
90. Neal JW, Gainor JF, Shaw AT. Developing biomarker-specific end points in lung cancer clinical trials. Nature reviews Clinical oncology. 2015;12(3):135-46.
91. Sozzi G, Boeri M, Rossi M, Verri C, Suatoni P, Bravi F, et al. Clinical utility of a plasma-based miRNA signature classifier within computed tomography lung cancer screening: a correlative MILD trial study. Journal of clinical oncology. 2014;32(8):768-73.
92. Sung HJ, Cho JY. Biomarkers for the lung cancer diagnosis and their advances in proteomics. BMB reports. 2008;41(9):615-25.
93. Goodacre R. Metabolomics of a superorganism. The Journal of nutrition. 2007;137(1 Suppl):259S-66S.
94. Spratlin JL, Serkova NJ, Eckhardt SG. Clinical applications of metabolomics in oncology: a review. Clinical cancer research. 2009;15(2):431-40.
95. Issaq HJ, Van QN, Waybright TJ, Muschik GM, Veenstra TD. Analytical and statistical approaches to metabolomics research. Journal of separation science. 2009;32(13):2183-99.
96. Patti GJ, Yanes O, Siuzdak G. Innovation: Metabolomics: the apogee of the omics trilogy. Nature reviews Molecular cell biology. 2012;13(4):263-9.
97. O'Connell TM. Recent advances in metabolomics in oncology. Bioanalysis. 2012;4(4):431-51.
98. Van QN, Veenstra TD. How close is the bench to the bedside? Metabolic profiling in cancer research. Genome medicine. 2009;1(1):5.
99. Cuperlovic-Culf M. NMR metabolomics in cancer research. 1 ed. Cambridge: Woodhead Publishing; 2013.
100. Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB. Metabolomics by numbers: acquiring and understanding global metabolite data. Trends in biotechnology. 2004;22(5):245-52.
101. Hollywood K, Brison DR, Goodacre R. Metabolomics: current technologies and future trends. Proteomics. 2006;6(17):4716-23.
102. Griffin JL, Shockcor JP. Metabolic profiles of cancer cells. Nature reviews Cancer. 2004;4(7):551-61.
103. Lindon JC, Nicholson JK. Spectroscopic and statistical techniques for information recovery in metabonomics and metabolomics. Annu Rev Anal Chem. 2008;1:45-69.
104. Collino S, Martin FP, Rezzi S. Clinical metabolomics paves the way towards future healthcare strategies. British journal of clinical pharmacology. 2013;75(3):619-29.
105. Fiehn O. Metabolomics--the link between genotypes and phenotypes. Plant molecular biology. 2002;48(1-2):155-71.

106. Lenz EM, Wilson ID. Analytical strategies in metabonomics. Journal of proteome research. 2007;6(2):443-58.
107. Holmes E, Wilson ID, Nicholson JK. Metabolic phenotyping in health and disease. Cell. 2008;134(5):714-7.
108. Bictash M, Ebbels TM, Chan Q, Loo RL, Yap IK, Brown IJ, et al. Opening up the "Black Box": metabolic phenotyping and metabolome-wide association studies in epidemiology. Journal of clinical epidemiology. 2010;63(9):970-9.
109. Kochhar S, Jacobs DM, Ramadan Z, Berruex F, Fuerholz A, Fay LB. Probing gender-specific metabolism differences in humans by nuclear magnetic resonance-based metabonomics. Analytical biochemistry. 2006;352(2):274-81.
110. Slupsky CM, Rankin KN, Wagner J, Fu H, Chang D, Weljie AM, et al. Investigations of the effects of gender, diurnal variation, and age in human urinary metabolomic profiles. Analytical chemistry. 2007;79(18):6995-7004.
111. Psihogios NG, Gazi IF, Elisaf MS, Seferiadis KI, Bairaktari ET. Gender-related and age-related urinalysis of healthy subjects by NMR-based metabonomics. NMR in biomedicine. 2008;21(3):195-207.
112. Lenz EM, Bright J, Wilson ID, Hughes A, Morrisson J, Lindberg H, et al. Metabonomics, dietary influences and cultural differences: a 1H NMR-based study of urine samples obtained from healthy British and Swedish subjects. Journal of pharmaceutical and biomedical analysis. 2004;36(4):841-9.
113. Stella C, Beckwith-Hall B, Cloarec O, Holmes E, Lindon JC, Powell J, et al. Susceptibility of human metabolic phenotypes to dietary modulation. Journal of proteome research. 2006;5(10):2780-8.
114. Emwas AH. The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. Methods Mol Biol. 2015;1277:161-93.
115. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0--The Human Metabolome Database in 2013. Nucleic acids research. 2013;41(Database issue):D801-7.
116. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, et al. HMDB: the Human Metabolome Database. Nucleic acids research. 2007;35(Database issue):D521-6.
117. Emwas AHM, Salek RM, Griffin JL, Merzaban J. NMR-based metabolomics in human disease diagnosis: applications, limitations, and recommendations. Metabolomics. 2013;9(5):1048-72.
118. Lindon JC, Nicholson JK. Analytical technologies for metabonomics and metabolomics, and multi-omic information recovery. Trac-Trend Anal Chem. 2008;27(3):194-204.
119. Dunn WB, Broadhurst DI, Atherton HJ, Goodacre R, Griffin JL. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. Chemical Society reviews. 2011;40(1):387-426.
120. Madsen R, Lundstedt T, Trygg J. Chemometrics in metabolomics--a review in human disease diagnosis. Analytica chimica acta. 2010;659(1-2):23-33.
121. Armitage EG, Barbas C. Metabolomics in cancer biomarker discovery: current trends and future perspectives. Journal of pharmaceutical and biomedical analysis. 2014;87:1-11.
122. Dumas ME, Maibaum EC, Teague C, Ueshima H, Zhou B, Lindon JC, et al. Assessment of analytical reproducibility of 1H NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP Study. Analytical chemistry. 2006;78(7):2199-208.
123. Keun HC, Ebbels TM, Antti H, Bollard ME, Beckonert O, Schlotterbeck G, et al. Analytical reproducibility in (1)H NMR-based metabonomic urinalysis. Chemical research in toxicology. 2002;15(11):1380-6.
124. Cacciatore S, Loda M. Innovation in metabolomics to improve personalized healthcare. Annals of the New York Academy of Sciences. 2015;1346(1):57-62.
125. Aboud OA, Weiss RH. New opportunities from the cancer metabolome. Clinical chemistry. 2013;59(1):138-46.

126. Nicholson JK, O'Flynn MP, Sadler PJ, Macleod AF, Juul SM, Sonksen PH. Proton-nuclear-magnetic-resonance studies of serum, plasma and urine from fasting normal and diabetic subjects. The Biochemical journal. 1984;217(2):365-75.
127. Nicholson JK, Higham DP, Timbrell JA, Sadler PJ. Quantitative high resolution 1H NMR urinalysis studies on the biochemical effects of cadmium in the rat. Molecular pharmacology. 1989;36(3):398-404.
128. Bernini P, Bertini I, Luchinat C, Nincheri P, Staderini S, Turano P. Standard operating procedures for pre-analytical handling of blood and urine for metabolomic studies and biobanks. Journal of biomolecular NMR. 2011;49(3-4):231-43.
129. Teahan O, Gamble S, Holmes E, Waxman J, Nicholson JK, Bevan C, et al. Impact of analytical bias in metabonomic studies of human blood serum and plasma. Analytical chemistry. 2006;78(13):4307-18.
130. Dona AC, Jimenez B, Schafer H, Humpfer E, Spraul M, Lewis MR, et al. Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. Analytical chemistry. 2014;86(19):9887-94.
131. Duarte IF, Diaz SO, Gil AM. NMR metabolomics of human blood and urine in disease research. Journal of pharmaceutical and biomedical analysis. 2014;93:17-26.
132. Lindon JC, Nicholson JK, Holmes E, Everett JR. Metabonomics: Metabolic processes studied by NMR spectroscopy of biofluids. Concept Magnetic Res. 2000;12(5):289-320.
133. Dallmann R, Viola AU, Tarokh L, Cajochen C, Brown SA. The human circadian metabolome. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(7):2625-9.
134. Beger RD. A review of applications of metabolomics in cancer. Metabolites. 2013;3(3):552-74.
135. Beckonert O, Keun HC, Ebbels TM, Bundy J, Holmes E, Lindon JC, et al. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. Nature protocols. 2007;2(11):2692-703.
136. Gebregiworgis T, Powers R. Application of NMR metabolomics to search for human disease biomarkers. Combinatorial chemistry & high throughput screening. 2012;15(8):595-610.
137. Meiboom S, Gill D. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. Rev Sci Instrum. 1958;29(8):688-91.
138. Pan Z, Raftery D. Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. Analytical and bioanalytical chemistry. 2007;387(2):525-7.
139. Duarte IF, Gil AM. Metabolic signatures of cancer unveiled by NMR spectroscopy of human biofluids. Prog Nucl Mag Res Sp. 2012;62:51-74.
140. Bertram HC, Malmendal A, Petersen BO, Madsen JC, Pedersen H, Nielsen NC, et al. Effect of magnetic field strength on NMR-based metabonomic human urine data. Comparative study of 250, 400, 500, and 800 MHz. Analytical chemistry. 2007;79(18):7110-5.
141. Gruetter R, Weisdorf SA, Rajanayagan V, Terpstra M, Merkle H, Truwit CL, et al. Resolution improvements in in vivo 1H NMR spectra with increased magnetic field strength. J Magn Reson. 1998;135(1):260-4.
142. Jeannerat D, Furrer J. NMR experiments for the analysis of mixtures: beyond 1D 1H spectra. Combinatorial chemistry & high throughput screening. 2012;15(1):15-35.
143. De Meyer T, Sinnaeve D, Van Gasse B, Tsiporkova E, Rietzschel ER, De Buyzere ML, et al. NMR-based characterization of metabolic alterations in hypertension using an adaptive, intelligent binning algorithm. Analytical chemistry. 2008;80(10):3783-90.
144. Beckwith-Hall BM, Nicholson JK, Nicholls AW, Foxall PJD, Lindon JC, Connor SC, et al. Nuclear magnetic resonance spectroscopic and principal components analysis investigations into biochemical effects of three model hepatotoxins. Chemical research in toxicology. 1998;11(4):260-72.
145. Bollard ME, Stanley EG, Lindon JC, Nicholson JK, Holmes E. NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition. NMR in biomedicine. 2005;18(3):143-62.
146. Spraul M, Neidig P, Klauck U, Kessler P, Holmes E, Nicholson JK, et al. Automatic reduction of NMR spectroscopic data for statistical and pattern recognition classification of samples. Journal of pharmaceutical and biomedical analysis. 1994;12(10):1215-25.

147. Craig A, Cloarec O, Holmes E, Nicholson JK, Lindon JC. Scaling and normalization effects in NMR spectroscopic metabonomic data sets. Analytical chemistry. 2006;78(7):2262-7.
148. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. Analytical chemistry. 2006;78(13):4281-90.
149. van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC genomics. 2006;7:142.
150. Gromski PS, Xu Y, Hollywood KA, Turner ML, Goodacre R. The influence of scaling metabolomics data on model classification accuracy. Metabolomics. 2015;11(3):684-95.
151. Trygg J, Holmes E, Lundstedt T. Chemometrics in metabonomics. Journal of proteome research. 2007;6(2):469-79.
152. Blekherman G, Laubenbacher R, Cortes DF, Mendes P, Torti FM, Akman S, et al. Bioinformatics tools for cancer metabolomics. Metabolomics. 2011;7(3):329-43.
153. Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). J Chemometr. 2002;16(3):119-28.
154. Palmnas MS, Vogel HJ. The future of NMR metabolomics in cancer therapy: towards personalizing treatment and developing targeted drugs? Metabolites. 2013;3(2):373-96.
155. Sciacovelli M, Gaude E, Hilvo M, Frezza C. The metabolic alterations of cancer cells. Methods in enzymology. 2014;542:1-23.
156. Kroemer G, Pouyssegur J. Tumor cell metabolism: cancer's Achilles' heel. Cancer cell. 2008;13(6):472-82.
157. Cairns RA, Harris IS, Mak TW. Regulation of cancer cell metabolism. Nature reviews Cancer. 2011;11(2):85-95.
158. Cantor JR, Sabatini DM. Cancer cell metabolism: one hallmark, many faces. Cancer discovery. 2012;2(10):881-98.
159. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100(1):57-70.
160. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74.
161. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science. 2009;324(5930):1029-33.
162. Bracke M, Lardon F, Vandenberghe P, Vanderkerken K. Kanker biomedisch bekeken. 1 ed. Antwerpen: Standaard Uitgeverij; 2011.
163. Warburg O. On the origin of cancer cells. Science. 1956;123(3191):309-14.
164. Warburg O, Wind F, Negelein E. The Metabolism of Tumors in the Body. The Journal of general physiology. 1927;8(6):519-30.
165. Amoedo ND, Valencia JP, Rodrigues MF, Galina A, Rumjanek FD. How does the metabolism of tumour cells differ from that of normal cells. Bioscience reports. 2013;33(6).
166. Wise DR, Thompson CB. Glutamine addiction: a new therapeutic target in cancer. Trends in biochemical sciences. 2010;35(8):427-33.
167. Icard P, Lincet H. A global view of the biochemical pathways involved in the regulation of the metabolism of cancer cells. Biochimica et biophysica acta. 2012;1826(2):423-33.
168. Weinberg F, Chandel NS. Mitochondrial metabolism and cancer. Annals of the New York Academy of Sciences. 2009;1177:66-73.
169. DeBerardinis RJ, Lum JJ, Hatzivassiliou G, Thompson CB. The biology of cancer: metabolic reprogramming fuels cell growth and proliferation. Cell metabolism. 2008;7(1):11-20.
170. DeBerardinis RJ, Mancuso A, Daikhin E, Nissim I, Yudkoff M, Wehrli S, et al. Beyond aerobic glycolysis: transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(49):19345-50.
171. Mullen AR, Wheaton WW, Jin ES, Chen PH, Sullivan LB, Cheng T, et al. Reductive carboxylation supports growth in tumour cells with defective mitochondria. Nature. 2012;481(7381):385-8.
172. Moreno-Sanchez R, Rodriguez-Enriquez S, Marin-Hernandez A, Saavedra E. Energy metabolism in tumor cells. The FEBS journal. 2007;274(6):1393-418.

173. Zu XL, Guppy M. Cancer metabolism: facts, fantasy, and fiction. Biochemical and biophysical research communications. 2004;313(3):459-65.
174. Ferreira LM, Hebrant A, Dumont JE. Metabolic reprogramming of the tumor. Oncogene. 2012;31(36):3999-4011.
175. Chen X, Qian Y, Wu S. The Warburg effect: evolving interpretations of an established concept. Free radical biology & medicine. 2015;79:253-63.
176. Deberardinis RJ, Sayed N, Ditsworth D, Thompson CB. Brick by brick: metabolism and tumor cell growth. Current opinion in genetics & development. 2008;18(1):54-61.
177. Iurlaro R, Leon-Annicchiarico CL, Munoz-Pinedo C. Regulation of cancer metabolism by oncogenes and tumor suppressors. Methods in enzymology. 2014;542:59-80.
178. Levine AJ, Puzio-Kuter AM. The control of the metabolic switch in cancers by oncogenes and tumor suppressor genes. Science. 2010;330(6009):1340-4.
179. Kim JW, Dang CV. Cancer's molecular sweet tooth and the Warburg effect. Cancer research. 2006;66(18):8927-30.
180. Elstrom RL, Bauer DE, Buzzai M, Karnauskas R, Harris MH, Plas DR, et al. Akt stimulates aerobic glycolysis in cancer cells. Cancer research. 2004;64(11):3892-9.
181. Mathupala SP, Ko YH, Pedersen PL. Hexokinase II: cancer's double-edged sword acting as both facilitator and gatekeeper of malignancy when bound to mitochondria. Oncogene. 2006;25(34):4777-86.
182. Pedersen PL. Warburg, me and Hexokinase 2: Multiple discoveries of key molecular events underlying one of cancers' most common phenotypes, the "Warburg Effect", i.e., elevated glycolysis in the presence of oxygen. Journal of bioenergetics and biomembranes. 2007;39(3):211-22.
183. Osthus RC, Shim H, Kim S, Li Q, Reddy R, Mukherjee M, et al. Deregulation of glucose transporter 1 and glycolytic gene expression by c-Myc. The Journal of biological chemistry. 2000;275(29):21797-800.
184. Gordan JD, Thompson CB, Simon MC. HIF and c-Myc: sibling rivals for control of cancer cell metabolism and proliferation. Cancer cell. 2007;12(2):108-13.
185. Wise DR, DeBerardinis RJ, Mancuso A, Sayed N, Zhang XY, Pfeiffer HK, et al. Myc regulates a transcriptional program that stimulates mitochondrial glutaminolysis and leads to glutamine addiction. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(48):18782-7.
186. Matoba S, Kang JG, Patino WD, Wragg A, Boehm M, Gavrilova O, et al. p53 regulates mitochondrial respiration. Science. 2006;312(5780):1650-3.
187. Bensaad K, Tsuruta A, Selak MA, Vidal MN, Nakano K, Bartrons R, et al. TIGAR, a p53-inducible regulator of glycolysis and apoptosis. Cell. 2006;126(1):107-20.
188. Kaelin WG. The von Hippel-Lindau tumor suppressor protein: roles in cancer and oxygen sensing. Cold Spring Harbor symposia on quantitative biology. 2005;70:159-66.
189. Semenza GL. HIF-1: upstream and downstream of cancer metabolism. Current opinion in genetics & development. 2010;20(1):51-6.
190. Lunt SY, Vander Heiden MG. Aerobic glycolysis: meeting the metabolic requirements of cell proliferation. Annual review of cell and developmental biology. 2011;27:441-64.
191. Semenza GL. Hypoxia-inducible factor 1 (HIF-1) pathway. Science's signal transduction knowledge environment. 2007;2007(407):cm8.
192. Semenza GL. Regulation of metabolism by hypoxia-inducible factor 1. Cold Spring Harbor symposia on quantitative biology. 2011;76:347-53.
193. Kim JW, Tchernyshyov I, Semenza GL, Dang CV. HIF-1-mediated expression of pyruvate dehydrogenase kinase: a metabolic switch required for cellular adaptation to hypoxia. Cell metabolism. 2006;3(3):177-85.
194. Papandreou I, Cairns RA, Fontana L, Lim AL, Denko NC. HIF-1 mediates adaptation to hypoxia by actively downregulating mitochondrial oxygen consumption. Cell metabolism. 2006;3(3):187-97.
195. Jones NP, Schulze A. Targeting cancer metabolism--aiming at a tumour's sweet-spot. Drug discovery today. 2012;17(5-6):232-41.
196. Laughner E, Taghavi P, Chiles K, Mahon PC, Semenza GL. HER2 (neu) signaling increases the rate of hypoxia-inducible factor 1alpha (HIF-1alpha) synthesis: novel

mechanism for HIF-1-mediated vascular endothelial growth factor expression. Molecular and cellular biology. 2001;21(12):3995-4004.
197. Simon MC. Mitochondrial reactive oxygen species are required for hypoxic HIF alpha stabilization. Advances in experimental medicine and biology. 2006;588:165-70.
198. Hes FJ, Hoppener JW, Luijt RB, Lips CJ. Von hippel-lindau disease. Hereditary cancer in clinical practice. 2005;3(4):171-8.
199. Pollard PJ, Wortham NC, Tomlinson IP. The TCA cycle and tumorigenesis: the examples of fumarate hydratase and succinate dehydrogenase. Annals of medicine. 2003;35(8):632-9.
200. Selak MA, Armour SM, MacKenzie ED, Boulahbel H, Watson DG, Mansfield KD, et al. Succinate links TCA cycle dysfunction to oncogenesis by inhibiting HIF-alpha prolyl hydroxylase. Cancer cell. 2005;7(1):77-85.
201. Gottlieb E, Tomlinson IP. Mitochondrial tumour suppressors: a genetic and biochemical update. Nature reviews Cancer. 2005;5(11):857-66.
202. Pouyssegur J, Dayan F, Mazure NM. Hypoxia signalling in cancer and approaches to enforce tumour regression. Nature. 2006;441(7092):437-43.
203. Hsu PP, Sabatini DM. Cancer cell metabolism: Warburg and beyond. Cell. 2008;134(5):703-7.
204. Swietach P, Vaughan-Jones RD, Harris AL. Regulation of tumor pH and the role of carbonic anhydrase 9. Cancer metastasis reviews. 2007;26(2):299-310.
205. Fischer K, Hoffmann P, Voelkl S, Meidenbauer N, Ammer J, Edinger M, et al. Inhibitory effect of tumor cell-derived lactic acid on human T cells. Blood. 2007;109(9):3812-9.
206. Gatenby RA, Gillies RJ. Why do cancers have high aerobic glycolysis? Nature reviews Cancer. 2004;4(11):891-9.
207. Pastorino JG, Hoek JB. Regulation of hexokinase binding to VDAC. Journal of bioenergetics and biomembranes. 2008;40(3):171-82.
208. Pfeiffer T, Schuster S, Bonhoeffer S. Cooperation and competition in the evolution of ATP-producing pathways. Science. 2001;292(5516):504-7.
209. Finley LW, Zhang J, Ye J, Ward PS, Thompson CB. SnapShot: cancer metabolism pathways. Cell metabolism. 2013;17(3):466- e2.
210. Fritz V, Fajas L. Metabolism and proliferation share common regulatory pathways in cancer cells. Oncogene. 2010;29(31):4369-77. Epub 2010/06/02.
211. Tong X, Zhao F, Thompson CB. The molecular determinants of de novo nucleotide biosynthesis in cancer cells. Current opinion in genetics & development. 2009;19(1):32-7.
212. Christofk HR, Vander Heiden MG, Harris MH, Ramanathan A, Gerszten RE, Wei R, et al. The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. Nature. 2008;452(7184):230-3.
213. Mazurek S, Boschek CB, Hugo F, Eigenbrodt E. Pyruvate kinase type M2 and its role in tumor growth and spreading. Seminars in cancer biology. 2005;15(4):300-8.
214. Gupta V, Bamezai RN. Human pyruvate kinase M2: a multifunctional protein. Protein science. 2010;19(11):2031-44.
215. Barger JF, Plas DR. Balancing biosynthesis and bioenergetics: metabolic programs in oncogenesis. Endocrine-related cancer. 2010;17(4):R287-304.
216. Wise DR, Ward PS, Shay JE, Cross JR, Gruber JJ, Sachdeva UM, et al. Hypoxia promotes isocitrate dehydrogenase-dependent carboxylation of alpha-ketoglutarate to citrate to support cell growth and viability. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(49):19611-6.
217. Metallo CM, Gameiro PA, Bell EL, Mattaini KR, Yang J, Hiller K, et al. Reductive glutamine metabolism by IDH1 mediates lipogenesis under hypoxia. Nature. 2012;481(7381):380-4.
218. Swinnen JV, Brusselmans K, Verhoeven G. Increased lipogenesis in cancer cells: new players, novel targets. Current opinion in clinical nutrition and metabolic care. 2006;9(4):358-65.
219. Menendez JA, Lupu R. Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. Nature reviews Cancer. 2007;7(10):763-77.

220. Kami K, Fujimori T, Sato H, Sato M, Yamamoto H, Ohashi Y, et al. Metabolomic profiling of lung and prostate tumor tissues by capillary electrophoresis time-of-flight mass spectrometry. Metabolomics. 2013;9(2):444-53.
221. Wen T, Gao L, Wen Z, Wu C, Tan CS, Toh WZ, et al. Exploratory investigation of plasma metabolomics in human lung adenocarcinoma. Molecular bioSystems. 2013;9(9):2370-8.
222. Miyagi Y, Higashiyama M, Gochi A, Akaike M, Ishikawa T, Miura T, et al. Plasma free amino acid profiling of five types of cancer patients and its application for early detection. PloS one. 2011;6(9):e24143.
223. Shingyoji M, Iizasa T, Higashiyama M, Imamura F, Saruki N, Imaizumi A, et al. The significance and robustness of a plasma free amino acid (PFAA) profile-based multiplex function for detecting lung cancer. BMC cancer. 2013;13:77.
224. Miyamoto S, Taylor SL, Barupal DK, Taguchi A, Wohlgemuth G, Wikoff WR, et al. Systemic Metabolomic Changes in Blood Samples of Lung Cancer Patients Identified by Gas Chromatography Time-of-Flight Mass Spectrometry. Metabolites. 2015;5(2):192-210.
225. Maeda J, Higashiyama M, Imaizumi A, Nakayama T, Yamamoto H, Daimon T, et al. Possibility of multivariate function composed of plasma amino acid profiles as a novel screening index for non-small cell lung cancer: a case control study. BMC cancer. 2010;10:690.
226. Fahrmann JF, Kim K, DeFelice BC, Taylor SL, Gandara DR, Yoneda KY, et al. Investigation of Metabolomic Blood Biomarkers for Detection of Adenocarcinoma Lung Cancer. Cancer epidemiology, biomarkers & prevention. 2015.
227. Hori S, Nishiumi S, Kobayashi K, Shinohara M, Hatakeyama Y, Kotani Y, et al. A metabolomic approach to lung cancer. Lung Cancer. 2011;74(2):284-92.
228. Li Y, Song X, Zhao X, Zou L, Xu G. Serum metabolic profiling study of lung cancer using ultra high performance liquid chromatography/quadrupole time-of-flight mass spectrometry. Journal of chromatography B, Analytical technologies in the biomedical and life sciences. 2014;966:147-53.
229. Mathe EA, Patterson AD, Haznadar M, Manna SK, Krausz KW, Bowman ED, et al. Noninvasive urinary metabolomic profiling identifies diagnostic and prognostic markers in lung cancer. Cancer research. 2014;74(12):3259-70.
230. Chen W, Zu Y, Huang Q, Chen F, Wang G, Lan W, et al. Study on metabonomic characteristics of human lung cancer using high resolution magic-angle spinning 1H NMR spectroscopy and multivariate data analysis. Magnetic resonance in medicine. 2011;66(6):1531-40.
231. Duarte IF, Rocha CM, Barros AS, Gil AM, Goodfellow BJ, Carreira IM, et al. Can nuclear magnetic resonance (NMR) spectroscopy reveal different metabolic signatures for lung tumours? Virchows Archiv. 2010;457(6):715-25.
232. Rocha CM, Carrola J, Barros AS, Gil AM, Goodfellow BJ, Carreira IM, et al. Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of blood plasma. Journal of proteome research. 2011;10(9):4314-24.
233. Carrola J, Rocha CM, Barros AS, Gil AM, Goodfellow BJ, Carreira IM, et al. Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of urine. Journal of proteome research. 2011;10(1):221-30.
234. Bothwell JH, Griffin JL. An introduction to biological nuclear magnetic resonance spectroscopy. Biological reviews of the Cambridge Philosophical Society. 2011;86(2):493-510.
235. Carr HY, Purcell EM. Effects of diffusion on free precession in nuclear magnetic resonance experiments. Physical Review. 1954;94(3):630.
236. Meiboom S, Gill D. Modified spin-echo method for measuring nuclear relaxation times. Rev Sci Instrum. 1958;29(8):688-91.
237. Edwards JC. Principles of NMR. 1998-2008 [12/08/2015]; Available from: http://www.process-nmr.com/nmr1.htm.
238. Barron AR, Boyd J. NMR Spin Coupling. 2014 [12/08/2015]; Available from: http://cnx.org/contents/ba27839d-5042-4a40-afcf-c0e6e39fb454@20.16:58.
239. Eriksson L, Byrne T, Johansson E, Trygg J, Vikstrom C. Multi-and megavariate data analysis. Sweden: MKS Umetrics AB; 2006.

240. Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laborary Systems. 2001;58(2):109-30.
241. Wiklund S, Johansson E, Sjostrom L, Mellerowicz EJ, Edlund U, Shockcor JP, et al. Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. Analytical chemistry. 2008;80(1):115-22.

CHAPTER 2

Optimization of the ^{1}H-NMR analysis protocol by spiking human blood plasma with known metabolites

Based on:

Evelyne Louis*, Liene Bervoets*, Gunter Reekmans, Eric De Jonge, Liesbet Mesotten, Michiel Thomeer, Peter Adriaensens. Phenotyping human blood plasma by ^{1}H-NMR: a robust protocol based on metabolite spiking and its evaluation in breast cancer. Metabolomics (2015), 11:225-236.

*These authors contributed equally to this work

Abstract

This study reports an accurate assignment of the resonance signals present in ^{1}H-NMR spectra of human blood plasma. Hereto, blood plasma was spiked with 37 different metabolites in relevant concentrations since reported chemical shift values show quite some variability depending on the biofluid under study and the applied experimental measuring conditions. The resulting information was used to rationally divide the ^{1}H-NMR spectrum in 110 well-defined integration regions for application in metabolomics. A case-control dataset of 53 breast cancer patients and 52 controls was investigated in order to demonstrate the proof of principle. After removal of noisy variables, i.e. variables exceeding a premised threshold for the coefficient of variation, the groups could be discriminated by OPLS-DA multivariate statistics with a sensitivity and specificity of 83% and 94%, respectively. In addition, the classification was validated in a small but independent cohort. The proposed methodology might pave the way towards a better understanding of disturbances in disease-related biochemical pathways and so, to the clinical relevance of study findings.

Introduction

The application of metabolomics as a tool in the search for new biomarkers has increased enormously over the past few years (1). The field of metabolomics encompasses the comprehensive and simultaneous analysis of small molecules within a given biological matrix, the so-called metabolites (2-4). Metabolites constitute the end products of cellular metabolism and changes in their concentrations reflect alterations in the metabolic phenotype (5, 6). The two major analytical tools used for metabolomics studies are mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy. In order to find patterns in the massive amount of information, the resulting data are subsequently analyzed by multivariate data analyses with the aim to retrieve diagnostic/theranostic information regarding diseases and to identify changes in biochemical pathways (1, 7).

Proton (^{1}H)-NMR spectroscopy enables a non-invasive identification and quantification of metabolites in complex mixtures such as plasma or urine in a single run. It is becoming a key tool for understanding metabolic processes in living systems and subsequently metabolic disorders (8, 9). The widespread use of this technique along with its exceptional capacity to handle complex mixtures made ^{1}H-NMR spectroscopy the preferred technology for launching the field of metabolomics (3, 10). While other techniques such as gas- and liquid chromatography-MS are also increasingly being used in metabolomics, ^{1}H-NMR spectroscopy has still a number of unique advantages. In particular, it is non-invasive, non-biased and easily quantifiable, needs no chemical derivatization, requires little or no sample preparation and permits the identification of novel compounds (11). However, there is still room for scientific improvement as the accurate interpretation of ^{1}H-NMR spectra in terms of metabolite identities and abundances can be challenging, in particular in crowded regions with severe signal overlap (6, 12). In current practice, signal assignment often relies on chemical shift values reported for different matrices or even non-human specimens (13-15). Additionally, most chemical shifts are dependent of the applied experimental measuring conditions such as temperature, ion content, pH and concentration (16-18). Further room for improvement can also be found in the assignment of noisy variables, which can be defined as signals showing unpredictable variation

in intensity from moment to moment and under identical experimental conditions (18). The identification and removal of these noisy variables should be optimized in order not to complicate the multivariate data analysis (19).

This paper describes the identification of the resonance signals observed in ^{1}H-NMR spectra of human blood plasma on the basis of spiking the plasma with 37 different metabolites. The proposed methodology offers a guidance to divide the ^{1}H-NMR spectrum rationally in well-defined integration regions, being the variables for the multivariate data analysis, and will contribute to a better understanding of the (disturbed) biochemical pathways involved in the disease under study. In order to evaluate the proposed analysis platform, it was applied to classify a case-control dataset consisting of 53 breast cancer patients and 52 controls (training cohort) and to examine the predictive accuracy of the classification in an independent validation cohort consisting of 20 breast cancer patients and 20 controls. In addition, noisy variables were identified and their influence on group differentiation is reported.

Material and methods

Ethics statement

The study was conducted in accordance with the ethical rules of the Helsinki Declaration and of Good Clinical Practice. The study protocol was approved by the Medical Ethics Committees of Ziekenhuis Oost-Limburg (Campus Sint-Jan, Genk, Belgium) and Hasselt University (Campus Diepenbeek, Hasselt, Belgium). Informed consent was obtained from all participants prior to their inclusion in the study.

Blood sampling and processing

Fasting venous blood samples were collected in 10 ml lithium-heparin tubes and stored at 4°C within 5 to 10 min. Within 8 h after blood collection, samples were centrifuged at 1,600 g for 15 min and plasma aliquots of 500 µl were transferred into sterile cryovials and stored at -80°C until NMR analysis within six months.

NMR sample preparation and analysis

After thawing, plasma aliquots were centrifuged at 13,000 g for 4 min at 4°C, followed by diluting 200 µl of the supernatant with 600 µl deuterium oxide (D2O, 99.9%) containing 0.3 µg/µl trimethylsilyl-2,2,3,3-tetradeuteropropionic acid (TSP, 98%) as a chemical shift reference. Samples were placed on ice until ^{1}H-NMR analysis. After mixing and transferring into 5 mm NMR tubes, the samples were acclimatized at 21.2°C during 7 min. The ^{1}H-NMR spectra were recorded on a 400 MHz (9.4 Tesla) Inova spectrometer (Agilent Technologies Inc.) at 21.2°C. Slightly T2-weighted spectra were acquired using the Carr-Purcell-Meiboom-Gill pulse sequence (total spin-echo time of 32 ms with an interpulse delay of 0.1 ms), preceded by an initial preparation delay of 0.5 s, followed by 3 s presaturation for water suppression (total relaxation delay of 3.5 s), 6,000 Hz spectral width, an acquisition time of 1.1 s, 13 K complex data points and 96 scans. Each free induction decay was zero-filled to 65 K points and multiplied by a 0.7 Hz exponential line-broadening function prior to Fourier transformation.

Spectral processing

Spectra were phased, baseline corrected and referenced to TSP (δ=0.015 parts per million (ppm)). The ^{1}H-NMR spectra were segmented into 110 variable-sized spectral regions, excluding the water region (4.7-5.2 ppm) and TSP (-0.3-0.3 ppm). Two sections of the ^{1}H-NMR spectrum (between 4.6-4.8 ppm and 3.7-3.85 ppm) were always double checked with respect to the location of the signals. These spectral regions were integrated and normalized relative to the total integrated area of all regions (except water and TSP), resulting in 110 integration values, being the variables for multivariate statistics. So, remark that an increase/decrease in the concentration of a metabolite reflects its relative concentration.

Subjects of the training cohort

A total of 53 untreated women with histologically proven breast cancer of varying stages of disease were included. A venous blood sample was taken at the day of primary surgery. Diagnosis was confirmed by means of a core needle biopsy. The control group consisted of 52 women who were referred to the department of Nuclear Medicine (Ziekenhuis Oost-Limburg, Campus Sint-Jan) for a stress examination of the heart. Exclusion criteria were defined as follows: (1) not fasted for at least 6 h; (2) medication intake at the day of blood sampling and (3) treatment or history of cancer in the preceding 5 years. Sample collection, preparation and NMR analysis was performed as described above.

Subjects of the independent validation cohort

A total of 20 untreated women with histologically proven breast cancer of varying stages of disease were included. A venous blood sample was taken at the day of primary surgery. Diagnosis was confirmed by means of a core needle biopsy. The control group consisted of 20 women who were referred to the department of Nuclear Medicine (Ziekenhuis Oost-Limburg, Campus Sint-Jan) for a stress examination of the heart. Exclusion criteria were identical to those for the case-control dataset. Sample collection, preparation and NMR analysis was performed as described above.

Metabolite spiking of reference plasma

Fasting venous blood was collected from a healthy 44-year-old female, with no family history of breast cancer and a recent negative mammography, and the plasma was obtained and processed as described above, and further referred to as reference plasma. Stock solutions for spiking were prepared by dissolving 1 mg of a known metabolite in 100 µl reference plasma. In a next step, 10 µl stock solution was added to a standard NMR sample (200 µl reference plasma and 600 µl D_2O containing TSP) and subsequently analyzed by ^{1}H-NMR spectroscopy as described above. This procedure was repeated for the 37 metabolites.

Multivariate statistical analyses

All multivariate statistical analyses were performed using SIMCA-P$^+$ (Version 13.0, Umetrics, Umea, Sweden). Identification of intrinsic clusters within the case-control dataset was accomplished via an unsupervised principal component analysis by which also outliers were detected on the basis of a Hotelling's T2 range plot and a distance to model plot. Orthogonal partial least squares discriminant analysis (OPLS-DA) was used to build models (statistical classifiers) to discriminate between breast cancer patients and controls. Models were compared on the basis of the total amount of explained variation (R^2X(cum) and R^2Y(cum)), predictive ability (Q^2(cum)), and the levels of sensitivity (the percentage of breast cancer patients that are actually classified as breast cancer patients) and specificity (the percentage of controls that are actually classified as controls) The predictive accuracy of the model was examined by means of classifying an independent validation cohort consisting of 20 breast cancer patients and 20 controls.

Results and discussion

^{1}H-NMR chemical shift assignment of human blood plasma metabolites via spiking

Reported chemical shift values of metabolites show quite some variability depending on the biofluid under study and the experimental measuring conditions employed (e.g. temperature, pH, ion strength and concentration) (16-18). To minimize these uncertainties in order to define a rational setting of the integration regions in NMR spectra, the metabolite chemical shifts have to be critically determined by spiking the biofluid with known metabolites in relevant concentrations. Although spiking is a commonly used analytical method to accurately identify chemical shifts in a ^{1}H-NMR spectrum (12, 20), the assignment of metabolite signals is often still based on existing literature and databases (13, 15, 21, 22).

Table 2.1 presents the ^{1}H-NMR chemical shift values and J-coupling patterns of metabolites present in human blood plasma as determined via spiking the plasma with known metabolites. The atom numbering of the metabolites follows the IUPAC-IUB nomenclature unless otherwise indicated in the structures included in **Table 2.1**. For the determination of the chemical shift values and J-coupling patterns, 37 aliquots of a reference plasma pool derived from a healthy 44-year-old female, with no family history of breast cancer and a recent negative mammography, were spiked with 37 known metabolites. The resulting information allowed us to rationally define 87 fixed integration regions in the spectrum having an identified metabolite composition. Including 23 additional integration regions emanating from broad lipid signals and non-identified substances, the ^{1}H-NMR spectrum could be divided into 110 well-defined integration regions, being the variables for the statistical OPLS-DA multivariate analysis. **Table 2.2** presents an overview of the start and end values (in ppm) of the 110 fixed integration regions and their contributing metabolites.

Table 2.1. ^{1}H-NMR chemical shifts (δ in ppm) of low molecular weight plasma metabolites and their J-coupling constants (in Hz).

Metabolite	Proton	δ (ppm)	Multiplicity	J (Hz)	Connectivity
Amino acids					
L-Alanine (Ala)	$^{\alpha}CH$	3.790	q	7.2	α-β
(CHEBI:57972)	$^{\beta}CH_3$	1.509	d	7.2	β-α
L-Arginine (Arg)	$^{\alpha}CH$	3.690	t	6.1	α-β; α-β′
(CHEBI:32682)	$^{\beta}CH_2$	1.700	m	-	-
	$^{\gamma}CH_2$	1.902	m	-	-
	$^{\delta}CH_2$	3.266	t	6.9	δ-γ
L-Asparagine (Asn)	$^{\alpha}CH$	3.999	dd	7.8; 4.3	α-β; α-β'
(CHEBI:58048)	$^{\beta}CH_2$	2.845	dd	16.7; 4.3	β-β'; β-α
	-	2.962	dd	16.7; 7.8	β'-β; β'-α
L-Aspartate (Asp)	$^{\alpha}CH$	3.930	dd	8.9; 3.7	α-β; α-β'
(CHEBI:29991)	$^{\beta}CH_2$	2.702	dd	17.5; 3.7	β-β'; β-α
	-	2.850	dd	17.5; 8.9	β'-β; β'-α
L-Cysteine (Cys)	$^{\alpha}CH$	3.973	dd	5.7; 4.3	α-β; α-β'
(CHEBI:35235)	$^{\beta}CH_2$	3.052	dd	14.7; 4.3	β-β'; β-α
		3.112	dd	14.7; 5.7	β'-β; β'-α
L-Glutamine (Gln)	$^{\alpha}CH$	3.786	t	6.2	α-β; α-β′
(CHEBI:58359)	$^{\beta}CH_2$	2.160	m	-	-
	$^{\gamma}CH_2$	2.480	m	-	-
L-Glutamate (Glu)	$^{\alpha}CH$	3.788	dd	7.1; 4.9	α-β; α-β'
(CHEBI:29985)	$^{\beta}CH_2$	2.120	m	-	-
	$^{\gamma}CH_2$	2.388	m	-	-
L-Glycine (Gly)	$^{\alpha}CH_2$	3.586	s	-	-
(CHEBI:57305)					
L-Histidine (His)	$^{\alpha}CH$	4.012	dd	8.0; 4.9	α-β; α-β'
(CHEBI:57595)	$^{\beta}CH_2$	3.150	dd	15.5; 8.0	β-β'; β-α
	-	3.260	dd	15.5; 4.9	β'-β; β′-α
	$^{\gamma}CH$	7.802	s	-	-
	$^{\delta}CH$	7.086	s	-	-
L-Isoleucine (Ile)	$^{\alpha}CH$	3.673	d	4.0	α-β
(CHEBI:58045)	$^{\beta}CH$	1.990	m	-	-
	$^{\gamma}CH_3$	1.015	d	7.0	γ-β
	$^{\delta}CH_2$	1.476	m	-	-
	$^{\epsilon}CH_3$	0.945	t	7.4	ε-δ
L-Leucine (Leu)	$^{\alpha}CH$	3.769	dd	7.0; 1.3	α-β; α-β′
(CHEBI:57427)	$^{\beta}CH_2$	1.742	m	-	-
	$^{\gamma}CH$	1.742	m	-	-
	$^{\delta}CH_3$	0.987	d	4.7	δ-γ
	-	1.003	d	4.7	δ'-γ
L-Lysine (Lys)	$^{\alpha}CH$	3.772	t	6.0	α-β; α-β′
(CHEBI:32551)	$^{\beta}CH_2$	1.928	m	-	-
	$^{\gamma}CH_2$	1.502	m	-	-
	$^{\delta}CH_2$	1.751	p	7.5	γ-δ; δ-ε
	$^{\epsilon}CH_2$	3.060	t	7.5	ε-δ
L-Methionine (Met)	$^{\alpha}CH$	3.875	dd	7.0; 5.4	α-β; α-β'
(CHEBI:57844)	$^{\beta}CH_2$	2.195	m	-	-
	$^{\gamma}CH_2$	2.673	t	7.6	γ-β; γ-β'
	$^{\delta}CH_3$	2.167	s	-	-
L-Phenylalanine (Phe)	$^{\alpha}CH$	3.998	dd	7.7; 5.2	α-β; α-β'
(CHEBI:58095)	$^{\beta}CH_2$	3.140	dd	14.4; 5.2	β-β'; β-α
	-	3.310	dd	14.4; 7.7	β'-β; β′-α
	$^{\gamma}CH$	7.353	d	7.2	γ-δ
	$^{\delta}CH$	7.454	t	7.2	δ-γ; δ-ε
	$^{\epsilon}CH$	7.414	t	7.2	ε-δ
L-Proline (Pro)	$^{\alpha}CH$	4.162	dd	8.9; 6.3	α-β; α-β'
(CHEBI:60039)	$^{\beta}CH_2$	2.382	m	-	-
	$^{\gamma}CH_2$	2.060	m	-	-
	$^{\delta}CH_2$	3.365	t	7.0	δ-γ
	-	3.441	t	7.0	δ'-γ

Table 2.1. continued.

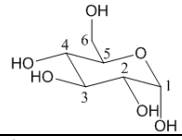
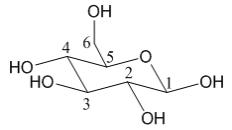
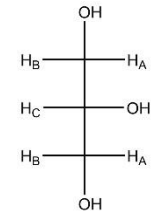
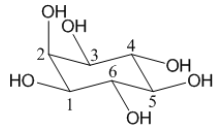
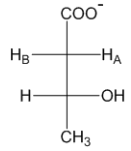
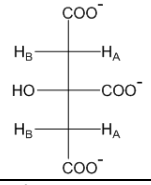
Metabolite	Proton	δ (ppm)	Multiplicity	J (Hz)	Connectivity
Amino acids					
L-Serine (Ser)	$^{\alpha}$CH	3.845	dd	5.6; 4.0	α-β; α-β'
(CHEBI:33384)	$^{\beta}$CH$_2$	3.953	dd	12.2; 5.6	β-β'; β-α
	-	4.012	dd	12.2; 4.0	β'-β; β'-α
L-Threonine (Thr)	$^{\alpha}$CH	3.596	d	4.9	α-β
(CHEBI:57926)	$^{\beta}$CH	4.276	dq	6.6; 4.9	β-α; β-γ
	$^{\gamma}$CH$_3$	1.358	d	6.6	γ-β
L-Tryptophan (Trp)	$^{\alpha}$CH	4.086	dd	8.1; 5.2	α-β; α-β'
(CHEBI:57912)	$^{\beta}$CH$_2$	3.338	dd	15.3; 8.1	β-β'; β-α
	-	3.224	dd	15.3; 5.2	β'-β; β'-α
	$^{\gamma}$CH	7.351	s	-	-
	$^{\delta}$CH	7.770	d	7.8	δ-ε
	$^{\varepsilon}$CH	7.229	t	7.8	ε-δ; ε-ζ
	$^{\zeta}$CH	7.310	t	7.8	ζ-ε; ζ-η
	$^{\eta}$CH	7.570	d	7.8	η-ζ
L-Tyrosine (Tyr)	$^{\alpha}$CH	3.957	dd	7.8; 5.0	α-β; α-β'
(CHEBI:58315)	$^{\beta}$CH$_2$	3.076	dd	14.2; 7.8	β-β'; β-α
	-	3.227	dd	14.2; 5.0	β'-β; β'-α
	$^{\gamma}$CH	6.924	d	8.4	γ-δ
	$^{\delta}$CH	7.222	d	8.4	δ-γ
L-Valine (Val)	$^{\alpha}$CH	3.635	d	4.3	α-β
(CHEBI:57762)	$^{\beta}$CH	2.305	m	-	-
	$^{\gamma}$CH$_3$	1.021	d	7.1	γ-β
	-	1.074	d	7.1	γ'-β
Carbohydrates					
D-glucose					
α-anomer	C1H	5.264	d	3.8	-
(CHEBI:17925)	C2H	3.563	dd	9.8; 3.8	-
	C3H	3.744	t	9.4	-
	C4H	3.439	t	9.4	-
	C5H	3.888	m	-	-
	C6H	3.858	dd	10; 2.2	-
	C6′H	3.792	dd	13.1; 6.3	-
β-anomer	C1H	4.678	d	7.8	-
(CHEBI:15903)	C2H	3.272	t	8.3	-
	C3H	3.518	t	9.2	-
	C4H	3.428	t	9.4	-
	C5H	3.492	m	-	-
	C6H	3.933	dd	12.2; 2.0	-
	C6′H	3.752	dd	12.2; 5.7	-
Glycerol	CH$_C$	3.814	m	-	-
(CHEBI:17754)	CH$_2$: H$_A$	3.591	dd	11.7; 6.6	-
	CH$_2$: H$_B$	3.682	dd	11.7; 4.4	-
Myo-Inositol	C5H	4.090	t	2.9	-
(CHEBI:17268)	C4H+C6H	3.562	dd	9.8; 2.9	-
	C1H+C3H	3.654	t	9.8	-
	C2H	3.306	t	9.3	-

Table 3.1. continued.

Metabolite	Proton	δ (ppm)	Multiplicity	J (Hz)	Connectivity
Organic acids					
Acetate (CHEBI:30089)	CH_3	1.948	s	-	-
Acetoacetate (CHEBI:13705)	CH_2	2.319	s	-	-
	CH_3	3.480	s	-	-
α-ketoglutarate (CHEBI:16810)	CH_2-CO	3.040	t	6.9	-
	CH_2-COO^-	2.470	t	6.9	-
D-β-hydroxybutyrate (CHEBI:10983) 	CH_A	2.400	dd	14.5; 7.3	-
	CH_B	2.300	dd	14.5; 7.3	-
	CH	4.184	m	-	-
	CH_3	1.231	d	6.3	-
Citrate (CHEBI:16947) 	CH_A	2.717	d	15.8	-
	CH_B	2.566	d	15.8	-
L-lactate (CHEBI:16651)	CH	4.138	q	6.9	-
	CH_3	1.354	d	6.9	-
Pyruvate (CHEBI:15361)	CH_3	2.402	s	-	-
Succinate (CHEBI:30031)	CH_2	2.439	s	-	-
Others					
Acetone (CHEBI:15347)	CH_3	2.264	s	-	-
Betaine (CHEBI:17750)	CH_3	3.300	s	-	-
	CH_2	3.939	s	-	-
Choline (CHEBI:133341)	CH_3	3.236	s	-	-
	CH_2	3.554	m	-	-
	CH_2OH	4.098	m	-	-
Creatine (CHEBI:57947)	CH_3	3.068	s	-	-
	CH_2	3.962	s	-	-
Creatinine (CHEBI:16737)	CH_3	3.075	s	-	-
	CH_2	4.087	s	-	-
Methanol (CHEBI:17790)	CH_3	3.402	s	-	-

Chemical shifts are expressed relatively to the singlet resonance of the trimethyl protons of TSP at δ 0.015 ppm and J-coupling patterns are described as: s, singlet; d, doublet; dd, double doublet; dq, double quadriplet; t, triplet; q, quadriplet; p, pentaplet; m, multiplet. Metabolite identifiers from the database of Chemical Entities of Biological Interest (ChEBI) are indicated. The atom numbering of the metabolites follows the IUPAC-IUB nomenclature unless otherwise indicated in the structures included in **Table 2.1.**

Table 2.2. Start and end values (in ppm) of the 110 integrations regions (variables) and their contributing metabolites, defined on the basis of metabolite spiking.

Variable	Start ppm	End ppm	Metabolite
1	7,9500	7,8200	NI
2	7,8200	7,7890	His
3	7,7890	7,7780	NI
4	7,7780	7,7480	His
5	7,7480	7,7200	NI
6	7,6800	7,5920	NI
7	7,5920	7,5480	NI
8	7,4840	7,3620	Phe
9	7,3620	7,3300	Phe, NI
10	7,3300	7,2820	NI
11	7,2820	7,2550	NI
12	7,2550	7,2390	Tyr, NI
13	7,2390	7,2000	Tyr, NI
14	7,1070	7,0656	His
15	6,9430	6,9050	Tyr
16	6,9050	6,8810	NI
17	6,7445	6,7020	NI
18	5,4300	5,2752	Lipids: -C**H**=C**H**- in FAC$^{\#}$
19	5,2752	5,2526	Glucose
20	5,2526	5,2030	C_2H in glycerol backbone of PL and TG$^{\#}$
21	4,6940	4,6620	Glucose
22	4,5560	4,5380	NI
23	4,5380	4,4100	NI
24	4,4100	4,3159	C_1H and C_3H in glycerol backbone of TG$^{\#}$
25	4,3159	4,2332	O-C**H**$_2$-CH_2-$N^+(CH_3)_3$ of PC and SM$^{\#}$, Thr
26	4,2000	4,1885	Pro, β-hydroxybutyrate
27	4,1885	4,1750	Pro, β-hydroxybutyrate, lactate
28	4,1750	4,1260	C_1H and C_3H in glycerol backbone of PL and TG$^{\#}$, lactate
29	4,1260	4,1110	NI
30	4,1110	4,1032	NI
31	4,1032	4,0700	Creatinine
32	4,0700	4,0570	NI
33	4,0570	4,0310	His, Ser
34	4,0310	4,0136	Asn, His, Phe, Ser
35	4,0136	4,0010	C_3H_2 in glycerol backbone of PL$^{\#}$, Asn, His, Phe, Ser
36	4,0010	3,9810	C_3H_2 in glycerol backbone of PL$^{\#}$, Asn, His, Phe, Ser
37	3,9810	3,9590	Creatine, Asn, His, Tyr, Ser
38	3,9590	3,8330	Glucose, Asp, Met, Ser, Tyr
39	3,8330	3,8100	Glucose, Ala, Ser
40	3,8100	3,7956	Glucose, Ala, Gln, Glu
41	3,7956	3,7820	Glucose, Ala, Gln, Glu, Leu, Lys
42	3,7820	3,7550	Glucose, Ala, Gln, Glu, Leu, Lys
43	3,7550	3,7390	Glucose, Ala, Leu
44	3,7390	3,7141	Glucose
45	3,7141	3,6680	O-CH_2-**CH_2**-$N^+(CH_3)_3$ of PC and SM$^{\#}$, glycerol, Ile
46	3,6680	3,6500	Glycerol
47	3,6500	3,6376	Glycerol, Val
48	3,6376	3,6240	Val
49	3,6240	3,6097	Thr
50	3,6097	3,5914	Thr
51	3,5914	3,5649	Glucose, glycerol, Gly, Thr
52	3,5649	3,5510	Glucose
53	3,5510	3,5360	Glucose, acetoacetate, Pro
54	3,5360	3,3980	Glucose, acetoacetate, Pro

Table 2.2. continued.

Variable	Start ppm	End ppm	Metabolite
55	3,3980	3,3765	Methanol, NI
56	3,3765	3,3430	Pro
57	3,3430	3,3230	Phe, Pro
58	3,3230	3,2186	O-CH_2-CH_2-N^+($C\mathbf{H}_3$)$_3$ of PC and SM$^{\#}$, glucose, His, Phe, Tyr
59	3,2186	3,1930	Tyr
60	3,1930	3,1760	NI
61	3,1760	3,1462	NI
62	3,1462	3,1090	His, Phe
63	3,1090	3,0860	Lys, Tyr
64	3,0860	3,0716	Creatinine, Lys, Tyr
65	3,0716	3,0640	Creatinine, creatine, Lys
66	3,0640	2,9950	α-ketoglutarate, Lys
67	2,9950	2,8860	Lipids: =CH-C$\mathbf{H}_2$-CH= in FAC$^{\#}$, Asn
68	2,8860	2,8550	Lipids: =CH-C$\mathbf{H}_2$-CH= in FAC$^{\#}$, Asn, Asp
69	2,8550	2,7500	Lipids: =CH-C$\mathbf{H}_2$-CH= in FAC$^{\#}$, Asn, Asp
70	2,7500	2,7360	Citrate, Asp
71	2,7360	2,6600	Citrate, Asp, Met
72	2,6600	2,6300	Met
73	2,5960	2,5340	Citrate
74	2,5340	2,5150	NI
75	2,5150	2,4920	Gln
76	2,4920	2,4500	α-ketoglutarate, β-hydroxybutyrate, Gln
77	2,4500	2,4324	α-ketoglutarate, β-hydroxybutyrate, succinate
78	2,4324	2,4148	β-hydroxybutyrate, Pro
79	2,4148	2,4050	β-hydroxybutyrate, Pro, Glu
80	2,4050	2,3990	Pyruvate, Pro, Glu
81	2,3990	2,3640	β-hydroxybutyrate, Pro, Glu
82	2,3640	2,3500	β-hydroxybutyrate, Pro, Glu
83	2,3500	2,3380	β-hydroxybutyrate, Pro, Val
84	2,3380	2,3170	β-hydroxybutyrate, acetoacetate, Pro, Val
85	2,3170	2,3040	β-hydroxybutyrate, acetoacetate, Val
86	2,3040	2,2915	Lipids: -C$\mathbf{H}_2$-C=O or -C$\mathbf{H}_2$-CH=CH- in FAC$^{\#}$, β-hydroxybutyrate, Val
87	2,2915	2,2690	Lipids: -C$\mathbf{H}_2$-C=O or -C$\mathbf{H}_2$-CH=CH- in FAC$^{\#}$, Met, Val
88	2,2690	2,2300	Lipids: -C$\mathbf{H}_2$-C=O or -C$\mathbf{H}_2$-CH=CH- in FAC$^{\#}$, aceton, Met, Val
89	2,2180	2,1970	Glu, Met
90	2,1970	2,1230	Gln, Glu, Met, Pro
91	2,1230	1,9720	Lipids: -C$\mathbf{H}_2$-CH=CH- in FAC$^{\#}$, C$\mathbf{H}_3$ of NAG$^{\$}$, Glu, Ile, Met, Pro
92	1,9720	1,9240	Acetate, Ile, Lys
93	1,9240	1,8800	Ile, Lys
94	1,8060	1,6860	Leu, Lys
95	1,6860	1,5600	Lipids: -C$\mathbf{H}_2$-CH_2-C=O or -C$\mathbf{H}_2$-CH_2-CH=CH- in FAC$^{\#}$, Lys
96	1,5400	1,4900	Ala, Ile, Lys
97	1,4900	1,4200	Ile, Leu, Lys
98	1,4200	1,3740	Lactate
99	1,3740	1,3450	Lactate, Thr
100	1,3450	1,2458	Lipids: CH_3-(C$\mathbf{H}_2$)$_{n-}$ in FAC$^{\#}$, Ile, Thr
101	1,2458	1,2180	β-hydroxybutyrate, Ile
102	1,2180	1,1300	NI
103	1,0930	1,0610	Val
104	1,0610	1,0400	Ile
105	1,0400	1,0220	Ile, Val
106	1,0220	1,0020	Ile, Leu, Val
107	1,0020	0,9860	Ile, Leu
108	0,9860	0,9760	Ile, Leu

Table 2.2. continued.

Variable	Start ppm	End ppm	Metabolite
109	0,9760	0,9660	Ile
110	0,9660	0,8000	Lipids: C**H**$_3$-(CH$_2$)$_n$- in FAC$^{\#}$

Some assignments are slightly adapted based on additional information obtained from spiking experiments on a 900 MHz NMR spectrometer (see Chapter 3). $^{\#}$The assignment of the lipid signals is based on literature (9, 23, 24). $^{\$}$The assignment of signals of N-acetylated glycoproteins is based on literature (25). Noisy variables with a coefficient of variation larger than the optimal threshold, i.e. 15%, are indicated in grey (see below). Amino acids are presented by their 3-letter code. FAC: fatty acid chain; NAG: N-acetylated glycoproteins; NI: non-identified; PC: phosphatidylcholine; PL: phospholipids; ppm: parts per million, SM: sphingomyelin; TG: triglycerides.

Application of the analysis procedure in a case-control training and validation study

In order to evaluate the analysis procedure, a multivariate statistical analysis was performed on a case-control dataset consisting of 53 breast cancer patients (female, mean age: 58 ± 12 years) and 52 controls (female, mean age: 63 ± 14 years). After identifying and excluding five outliers (4 controls and 1 breast cancer patient), a statistical classifier was built by means of OPLS-DA. By using all variables (the normalized integration values of the fixed 110 integration regions), the model was able to discriminate between breast cancer patients and controls with a sensitivity and specificity of 79% and 94%, respectively. This model explains 68.3% of the variation observed within the groups (R^2X(cum)) and 52.6% of the variation observed between the groups (R^2Y(cum)) (**Figure 2.2A** and **Table 2.3**). In addition, the predictive accuracy of this model was evaluated by classifying an independent cohort consisting of 20 women with breast cancer (mean age: 57 ± 11 years) and 20 controls (female, mean age: 66 ± 12 years), resulting in a correct classification of 80% of the breast cancer patients and 70% of the controls. Subject characteristics of the case-control training and validation cohorts, as well as histopathology, stage and hormone receptor status of the breast tumors are presented in **Table 2.4**.

Table 2.3. The number of latent variables, the total explained variation in X and Y ($R^2X(cum)$ and $R^2Y(cum)$), predictive ability ($Q^2(cum)$, and sensitivity and specificity levels for OPLS-DA models constructed with a decreasing number of variables.

	Threshold for CV					
	None	**30%**	**25%**	**20%**	**15%**	**10%**
Remaining variables	110	89	83	80	70	59
LV	3	3	3	3	3	2
$R^2X(cum)$	0.683	0.695	0.697	0.699	0.706	0.674
$R^2Y(cum)$	0.526	0.523	0.522	0.516	0.515	0.263
$Q^2(cum)$	0.416	0.418	0.417	0.414	0.419	0.078
Sensitivity, %	79	81	81	81	83	77
Specificity, %	94	94	94	94	94	67

A decreasing threshold limit for the coefficient of variation, from 30% to 10%, was used as an exclusion criterion for noisy variables. CV: coefficient of variation; LV: latent variable.

Table 2.4. Subject characteristics of the case-control training and validation cohorts, together with histopathology, stage and hormone receptor status of the breast tumors.

	Training cohort		Validation cohort	
	BC	**C**	**BC**	**C**
Number of female subjects, n	53	52	20	20
Age, yrs (range)	59 ± 12 (35–85)	63 ± 14 (23–84)	57 ± 11 (40–78)	66 ± 12 (49–89)
BMI, kg/m^2 (range)	25.9 ± 5.2 (18.6–43.7)	26.8 ± 5.6 (16.5-41.1)	25.3 ± 4.5 (16.2-35.0)	27.3 ± 4.9 (18.7-35.9)
Smoking habits				
Non-smoker, n (%)	40 (75)	42 (81)	18 (90)	18 (90)
Smoker, n (%)	10 (19)	10 (19)	2 (10)	2 (10)
Unknown, n (%)	3 (6)	0 (0)	0 (0)	0 (0)
Histopathology				
IDA, n (%)	46 (87)		15 (75)	
ILA, n (%)	6 (11)		5 (25)	
DCIS, n (%)	1 (2)		0 (0)	
Stage				
0, n (%)	1 (2)		0 (0)	
I, n (%)	18 (34)		10 (50)	
IIA, n (%)	21 (40)		7 (35)	
IIB, n (%)	6 (11)		0 (0)	
IIIA, n (%)	4 (8)		3 (15)	
IIIC, n (%)	3 (6)		0 (0)	
Receptor status				
ER				
Positive, n (%)	45 (82)		17 (85)	
Negative, n (%)	10 (18)		3 (15)	
PR				
Positive, n (%)	37 (67)		15 (75)	
Negative, n (%)	18 (23)		5 (25)	

Data are presented as means ± standard deviation (SD) and range. BC: breast cancer patients; BMI: body mass index; C: controls; DCIS; ductal carcinoma in situ; ER: estrogen receptor; IDA: invasive ductal adenocarcinoma; ILA: invasive lobular adenocarcinoma; PR: progesterone receptor.

Identification and influence of noisy variables

Although the discriminative power of the model built by using all variables is already fairly good, we have to keep in mind that experimental data are always subjected to a certain degree of noise. In a fixed NMR set-up (with a defined magnetic field strength and probe-head), the signal-to-noise ratio (S/N) mainly depends on the chosen plasma concentration and the number of accumulations (number of excitations by the radiofrequency-pulse) to acquire the spectrum. Although a higher plasma concentration will result in a higher S/N, starting from a certain level it will unfortunately also reduce the spectral resolution (an increase of line-broadening due to a decrease of the T_2-relaxation decay times with increasing concentration). So, once the plasma concentration is chosen in the analysis protocol (in our study 200 µl plasma diluted with 600 µl D_2O) the S/N will be determined by the number of accumulations. However, also a reasonable time

frame per sample has to be taken into account, especially if a high sample throughput is desirable as for metabolomics. In our analysis protocol, we have chosen to acquire 96 accumulations per sample, resulting in an acquisition time of 7'44''. The fact that some degree of noise can never be excluded from experimental data and that noisy variables can complicate the multivariate data analysis, explains our effort to implement a protocol that allows to identify and classify noisy variables and subsequently to study the impact of their removal from the dataset on group differentiation.

In order to identify noisy variables, the following three series of ^{1}H-NMR experiments were performed by using three aliquots of the reference plasma pool: (1) series A: five consecutive measurements on a sample prepared from a first aliquot (after initial spectrometer conditioning, i.e. locking, shimming and optimization of water suppression); (2) series B: five similar measurements at another time point using a second aliquot and (3) series C: five measurements using a third aliquot but with full spectrometer conditioning before each acquisition, i.e. sample insertion, locking, shimming and optimization of water suppression. For each series of acquisitions, the data analysis was accomplished by three independent researchers familiar with ^{1}H-NMR metabolomics. This results, for each of the series, in 15 normalized integration values for all (110) variables by means of which the coefficients of variation (%) were determined. For the three series A-C, **Figure 2.1A-C** presents a plot of the resulting coefficients of variation for all variables which are, for clarity, divided into 3 groups on the basis of their mean normalized integration value, i.e. between 1-20 (60 variables); 20-100 (40 variables) and 100-1000 (10 variables). Based on these plots, five visual thresholds were defined, i.e. coefficients of variation below 10, 15, 20, 25 and 30%. If a variable exceeded the predefined threshold in at least two of the three series it was assigned as 'noisy'. A variable for which the coefficient of variation was higher than the threshold in only one of the series was classified as noisy only if the variation was not the result of a single outlier. Remark that these threshold settings might be dependent on the NMR set-up, and consequently can be even lower for higher magnetic field strengths or when cryo-probes are used.

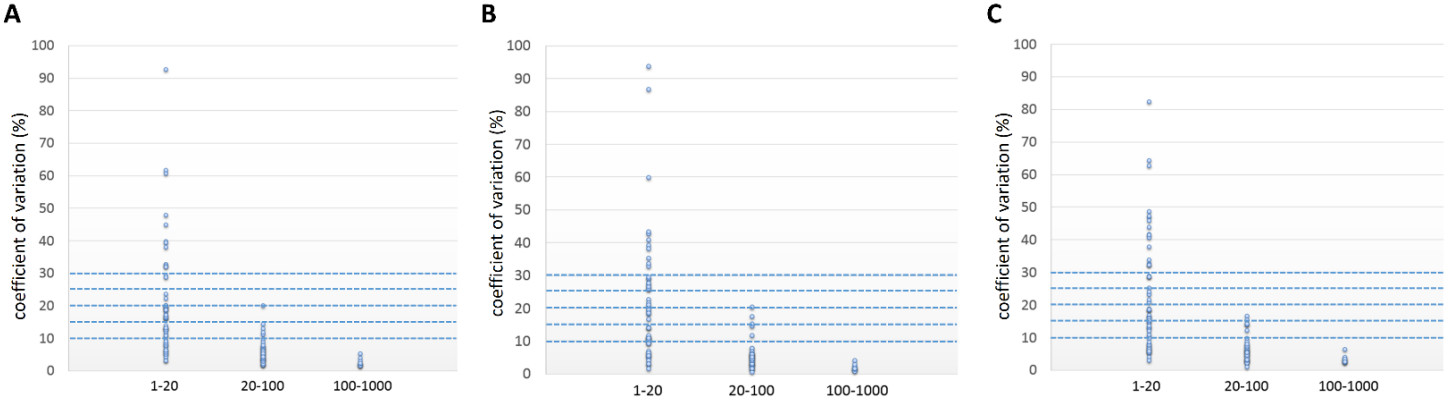


Figure 2.1. Coefficient of variation (%) for the 110 variables which are divided into three groups on the basis of their mean normalized integration values, i.e. those with a mean normalized integration value between 1-20, between 20-100 and between 100-1000. In order to determine the coefficients of variation, three series (A-C) of NMR measurements were accomplished using samples prepared from three different aliquots of a reference plasma pool: (A) 5 consecutive measurements (after initial spectrometer conditioning, i.e. locking, shimming and optimization of water suppression); (B) 5 similar measurements at another time point and (C) 5 measurements with full spectrometer conditioning before the acquisition, i.e. sample insertion, locking, shimming and optimization of water suppression. Data analysis of each series by three independent researchers resulted in 15 normalized integration values for each variable, from which the coefficients of variation (%) were determined.

By means of this information, the influence of noisy variables on the classification was investigated. Hereto, OPLS-DA classifiers were built and compared after removal of the noisy variables according the above defined criteria. An overview of the results is presented in **Table 2.3** and **Figure 2.2**. **Table 2.3** shows that removing variables with a coefficient of variation above 15% has a beneficial effect on group differentiation. In more detail, deleting 30 variables with a coefficient of variation above 20% (**Figure 2B-D**) results in an increase of the sensitivity with 2% (to 81%) as compared to the classifier constructed with all 110 variables (**Figure 2.2A** and **Table 2.3**) while the specificity remains unchanged. If the allowed coefficient of variation was further reduced to 15%, 70 variables remained and the sensitivity further increased with 2% (to 83%), again with no change in specificity (**Figure 2.2E** and **Table 2.3**). This is an indication that the discriminating power of the remaining 70 variables, and so constituting metabolites, is very strong. However, when the allowed coefficient of variation was further reduced to 10%, only 59 variables remained and the discriminative power of the OPLS-DA model declined strongly to a specificity of 67% and a sensitivity of 77% (**Figure 2.2F** and **Table 2.3**). This loss of discriminative power also becomes clear from the other model parameters: while $R^2X(cum)$, $R^2Y(cum)$ and $Q^2(cum)$ improve or remain quasi stable in going from a threshold setting of 30% to 15%, they strongly decline upon setting the threshold to 10%. This severe loss of discriminative power can be explained by the fact that only about half of the variables (59 out of 110) still contribute to the model and indicates that the 11 variables, which were additionally removed in going from a threshold setting of 15% to 10%, are important for group differentiation. All above indicates that a threshold of 15% for the coefficient of variation is optimal for the identification and exclusion of noisy variables. In a next step, the influence of removing these noisy variables on the predictive accuracy of the classification was examined in an independent validation cohort. Although the size of the cohort is still limited, removal of the noisy variables leads to an increase in sensitivity from 80% to 85% while the specificity slightly drops from 70% to 65%. Moreover, and as also observed for the training cohort, a further lowering of the threshold for the coefficient of variation to 10% results in a strong decline in discriminative power.

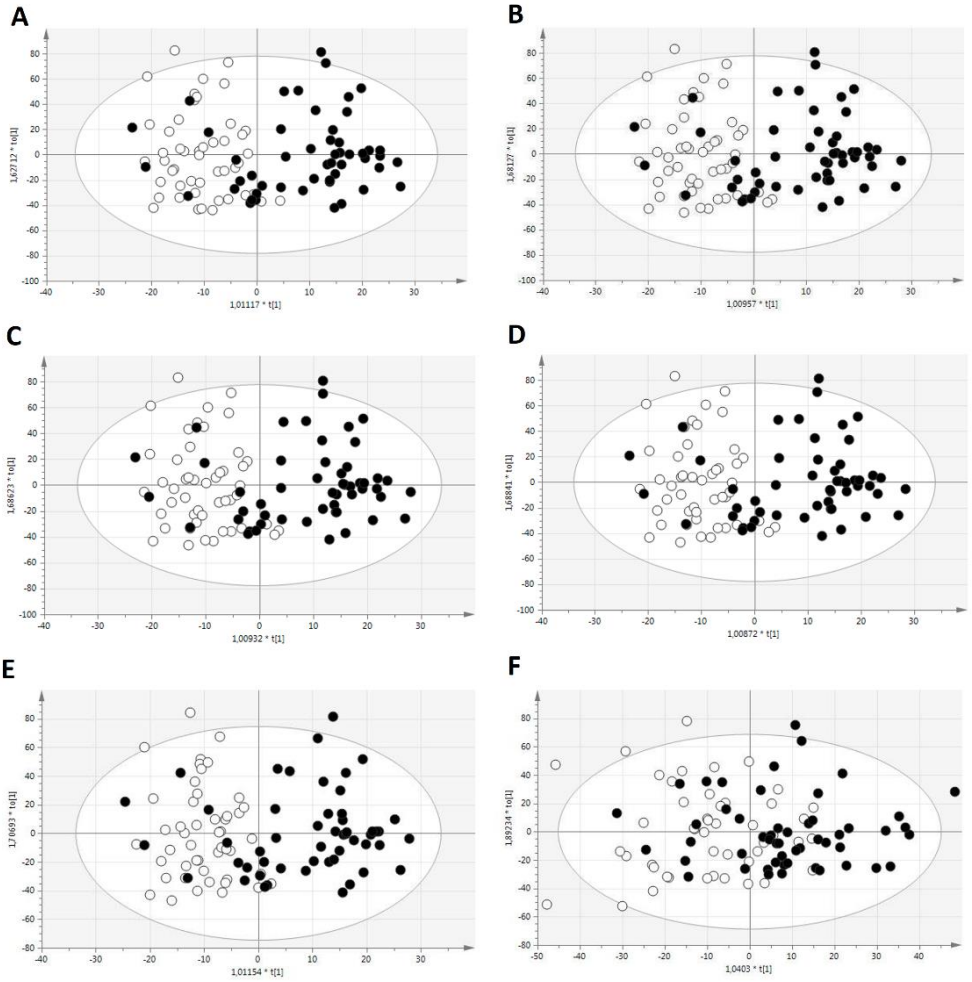


Figure 2.2. OPLS-DA score plots of the models built with all 110 variables (A), with 89 variables (B), with 83 variables (C), with 80 variables (D), with 70 variables (E) and with 59 variables (F). Variables exceeding a predefined threshold for the coefficient of variation, i.e. 30% (B), 25% (C), 20% (D), 15% (E) and 10% (F) were assigned as 'noisy' and excluded for the building of the classifier. Breast cancer patients are marked as ● and controls as ○.

To double check that none of the 40 excluded variables are highly important for group differentiation, they were also excluded one by one, i.e. 40 OPLS-DA classifiers were built with the remaining 109 variables, and evaluated. The resulting classifiers performed evenly or even slightly better than the one constructed with all 110 variables. This result confirms that the 40 variables with a coefficient of variation above 15% can be defined as noisy and that they are not important for group differentiation. Important to notice is that these noisy variables, marked in gray in **Table 2.2**, all represent very low intensity signals. Among them, 11 are not identified, 13 are specific for a single metabolite, and 16 are specific for 2 or more metabolites. Despite of this reduction in variables, all metabolites are still represented, either via a unique signal or via a composite signal.

Conclusions

By spiking 37 aliquots of a reference human plasma pool with known metabolites, the chemical shift values of these metabolites were determined with high accuracy in the biofluid under study. Hereby, the ^{1}H-NMR spectrum of human blood plasma could be rationally divided into 110 well-defined integration regions, paving the way towards a better understanding of disturbances in the underlying biochemical pathways. After removal of noisy variables, defined as variables exceeding a premised threshold for the coefficient of variation, the proposed methodology allowed to discriminate a case-control dataset of 53 breast cancer patients and 52 controls with a sensitivity and specificity of respectively 83% and 94% using OPLS-DA multivariate statistics. In addition, we applied the methodology to a small but independent validation cohort for which the same trends are observed.

References

1. Craig A, Cloarec O, Holmes E, Nicholson JK, Lindon JC. Scaling and normalization effects in NMR spectroscopic metabonomic data sets. Analytical chemistry. 2006;78(7):2262-7.
2. Lindon JC, Nicholson JK. Spectroscopic and statistical techniques for information recovery in metabonomics and metabolomics. Annu Rev Anal Chem. 2008;1:45-69.
3. Nicholson JK, Wilson ID. Opinion: understanding 'global' systems biology: metabonomics and the continuum of metabolism. Nature reviews Drug discovery. 2003;2(8):668-76.
4. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, et al. HMDB: the Human Metabolome Database. Nucleic acids research. 2007;35:D521-6.
5. Fiehn O. Metabolomics--the link between genotypes and phenotypes. Plant molecular biology. 2002;48(1-2):155-71.
6. O'Connell TM. Recent advances in metabolomics in oncology. Bioanalysis. 2012;4(4):431-51.
7. Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica; the fate of foreign compounds in biological systems. 1999;29(11):1181-9.
8. Engelke UF, Liebrand-van Sambeek ML, de Jong JG, Leroy JG, Morava E, Smeitink JA, et al. N-acetylated metabolites in urine: proton nuclear magnetic resonance spectroscopic study on patients with inborn errors of metabolism. Clinical chemistry. 2004;50(1):58-66.
9. Oostendorp M, Engelke UF, Willemsen MA, Wevers RA. Diagnosing inborn errors of lipid metabolism with proton nuclear magnetic resonance spectroscopy. Clinical chemistry. 2006;52(7):1395-405.
10. German JB, Hammock BD, Watkins SM. Metabolomics: building on a century of biochemistry to guide human health. Metabolomics. 2005;1(1):3-9.
11. Wishart DS. Quantitative metabolomics using NMR. Trac-Trend Anal Chem. 2008;27(3):228-37.
12. Zheng C, Zhang S, Ragg S, Raftery D, Vitek O. Identification and quantification of metabolites in (1)H NMR spectra by Bayesian model selection. Bioinformatics. 2011;27(12):1637-44.
13. Salek RM, Maguire ML, Bentley E, Rubtsov DV, Hough T, Cheeseman M, et al. A metabolomic comparison of urinary changes in type 2 diabetes in mouse, rat, and human. Physiological genomics. 2007;29(2):99-108.
14. Wang L, Chen J, Chen L, Deng P, Bu Q, Xiang P, et al. 1H-NMR based metabonomic profiling of human esophageal cancer tissue. Molecular cancer. 2013;12:25.
15. Garcia E, Andrews C, Hua J, Kim HL, Sukumaran DK, Szyperski T, et al. Diagnosis of early stage ovarian cancer by 1H NMR metabonomics of serum explored by use of a microflow NMR probe. Journal of proteome research. 2011;10(4):1765-71.
16. Staab JM, O'Connell TM, Gomez SM. Enhancing metabolomic data analysis with Progressive Consensus Alignment of NMR Spectra. BMC bioinformatics. 2010;11:123.
17. Chang D, Weljie A, Newton J. Leveraging latent information in NMR spectra for robust predictive models. Pacific Symposium on Biocomputing. 2007:115-26.
18. Kohl SM, Klein MS, Hochrein J, Oefner PJ, Spang R, Gronwald W. State-of-the art data normalization methods improve NMR-based metabolomic analysis. Metabolomics. 2012;8(Suppl 1):146-60.
19. van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC genomics. 2006;7:142.
20. Yang WJ, Wang YW, Zhou QF, Tang HR. Analysis of human urine metabolites using SPE and NMR spectroscopy. Sci China Ser B. 2008;51(3):218-25.
21. Markley JL, Ulrich EL, Berman HM, Henrick K, Nakamura H, Akutsu H. BioMagResBank (BMRB) as a partner in the Worldwide Protein Data Bank (wwPDB): new

policies affecting biomolecular NMR depositions. Journal of biomolecular NMR. 2008;40(3):153-5.
22. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0--The Human Metabolome Database in 2013. Nucleic acids research. 2013;41:D801-7.
23. Kriat M, Vion-Dury J, Confort-Gouny S, Favre R, Viout P, Sciaky M, et al. Analysis of plasma lipids by NMR spectroscopy: application to modifications induced by malignant tumors. Journal of lipid research. 1993;34(6):1009-19.
24. Papathanasiou A, Kostara C, Cung MT, Seferiadis K, Elisaf M, Bairaktari E, et al. Analysis of the composition of plasma lipoproteins in patients with extensive coronary heart disease using 1H NMR spectroscopy. Hellenic journal of cardiology. 2008;49(2):72-8.
25. Lenz EM, Bright J, Wilson ID, Morgan SR, Nash AF. A 1H NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects. Journal of pharmaceutical and biomedical analysis. 2003;33(5):1103-15.

CHAPTER 3

Advantages and disadvantages of 900 MHz vs. 400 MHz spectrometers for NMR metabolomics

Based on:

Evelyne Louis, Francois-Xavier Cantrelle, Liesbet Mesotten, Gunter Reekmans, Karolien Vanhove, Michiel Thomeer, Guy Lippens, Peter Adriaensens. Metabolic phenotyping of human plasma by NMR at higher magnetic field strengths: advantages and disadvantages of 900 versus 400 MHz. Under review.

Abstract

Accurate identification and quantification of human plasma metabolites and subsequent interpretation of underlying, disease disturbed, biochemical pathways can be challenging in crowded NMR regions with severe signal overlap. Therefore, this study describes metabolite spiking experiments on the basis of which the NMR spectrum can be rationally segmented into well-defined integration regions, and this for magnetic field strengths corresponding to 400 MHz and 900 MHz spectrometers. The integration data of a case-control dataset of 69 lung cancer patients and 74 controls were then used to train a multivariate statistical classification model for the data at both field strengths. In this way, the advantages/disadvantages of high versus medium magnetic field strengths could be evaluated. On the one hand, the discriminative power of the integration data collected at both magnetic field strengths is rather similar, i.e. a sensitivity and specificity of 94% and 97% respectively by means of the 400 MHz data versus 90% and 100% for the 900 MHz data. Taking the housing and the cost of a high-field spectrometer into account, a medium-field 400-600 MHz NMR spectrometer looks to be most appropriate for clinical metabolomics. On the other hand, the increased spectral resolution and signal-to-noise ratio of 900 MHz spectra allow a more accurate delineation of the integration regions, resulting in an increased number of integration regions that represent a single metabolite, being beneficial for the understanding of the underlying biochemical pathways.

Introduction

Metabolomics is a powerful discipline which allows to detect hundreds to thousands of metabolites in biological samples such as plasma or urine (1, 2). The two major high-throughput analytical platforms used for metabolite analysis are mass spectrometry (MS) and proton-nuclear magnetic resonance (^{1}H-NMR) spectroscopy (3). Subsequently, the large amount of resulting data is analyzed by multivariate pattern recognition methods in an effort to reduce the complexity of the data and to recover diagnostic information regarding diseases and to identify disturbed biochemical pathways (4-6). Although ^{1}H-NMR spectroscopy is less sensitive than MS, it offers various distinct advantages. More specifically, it allows to identify and quantify metabolites in biological samples in a non-destructive manner with minimal sample preparation as well as an exceptional analytical reproducibility (± 98%) (3, 7, 8).

The application of ^{1}H-NMR-based metabolomics in the search for cancer biomarkers has increased enormously over the past decade (9-15). Most of these studies used NMR spectrometers with a magnetic field strength between 400 and 600 MHz to study the disturbed metabolism of cancer cells (9, 11, 13-15). However, in the past two years, high-field NMR spectrometers up to 800 MHz have also been employed in this research field (10, 12). Although NMR spectra obtained at these high field strengths have improved spectral resolution and signal-to-noise ratio (S/N), the cost and housing facilities also raise strongly (16).

Recently, our research group has performed metabolite spiking experiments on a 400 MHz spectrometer in order to rationally segment the human plasma ^{1}H-NMR spectrum into 110 variable-sized integration regions, whereby the latter integrals are the variables for multivariate statistics and represent the metabolic phenotype (17). However, it has to be taken into account that accurate identification and quantification of metabolites can be challenging on a 400 MHz spectrometer in crowded regions with severe signal overlap (18, 19).

In the present study, spiking experiments were performed on a high-field 900 MHz spectrometer in order to rationally divide the spectrum into well-defined integration regions. The 900 MHz spectra have less signal overlap, an increased S/N and a reduced integration error as compared to the 400 MHz spectra.

Moreover, the integration data of a case-control dataset of 69 lung cancer patients and 74 controls were used to train a classifier in differentiating between both groups, and this for the 400 MHz as well as for the 900 MHz data in order to find out the advantages and disadvantages of both magnetic field strengths.

Patients and Methods

Subjects

Lung cancer patients (n=69) were included in the Limburg Positron Emission Tomography center (Hasselt, Belgium) from March 2011 to January 2012. The diagnosis of lung cancer was confirmed by a pathological biopsy or a clinician specialized in interpreting radiological and clinical lung cancer data. Clinical staging of the tumors was performed according to the 7th edition of the tumor, node, metastasis classification of malignant tumors (20) and independently checked by two of the authors (EL and KV). Controls (n=74) were patients with non-cancer diseases who were included at Ziekenhuis Oost-Limburg (Genk, Belgium) between December 2011 and April 2012.

Exclusion criteria were: 1) not fasted for at least 6 h; 2) fasting blood glucose concentration ≥ 200 mg/dl; 3) medication intake on the morning of blood sampling and 4) treatment or history of cancer in the past 5 years. The study was conducted in accordance with the ethical rules of the Helsinki Declaration and Good Clinical Practice and was approved by the ethical committees of Ziekenhuis Oost-Limburg and Hasselt University (Hasselt, Belgium). All study participants provided written informed consent.

Blood sampling and processing

See **Chapter 2**: Blood sampling and processing.

NMR sample preparation and analysis

NMR sample preparation was performed as described in **Chapter 2**: NMR sample preparation and analysis. ^{1}H-NMR measurements were performed at 21.2°C on a 400 MHz spectrometer (9.4 Tesla; Varian Inova; Agilent Technologies Inc.) and a 900 MHz spectrometer (21.1 Tesla; Bruker Avance; Bruker Biospin). The 400 MHz spectrometer is equipped with an Agilent OneNMR 5mm probe, whereas the 900 MHz spectrometer has a triple resonance cryoprobe. Slightly T2-weighted spectra were acquired using the Carr-Purcell-Meiboom-Gill pulse sequence (total spin-echo time of 32 ms), preceded by presaturation for water suppression. Other parameters were: a spectral width of 6,000 Hz/14,423 Hz (400 MHz/900 MHz), a preparation delay of 3.5 s, and 96/64 scans (400 MHz/900 MHz). Each free

induction decay was zero-filled to 65 K points and multiplied by a 0.7 Hz exponential line-broadening function prior to Fourier transformation.

Signal assessment by spiking reference plasma with known metabolites on the 400 MHz and 900 MHz spectrometer

Fasting venous blood was collected from a healthy 44-year-old female. The plasma was obtained and processed as described above, and further referred to as reference plasma. Stock solutions for spiking were prepared by dissolving 1 mg of a known metabolite in 100 µl reference plasma. In a next step, 10 µl stock solution was added to a standard NMR sample (200 µl reference plasma and 600 µl D_2O containing TSP) and subsequently analyzed on the 400 MHz and 900 MHz spectrometer as described above. This procedure was repeated for 37 metabolites, i.e. alanine, arginine, asparagine, aspartate, cysteine, glutamine, glutamate, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, valine, glucose, myo-inositol, acetate, acetoacetate, α-ketoglutarate, β-hydroxybutyrate, citrate, lactate, pyruvate, succinate, creatine, creatinine, aceton, betaine, choline, glycerol and methanol (17).

Spectral processing of 400 MHz spectra

See **Chapter 2**: Spectral processing.

Spectral processing of 900 MHz spectra

Spectra were phased, baseline corrected and referenced to TSP (δ=0.015 ppm). Based on the metabolite spiking, the ^{1}H-NMR spectra were segmented into 105 variable-sized spectral regions for integration, excluding the water region (4.7-5.1 ppm) and TSP (-0.3-0.3 ppm). Two sections of the ^{1}H-NMR spectrum (between 7.1-7.3 ppm and 3.45-3.9 ppm) were always checked with respect to the integration settings. The integrated regions were then normalized relatively to the total integrated area of all regions (except water and TSP), resulting in 105 normalized integration values, being the variables for multivariate statistics.

Statistical analysis

Multivariate statistics was performed using SIMCA-P+ (Version 14, Umetrics, Umea, Sweden). After mean-centering and Pareto scaling of both 400 MHz and

900 MHz integration data, supervised orthogonal partial least squares discriminant analysis (OPLS-DA) was used to train a classification model in discriminating between lung cancer patients and controls (21). The robustness of the classification models trained by means of the 400 and 900 MHz integration data respectively was further evaluated using receiver operating characteristic curve explorer & tester (22). Via an S-plot, the most discriminating variables of the models were identified together with their corresponding variable importance for the projection (VIP) values (23).

Results and Discussion

Figure 3.1 shows ^{1}H-NMR spectra of a representative plasma sample acquired on a 400 MHz and a 900 MHz spectrometer. An improved spectral resolution as well as S/N can be observed for the 900 MHz spectrum compared to the 400 MHz spectrum. To assign the resonance signals in both spectra, spiking experiments were performed with known metabolites. Our research group already reported the spiking of reference plasma with 37 different metabolites in relevant concentrations in order to assign the signals and rationally divide the 400 MHz spectrum into 110 well-defined integration regions (**Table 3.1, right half**) (17). In this study, reference plasma was spiked with the same metabolites in relevant concentrations and analyzed on a 900 MHz spectrometer. The resulting information allows to rationally divide the 900 MHz spectrum into 105 well-defined integration regions (**Table 3.1, left half**). On the basis of the spiking experiments and even despite of the improved S/N of a 900 MHz spectrometer, it can be concluded that the plasma levels of arginine, betaine, choline, cysteine, tryptophan and myo-inositol are below the detection limit, explaining why they are not taken up in **Table 3.1**. Furthermore, the increased spectral resolution and S/N of the 900 MHz spectra enable to define the integration regions more accurately, resulting in a larger number of integration regions that represent a single metabolite, thereby contributing to the identification of the most discriminating metabolites and to the understanding of the underlying disturbed biochemical pathways of disease.

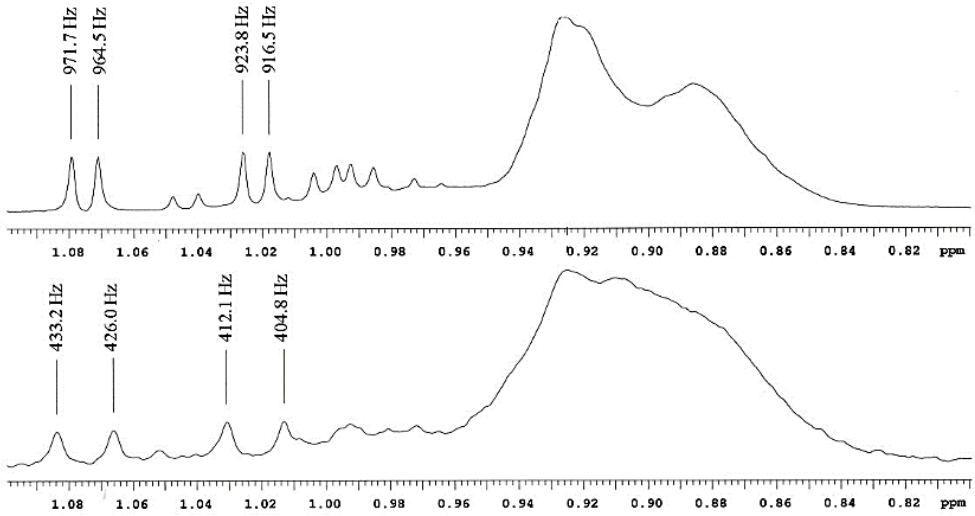


Figure 3.1. Zoom-in between 0.80-1.10 ppm of the ^{1}H-NMR spectrum of human plasma acquired at 900 (top) and 400 MHz (bottom). Remark that the J-coupling is independent of the field strength when expressed in Hz (see marked resonance frequencies in the figure), but not when expressed in ppm. ppm: parts per million.

Table 3.1. Overview of the 105 rationally defined integration regions of the 900 MHz NMR spectrum and their contributing metabolites (left half) versus the 110 rationally defined integration regions and contributing metabolites of the 400 MHz NMR spectrum (right half).

900 MHz					400 MHz		
VAR	**VIP**	**Contributing metabolites**	**Start**	**End**	**VAR**	**VIP**	**Contributing metabolites**
1		NI	8,4914	8,4796			
2		Formate	8,3702	8,3602			
3		NI	8,2601	8,2500			
4		NI	8,2300	8,2050			
5		NI	7,8561	7,8104	**1*[i]**	0.68	NI
					2		His
6		His	7,7812	7,7544	**3**		NI
					4*[i]	0.53	His
					5		NI
					6		NI
					7		NI
7		Phe	7,4677	7,4380	**8**		Phe
8		Phe, NI	7,4162	7,3755			
9		Phe	7,3675	7,3484	**9**		Phe, NI
10		NI	7,3484	7,3227	**10**		NI
					11		NI
					12		Tyr, NI
11		Tyr	7,2327	7,2046	**13**		Tyr, NI
12		NI	7,1894	7,1591			
13		His	7,0792	7,0597	**14**		His
14		NI	7,0201	6,9652			
15		Tyr	6,9355	6,9056	**15**		Tyr
					16		NI
16		NI	6,7460	6,7004	**17**		NI
17*[d]	1.62	Lipids: -C**H**=C**H**- in FAC[#]	5,4422	5,2833	**18*[d]**	1.96	Lipids: -C**H**=C**H**- in FAC[#]
18*[i]	0.88	Glucose	5,2751	5,2542	**19*[i]**	0.73	Glucose
19		C_2H in glycerol backbone of PL and TG[#]	5,2542	5,2301	**20**		C_2H in glycerol backbone of PL and TG[#]
20		C_2H in glycerol backbone of PL and TG[#]	5,2186	5,2038			
21		NI	5,1525	5,1187			
22*[i]	0.98	Glucose	4,7088	4,6421	**21*[i]**	1.00	Glucose
					22		NI
					23*[i]	0.79	NI
23		C_1H and C_3H in glycerol backbone of TG[#]	4,3579	4,2902	**24*[i]**	0.70	C_1H and C_3H in glycerol backbone of TG[#]
24		O-C**H**$_2$-CH$_2$-N$^+$(CH$_3$)$_3$ of PC and SM[#], Thr	4,2852	4,2536	**25**	0.73	O-C**H**$_2$-CH$_2$-N$^+$(CH$_3$)$_3$ of PC and SM[#], Thr
25		β-hydroxybutyrate	4,2000	4,1607	**26**		β-hydroxybutyrate, Pro
					27		β-hydroxybutyrate, lactate, Pro

Table 3.1. continued.

VAR	VIP	Contributing metabolites	Start	End	VAR	VIP	Contributing metabolites
26*[d]	1.47	C_1H and C_3H in glycerol backbone of PL and TG[#], lactate	4,1570	4,1276	**28*[d]**	1.49	C_1H and C_3H in glycerol backbone of PL and TG[#], lactate
27*[d]	0.54	NI	4,1276	4,0942	**29*[d]**	0.83	NI
					30		NI
28		Creatinine	4,0904	4,0780	**31**		Creatinine
					32		NI
					33		His, Ser
29		C_3H_2 in glycerol backbone of PL[#], Asn, His, Phe, Ser	4,0400	3,9913	**34**		Asn, His, Phe, Ser
					35		C_3H_2 in glycerol backbone of PL[#], Asn, His, Phe, Ser
					36*[i]	1.01	C_3H_2 in glycerol backbone of PL[#], Asn, His, Phe, Ser
30*[i]	0.74	Asn, His, Ser, Tyr	3,9903	3,9644	**37*[i]**	1.34	Creatine, Asn, His, Tyr, Ser
31*[i]	0.51	Creatine	3,9644	3,9586			
32		Tyr	3,9586	3,9527	**38*[i]**	3.61	Glucose, Asp, Met, Ser, Tyr
33*[i]	2.14	Glucose	3,9527	3,9120			
34*[i]	1.16	Glucose	3,9120	3,8957			
35*[i]	1.84	Glucose	3,8881	3,8306			
36		Glucose, Ala, Gln, Glu, Ser	3,8286	3,8097	**39*[i]**	0.90	Glucose, Ala, Ser
					40*[d]	0.60	Glucose, Ala, Gln, Glu
37*[i]	1.21	Glucose, Ala, Gln	3,8097	3,7794	**41*[i]**	0.55	Glucose, Ala, Gln, Glu, Leu, Lys
38*[i]	1.90	Glucose	3,7776	3,7275	**42*[i]**	1.69	Glucose, Ala, Gln, Glu, Leu, Lys
					43*[i]	1.11	Glucose, Ala, Leu
					44*[i]	1.38	Glucose
39*[i]	2.29	Glycerol	3,7204	3,6453	**45*[i]**	1.93	O-CH_2-**CH_2**-$N^+(CH_3)_3$ of PC and SM[#], glycerol, Ile
					46*[i]	0.96	Glycerol
					47*[i]	0.51	Glycerol, Val
40*[i]	0.74	Val	3,6453	3,6212	**48*[i]**	0.87	Val
41*[i]	1.16	Thr	3,6163	3,5861	**49*[i]**	0.89	Thr
					50*[i]	0.90	Thr
42		Glycerol	3,5861	3,5771	**51*[i]**	1.18	Glucose, glycerol, Gly, Thr
43*[i]	1.02	Glucose	3,5771	3,5481	**52*[i]**	0.53	Glucose
					53*[i]	0.78	Glucose, acetoacetate, Pro
44*[i]	1.76	Glucose	3,5355	3,4798	**54*[i]**	2.17	Glucose, acetoacetate, Pro
45		Pro	3,4772	3,4576			
46*[i]	1.45	Glucose	3,4576	3,4093			
47		Methanol	3,3964	3,3924	**55*[d]**	0.52	Methanol, NI
48		NI	3,3924	3,3746			
49		Pro	3,3746	3,3465	**56**		Pro
					57		Phe, Pro

Table 3.1. continued.

VAR	VIP	Contributing metabolites	Start	End	VAR	VIP	Contributing metabolites
50		Phe	3,3256	3,3132	**58*[d]**	2.55	O-CH_2-CH_2-N^+(CH_3)$_3$ of PC and SM[#], glucose, His, Phe, Tyr
51		Phe, NI	3,3132	3,3030			
52		NI	3,3030	3,2956			
53		NI	3,2956	3,2909			
54*[i]	1.11	Glucose	3,2909	3,2616			
55*[d]	2.21	O-CH_2-CH_2-N^+(CH_3)$_3$ of PC and SM[#]	3,2616	3,2085			
56		Tyr, NI	3,2085	3,1895	**59**		Tyr, NI
57		NI	3,1881	3,1821	**60**		NI
58		NI	3,1821	3,1724			
59		NI	3,1707	3,1571	**61**		NI
60		His, Phe	3,1541	3,1378	**62*[d]**	0.51	His, Phe
61		Tyr	3,0921	3,0769	**63*[d]**	0.55	Lys, Tyr
62		Creatinine	3,0769	3,0699	**64*[d]**	0.56	Creatinine, Lys, Tyr
63		Creatine	3,0699	3,0635	**65**		Creatinine, creatine, Lys
64*[i]	0.64	α-ketoglutrate, Lys	3,0635	3,0047	**66*[i]**	0.84	α-ketoglutarate, Lys
65		Lipids: =CH-C**H**$_2$-CH= in FAC[#]	3,0047	2,9655	**67*[d]**	1.00	Lipids: =CH-C**H**$_2$-CH= in FAC[#], Asn
66*[d]	0.50	Asn	2,9597	2,9201			
67		Asn, Asp	2,8874	2,8465	**68*[d]**	0.60	Lipids: =CH-C**H**$_2$-CH= in FAC[#], Asn, Asp
68*[d]	0.55	Lipids: =CH-C**H**$_2$-CH= in FAC[#]	2,8465	2,7623	**69*[d]**	1.35	Lipids: =CH-C**H**$_2$-CH= in FAC[#], Asn, Asp
69		Citrate	2,7571	2,7493	**70**		Citrate, Asp
70		NI	2,7472	2,7390			
71		Citrate	2,7368	2,7251	**71*[i]**	0.55	Citrate, Asp, Met
72*[i]	0.76	Asp	2,7237	2,6768			
73		Met	2,6768	2,6597	**72*[i]**	0.52	Met
74*[d]	0.52	Citrate	2,5865	2,5426	**73*[d]**	0.75	Citrate
					74		NI
75*[d]	0.76	Gln	2,5183	2,4428	**75**		Gln
					76		β-hydroxybutyrate, α-ketoglutarate, Gln
76		β-hydroxybutyrate	2,4428	2,4280	**77**		β-hydroxybutyrate, α-ketoglutarate, succinate
					78		β-hydroxybutyrate, Pro
					79		β-hydroxybutyrate, Glu, Pro
77		Pyruvate	2,4060	2,3978	**80**		Pyruvate, Pro, Glu
78		Glu	2,3978	2,3648	**81**		β-hydroxybutyrate, Pro, Glu
					82		β-hydroxybutyrate, Pro, Glu
79		β-hydroxybutyrate	2,3540	2,3194	**83**		β-hydroxybutyrate, Pro, Val
					84		β-hydroxybutyrate, acetoacetate, Pro, Val

Table 3.1. continued.

VAR	VIP	Contributing metabolites	Start	End	VAR	VIP	Contributing metabolites
80		Acetoacetate	2,3134	2,3067	**85**		β-hydroxybutyrate, acetoacetate, Val
81*[d]	0.64	Lipids: -C**H**$_2$-C=O or -C**H**$_2$-CH=CH- in FAC[#]	2,3067	2,2630	**86**		Lipids: -C**H**$_2$-C=O or -C**H**$_2$-CH=CH- in FAC[#], β-hydroxybutyrate, Val
					87*[d]	0.84	Lipids: -C**H**$_2$-C=O or -C**H**$_2$-CH=CH- in FAC[#], Met, Val
82		Aceton	2,2630	2,2563	**88*[d]**	0.66	Lipids: -C**H**$_2$-C=O or -C**H**$_2$-CH=CH- in FAC[#], aceton, Met, Val
					89		Glu, Met
83		NI	2,1975	2,1814	**90*[i]**	0.81	Gln, Glu, Pro, Met
84		Gln	2,1777	2,1670			
85		Met	2,1670	2,1919			
86*[i]	0.89	Gln	2,1619	2,1311			
87*[i]	1.80	Lipids: -C**H**$_2$-CH=CH- in FAC[#]	2,1289	2,0993	**91*[i]**	2.72	Lipids: -C**H**$_2$-CH=CH- in FAC[#], C**H**$_3$ of NAG[$], Glu, Ile, Met, Pro
88*[i]	2.70	Lipids: -C**H**$_2$-CH=CH- in FAC[#], C**H**$_3$ of NAG[$]	2,0993	1,9889			
89		Acetate	1,9547	1,9421	**92**		Acetate, Ile, Lys
90		Lys	1,9421	1,9028	**93*[i]**	0.50	Ile, Lys
91*[i]	1.06	Leu	1,8006	1,6758	**94*[i]**	0.84	Leu, Lys
92		Lipids: -C**H**$_2$-CH$_2$-C=O or -C**H**$_2$-CH$_2$-CH=CH- in FAC[#]	1,6530	1,5770	**95*[d]**	1.36	Lipids: -C**H**$_2$-CH$_2$-C=O or -C**H**$_2$-CH$_2$-CH=CH- in FAC[#], Lys
93*[d]	0.56	Ala	1,5226	1,4919	**96*[d]**	0.85	Ala, Ile, Lys
94*[i]	0.75	Lys	1,4587	1,4201	**97**		Ile, Leu, Lys
95*[d]	1.50	Lactate	1,4169	1,3675	**98*[d]**	1.60	Lactate
96*[d]	2.80	Lactate	1,3675	1,3516	**99*[d]**	3.11	Lactate, Thr
97*[d]	4.52	Lipids: -CH$_3$-(C**H**$_2$)$_n$- in FAC[#]	1,3516	1,2366	**100*[d]**	3.54	Lipids: CH$_3$-(C**H**$_2$)$_n$- in FAC[#], Ile, Thr
98*[i]	0.76	β-hydroxybutyrate	1,2366	1,2240	**101*[i]**	0.64	β-hydroxybutyrate, Ile
99*[i]	1.99	NI	1,2240	1,1766	**102**		NI
100		Val	1,0860	1,0592	**103**		Val
101*[i]	0.56	Ile	1,0513	1,0340	**104**		Ile
102*[i]	0.78	Val	1,0396	1,0106	**105**		Ile, Val
					106		Ile, Leu, Val
103*[i]	0.72	Leu	1,0083	0,9766	**107**		Ile, Leu
					108		Ile, Leu
104*[i]	0.68	Ile	0,9766	0,9663	**109*[i]**	0.54	Ile
105*[d]	2.34	Lipids: C**H**$_3$-(CH$_2$)$_n$- in FAC[#]	0,9663	0,7961	**110*[d]**	2.39	Lipids: C**H**$_3$-(CH$_2$)$_n$- in FAC[#]

The assignment of the resonance signals present in the 400 MHz spectra was based on spiking experiments (17). Nonetheless, some assignments are slightly adapted based on additional information

obtained from spiking experiments on the 900 MHz spectrometer. [#]The assignment of the lipid signals is based on literature (24-26). [$]The assignment of signals of N-acetylated glycoproteins is based on literature (27). Amino acids are presented by their 3-letter code. *[d]: Variables with a VIP value exceeding 0.5 and which are decreased in plasma of lung cancer patients, *[i]: Variables with a VIP value exceeding 0.5 and which are increased in plasma of lung cancer patients. FAC: fatty acid chain; NAG: N-acetylated glycoproteins; NI: non-identified; PC: phosphatidylcholine; PL: phospholipids; ppm: parts per million; SM: sphingomyelin; TG: triglycerides; VAR: variable; VIP: variable importance for the projection.

An interesting example is VAR58 of the 400 MHz spectrum, which can be divided into 6 regions (ranging from VAR50-55) in the 900 MHz spectrum, revealing that the plasma concentration of glucose (VAR54) is increased whereas the level of sphingomyelin and phosphatidylcholine (VAR55) is decreased for lung cancer patients. Remark that sphingomyelin and phosphatidylcholine can be discriminated from the other lipids on the basis of the strong singlet NMR signal of the nine protons of the three methyl groups of the choline head group. On the other hand, adjacent regions which encompass the same dominant metabolite are combined into a single integration region, e.g. VAR42-44 in the 400 MHz spectrum (composed of very strong signals of glucose next to very weak signals of the α-CH group of several amino acids) is replaced by VAR38 (a single glucose signal) in the 900 MHz spectrum. Taken together, 68% of the variables in the 900 MHz spectrum (71 out of the 105) represent a single metabolite as opposed to only 24% in the 400 MHz spectrum (26 out of 110). The number of non-identified signals however is rather independent of the applied magnetic field strength, i.e. 20% (21 out of 105) of the variables in the 900 MHz spectrum and 17% (19 out of 110) in the 400 MHz spectrum.

Application of the analysis protocols in a case-control dataset

In order to investigate the discriminative power of plasma metabolic phenotyping data derived from the integration data collected at different magnetic field strengths, a case-control dataset of 69 lung cancer patients and 74 controls was analyzed on a 400 MHz and a 900 MHz spectrometer. Subject characteristics of the case-control dataset are presented in **Table 3.2**. Supervised OPLS-DA analysis was conducted to train robust classification models in discriminating between lung cancer patients and controls. Using the 400 MHz integration data, the best model having 1 predictive and 4 orthogonal components allows to classify 94% (65 out of 69) of the lung cancer patients and 97% (72 out of 74) of the

controls correctly with an area under the curve (AUC) of 0.93 (**Figure 3.2A-B, Table 3.3**). By means of the 900 MHz integration data, the model having 1 predictive and 4 orthogonal components allows to classify 86% (59 out of 69) of the lung cancer patients and 95% (70 out of 74) of the controls correctly (**Table 3.3**). However, the discriminative power of the 900 MHz model still increases significantly when 9 orthogonal components were used to train the model, resulting in a sensitivity of 90%, a specificity of 100% and an AUC of 0.90 (**Figure 3.2C-D, Table 3.3**).

Table 3.2. Characteristics of the subjects included in the study.

	LC	C
Number of subjects, n	69	74
Gender, n (%)		
Male	46 (66.7)	44 (59.5)
Female	23 (33.3)	30 (40.5)
Age, yrs	68 ± 10	64 ± 13
(range)	(36 – 88)	(23 – 84)
BMI, kg/m^2	25.3 ± 4.6	26.3 ± 4.6
(range)	(17.5 – 38.5)	(16.5 – 39.0)
Smoking habits		
Smoker, n (%)	40 (58.0)	19 (25.7)
Ex-smoker, n (%)	26 (37.7)	28 (37.8)
Non-smoker, n (%)	3 (4.3)	27 (36.5)
Pack years	34 ± 21	18 ± 28
(range)	(0-125)	(0-175)
Laterality		
Left, n (%)	23 (33.3)	
Right, n (%)	39 (56.5)	
Bilateral, n (%)	5 (7.2)	
Unknown, n (%)	2 (2.9)	
Amount of tumors, n	74	
Histological subtype		
NSCLC-Adenocarcinoma, n (%)	27 (36.5)	
NSCLC-Spinocellular carcinoma, n (%)	18 (24.3)	
NSCLC-Adenosquamous carcinoma, n (%)	3 (4.1)	
NSCLC-Carcinoid, n (%)	1 (1.3)	
NSCLC-NOS, n (%)	3 (4.1)	
SCLC, n (%)	12 (16.2)	
Unknown	10 (13.5)	
Clinical stage according to 7th TNM edition		
IA, n (%)	18 (24.3)	
IB, n (%)	5 (6.7)	
IIA, n (%)	4 (5.4)	
IIB, n (%)	2 (2.7)	
IIIA, n (%)	15 (20.3)	
IIIB, n (%)	11 (14.9)	
IV, n (%)	19 (25.7)	

Data are presented as mean ± standard deviation and range, unless otherwise indicated. BMI: body mass index; C: controls; LC: lung cancer; NOS: not otherwise specified; NSCLC: non-small cell lung cancer; SCLC: small cell lung cancer; TNM: tumor, node; metastasis.

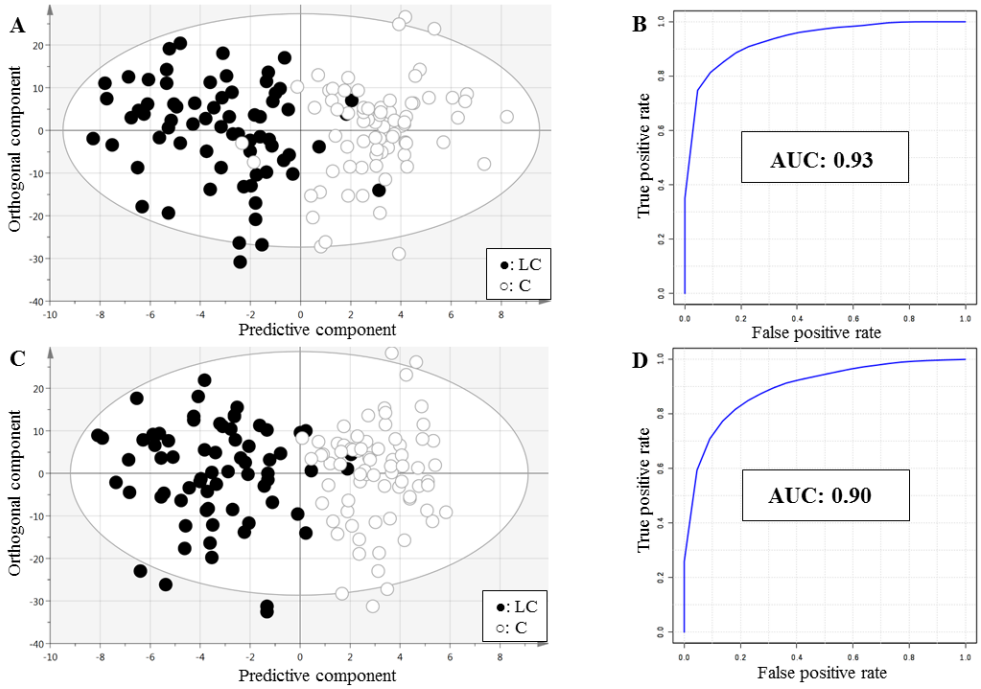


Figure 3.2. (A) OPLS-DA score plot derived from the 400 MHz data, (B) receiver operating characteristic curve derived from the 400 MHz data, (C) OPLS-DA score plot derived from the 900 MHz data, (D) receiver operating characteristic curve derived from the 900 MHz data. AUC: area under the curve; C: controls; LC: lung cancer patients; OPLS-DA: orthogonal partial least squares discriminant analysis.

Table 3.3. Characteristics of the trained OPLS-DA classification models resulting from the 400 MHz and 900 MHz data.

	LV (P+O)	R^2X (cum)	R^2Y (cum)	Q^2 (cum)	Sens (%)	Spec (%)	AUC
LC vs. C (400 MHz)	5 (1+4)	0.83	0.67	0.54	94	97	0.93
LC vs. C (900 MHz)	5 (1+4)	0.81	0.58	0.38	86	95	/
	10 (1+9)	0.90	0.75	0.49	90	100	0.90

AUC: area under the curve; C: controls; LC: lung cancer patients; LV: latent variable; MHz: megahertz; O: number of orthogonal components; OPLS-DA: orthogonal partial least squares discriminant analysis; P: number of predictive components; R^2X(cum): total explained variation in X; R^2Y(cum): total explained variation in Y; Sens: sensitivity; Spec: specificity; Q^2(cum): predicted variation.

When a closer look is taken at the model characteristics (**Table 3.3**), it can be concluded that on the one hand, the best 900 MHz model contains twice as much latent variables, clarifying why this model explains 90% of the intra-group variation ($R^2X(cum)$) relatively to 83% in the 400 MHz model. In addition, the 900 MHz model explains 75% of the inter-group variation ($R^2Y(cum)$) compared to 67% in the 400 MHz model. On the other hand, the predictive ability of the 400 MHz model is slightly better, i.e. 54% relatively to 49%. Taken all above in consideration, it can be concluded that the discriminative power of the models obtained on the basis of the 400 MHz data and the 900 MHz data is comparable. This finding is in line with the results of Bertram et al., who examined the impact of varying magnetic field strengths, i.e. 250, 400, 500 and 800 MHz, on the urinary metabolic phenotype before and after a dietary intervention (16). They concluded that although the power of the urinary metabolic phenotype to discriminate between pre- and post-intervention samples significantly improved when increasing the magnetic field strength from 250 to 500 MHz, it remained quasi stable when the magnetic field strength was further increased from 500 to 800 MHz.

In order to explain the disturbed biochemical pathways in lung cancer, only variables with a VIP value exceeding 0.5 were considered. This corresponds to 49% (54 out of 110) and 41% (43 out of 105) of the variables for the 400 MHz and 900 MHz data, respectively. All these variables (indicated with an asterisk in **Table 3.1**) are situated far out on the wings of the respective S-plots as demonstrated in **Figure 3.3A-B**. For both 400 MHz and 900 MHz data and with respect to the controls, the plasma concentration of aspartate, β-hydroxybutyrate, creatine, glutamine, glucose, glycerol, isoleucine, leucine, lysine, N-acetylated glycoproteins, threonine and valine is increased, whereas the concentration of alanine, asparagine, citrate, lactate, non-cholinated lipids, phosphatidylcholine and sphingomyelin is decreased. Only a few (4) additional altered variables were found for the 400 MHz data, but all of them with low VIP values between 0.5 and 0.6. Similar findings were reported by Bertram et al., demonstrating that the same variables were found in the discrimination between pre- and post-intervention samples at different magnetic field strengths (16).

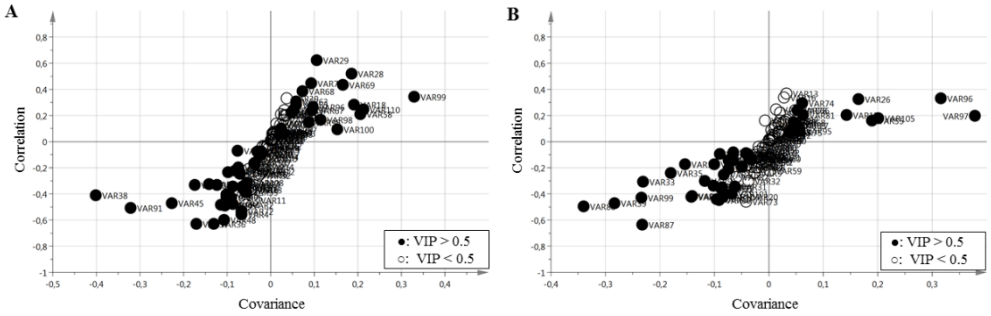


Figure 3.3. (A) S-plot of the OPLS-DA model derived from the 400 MHz data showing the variables contributing most to group discrimination. Variables situated at the right end are increased in the plasma of controls, while those situated at the left end are increased for the lung cancer patients, **(B) S-plot of the OPLS-DA model derived from the 900 MHz data.** Variables used to explain the disturbed biochemical pathways in lung cancer (VIP > 0.5) are marked (●). Var: variable; VIP: variable importance for the projection.

Advantages and disadvantages of high (900 MHz) versus medium (400 MHz) field strengths

High-field spectra (e.g. 900 MHz spectra) have an increased spectral resolution and S/N compared to 400 MHz spectra, allowing a more accurate setting of the integration regions, resulting in an increase in integration regions which represent a single metabolite and thereby facilitate the explanation of the underlying disturbed biochemical pathways. However, high-field spectrometers also have some disadvantages: i) they need an isolated environment, i.e. a separate building, making them less practical for clinical practice and ii) the cost of the instrumentation raises strongly with the magnetic field strength. Taking all into account, medium-field (400-600 MHz) spectrometers seem to be most suitable for clinical metabolomics.

Conclusions

On the basis of the information obtained from spiking experiments, 400 MHz and 900 MHz spectra of human blood plasma were rationally divided into 110 and 105 well-defined integration regions, respectively. The increased spectral resolution and S/N of the 900 MHz spectra enable to define the integration regions more accurately, resulting in a larger number of integration regions that represent a single metabolite. In this way, it becomes more convenient to unravel the underlying disturbed biochemical pathways. Nonetheless, when the integration data collected at a 400 MHz and a 900 MHz spectrometer were applied to classify a case-control dataset of 69 lung cancer patients and 74 controls, the discriminative power was quasi comparable. Therefore, it can be concluded that medium-field NMR spectrometers, as a 400 MHz, are satisfactory if group discrimination is the only aim. High-field NMR spectrometers on the other hand are indispensable if the goal includes the unraveling of the disturbed biochemical pathways. The global findings of this study indicate that medium-field (400-600 MHz) spectrometers, commonly available in research institutes, are most suitable for clinical metabolomics studies as the increase in spectral resolution and S/N does not outweigh the rise in equipment, housing and maintenance costs.

References

1. Hollywood K, Brison DR, Goodacre R. Metabolomics: current technologies and future trends. Proteomics. 2006;6(17):4716-23.
2. Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica; the fate of foreign compounds in biological systems. 1999;29(11):1181-9.
3. Lindon JC, Nicholson JK. Spectroscopic and statistical techniques for information recovery in metabonomics and metabolomics. Annu Rev Anal Chem. 2008;1:45-69.
4. Eliasson M, Rannar S, Trygg J. From data processing to multivariate validation--essential steps in extracting interpretable information from metabolomics data. Current pharmaceutical biotechnology. 2011;12(7):996-1004.
5. Madsen R, Lundstedt T, Trygg J. Chemometrics in metabolomics--a review in human disease diagnosis. Analytica chimica acta. 2010;659(1-2):23-33.
6. Trygg J, Holmes E, Lundstedt T. Chemometrics in metabonomics. Journal of proteome research. 2007;6(2):469-79.
7. Dumas ME, Maibaum EC, Teague C, Ueshima H, Zhou B, Lindon JC, et al. Assessment of analytical reproducibility of 1H NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP Study. Analytical chemistry. 2006;78(7):2199-208.
8. Emwas AH, Al-Talla ZA, Guo X, Al-Ghamdi S, Al-Masri HT. Utilizing NMR and EPR spectroscopy to probe the role of copper in prion diseases. Magnetic resonance in chemistry. 2013;51(5):255-68.
9. Garcia E, Andrews C, Hua J, Kim HL, Sukumaran DK, Szyperski T, et al. Diagnosis of early stage ovarian cancer by 1H NMR metabonomics of serum explored by use of a microflow NMR probe. Journal of proteome research. 2011;10(4):1765-71.
10. Jobard E, Pontoizeau C, Blaise BJ, Bachelot T, Elena-Herrmann B, Tredan O. A serum nuclear magnetic resonance-based metabolomic signature of advanced metastatic human breast cancer. Cancer letters. 2014;343(1):33-41.
11. Kline EE, Treat EG, Averna TA, Davis MS, Smith AY, Sillerud LO. Citrate concentrations in human seminal fluid and expressed prostatic fluid determined via 1H nuclear magnetic resonance spectroscopy outperform prostate specific antigen in prostate cancer detection. The Journal of urology. 2006;176(5):2274-9.
12. Kumar D, Gupta A, Mandhani A, Sankhwar SN. Metabolomics-Derived Prostate Cancer Biomarkers: Fact or Fiction? Journal of proteome research. 2015.
13. Rocha CM, Carrola J, Barros AS, Gil AM, Goodfellow BJ, Carreira IM, et al. Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of blood plasma. Journal of proteome research. 2011;10(9):4314-24.
14. Zhang L, Jin H, Guo X, Yang Z, Zhao L, Tang S, et al. Distinguishing pancreatic cancer from chronic pancreatitis and healthy individuals by (1)H nuclear magnetic resonance-based metabonomic profiles. Clinical biochemistry. 2012;45(13-14):1064-9.
15. Fedele TA, Galdos-Riveros AC, Jose de Farias e Melo H, Magalhaes A, Maria DA. Prognostic relationship of metabolic profile obtained of melanoma B16F10. Biomedicine & pharmacotherapy. 2013;67(2):146-56.
16. Bertram HC, Malmendal A, Petersen BO, Madsen JC, Pedersen H, Nielsen NC, et al. Effect of magnetic field strength on NMR-based metabonomic human urine data. Comparative study of 250, 400, 500, and 800 MHz. Analytical chemistry. 2007;79(18):7110-5.
17. Louis E, Bervoets L, Reekmans G, De Jonge E, Mesotten L, Thomeer M, et al. Phenotyping human blood plasma by ^{1}H-NMR: a robust protocol based on metabolite spiking and its evaluation in breast cancer. Metabolomics. 2015;11:225-36.
18. O'Connell TM. Recent advances in metabolomics in oncology. Bioanalysis. 2012;4(4):431-51.
19. Zheng C, Zhang S, Ragg S, Raftery D, Vitek O. Identification and quantification of metabolites in (1)H NMR spectra by Bayesian model selection. Bioinformatics. 2011;27(12):1637-44.

20. Goldstraw P, Crowley J, Chansky K, Giroux DJ, Groome PA, Rami-Porta R, et al. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours. J Thorac Oncol. 2007;2(8):706-14.
21. Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). J Chemometr. 2002;16(3):119-28.
22. Xia J, Broadhurst DI, Wilson M, Wishart DS. Translational biomarker discovery in clinical metabolomics: an introductory tutorial. Metabolomics. 2013;9(2):280-99.
23. Wiklund S, Johansson E, Sjostrom L, Mellerowicz EJ, Edlund U, Shockcor JP, et al. Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. Analytical chemistry. 2008;80(1):115-22.
24. Kriat M, Vion-Dury J, Confort-Gouny S, Favre R, Viout P, Sciaky M, et al. Analysis of plasma lipids by NMR spectroscopy: application to modifications induced by malignant tumors. Journal of lipid research. 1993;34(6):1009-19.
25. Oostendorp M, Engelke UF, Willemsen MA, Wevers RA. Diagnosing inborn errors of lipid metabolism with proton nuclear magnetic resonance spectroscopy. Clinical chemistry. 2006;52(7):1395-405.
26. Papathanasiou A, Kostara C, Cung MT, Seferiadis K, Elisaf M, Bairaktari E, et al. Analysis of the composition of plasma lipoproteins in patients with extensive coronary heart disease using 1H NMR spectroscopy. Hellenic journal of cardiology. 2008;49(2):72-8.
27. Lenz EM, Bright J, Wilson ID, Morgan SR, Nash AF. A 1H NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects. Journal of pharmaceutical and biomedical analysis. 2003;33(5):1103-15.

CHAPTER 4

Influence of preanalytical conditions on the metabolic phenotype of human plasma and the added value of the Standard PREanalytical Code

Liene Bervoets*, Evelyne Louis*, Gunter Reekmans, Liesbet Mesotten, Michiel Thomeer, Peter Adriaensens, Loes Linsen. Influence of preanalytical sampling conditions on the ^{1}H-NMR metabolic profile of human blood plasma and introduction of the Standard PREanalytical Code used in biobanking. Metabolomics (2015), 11:1197-1207.

*These authors contributed equally to this work

Abstract

Variations in sample collection, processing and storage within the field of clinical metabolomics might hamper its effective implementation. In this study, the impact of relevant preanalytical conditions on the plasma ^{1}H-NMR metabolic profile was examined. The biobanking community recently developed a method for coding preanalytical conditions called the Standard PREanalytical Code (SPREC). It is envisaged that SPREC will ultimately identify which samples are fit for a particular analysis, based on prior validation by a panel of experts in the respective field. In an effort to validate SPREC for ^{1}H-NMR plasma metabolomics, we have coded the conditions used here, when possible, according to SPREC and evaluated its power to identify preanalytical conditions that affect the plasma ^{1}H-NMR metabolic profile. From all preanalytical conditions studied, only prolonged processing delays (3 h and 8 h) have a significant impact on the plasma ^{1}H-NMR metabolic profile as compared to the reference condition (30 min). Principal component analysis shows a clear systematic shift as a function of increasing processing delay. Nevertheless, the inter-individual variation is clearly much larger than this preanalytical variation, indicating that the impact on multivariate group classification will be minimal. Nonetheless, we recommend to keep the time gap between blood collection and centrifugation similar for all samples within a study. The implementation of SPREC within clinical metabolomics allows for an appropriate sample encoding and exclusion of samples that were subjected to unwanted, interfering preanalytical conditions. Without doubt, it will contribute to the validation of ^{1}H-NMR metabolomics in clinical, biobank and multicenter research settings.

Introduction

Metabolomics is a powerful tool to detect metabolites in biological samples. It has great promise for the discovery of novel clinical biomarkers and the elucidation of disease-specific pathways to improve prognosis, diagnosis and therapy (1-5). Metabolomics research is mainly based on mass spectrometry (MS) and/or nuclear magnetic resonance (NMR) spectroscopy combined with multivariate statistics in order to understand and interpret the resulting data (6-8). NMR-based metabolomics has several advantages which make it highly suitable for clinical implementation. First, it can be used to study biofluids (e.g. plasma, serum or urine) and second, requires only limited sample preparation and processing (9). While serum traditionally makes up the majority of samples in clinical laboratories, plasma is preferentially used in laboratory medicine because it is more time-saving, has a higher yield and prevents coagulation-induced interferences (10). Moreover, lithium-heparin (LiHe) is described as the most suitable anticoagulant for proton (^{1}H)-NMR analysis (11, 12).

Nevertheless, due to the nature of the clinical setting, samples can be subjected to preanalytical variations in collection, processing and storage procedures. Additionally, the high number of samples needed for discovery and validation metabolomics is often gathered from multiple research centers, clinics or biobanks, increasing the likelihood of discrepancies between sample handling (9). It has been described for liquid chromatography-MS-based metabolomics in particular that preanalytical changes can have a major impact on the quality of samples, impeding interpretation of analytical results and decreasing the credibility of research outcomes (13-17). As this can potentially affect clinical implementation of metabolomics, it is clear that the impact of clinical sources of preanalytical bias on the plasma ^{1}H-NMR metabolome needs to be elucidated.

Although efforts are currently made within the field of metabolomics to move towards defining standard operation procedures for preanalytical handling (18), complete standardization of the preanalytical processing is not yet feasible between and within clinical settings. Alternatively, application of a preanalytical sample code that traces and manages these variations would allow sample harmonization. Hereto, the Standard PREanalytical Code (SPREC) was developed within the field of biobanking (19, 20). This is an easy to implement and

comprehensive tool, consisting of seven elements that document the critical preanalytical details of biospecimens. However, its value within clinical metabolomics remains to be evaluated.

To investigate the potential of metabolomics for implementation in the clinic, we examined the impact of relevant preanalytical conditions on the plasma ^{1}H-NMR metabolic profile. The preanalytical protocols used in this study were encoded according to SPREC in order to evaluate their power with respect to the identification of preanalytical conditions that affect the plasma ^{1}H-NMR metabolic profile. This was done in order to contribute to the validation of SPREC in clinical metabolomics.

Materials and Methods

Subjects

Twenty volunteers (8 males and 12 females) aged between 21 and 72 (median age: 32) were included in the study. The study participants consisted of university staff and controls, who were recruited in 2013 participating in another metabolomics study conducted at Ziekenhuis Oost-Limburg (Genk, Belgium). All subjects had fasted at least 12 h before blood collection. The study was conducted in accordance with the ethical rules of the Helsinki Declaration and Good Clinical Practice, and was approved by the ethical committees of Jessa Hospital and Hasselt University. All participants gave written informed consent prior to inclusion in the study.

Sample collection, preparation and storage

Blood samples were collected into 10 ml LiHe tubes at 9 am according to the World Health Organization guidelines on drawing blood (21). An overview of the entire study protocol is presented in **Figure 4.1**. The reference processing protocol consisted of keeping the freshly drawn blood for 5 min at room temperature (RT), followed by a 30 min incubation on ice, 15 min centrifugation at 1,600 g at 4°C and subsequent storage at -80°C in 1 ml cryovials as 350 µl aliquots (**Figure 4.1**).

In study group 1 (n=6), a double concentration of LiHe was obtained by drawing only 5 ml blood in the 10 ml tube. Short-term exposure to an oxidative atmosphere was acquired by transferring blood from the primary tube to another. Blood samples subjected to both preanalytical conditions were also subjected to a subsequent 3 h and 8 h processing delay at 4°C to determine potential cumulative effects. In study group 2 (n=6), hemolysis grade 1 (moderate) and 2 (strong) was induced by putting the blood samples directly on dry ice for three and six minutes, respectively (22). The free hemoglobin concentration was determined on a Roche/Hitachi MODULAR P analyzer (D-BIL Cobas, Roche Diagnostics, Mannheim, Germany). Hemolysis grade 1 corresponds to a free hemoglobin concentration ≥ 10 mg/dl and hemolysis grade 2 to a concentration ≥ 100 mg/dl (H index, manual MODULAR P analyzer). The effect of a variable processing delay was examined as described for study group 1. To examine the effect of centrifugation temperature, blood samples were centrifuged at 4°C or RT after a standard processing delay of

30 min on ice. To examine freezing procedure effects, plasma was aliquoted into 1 ml cryovials and immediately stored at -80°C or kept for 8 h on dry ice before storage at -80°C or kept in liquid nitrogen (LN_2) for 8 h before storage at -80°C. The effect of plasma storage duration at -80°C in study group 3 (n=10) was determined by storing the samples at -80°C for 2 and 10 months.

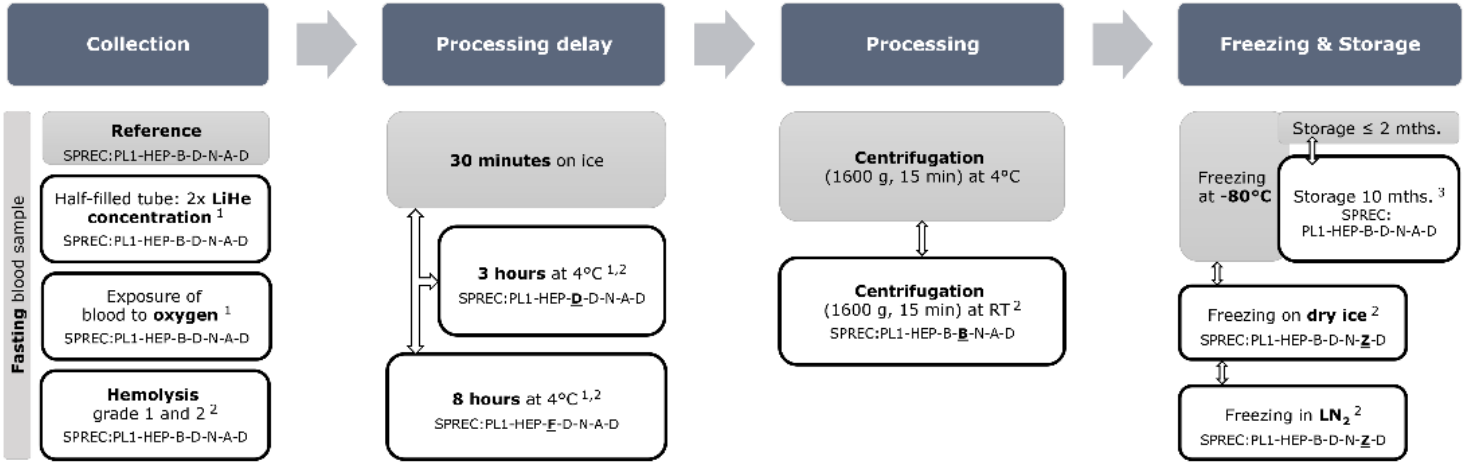


Figure 4.1. Overview of the study protocol. Fasting blood samples were obtained and handled according to the reference protocol (grey boxes) or subjected to several preanalytical conditions: (1) a double LiHe concentration or an oxidative atmosphere, in combination with a processing delay of 3 h or 8 h at 4°C was investigated in study group 1 (n=6); (2) hemolysis grade 1 and 2, a processing delay of 3 h and 8 h at 4°C, centrifugation at RT, and 8 h freezing on dry ice or in LN_2 was examined in study group 2 (n=6); (3) storage of plasma aliquots during 10 months at -80°C was studied in study group 3 (n=10). SPREC annotations that are different from the reference are indicated in bold and are underlined. LiHe: lithium-heparin; LN_2: liquid nitrogen; mths: months; SPREC: Standard PREanalytical Code; RT: room temperature.

NMR sample preparation and analysis

To account for experimental variability, samples were measured in random order and in duplicate. To exclude inter-operator variability, samples were prepared, measured and post-processed by the same operator. NMR sample preparation and ^{1}H-NMR analysis was performed as described in **Chapter 2**: NMR sample preparation and analysis.

Spectral processing

See **Chapter 2**: Spectral processing.

Statistical analysis

Multivariate statistics was performed on the average value of the variables obtained from plasma analyses in duplicate (SIMCA-P^{+}; Version 13.0.3, Umetrics, Umea, Sweden). After mean-centering and Pareto scaling of the variables, unsupervised principal component analysis (PCA) was performed to identify patterns or clusters. The variance structure of the data is explained through linear combinations of the variables, i.e. the so-called principal components (PCs). The first PC explains the largest variance within the dataset, followed by the second and third PCs. This multivariate analysis was performed by means of the 110 integration values (variables) which were defined and numbered as previously described in literature (23). Additionally, as 40 of these 110 variables were assigned as 'noisy', i.e. having a coefficient of variation exceeding 15% (23), the multivariate analysis was also performed by using only the remaining 70 variables. However, no significant differences were detected as can be seen, for the processing delay as an example, in **Figure 4.2**. Therefore, multivariate analyses were performed by using all 110 variables.

To explore the discriminating variables more in detail, univariate statistics was performed on data obtained from all measurements, and not on average values of the duplo measurements (IBM SPSS Statistics; Version 22, IBM Corp., Armonk, NY, USA). Since univariate statistics is more prone to noise compared to multivariate statistics, the 40 noisy variables were excluded. The analysis was accomplished by non-parametric testing (Kruskal-Wallis test for more than two groups or Mann-Whitney U test for two groups, $p<0.05$).

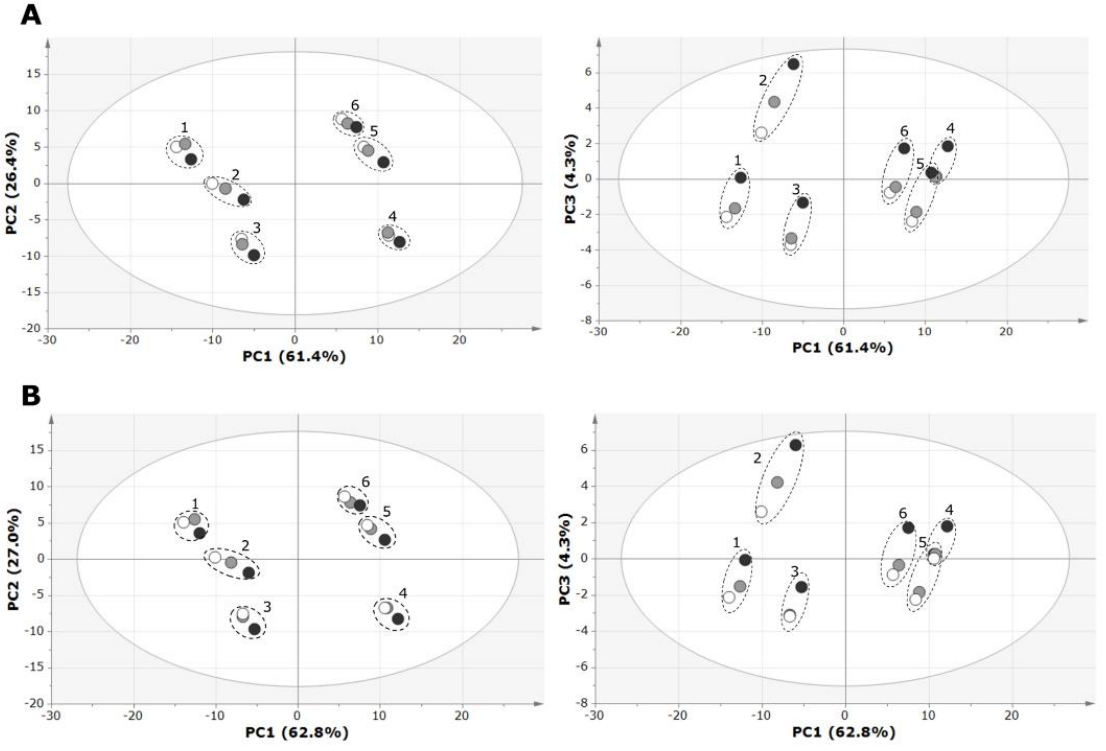


Figure 4.2. PCA score plots showing the influence of processing delay (time between blood collection and centrifugation) and made by using (A) all 110 integration values and (B) only the 70 'non-noisy' integration values of the plasma ^{1}H-NMR spectra. Plasma samples originating from blood processed after a delay of 30 min (B; ○, reference), 3 h (B; ●) and 8 h (B; ●). No significant differences are observed between the two plots. PC: principal component.

Results and Discussion

Nowadays, the powerful combination of analytical techniques and multivariate statistics is increasingly used to study differences between healthy and diseased subjects and to discover disease-related biomarkers in clinical metabolomics (24). Because clinical metabolomics seems to shed new light on (1) biochemical pathways involved in the etiology of diseases, (2) disease diagnosis and (3) new markers to judge therapy response (25), it is crucial to ensure its robustness, i.e. its reproducibility and accuracy. In other words, the variability should be under tight control in a clinical setting, ensuring that differences in the metabolic profile are resulting from the physiological status and not from differences in preanalytical sampling conditions such as collection, preparation and storage procedures (16, 26, 27). Preanalytical conditions which induce variation surpassing the inter-individual variation should be avoided (or samples collected under such conditions removed from the study).

Impact of a double LiHe concentration on the plasma metabolome

Due to time constraints or incorrect blood drawing in routine clinical practice, blood collection tubes can be under-filled leading to an increased concentration of the anticoagulant in the blood. To our knowledge, a possible influence on the plasma metabolome has not been investigated before. To this end, ^{1}H-NMR spectra of plasma obtained from half-filled blood tubes (double LiHe concentration) were compared with those from plasma obtained via a reference protocol (reference plasma) as described in **Figure 4.1**. Furthermore, we investigated a possible confounding time-dependent effect by subjecting both the reference and conditioned sample to a processing delay of 3 h and 8 h at 4°C. As shown in the PCA score plots of **Figure 4.3A**, the plasma metabolome of blood subjected to a double LiHe concentration could not be discriminated from that of reference plasma, also not after a processing delay of 8 h. As the inter-individual variation is clearly exceeding this preanalytical variation, it can be concluded that plasma samples originating from blood collected in LiHe tubes which are only half-filled are still reliable for ^{1}H-NMR metabolomics. On the other hand, a clear and systematic change is observed as a function of increasing processing delay which

can be attributed to alterations in the concentration of pyruvate and lactate (see below).

Impact of short-term exposure of blood to an oxidative atmosphere on the plasma metabolome

While blood collection using a vacuum collection tube holder is standard practice, it sometimes cannot be applied (e.g. in pediatric setting). Consequently, blood will be exposed to an oxidative atmosphere during transfer from syringe to tube, which might initiate specific enzymatic/chemical reactions (18). Here, we examined whether a short-term exposure of blood to an oxidative atmosphere induces changes in the plasma metabolome as compared to reference plasma (**Figure 4.1**). Furthermore, we investigated a possible confounding time-dependent effect by subjecting both the reference and conditioned sample to a processing delay of 3 h and 8 h at 4°C. As shown in the PCA score plots of **Figure 4.3B**, the plasma metabolome of blood subjected to an oxidative atmosphere could not be discriminated from that of reference plasma, also not after a processing delay of 8 h. As the inter-individual variation is clearly exceeding this preanalytical variation, it can be concluded that a short exposure of blood to an oxidative atmosphere does not affect the plasma ^{1}H-NMR metabolic profile. On the other hand, again a clear and systematic change is observed as a function of increasing processing delay which can be attributed to alterations in the concentration of pyruvate and lactate (see below).

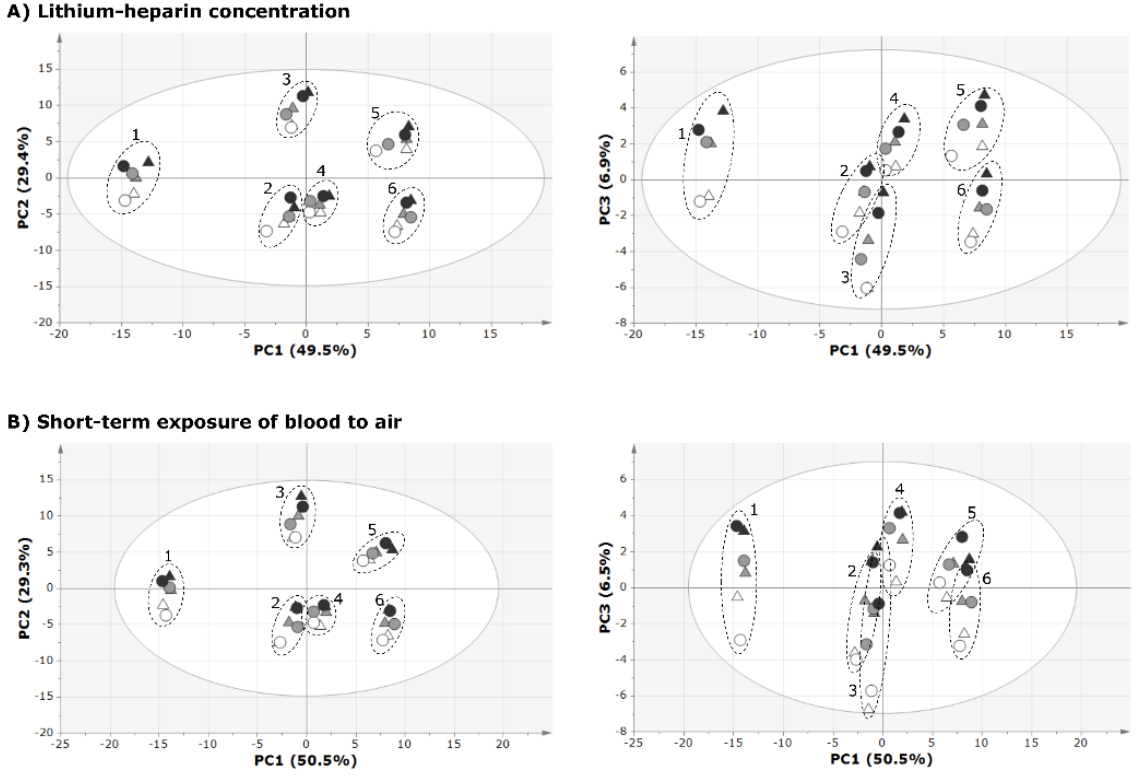


Figure 4.3. PCA score plots showing the influence of a double lithium-heparin (LiHe) concentration (A) and oxidative atmosphere (B), made by the 110 integration regions of the plasma ^{1}H-NMR spectra of study group 1 (n=6). Plasma samples originating from blood exposed to a double concentration of LiHe or oxidative atmosphere (Δ) as compared to reference plasma (○). Blood samples processed after a processing delay of 30 min (white; reference), 3 h (grey) and 8 h (black). PC: principal component.

Effect of hemolysis on the plasma metabolome

Hemolysis frequently occurs in clinical routine because of incorrect blood drawing techniques, e.g. improper choice of venipuncture site, prolonged tourniquet time, blood collection through a peripheral IV catheter or exposure to excessive heat or cold (22, 28, 29). The release of hemoglobin and intracellular species in the plasma due to red blood cell lysis affects several biochemical laboratory tests (30-32). Nevertheless, the effect on the human plasma ^{1}H-NMR metabolome has to our knowledge not yet been investigated. Hereto, moderate hemolysis (grade 1) and severe hemolysis (grade 2) were induced by exposing blood to excessive cold and the degree of hemolysis was defined on the basis of the concentration of free hemoglobin. No significant differences were observed between the reference and hemolytic plasma metabolomes (**Figure 4.4A**). This in contrast to Yin et al. who found 69 species to be significantly altered in moderate and severe hemolytic plasma by non-targeted LC-MS (16). However, as LC-MS requires a more extensive sample preparation and has increased sensitivity compared to NMR (16, 33, 34), it is consequently also more prone to preanalytical variation. In summary, we can conclude that hemolysis does not affect the plasma ^{1}H-NMR metabolic profile.

Impact of processing delay on the plasma metabolome

Because of clinical sample flow, it is often not possible to process blood immediately after collection. Moreover, samples have to be transferred to on- or offsite laboratories for analysis, which can take up to several hours. Therefore, the effect of an increasing processing delay between blood collection and centrifugation was examined, i.e. delays of 3 h and 8 h at 4°C were compared to the reference protocol (i.e. 30 min on ice). The PCA score plot shows a clear and systematic change as a function of increasing processing delay **(Figure 4.4B)**, which was already observed previously (cfr. **Figure 4.3A and 4.3B**). In order to find out which metabolites are responsible, univariate statistics was performed of which the outcome demonstrates that processing after 3 h instead of 30 min results in a significant decrease of the pyruvate signal between 2.405-2.399 ppm (VAR80: $p = 0.023$) (**Figure 4.5**). Processing after 8 h induces an additional rise of lactate signals between 4.175-4.111 ppm (VAR28: $p = 0.019$ and VAR29: $p = 0.028$) and between 1.374-1.345 ppm (VAR99: $p = 0.016$) next to a downward

trend in the glucose signal between 3.536-3.398 ppm (VAR54: p = 0.049). Presumably, these changes are attributable to a continued anaerobic cell metabolism due to contact with erythrocytes (35). For serum, Fliniaux et al. report no impact when blood is kept at 4°C during a processing delay from 4 h to 24 h but report changes in lactate and glucose concentrations upon storage at RT (26). In agreement with our findings, Bernini et al. report a decreased plasma concentration of pyruvate when blood is kept at 4°C and further confirm that preservation at 4°C causes less profound changes as compared to RT (18).

Nevertheless our results show that increasing the processing delay affects lactate, pyruvate and glucose concentrations, the inter-individual variation is clearly much larger than this preanalytical variation, indicating that the impact of a processing delay at 4°C of up to 8 h on multivariate cluster analysis will be minimal (for a variable to contribute significantly to the differentiating power of a statistical classifier which differentiates between groups of healthy and diseased subjects, its variation between the groups has to be larger than this within the groups). Nonetheless, we recommend to keep the time gap between blood collection and centrifugation similar for all samples within a study.

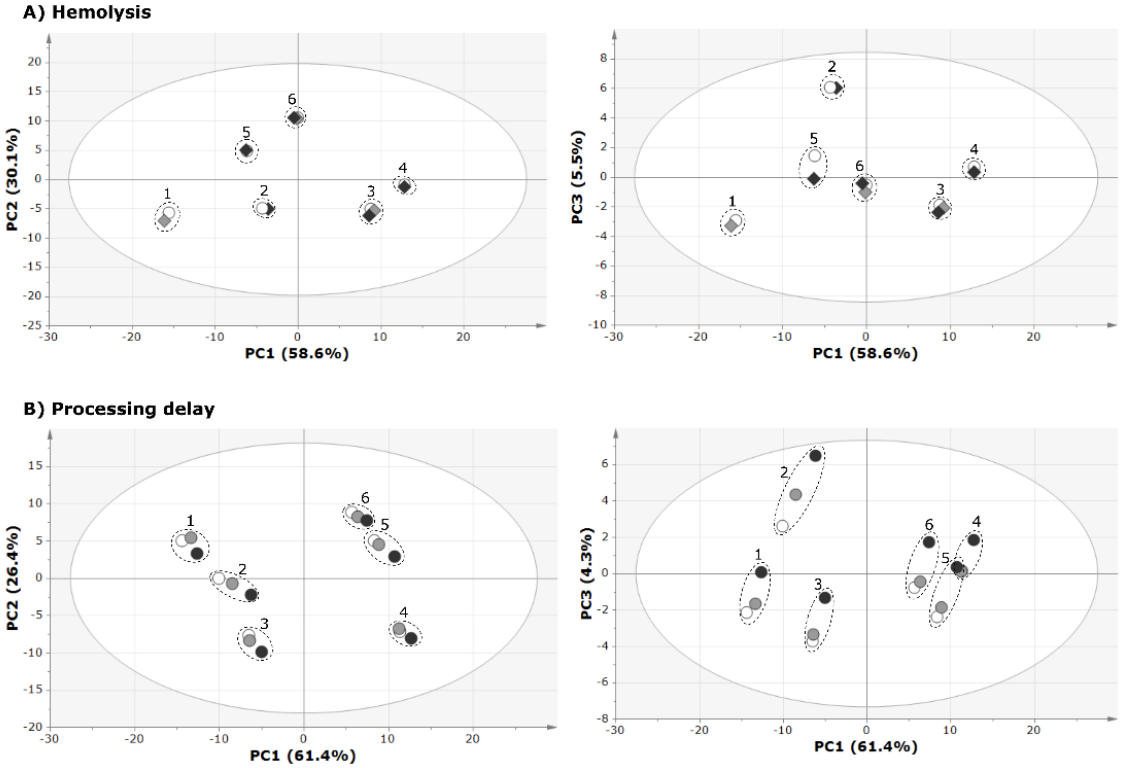


Figure 4.4. PCA score plots showing the influence of hemolysis (A) and processing delay (B), made by the 110 integration regions of the plasma ^{1}H-NMR spectra from study group 2 (n=6). Plasma samples without hemolysis (A; ○, reference), with hemolysis grade 1 (A; ♦) and hemolysis grade 2 (A; ♦). Plasma samples originating from blood processed after a delay of 30 min (B; ○, reference), 3 h (B; ●) and 8 h (B; ●). PC: principal component.

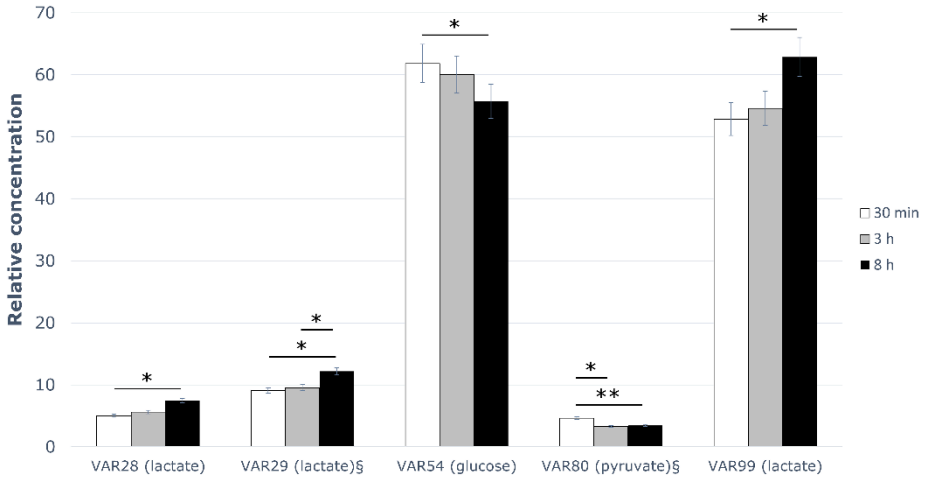


Figure 4.5. Relative concentrations of lactate, pyruvate and glucose in plasma originating from a cooled blood sample which is processed after 30 min, 3 h and 8 h. The mean value, obtained from all measurements per condition, are presented with an error bar 95% CI. §Relative concentrations of lactate (VAR29) and pyruvate (VAR80) were multiplied by 10 for a better representation. VAR: variable. * $P < 0.05$; ** $P = 0.005$.

Effect of centrifugation temperature on the plasma metabolome

To slow down enzymatic activity, it is standard practice in metabolomics to cool samples around 4°C during processing (18). However, since a refrigerated centrifuge is not always accessible, the impact of centrifugation at RT was examined. No significant difference was found between centrifugation at RT or 4°C (**Figure 4.6A**), indicating that 15 min centrifugation time is too short to induce changes in pre-cooled samples and therefore has no significant impact on the plasma ^{1}H-NMR metabolic profile.

Impact of freezing procedure on the plasma metabolome

In metabolomics, it is common practice to store plasma aliquots immediately at -80°C to ensure a quench of the metabolism and to allow ^{1}H-NMR measurements in larger sample series. However, since not all laboratories have a -80°C freezer, samples are often transported to another laboratory for processing and storage. Hereto, plasma samples are kept temporarily on dry ice (-78.5°C) or in LN_2 (-196°C). Alternatively, it is common practice in biobanks to snap-freeze plasma samples in LN_2 to quickly attenuate biochemical activity and to preserve structural

integrity (20). Nevertheless, and to our knowledge, the effect of different freezing conditions on the plasma ^{1}H-NMR metabolic profile has not been examined before. As shown in **Figure 4.6B**, our experimental data show no significant differences between plasma samples immediately stored at -80°C (reference plasma) and samples kept for 8 h on dry ice or in LN_2 before storage at -80°C.

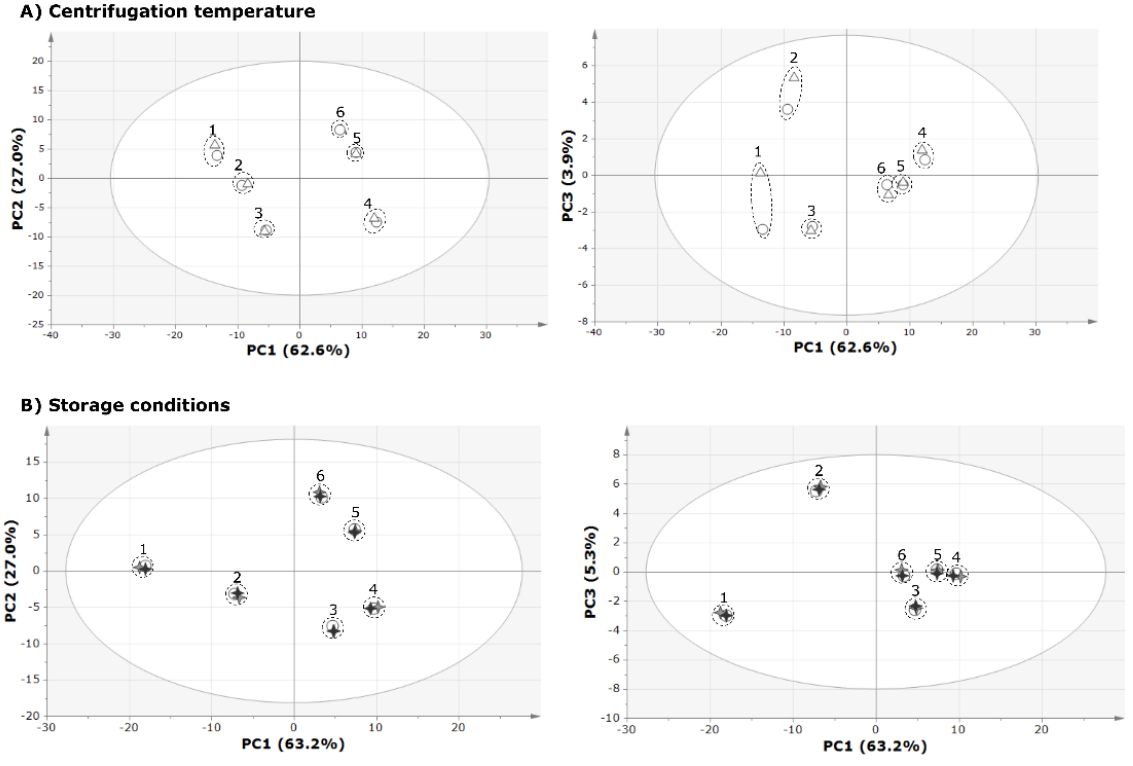


Figure 4.6. PCA score plots showing the influence of centrifugation temperature (A) and initial plasma freezing on dry ice or in LN_2 (B), made by the 110 integration regions of the plasma ^{1}H-NMR spectra from study group 2 (n=6). Plasma samples originating from blood centrifuged at 4°C (A; ○, reference) or at RT (A; Δ). Plasma samples stored directly at -80°C (B; ○, reference), and after a delay of 8 h on dry ice (B; ✦) or 8 h in LN_2 (B; ✦). PC: principal component.

Effect of storage duration at -80°C on the plasma metabolome

Often, the delay between sample storage and effective measurement can exceed several months, especially when samples are stored in biobanks. However, long-term storage might modify interactions between macromolecules and small molecules and consequently affect their plasma ^{1}H-NMR signals, introducing non-disease related artefacts (36). Additionally, long-term storage might induce a shift of some of the metabolite signals due to a change in pH (36). To investigate whether the duration of plasma storage at -80°C has an impact on ^{1}H-NMR results, we compared plasma of 10 controls stored at -80°C for two and ten months, respectively. No significant differences were detected as shown in **Figure 4.7**, indicating that plasma is stable at -80°C for at least 10 months, being in line with the findings of Deprez et al. (36) who demonstrated that rat plasma is stable for up to 6 months at -80°C. Furthermore, Pinto et al. showed that human plasma is stable for up to 30 months when stored at -80°C (27).

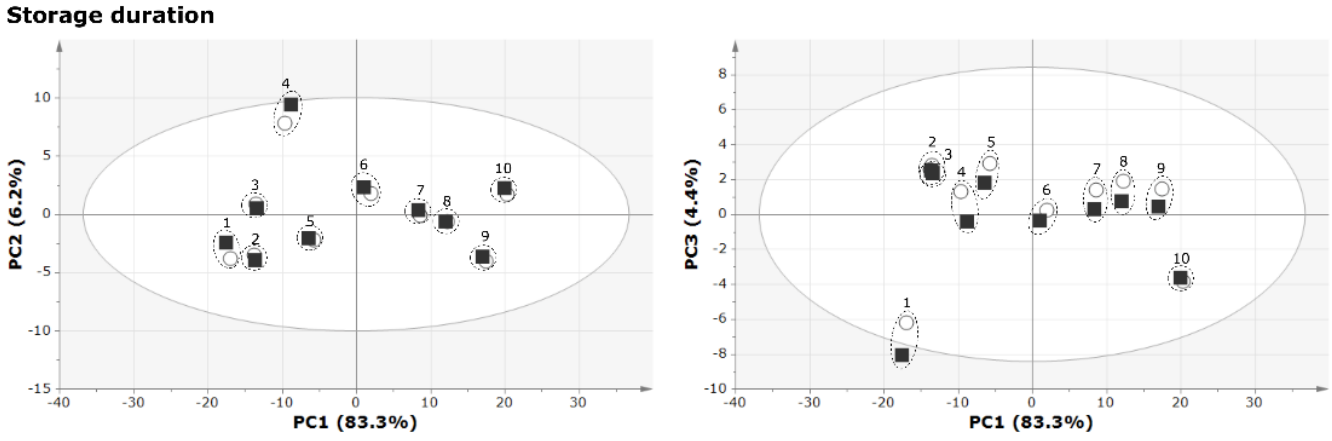


Figure 4.7. PCA score plots showing the influence of storage duration at -80°C, made by the 110 integration regions of the plasma ^{1}H-NMR spectra from study group 3 (n=10). Plasma samples stored for maximum 2 months (○, reference) and for 10 months (■) at -80°C. PC: principal component.

Evaluation of SPREC to document preanalytical variation in clinical ^{1}H-NMR-based metabolomics

SPREC was developed within the biobank environment to facilitate documentation and communication of the most important preanalytical quality parameters for different types of research biospecimens (19, 20). It allows to exclusively select samples that are fit for the purpose aimed for, while excluding samples that were subjected to unwanted, interfering preanalytical conditions. While it has become a standard within the biobank field, it is relatively unknown in other clinical settings and its usefulness therein remains to be determined. However, its implementation in clinical laboratories is rather straightforward as the 7 elements of SPREC can easily be extracted from the Laboratory Information Management System.

We annotated the samples from our study with SPREC version 2 and evaluated if this is in agreement with our experimental findings and of practical value in clinical metabolomics (20). As shown in **Figure 4.1** and **Table 4.1**, the reference condition is translated as single spun plasma samples – primary container with LiHe (without gel) – pre-centrifugation delay 2-10°C < 2 h – centrifugation at 2-10°C 10-15 min < 3000 g (with braking) – no second centrifugation – post-centrifugation delay < 1 h 2-10°C – long-term storage in cryovial 1 to 2 ml at -85°C to -65°C, which is encoded as PL1-HEP-B-D-N-A-D. Hemolysis, exposure to oxygen and insufficient tube filling (or increased LiHe concentration) are not contained within the current version of SPREC, resulting in an identical code as the reference condition. This is not disadvantageous, however, as these variations did not introduce significant differences in the plasma ^{1}H-NMR metabolic profile. Differences in centrifugation temperature and freezing method are encoded within SPREC version 2, but also did not induce differences in the ^{1}H-NMR metabolome. Interestingly, the processing delays of 3 h and 8 h, which induce a systematic variation in the plasma metabolome, are discriminated by the SPREC codes, i.e. PL1-HEP-**D**-D-N-A-D and PL1-HEP-**F**-D-N-A-D, respectively. This means that SPREC allows to select plasma samples with a known time gap between blood collection and centrifugation for metabolomics applications. To further test the potential of SPREC, we expanded this evaluation to samples which were subjected to other preanalytical conditions as described by Pinto et al. (27). Again, SPREC

identifies the conditions that affect the plasma ^{1}H-NMR metabolic profile (**Table 4.1**). In addition to plasma, Fliniaux et al. previously introduced SPREC into the field of serum ^{1}H-NMR analysis for biobanks (26). These combined plasma and serum studies clearly illustrate the value of SPREC to encode relevant preanalytical conditions and fully support its implementation in clinical metabolomics.

Table 4.1. SPREC annotation for preanalytical conditions tested in plasma samples by ^{1}H-NMR spectroscopy.

Condition	SPREC	Impact on profile?
Reference sample	PL1-HEP-B-D-N-A-D	No
Double LiHe concentration	PL1-HEP-B-D-N-A-D	No
Oxygen exposure	PL1-HEP-B-D-N-A-D	No
Hemolysis	PL1-HEP-B-D-N-A-D	No
Processing delay 3 h at 4°C	PL1-HEP-**D**-D-N-A-D	Yes
Processing delay 8 h at 4°C	PL1-HEP-**F**-D-N-A-D	Yes
Centrifugation at RT	PL1-HEP-B-**B**-N-A-D	No
Freezing on dry ice	PL1-HEP-B-D-N-Z-D	No
Freezing in LN_2	PL1-HEP-B-D-N-Z-D	No
Storage at -80°C (for 10 months)	PL1-HEP-B-D-N-A-D	No (storage time$^{\$}$)
Reference sample Pinto et al.*	PL1-HEP-A^{1}-D-N-X-A^{2}	No
EDTA additive	PL1-**SED**-A^{1}-D-N-X-A^{2}	Yes
Processing delay 2,5 h - 21 h at RT*	PL1-HEP-**C/E/G/I**-D-N-X- A^{2}	Yes
Storage at -20°C (for 1 month)*	PL1-HEP-A^{1}-D-N-X-**B**^{2}	Yes: storage temperature (storage time$^{\$}$)
Freeze/thaw cycles*	PL1-HEP-A^{1}-D-N-X-A^{2}	Yes$^{\$}$
Non-fasting donor*	PL1-HEP-A^{1}-D-N-X-A^{2}	No$^{\$}$

*: conditions obtained from Pinto et al. (27); ^{1}: assumption preprocessing delay at RT; ^{2}: assumption storage in standard polypropylene tube; $^{\$}$: beyond scope of SPREC as no true preanalytical condition; impacting elements of SPREC are indicated in bold. EDTA: ethylenediaminetetraacetic acid; LiHe: lithium-heparin; LN_2: liquid nitrogen; RT: room temperature; SPREC: Standard PREanalytical Code.

Conclusions

Since metabolomics is gaining increasing interest for clinical biomarker research, knowledge of its sensitivity to variations in sample collection, processing and storage becomes essential in the discussions regarding its clinical implementation. Regarding human blood plasma, our results show no significant impact of a double LiHe concentration, a short-term exposure to an oxidative atmosphere, hemolysis, centrifugation temperature, freezing procedure and storage duration at -80°C on the ^{1}H-NMR metabolic profile. Only increasing the processing delay from 30 min to 3 h and 8 h has a significant impact on the plasma concentration of pyruvate, lactate and glucose. Nevertheless, as the inter-individual variation is much larger than this preanalytical variation, the impact on multivariate group classification will be minimal. Nonetheless, we recommend to keep the time gap between blood collection and centrifugation similar for all samples within a study. Hereto, the implementation of SPREC within clinical metabolomics allows for an appropriate sample encoding and exclusion of samples that were subjected to unwanted, interfering preanalytical conditions. Without doubt, it will contribute to the validation of ^{1}H-NMR metabolomics in clinical, biobank and multicenter research settings.

References

1. Brindle JT, Antti H, Holmes E, Tranter G, Nicholson JK, Bethell HW, et al. Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics. Nature medicine. 2002;8(12):1439-44.
2. Fiehn O. Metabolomics--the link between genotypes and phenotypes. Plant molecular biology. 2002;48(1-2):155-71.
3. Madsen RK, Lundstedt T, Gabrielsson J, Sennbro CJ, Alenius GM, Moritz T, et al. Diagnostic properties of metabolic perturbations in rheumatoid arthritis. Arthritis research & therapy. 2011;13(1):R19.
4. Shen J, Yan L, Liu S, Ambrosone CB, Zhao H. Plasma metabolomic profiles in breast cancer patients and healthy controls: by race and tumor receptor subtypes. Translational oncology. 2013;6(6):757-65.
5. Zhang A, Sun H, Wang X. Power of metabolomics in biomarker discovery and mining mechanisms of obesity. Obesity reviews. 2013;14(4):344-9.
6. Van der Greef J, Tas A, Bouwman J, Ten Noever de Brauw M, Schreurs W. Evaluation of field-desorption and fast atom-bombardment mass spectrometric profiles by pattern recognition techniques. Analytica Chimica Acta. 1983;150:45-52.
7. Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica. 1999;29(11):1181-9.
8. Lindon JC, Nicholson JK. Spectroscopic and statistical techniques for information recovery in metabonomics and metabolomics. Annu Rev Anal Chem. 2008;1:45-69.
9. Teahan O, Gamble S, Holmes E, Waxman J, Nicholson JK, Bevan C, et al. Impact of analytical bias in metabonomic studies of human blood serum and plasma. Anal Chem. 2006;78(13):4307-18.
10. World Health Organization. Use of anticoagulants in diagnostic laboratory investigations. Geneva: 2002.
11. Nicholson JK, Buckingham MJ, Sadler PJ. High resolution 1H-NMR studies of vertebrate blood and plasma. Biochem J. 1983;211:605-15.
12. Nicholson JK, Foxall PJ, Spraul M, Farrant RD, Lindon JC. 750 MHz 1H and 1H-13C NMR spectroscopy of human blood plasma. Anal Chem. 1995;67(5):793-811.
13. Dunn WB, Broadhurst D, Ellis DI, Brown M, Halsall A, O'Hagan S, et al. A GC-TOF-MS study of the stability of serum and urine metabolomes during the UK Biobank sample collection and preparation protocols. Int J Epidemiol. 2008;37(suppl 1):i23-i30.
14. Wood JT, Williams JS, Pandarinathan L, Courville A, Keplinger MR, Janero DR, et al. Comprehensive profiling of the human circulating endocannabinoid metabolome: clinical sampling and sample storage parameters. Clin Chem Lab Med. 2008;46(9):1289-95.
15. Yang W, Chen Y, Xi C, Zhang R, Song Y, Zhan Q, et al. Liquid Chromatography–Tandem Mass Spectrometry-Based Plasma Metabonomics Delineate the Effect of Metabolites' Stability on Reliability of Potential Biomarkers. Anal Chem. 2013;85(5):2606-10.
16. Yin P, Peter A, Franken H, Zhao X, Neukamm SS, Rosenbaum L, et al. Preanalytical aspects and sample quality assessment in metabolomics studies of human blood. Clin Chem. 2013;59(5):833-45.
17. Kamlage B, Maldonado SG, Bethan B, Peter E, Schmitz O, Liebenberg V, et al. Quality Markers Addressing Preanalytical Variations of Blood and Plasma Processing Identified by Broad and Targeted Metabolite Profiling. Clin Chem. 2014;60(2):399-412.
18. Bernini P, Bertini I, Luchinat C, Nincheri P, Staderini S, Turano P. Standard operating procedures for pre-analytical handling of blood and urine for metabolomic studies and biobanks. J Biomol NMR. 2011;49(3-4):231-43.
19. Betsou F, Lehmann S, Ashton G, Barnes M, Benson EE, Coppola D, et al. Standard preanalytical coding for biospecimens: defining the sample PREanalytical code. Cancer Epidem Biomar. 2010;19(4):1004-11.

20. Lehmann S, Guadagni F, Moore H, Ashton G, Barnes M, Benson E, et al. Standard preanalytical coding for biospecimens: review and implementation of the Sample PREanalytical Code (SPREC). Biopreservation and biobanking. 2012;10(4):366-74.
21. World Health Organization. WHO guidelines on drawing blood: best practices in phlebotomy. Geneva: 2010.
22. Kroll MH, Elin RJ. Interference with clinical laboratory analyses. Clin Chem. 1994;40(11):1996-2005.
23. Louis E, Bervoets L, Reekmans G, De Jonge E, Mesotten L, Thomeer M, et al. Phenotyping human blood plasma by ^{1}H-NMR: a robust protocol based on metabolite spiking and its evaluation in breast cancer. Metabolomics. 2015;11:225-36.
24. Issaq HJ, Van QN, Waybright TJ, Muschik GM, Veenstra TD. Analytical and statistical approaches to metabolomics research. J Sep Sci. 2009;32(13):2183-99.
25. Collino S, Martin FP, Rezzi S. Clinical metabolomics paves the way towards future healthcare strategies. Br J Clin Pharmacol. 2013;75(3):619-29.
26. Fliniaux O, Gaillard G, Lion A, Cailleu D, Mesnard F, Betsou F. Influence of common preanalytical variations on the metabolic profile of serum samples in biobanks. J Biomol NMR. 2011;51(4):457-65.
27. Pinto J, Domingues MRM, Galhano E, Pita C, do Céu Almeida M, Carreira IM, et al. Human plasma stability during handling and storage: impact on NMR metabolomics. Analyst. 2014:1168-77.
28. Kennedy C, Angermuller S, King R, Noviello S, Walker J, Warden J, et al. A comparison of hemolysis rates using intravenous catheters versus venipuncture tubes for obtaining blood samples. J Emerg Nurs. 1996;22(6):566-9.
29. Burns ER, Yoshikawa N. Hemolysis in serum samples drawn by emergency department personnel versus laboratory phlebotomists. Laboratory Medicine. 2002;33(5):378-80.
30. Koseoglu M, Hur A, Atay A, Cuhadar S. Effects of hemolysis interferences on routine biochemistry parameters. Biochem Med. 2011;21(1):79-85.
31. Lippi G, Salvagno GL, Montagnana M, Brocco G, Guidi GC. Influence of hemolysis on routine clinical chemistry testing. Clin Chem Lab Med. 2006;44(3):311-6.
32. Yucel D, Dalva K. Effect of in vitro hemolysis on 25 common biochemical tests. Clin Chem. 1992;38(4):575-7.
33. Kentgens A, Bart J, Van Bentum P, Brinkmann A, Van Eck E, Gardeniers J, et al. High-resolution liquid-and solid-state nuclear magnetic resonance of nanoliter sample volumes using microcoil detectors. J Chem Phys. 2008;128(5):052202.
34. Felli IC, Brutscher B. Recent advances in solution NMR: fast methods and heteronuclear direct detection. Chemphyschem. 2009;10(9-10):1356-68.
35. Baynes JW, Dominiczak MH. Medical Biochemistry. Amsterdam: Elsevier; 2010.
36. Deprez S, Sweatman BC, Connor SC, Haselden JN, Waterfield CJ. Optimisation of collection, storage and preparation of rat plasma for 1H NMR spectroscopic analysis in toxicology studies to determine inherent variation in biochemical profiles. J Pharm Biomed Anal. 2002;30(4):1297-310.

CHAPTER 5

^{1}H-NMR-based metabolomics of blood plasma allows to detect lung cancer

Based on:

Evelyne Louis, Peter Adriaensens, Wanda Guedens, Theophile Bigirumurame, Kurt Baeten, Karolien Vanhove, Kurt Vandeurzen, Karen Darquennes, Johan Vansteenkiste, Christophe Dooms, Ziv Shkedy, Liesbet Mesotten, Michiel Thomeer. Detection of lung cancer via metabolic changes in blood plasma. Under revision.

Abstract

Background. Lung cancer constitutes a major public health problem. Low-dose computed tomography, the currently used tool for lung cancer screening, is characterized by a high rate of false positive results. Accumulating evidence has shown that cancer cell metabolism differs from that of normal cells. Therefore, this study aims to evaluate whether the metabolic phenotype of blood plasma allows to detect lung cancer.

Patients and methods. The proton nuclear magnetic resonance spectrum of plasma is divided into 110 integration regions, representing the metabolic phenotype. These integration regions reflect the relative metabolite concentrations and were used to train a classification model in discriminating between 233 lung cancer patients and 226 controls. The validity of the model was examined by permutation testing and by classifying an independent validation cohort of 98 lung cancer patients and 89 controls.

Results. The model allows to classify 78% of the lung cancer patients and 92% of the controls correctly with an area under the curve of 0.88. Important moreover is that the model is convincing which is demonstrated by validation in the independent cohort with a sensitivity of 71%, a specificity of 81% and an area under the curve of 0.84. The most discriminating variables indicate that the disturbed biochemical pathways include i) an elevated hepatic glycogenolysis, gluconeogenesis and ketogenesis, ii) an impaired Krebs cycle in hepatocytes and cancer cells and iii) an enhanced membrane synthesis in cancer cells. The limited number of patients in the subgroups does not (yet) enable to differentiate between histological subtypes and tumor stages.

Conclusion. Metabolic phenotyping of plasma allows to detect lung cancer and to identify the metabolic changes involved.

Introduction

Lung cancer is the leading cause of cancer death worldwide with an overall five-year survival of only circa 15%, mainly because it is often only diagnosed in a late stage of the disease course (1-4). This highlights the importance of screening tools that allow to detect lung cancer as early as possible. A promising screening tool is low-dose computed tomography (LDCT), which has been shown to reduce lung cancer mortality by 20% as compared to chest radiography screening (5). However, LDCT screening has some disadvantages such as the high cost associated with screening all patients at risk according to current risk models, radiation exposure and the low positive predictive value (PPV) (4, 6). Because of these limitations, other non-invasive detection platforms are being evaluated, all with their advantages and shortcomings (7).

This study aims to evaluate the role of metabolomics, an upcoming tool in the field of oncology, in the diagnosis of lung cancer (8-10). Over the past decade, accumulating evidence has shown that cancer cell metabolism differs from that of normal cells (11-13). More specifically, the entire metabolism of cancer cells is reprogrammed to promote cell proliferation and survival. Metabolic reprogramming during cancer development is driven by aberrant signaling pathways induced by the activation of oncogenes or the inactivation of tumor suppressor genes (14). One of the main adaptations of cancer cells is that, even in the presence of normal oxygen levels, they rely on anaerobic energy production through glycolysis, a hallmark known as the Warburg effect (15). The main advantage of predominantly relying on glycolysis for production of energy and essential building blocks is that many glycolytic intermediates can be shunted to anabolic pathways, thereby preserving cancer cell proliferation (13). As metabolites are the end products of cellular processes, changes in their concentration reflect alterations in the metabolic phenotype (16). This explains the interest in metabolites as attractive cancer biomarkers (17). Proton nuclear magnetic resonance (^{1}H-NMR) spectroscopy allows a fast (<10 min), non-invasive identification and quantification of complex mixtures of metabolites, as appearing in plasma (9, 10, 18).

The present study aims to 1) investigate whether the ^{1}H-NMR derived metabolic phenotype of blood plasma allows to discriminate between lung cancer patients

and controls, 2) evaluate the predictive accuracy of the trained classification model in an independent cohort, 3) improve the understanding of the disturbed biochemical pathways in lung cancer and 4) examine whether the metabolic phenotype allows to discriminate between histological subtypes and clinical tumor stages.

Materials and Methods

Subjects

Lung cancer patients (n=357) were included in the Limburg Positron Emission Tomography center (n=273) (Hasselt, Belgium) and at the Department of Respiratory Medicine of University Hospitals Leuven (n=84) (Leuven, Belgium) from March 2011 to June 2014. The diagnosis was confirmed by a pathological biopsy or a clinician specialized in interpreting radiological and clinical lung cancer data. Clinical staging of the tumors was performed according to the 7th edition of the tumor, node, metastasis classification (19). Controls (n=347) were patients with non-cancer diseases who were included at Ziekenhuis Oost-Limburg (Genk, Belgium) between March 2012 and June 2014. For both groups, blood sampling and sample preparation was done according to a fixed protocol and by trained staff.

Exclusion criteria were: 1) not fasted for at least 6 h; 2) fasting blood glucose concentration ≥ 200 mg/dl; 3) medication intake on the morning of blood sampling and 4) treatment or history of cancer in the past 5 years. The study was conducted in accordance with the ethical rules of the Helsinki Declaration and Good Clinical Practice and was approved by the ethical committees of Ziekenhuis Oost-Limburg, Hasselt University (Hasselt, Belgium) and University Hospitals Leuven. All study participants provided written informed consent. The study is registered at clinical trials.gov (NCT02024113).

Both groups were subdivided into a training cohort and a validation cohort (**Figure 5.1**). More specifically, 250 out of the 357 lung cancer patients and 250 out of the 347 controls were randomly assigned to the training cohort, leaving a validation cohort of 107 lung cancer patients and 97 controls. Forty-one statistical outliers of the training cohort (17 lung cancer patients and 24 controls) and 17 of the validation cohort (9 lung cancer patients and 8 controls) were excluded. According to their medical files, they showed abnormal concentrations of glucose, lipids or ketone bodies. The individuals with high glucose level were diagnosed with diabetes or had an increased fasting glucose while those with high lipid levels suffered from obesity, hyperlipidemia or took cholesterol-lowering medication. Most of the individuals with high ketone body levels had a low BMI.

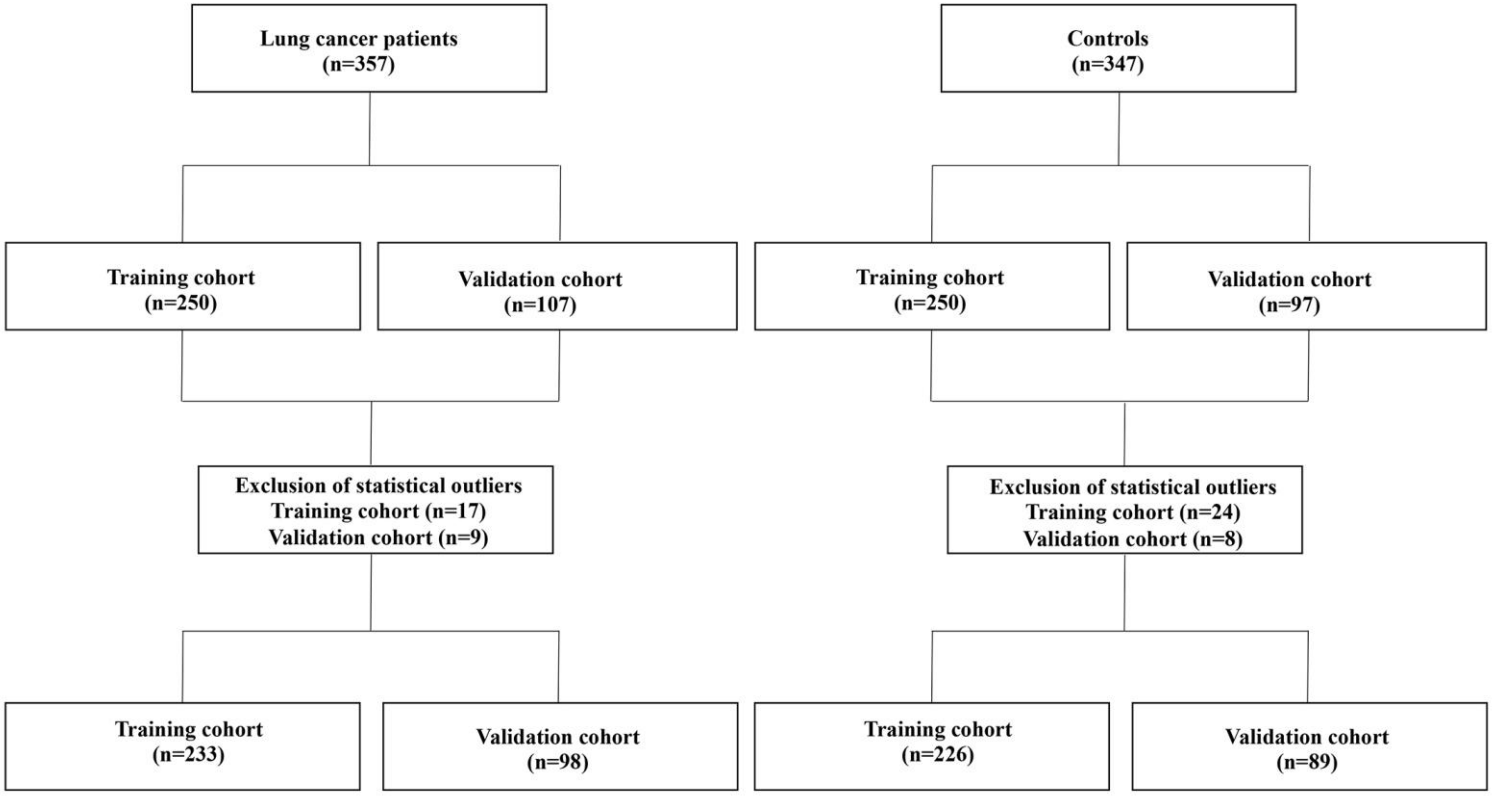


Figure 5.1. CONSORT diagram of the study.

Blood sampling and processing

See **Chapter 2**: Blood sampling and processing.

NMR sample preparation and analysis

See **Chapter 2**: NMR sample preparation and analysis.

Spectral processing

See **Chapter 2**: Spectral processing.

Statistical analysis

Multivariate statistics was performed using SIMCA-P+ (Version 14, Umetrics, Umea, Sweden). After mean-centering and Pareto scaling of the variables, unsupervised principal component analysis (PCA) was performed to identify outliers by means of a Hotelling's T2 range test and a distance to model plot. After removing outliers, supervised orthogonal partial least squares discriminant analysis (OPLS-DA) was used to train a classification model (20). The validity of the model was confirmed by i) permutation testing, ii) classifying an independent cohort with a classification cut-off value of 0.5, iii) receiver operating characteristic curve explorer & tester (21) and last but not least by comparison with the outcome of an independent model constructed by means of partial least squares discriminant analysis (PLS-DA) (R Version 3.1.2, Vienna, Austria). Via an S-plot, the most discriminating variables and their variable importance for the projection (VIP) values were identified (22). Variables with a VIP value exceeding 0.5 were used to describe the disturbed biochemical pathways. Additionally, a student t-test with correction for multiple testing by the Benjamini-Hochberg method (23) was applied (IBM SPSS Version 22.0, Chicago, Illinois, USA). Note that noisy variables, which were identified in **Chapter 2**, were not excluded since none of the noisy variables had a VIP value exceeding 0.5 and therefore, were not selected to explain the disturbed biochemical pathways.

Results

Multivariate OPLS-DA statistics was used to train a classification model (classifier) in discriminating between lung cancer patients and controls based on data input from their metabolic phenotype. The resulting model was validated in an independent cohort. **Table 5.1** shows the characteristics of the training and validation cohorts. However, before starting with the OPLS-DA statistics, a PCA analysis was conducted to look for possible confounders. **Figure 5.2** shows PCA score plots stained for disease (A), gender (B), smoking habits (C) and chronic obstructive pulmonary disease (COPD) (D), demonstrating that smoking habits and gender are no confounders. Regarding COPD, the PCA results cannot be interpreted straightforward but OPLS-DA confirmed that it is no confounder either (see later).

Table 5.1. Characteristics of the subjects included in the study (without outliers).

	Training cohort		Validation cohort	
	C	**LC**	**C**	**LC**
Number of subjects, n	226	233	89	98
Gender, n (%)				
Male	119 (53)	160 (69)	44 (49)	66 (67)
Female	107 (47)	73 (31)	45 (51)	32 (33)
Age, yrs	67 ± 11	68 ± 10	69 ± 10	64 ± 9
(range)	(38-88)	(36-88)	(47-89)	(45-83)
BMI, kg/m^2	28.3 ± 5.0	25.8 ± 4.5	28.4 ± 5.7	26.2 ± 4.7
(range)	(18.7-46.7)	(17.5-41.8)	(16.2-52.0)	(16.8-38.5)
COPD, n (%)	39 (17)	119 (51)	9 (10)	35 (36)
Taking lipid-lowering medication, n (%)	124 (55)	122 (52)	56 (63)	39 (40)
Diabetes, n (%)	23 (10)	40 (17)	20 (22)	12 (12)
Smoking habits				
Smoker, n (%)	47 (21)	113 (49)	15 (17)	48 (49)
Ex-smoker, n (%)	102 (45)	110 (47)	36 (40)	46 (47)
Non-smoker, n (%)	77 (34)	10 (4)	38 (43)	4 (4)
Pack years	16 ± 24	33 ± 21	13 ± 18	38 ± 21
(range)	(0-175)	(0-125)	(0-60)	(0-150)
Laterality				
Left, n (%)		103 (44)		40 (41)
Right, n (%)		119 (51)		54 (55)
Bilateral, n (%)		6 (3)		4 (4)
Unknown, n (%)		5 (2)		0 (0)
Amount of tumors, n		239		102
Histological subtype				
NSCLC-Adenocarcinoma, n (%)		91 (38)		46 (45)
NSCLC-Spinocellular carcinoma, n (%)		66 (28)		29 (28)
NSCLC-Adenosquamous carcinoma, n (%)		5 (2)		1 (1)
NSCLC-Carcinoid, n (%)		5 (2)		0 (0)
NSCLC-NOS, n (%)		8 (3)		6 (6)
SCLC, n (%)		30 (13)		15 (15)
Unknown, n (%)		34 (14)		5 (5)
Clinical stage according to 7th TNM edition				
IA, n (%)		55 (23)		12 (12)
IB, n (%)		21 (9)		5 (5)
IIA, n (%)		11 (5)		7 (7)
IIB, n (%)		15 (6)		4 (4)
IIIA, n (%)		48 (20)		17 (16)
IIIB, n (%)		26 (11)		12 (12)
IV, n (%)		63 (26)		45 (44)

Data are presented as mean ± standard deviation and range, unless otherwise indicated. The outliers (41 of the training cohort and 24 of the validation cohort) are not included in this table. BMI: body mass index; C: controls; COPD, chronic obstructive pulmonary disease; LC: lung cancer; NOS: not otherwise specified; NSCLC: non-small cell lung cancer; SCLC: small cell lung cancer; TNM: tumor, node, metastasis.

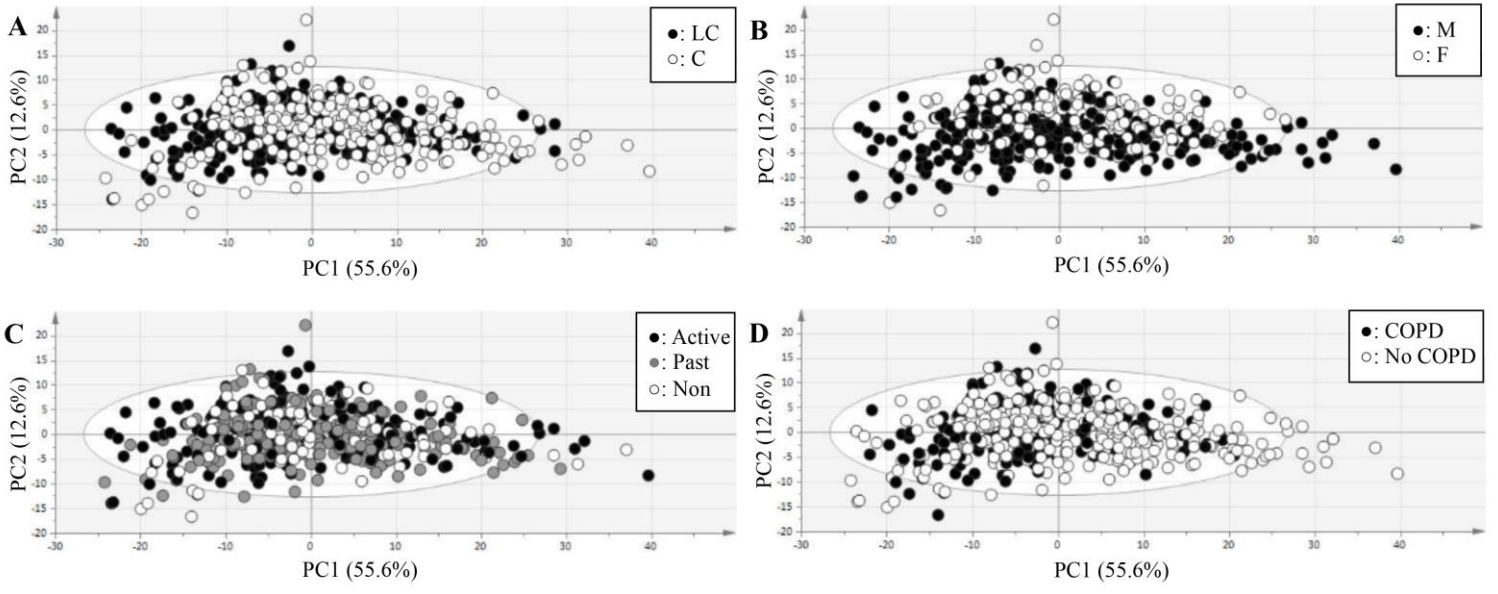


Figure 5.2. PCA score plot of all subjects (357 lung cancer patients and 347 controls) stained for (A) disease, (B) gender, (C) smoking habits, and (D) COPD. C: controls; COPD: chronic obstructive pulmonary disease; F: females; LC: lung cancer; M: males; PC: principal component.

Applying OPLS-DA statistics on the training set resulted in a model that allows to classify 78% of the 233 lung cancer patients and 92% of the 226 controls correctly (**Figure 5.3A-B**). An independent permutation test confirms that the classification is not due to overfitting (**Figure 5.4**). The predictive accuracy of the model was assessed by applying it to the independent validation cohort consisting of 98 lung cancer patients and 89 controls (**Figure 5.3B-C**), resulting in a sensitivity of 71% and a specificity of 81%. The sensitivity, specificity, area under the curve (AUC), positive predictive value (PPV) and negative predictive value (NPV) of this model as well as all other models are shown in **Table 5.2**. To further confirm, the dataset was analyzed independently (by the research group of biostatistics) using PLS-DA. The resulting classifier permits to classify 82% of the lung cancer patients and 89% of the controls correctly. Regarding the independent validation, the classifier was able to discriminate between the lung cancer patients and controls with a sensitivity of 75% and a specificity of 82%. In order to exclude COPD as a confounder, an OPLS-DA model was trained in discriminating between subjects with (n=158) and without (n=301) COPD (**Figure 5.5**). The extremely poor predictive accuracy of the resulting model (Q2: 0.08) confirms that COPD is no confounder.

Table 5.2. Characteristics of the trained (O)PLS-DA classification models.

	LV (P+O)	R^2X (cum)	R^2Y (cum)	Q^2 (cum)	Sens (%)	Spec (%)	PPV (%)	NPV (%)	AUC
Training cohort									
LC vs. C (OPLS-DA)	6 (1+5)	0.864	0.477	0.391	78	92	91	80	0.88
LC vs. C (PLS-DA)	6 (1+5)	/	/	/	82	89	89	82	/
Stage I LC vs. C	3 (1+2)	0.651	0.378	0.286	74	78	75	77	0.79
Validation cohort									
LC vs. C (OPLS-DA)	/	/	/	/	71	81	80	72	0.84
LC vs. C (PLS-DA)	/	/	/	/	75	82	82	75	/

AUC: area under the curve; C: controls; LC: lung cancer; LV: latent variable; NPV: negative predictive value; O: number of orthogonal components; OPLS-DA: orthogonal partial least squares-discriminant analysis; P: number of predictive components; PLS-DA: partial least squares-discriminant analysis; PPV: positive predictive value; R^2X(cum): total explained variation in X; R^2Y(cum): total explained variation in Y; Sens: sensitivity; Spec: specificity; Q^2(cum): predicted variation as determined by 7-fold cross-validation.

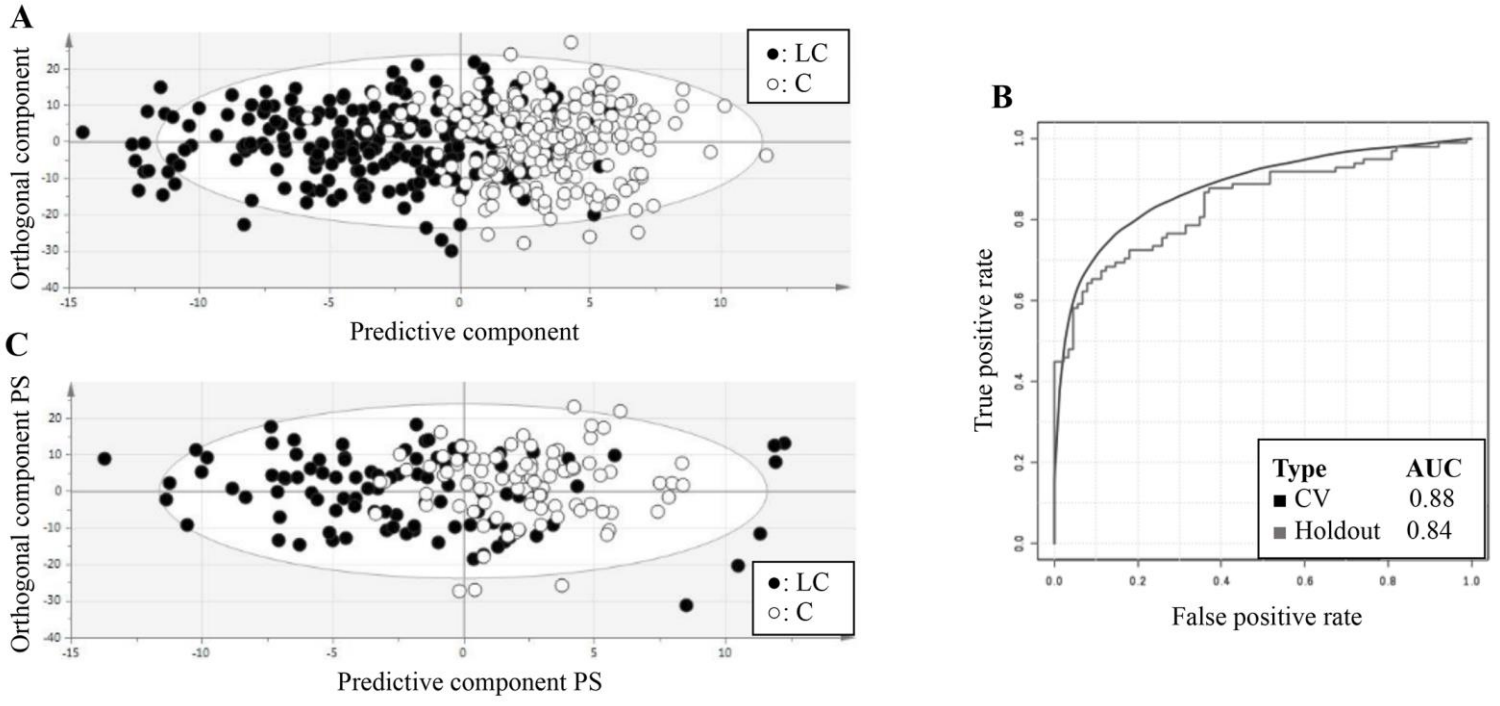


Figure 5.3. (A) OPLS-DA score plot derived from the training cohort of 226 controls and 233 lung cancer patients, (B) ROC curves showing for the cross-validation (CV) of the training cohort as well as for the independent validation the high predictive accuracy of the OPLS-DA model. Regarding the CV, sub-sampling CV was used, i.e. 2/3 of the observations were used for model training and the remaining 1/3 of the observations were used for validation. This procedure was repeated 50 times. (C) OPLS-DA score plot for the classification of the independent cohort of 89 controls and 98 lung cancer patients by means of the trained classifier. AUC: area under the curve; C: controls; CV: cross-validation; LC: lung cancer; PS: predicted scores.

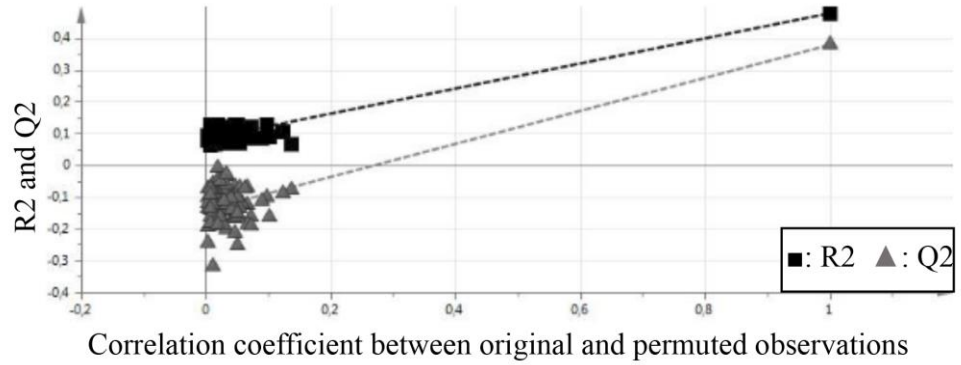


Figure 5.4. Result of the permutation test demonstrating that the obtained classification model is not the result of overfitting. All resulting R^2 and Q^2 values (at the left) are lower than these of the model (at the right), indicative for a valid model. R^2: explained variation; Q^2: predicted variation as determined by 7-fold cross-validation.

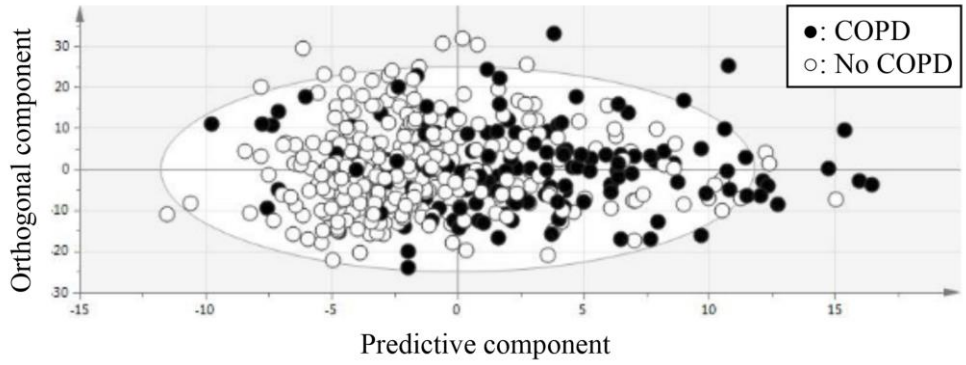


Figure 5.5. **OPLS-DA score plot of patients with and without COPD**. COPD: chronic obstructive pulmonary disease.

In an attempt to explain the disturbed biochemical pathways in lung cancer, only the 45 most discriminating variables, i.e. these having a VIP value exceeding 0.5, were selected (**Figure 5.6**). **Tables 5.3 and 5.4** show the variables (and contributing metabolites) which are increased versus decreased for lung cancer patients together with their univariate p-value and corresponding fold change (FC). A positive/negative FC denotes a relatively higher/lower concentration in plasma of lung cancer patients. The metabolites of which the concentration is increased are glucose, glycerol, N-acetylated glycoproteins, β-hydroxybutyrate, leucine, tyrosine, threonine, glutamine, valine and aspartate whereas the metabolites of which the concentration is decreased include alanine, lactate, sphingomyelin and phosphatidylcholine (and other cholinated phospholipids),

citrate and other phospholipids. Remark that sphingomyelin and phosphatidylcholine can be discriminated from the other phospholipids on the basis of the strong singlet signal in the NMR spectrum of the nine protons of the three methyl groups of the choline head group.

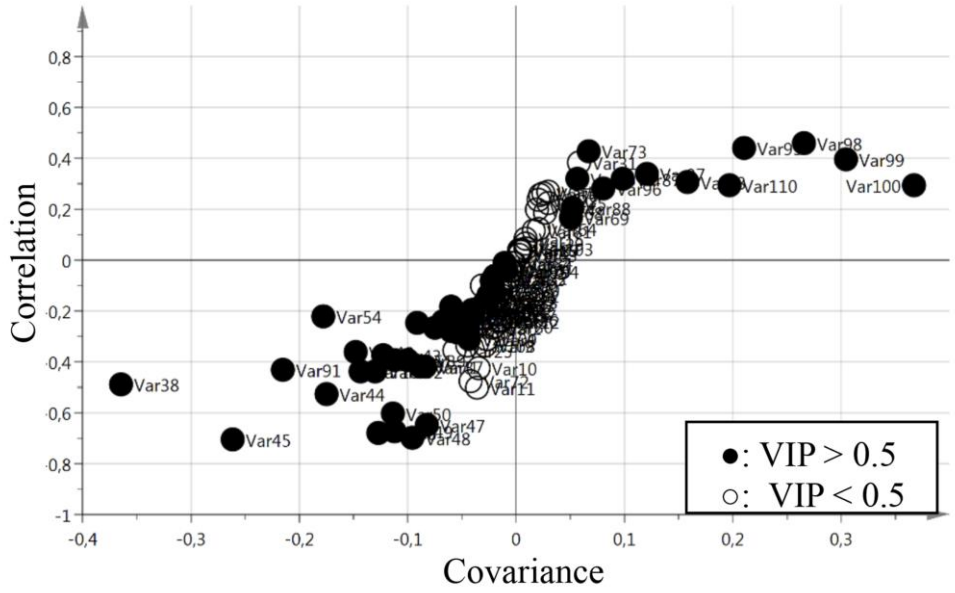


Figure 5.6. S-plot of the OPLS-DA model showing the variables contributing most to group discrimination. Variables situated at the right end are increased in the plasma of controls, whereas those situated at the left end are increased for the lung cancer patients. Variables used to explain the disturbed biochemical pathways (VIP > 0.5) in lung cancer are marked (●). Var: variable; VIP: variable importance for the projection.

Table 5.3. Integration regions (variables) with a VIP value > 0.5 which are increased in the NMR spectra of blood plasma of lung cancer patients.

VAR	Integration region (ppm)	VIP	p-value	FC	Contributing metabolites	Assignment
38#	3.9590 – 3.8330	3.34	$1.01x10^{-14}$	1.15	Asp, Met, Ser, Tyr, glucose	Glucose
54#	3.5360 – 3.3980	2.46	$4.09x10^{-4}$	1.08	Glucose, Pro, acetoacetate	Glucose
45*	3.7141 – 3.6680	2.16	$4.05x10^{-28}$	1.27	Ile, glycerol,	Glycerol
91*	2.1230 – 1.9720	2.05	$2.59x10^{-10}$	1.05	C**H**$_2$-$N^+(CH_3)_3$ of SM & PC Glu, Met, Pro, Ile, CH_3 of NAG, lipids$	CH_3 of NAG
44	3.7390 – 3.7141	1.56	$1.00x10^{-16}$	1.17	Glucose	Glucose
42#	3.7820 – 3.7550	1.53	$4.86x10^{-8}$	1.10	Ala, Glu, Gln, Leu, Lys, glucose	Glucose
101	1.2458 – 1.2180	1.26	$7.18x10^{-12}$	1.16	Ile, β-hydroxybutyrate	β-hydroxybutyrate
43#	3.7550 – 3.7390	1.23	$3.85x10^{-8}$	1.11	Ala, Leu, glucose	Glucose
102	1.2180 – 1.1300	1.16	$1.34x10^{-11}$	1.20	NI	NI
21	4.6940 – 4.6620	1.15	$1.01x10^{-4}$	1.09	Glucose	Glucose
51*	3.5914 – 3.5649	1.11	$1.40x10^{-9}$	1.11	Glycerol, Gly, Thr, glucose	Glycerol
46	3.6680 – 3.6500	1.04	$3.63x10^{-26}$	1.29	Glycerol	Glycerol
39#	3.8330 – 3.8100	0.98	$1.38x10^{-11}$	1.13	Ala, Ser, glucose	Glucose
50	3.6097 – 3.5914	0.96	$1.34x10^{-22}$	1.24	Thr	Thr
49	3.6240 – 3.6097	0.94	$7.38x10^{-27}$	1.29	Thr	Thr
66	3.0640 – 2.9950	0.90	$1.90x10^{-2}$	1.04	Lys, α-ketoglutarate	?
19	5.2752 – 5.2526	0.89	$4.31x10^{-5}$	1.07	Glucose	Glucose
94*	1.8060 – 1.6860	0.86	$4.76x10^{-4}$	1.07	Leu, Lys	Leu

Table 5.3 continued.

VAR	Integration region (ppm)	VIP	p-value	FC	Contributing metabolites	Assignment
71*	2.7360 – 2.6600	0.83	$1.70x10^{-9}$	1.12	Asp, Met, citrate	Asp
48	3.6376 – 3.6240	0.79	$2.98x10^{-25}$	1.25	Val	Val
37	3.9810 – 3.9590	0.78	$1.80x10^{-11}$	1.18	Asn, His, Ser, Tyr, creatine	?
90*	2.1970 – 2.1230	0.76	$4.57x10^{-2}$	1.03	Glu, Gln, Pro, Met	Gln
53#	3.5510 – 3.5360	0.72	$1.19x10^{-5}$	1.10	Glucose, Pro, acetoacetate	Glucose
67	2.9950 – 2.8860	0.69	$7.85x10^{-1}$	1.01	Asn, lipids$	?
47*	3.6500 – 3.6376	0.68	$1.56x10^{-22}$	1.23	Val, glycerol	Glycerol
40#	3.8100 – 3.7956	0.63	$3.80x10^{-5}$	1.08	Ala, Glu, Gln, glucose	Glucose
59	3.2186 – 3.1930	0.61	$4.63x10^{-5}$	1.17	Tyr	Tyr
52	3.5649 – 3.5510	0.55	$6.16x10^{-6}$	1.09	Glucose	Glucose
76*	2.4920 – 2.4500	0.54	$4.85x10^{-1}$	1.01	Gln, α-ketoglutarate, β-hydroxybutyrate	Gln
106	1.0220 – 1.0020	0.54	$4.11x10^{-2}$	1.04	Val, Ile, Leu	Val
41#	3.7956 – 3.7820	0.52	$2.31x10^{-2}$	1.04	Ala, Glu, Gln, Leu, Lys, glucose	Glucose
107	1.0020 – 0.9860	0.51	$2.56x10^{-2}$	1.03	Ile, Leu	Leu

#It can be deduced that these composite regions (variables) emanate from glucose, because variables 19, 21, 44 and 52 representing a unique signal of glucose, are also increased. *The assignment of metabolites in crowded regions with signal overlap was improved by measuring plasma samples spiked with known metabolites in relevant concentrations on a 900 MHz NMR spectrometer (Lille, France), having a higher spectral resolution and signal-to-noise ratio. $Common signals for all lipids. Amino acids are presented by their 3-letter code. FC: fold change; NAG: N-acetylated glycoproteins; NI: non-identified; PC: phosphatidylcholine; ppm: parts per million; SM: sphingomyelin; VAR: variable; VIP: variable importance for the projection.

Table 5.4. Integration regions (variables) with a VIP value > 0.5 which are decreased in the NMR spectra of blood plasma of lung cancer patients.

VAR	Integration region (ppm)	VIP	p-value	FC	Contributing metabolites	Assignment
100	1.3450 – 1.2458	4.28	$1.21x10^{-5}$	-1.08	Thr, Ile, lipids[$]	Lipids[$]
99[*]	1.3740 – 1.3450	2.95	$1.47x10^{-10}$	-1.12	Lactate, Thr	Lactate
98	1.4200 – 1.3740	2.47	$7.70x10^{-12}$	-1.27	Lactate	Lactate
110	0.9660 – 0.8000	2.26	$2.41x10^{-5}$	-1.04	Lipids[$]	Lipids[$]
58[*]	3.3230 – 3.2186	2.06	$6.89x10^{-1}$	-1.01	His, Phe, Tyr, glucose, $N^+(CH_3)_3$ of SM & PC	$N^+(CH_3)_3$ of SM & PC
95[#]	1.6860 – 1.5600	2.00	$2.20x10^{-11}$	-1.24	Lys, lipids[$]	Lipids[$]
18	5.4300 – 5.2752	1.68	$7.10x10^{-6}$	-1.08	Lipids[$]	Lipids[$]
87[#]	2.2915 – 2.2690	1.08	$6.84x10^{-7}$	-1.18	Met, Val, lipids[$]	Lipids[$]
96[*]	1.5400 – 1.4900	0.87	$1.91x10^{-5}$	-1.11	Ala, Ile, Lys	Ala
88[#]	2.2690 – 2.2300	0.73	$2.77x10^{-3}$	-1.08	Met, Val, aceton, lipids[$]	Lipids[$]
69[#]	2.8550 – 2.7500	0.67	$1.57x10^{-2}$	-1.04	Lipids[$], Asn, Asp	Lipids[$]
86[#]	2.3040 – 2.2915	0.61	$3.08x10^{-7}$	-1.15	Val, β-hydroxybutyrate, lipids[$]	Lipids[$]
73	2.5960 – 2.5340	0.59	$3.14x10^{-8}$	-1.19	Citrate	Citrate

[#]It can be deduced that these composite regions (variables) emanate from lipids, because variables 18 and 110 representing a unique signal of lipids, are also decreased. [*]The assignment of metabolites in crowded regions with signal overlap was improved by measuring plasma samples spiked with known metabolites in relevant concentrations on a 900 MHz NMR spectrometer (Lille, France), having a higher spectral resolution and signal-to-noise ratio. [$]Common signals for all lipids. Amino acids are presented by their 3-letter code. FC: fold change; ppm: parts per million; VAR: variable; VIP: variable importance for the projection.

In a next step, it was evaluated whether histological subtypes and tumor stages can be discriminated based on the metabolic phenotype. PCA is not able to cluster the histological subtypes and the same holds for OPLS-DA (**Figure 5.7**). Since adenocarcinomas (n=91) and spinocellular carcinomas (n=66) are the most common histological subtypes (**Table 5.1**), an OPLS-DA model was trained in discriminating between them. The resulting model classifies 81% of the adenocarcinomas correctly but only 38% of the spinocellular carcinomas (**Figure 5.8**). Moreover, the predictive ability of the model was extremely poor (Q2: 0.04). Concerning tumor stage, no significant clustering was obtained by PCA nor by OPLS-DA (**Figure 5.9**). In an attempt to further discriminate between early stage patients (stage I, n=76) and metastatic patients (stage IV, n=63), a model was trained which classifies 79% of the early stage patients and 52% of the metastatic patients correctly (**Figure 5.10A**), but the predictive ability of this model was very low (Q2: 0.06). On the other hand, an OPLS-DA model was able to discriminate between early stage patients (stage I, n=76) and a randomly selected equally populated group of controls (n=76) with a sensitivity of 74% and a specificity of 78% (**Figure 5.10B**). Most of the discriminating variables of this model were also found for the full model.

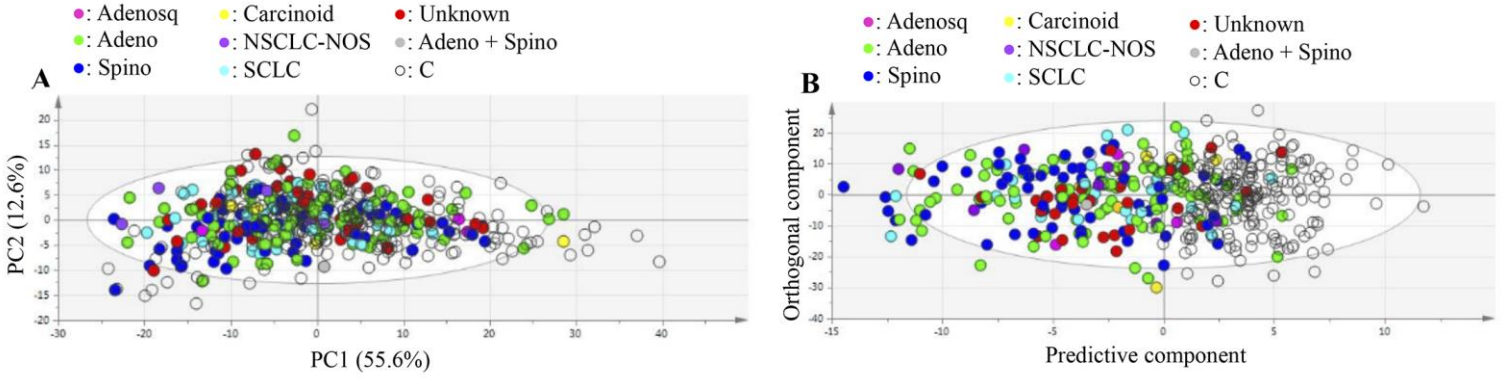


Figure 5.7. Discrimination according to histological subtype. (A) PCA score plot, (B) OPLS-DA score plot. Adeno: adenocarcinoma; Adenosq:adenosquamous carcinoma; C: controls; NOS: not otherwise specified; NSCLC: non-small cell lung cancer; PC: principal component; Spino: spinocellular carcinoma.

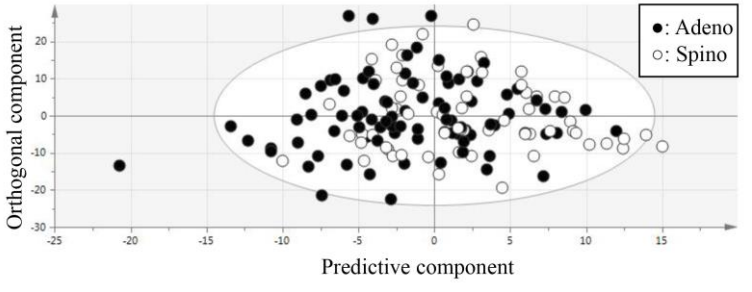


Figure 5.8. **Discrimination according to histological subtype.** OPLS-DA score plot of patients with an adenocarcinoma (n=91) and a spinocellular carcinoma (n=66). Adeno: adenocarcinoma; Spino: spinocellular carcinoma.

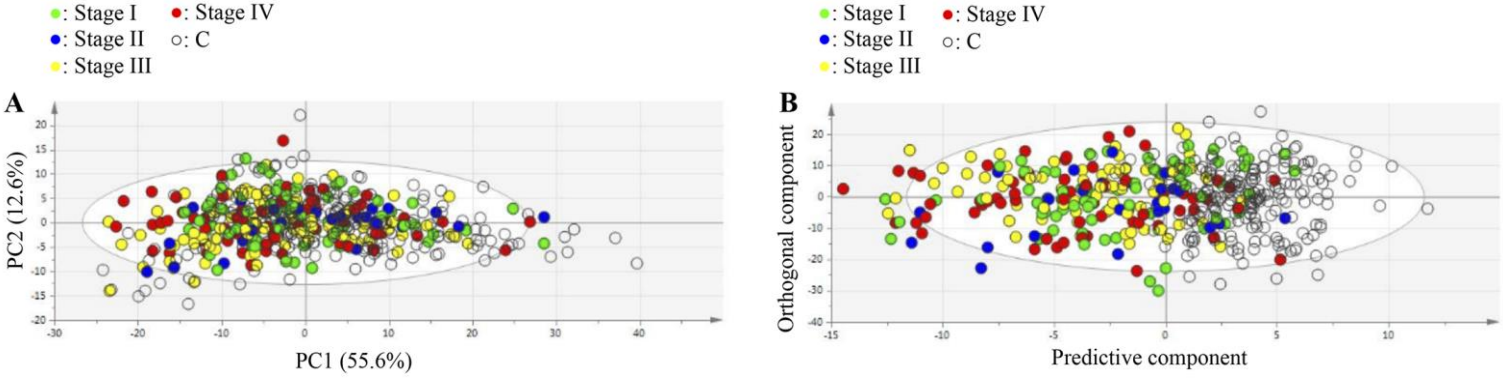


Figure 5.9. Discrimination according to tumor stage. (A) PCA score plot, (B) OPLS-DA score plot. Abbreviations: C: controls; PC: principal component.

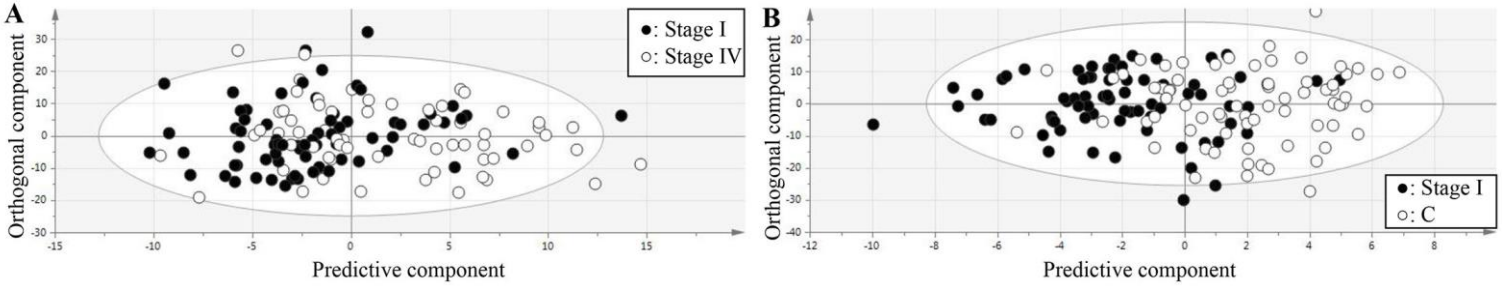


Figure 5.10. Discrimination according to tumor stage. (A) OPLS-DA score plot of early stage patients (n=76) and metastatic patients (n=63), (B) OPLS-DA score plot of early stage patients (n=76) and a randomly chosen but equally populated group of controls (n=76). C: controls.

Discussion

This study demonstrates that 1) the model allows to classify 78% of the lung cancer patients and 92% of the controls correctly with an AUC of 0.88; 2) the model discriminates between lung cancer patients and controls of the independent cohort with a sensitivity of 71%, a specificity of 81% and an AUC of 0.84; 3) plasma of lung cancer patients is characterized by increased levels of glucose, glycerol, N-acetylated glycoproteins, β-hydroxybutyrate, leucine, tyrosine, threonine, glutamine, valine and aspartate and decreased levels of alanine, lactate and 4) the limited numbers of patients in the subgroups do not (yet) allow to discriminate between histological subtypes and clinical tumor stages, but it looks that stage I patients can be reasonably well differentiated from controls.

The metabolic phenotype, which is represented by the relative abundance of plasma metabolites, has to be seen as a single biomarker that cannot be defined based on a cut-off value. It is demonstrated that the combination of a series of subtle metabolic alterations (metabolites of which the plasma concentration is in- or decreased in lung cancer patients compared to controls), detected by means of ^{1}H-NMR spectroscopy and presented by OPLS-DA, enables to diagnose lung cancer.

Recently, many studies have explored lung cancer metabolism, but mostly by mass spectrometry (MS) techniques rather than by ^{1}H-NMR spectroscopy (24-26). Although MS is without doubt more sensitive, ^{1}H-NMR spectroscopy requires no invasive extraction procedures, and so minimal sample preparation (18). Both techniques are therefore complementary and of importance in the field of metabolomics. Moreover, most published NMR studies focused on the metabolic composition of the lung cancer tissue notwithstanding the fact that metabolic phenotyping of blood plasma has the advantage to assess more directly the complex interaction between tumor and host (27, 28). According to a review of Duarte et al. (29), only Rocha et al. investigated lung cancer-induced metabolic alterations in plasma by ^{1}H-NMR spectroscopy, demonstrating a discrimination between 85 lung cancer patients and 78 controls with a sensitivity and specificity of approximately 90% (30). However, the findings were not validated in an independent cohort.

^{1}H-NMR metabolic phenotyping of plasma as a complementary tool to identify high-risk individuals

Currently, low-dose computed tomography (LDCT) is the most studied tool to screen for lung cancer. The NELSON trial demonstrates that LDCT screening has a sensitivity of 84.6% and a specificity of 98.6% in comparison to no screening, (31). However, a major limitation of LDCT is the low PPV ranging from 3.8% in the National Lung Screening Trial to 40.4% in the NELSON trial. This means that more than half of the study participants were referred for further investigations, being not without cost and risk, on the basis of false-positive results (5, 31). A recent cost-effectiveness analysis of LDCT screening showed a cost of $81,000 per quality-adjusted life-year gained (32). By means of current risk models, many patients are selected as eligible for LDCT screening. Strengthening of these models by incorporating (fast and cheap) metabolic phenotype information might be the way to better identify patients eligible for screening, leading moreover to a strong decrease of the financial burden. In this respect, ^{1}H-NMR-based metabolomics seems to be reasonably able to discriminate between early stage patients and a randomly selected equally populated group of controls. This indicates that the metabolic alterations present in the initial phase of cancer development can already be detected by ^{1}H-NMR-based metabolomics. Although these results look promising, the number of early stage patients evidently needs to be increased to confirm.

Looking at the disturbed metabolism associated with lung cancer

Because cancer cells often outgrow the surrounding vasculature, they have adopted an oxygen-independent metabolism to produce sufficient energy for proliferation. Even in the presence of normal oxygen levels, cancer cells prefer to convert glucose into lactate rather than oxidizing it via oxidative phosphorylation, a hallmark known as the Warburg effect or "aerobic glycolysis" (15). This latter term is somewhat confusing, because oxygen is not involved in the fermentation of pyruvate to lactate. Due to the Warburg effect, a decreased glucose level and increased lactate level could be expected in plasma of lung cancer patients. However, it should be kept in mind that the human body has an intrinsic tendency to maintain anabolic and catabolic pathways in balance. Owing to the elevated glycolytic rate in cancer cells, other cells are deprived from the fuels normally

provided by glucose-derived carbon. In response to this, the body will attempt to maintain homeostasis and provide healthy cells with an adequate energy supply to survive.

This study reveals that lung cancer metabolism is characterized by increased plasma levels of glucose and decreased levels of alanine and lactate, in line with the inherent nature of the body to supply all tissues with sufficient metabolic fuel to function properly. More specifically, liver glycogen will be degraded in response to the Warburg effect in order to compensate for the lack of glucose as energy source for the normal cells. But once depleted, gluconeogenic substrates will be fed into the corresponding pathway. Consequently, less oxaloacetate is available to condense with acetyl-CoA, so the latter is diverted to form ketone bodies rather than fueling oxidative phosphorylation via the Krebs cycle (33). The metabolic state of ketosis in lung cancer patients is confirmed by increased plasma levels of the ketone body β-hydroxybutyrate and the ketogenic amino acid leucine. The decreased levels of alanine and lactate and the increased levels of glucose and β-hydroxybutyrate are in line with the findings of Chen et al., who examined the serum metabolome of lung cancer patients by means of MS (**Table 5.5**) (25).

Table 5.5. Overview of the reported concentration change of glucose, lactate and lipids in studies which investigated the disturbed lung cancer metabolism in plasma/serum.

Authors	Technique	Body fluid	Glucose	Lactate	Lipids
Louis et al.	^{1}H-NMR spectroscopy	Plasma	↑	↓	↓
Rocha et al.	^{1}H-NMR spectroscopy	Plasma	↓	↑	↓ HDL ↑ VLDL ↑ LDL ↑ TG
Chen et al.	LC-MS and GC-MS	Serum	↑	↓	↓

^{1}H-NMR; proton-nuclear magnetic resonance; LC: liquid chromatography; LDL: low-density lipoprotein; GC: gas chromatography; HDL: high-density lipoprotein; MS: mass spectrometry; TG: triglycerides; VLDL: very-low density lipoprotein.

Although the increased gluconeogenesis in lung cancer patients is in accordance with the decreased levels of alanine and lactate, a lower concentration of the gluconeogenic precursor pyruvate could not be confirmed since the pyruvate signal coincides with one of the β-hydroxybutyrate signals in the ^{1}H-NMR spectrum. Since another, free-lying signal of β-hydroxybutyrate is increased (**Table 5.3**), the increased β-hydroxybutyrate level might mask the decrease of pyruvate. In contrast with our findings, Rocha and co-workers reported increased

lactate and decreased glucose levels in plasma of lung cancer patients (**Table 5.5**) (30). A possible explanation might be found in that the majority of patients in their study had early stage lung cancer, whereas in our study, patients in early stage as well as in more advanced stages are included. This might indicate that gluconeogenesis and glycogenolysis, which counteract the Warburg effect occurring in cancer cells, are more pronounced in advanced stages.

Besides hepatocytes, which display an impaired Krebs cycle in response to the Warburg effect, also cancer cells are characterized by a disturbed Krebs cycle (34). This lung cancer-induced metabolic derangement is affirmed by decreased levels of citrate and further supported by increased levels of amino acids which normally would be consumed for the production of Krebs cycle intermediates (aspartate, threonine, valine, glutamine and tyrosine) (35). The plasma concentration of citrate is decreased since, once it is shuttled to the cytoplasm, it becomes converted to acetyl-CoA, being a critical precursor of fatty acids and cholesterol, i.e. fatty acids and cholesterol are synthesized from the carbons provided by acetyl-CoA (36). After conversion to phospholipids, the fatty acids are incorporated in the plasma membrane of new daughter cells together with cholesterol. Our study confirms an enhanced synthesis of membranes in cancer cells by measuring decreased plasma levels of sphingomyelin and phosphatidylcholine and of other phospholipids. Note that although the increased plasma levels of glycerol seem to be in contrast with the enhanced synthesis of membranes in cancer cells, this discrepancy might be explained by the fact that cancer cells do not use glycerol obtained from plasma to synthesize their membranes but rather glycerol derived from the glycolytic pathway. Finally, blood plasma of lung cancer patients is characterized by increased levels of N-acetylated glycoproteins. This finding is in line with emerging evidence which reveals that the enzyme catalyzed addition of N-acetylglucosamine to serine or threonine residues of proteins plays a role in the metabolic reprogramming of cancer cells (37).

Conclusions

This paper validates ^{1}H-NMR derived metabolic phenotyping of blood plasma as a complementary tool to discriminate between lung cancer patients and controls. Our findings indicate that metabolic alterations can already be detected in an early stage. Remark however that our intent is not to use the metabolome as a separate screening tool but to complement current risk models with additional parameters which reflect the metabolic phenotype to better identify and select high-risk individuals eligible for LDCT screening. As a result, it might be anticipated that the PPV of LDCT screening will increase, thereby improving its cost-effectiveness. Due to the limited subgroup numbers, it is not possible yet to decide whether metabolic phenotyping can further discriminate between histological subtypes or clinical tumor stages.

References

1. Boyle P, Levin B. World Cancer Report 2008. Lyon: World Health Organization, 2008.
2. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. International journal of cancer. 2015;136(5):E359-86.
3. Mulshine JL, Sullivan DC. Clinical practice. Lung cancer screening. The New England journal of medicine. 2005;352(26):2714-20.
4. Tammemagi MC, Lam S. Screening for lung cancer using low dose computed tomography. BMJ. 2014;348.
5. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. The New England journal of medicine. 2011;365(5):395-409.
6. Bach PB, Mirkin JN, Oliver TK, Azzoli CG, Berry DA, Brawley OW, et al. Benefits and harms of CT screening for lung cancer: a systematic review. Jama. 2012;307(22):2418-29.
7. Hasan N, Kumar R, Kavuru MS. Lung cancer screening beyond low-dose computed tomography: the role of novel biomarkers. Lung. 2014;192(5):639-48.
8. O'Connell TM. Recent advances in metabolomics in oncology. Bioanalysis. 2012;4(4):431-51.
9. Bervoets L, Louis E, Reekmans G, Mesotten L, Thomeer M, Adriaensens P, et al. Influence of preanalytical sampling conditions on the ^{1}H NMR metabolic profile of human blood plasma and introduction of the standard preanalytical code used in biobanking. Metabolomics. 2015;11:1197-1207.
10. Louis E, Bervoets L, Reekmans G, De Jonge E, Mesotten L, Thomeer M, et al. Phenotyping human blood plasma by ^{1}H-NMR: a robust protocol based on metabolite spiking and its evaluation in breast cancer. Metabolomics. 2015;11:225-36.
11. Cantor JR, Sabatini DM. Cancer cell metabolism: one hallmark, many faces. Cancer discovery. 2012;2(10):881-98.
12. Munoz-Pinedo C, El Mjiyad N, Ricci JE. Cancer metabolism: current perspectives and future directions. Cell death & disease. 2012;3:e248.
13. Sciacovelli M, Gaude E, Hilvo M, Frezza C. The metabolic alterations of cancer cells. Methods in enzymology. 2014;542:1-23.
14. Iurlaro R, Leon-Annicchiarico CL, Munoz-Pinedo C. Regulation of cancer metabolism by oncogenes and tumor suppressors. Methods in enzymology. 2014;542:59-80.
15. Upadhyay M, Samal J, Kandpal M, Singh OV, Vivekanandan P. The Warburg effect: insights from the past decade. Pharmacology & therapeutics. 2013;137(3):318-30.
16. Holmes E, Wilson ID, Nicholson JK. Metabolic phenotyping in health and disease. Cell. 2008;134(5):714-7.
17. Mamas M, Dunn WB, Neyses L, Goodacre R. The role of metabolites and metabolomics in clinically applicable biomarkers of disease. Archives of toxicology. 2011;85(1):5-17.
18. Lindon JC, Nicholson JK. Spectroscopic and statistical techniques for information recovery in metabonomics and metabolomics. Annu Rev Anal Chem 2008;1:45-69.
19. Goldstraw P, Crowley J, Chansky K, Giroux DJ, Groome PA, Rami-Porta R, et al. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours. Journal of thoracic oncology. 2007;2(8):706-14.
20. Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). J Chemometr. 2002;16(3):119-28.
21. Xia J, Broadhurst DI, Wilson M, Wishart DS. Translational biomarker discovery in clinical metabolomics: an introductory tutorial. Metabolomics. 2013;9(2):280-99.
22. Wiklund S, Johansson E, Sjostrom L, Mellerowicz EJ, Edlund U, Shockcor JP, et al. Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. Analytical chemistry. 2008;80(1):115-22.
23. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met. 1995;57(1):289-300.

24. Wen T, Gao L, Wen Z, Wu C, Tan CS, Toh WZ, et al. Exploratory investigation of plasma metabolomics in human lung adenocarcinoma. Molecular bioSystems. 2013;9(9):2370-8.
25. Chen Y, Ma Z, Li A, Li H, Wang B, Zhong J, et al. Metabolomic profiling of human serum in lung cancer patients using liquid chromatography/hybrid quadrupole time-of-flight mass spectrometry and gas chromatography/mass spectrometry. Journal of cancer research and clinical oncology. 2015;141(4):705-18.
26. Hori S, Nishiumi S, Kobayashi K, Shinohara M, Hatakeyama Y, Kotani Y, et al. A metabolomic approach to lung cancer. Lung Cancer. 2011;74(2):284-92.
27. Chen W, Zu Y, Huang Q, Chen F, Wang G, Lan W, et al. Study on metabonomic characteristics of human lung cancer using high resolution magic-angle spinning 1H NMR spectroscopy and multivariate data analysis. Magnetic resonance in medicine. 2011;66(6):1531-40.
28. Duarte IF, Rocha CM, Barros AS, Gil AM, Goodfellow BJ, Carreira IM, et al. Can nuclear magnetic resonance (NMR) spectroscopy reveal different metabolic signatures for lung tumours? Virchows Archiv. 2010;457(6):715-25.
29. Duarte IF, Rocha CM, Gil AM. Metabolic profiling of biofluids: potential in lung cancer screening and diagnosis. Expert review of molecular diagnostics. 2013;13(7):737-48.
30. Rocha CM, Carrola J, Barros AS, Gil AM, Goodfellow BJ, Carreira IM, et al. Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of blood plasma. Journal of proteome research. 2011;10(9):4314-24.
31. Horeweg N, Scholten ET, de Jong PA, van der Aalst CM, Weenink C, Lammers JW, et al. Detection of lung cancer through low-dose CT screening (NELSON): a prespecified analysis of screening test performance and interval cancers. The Lancet Oncology. 2014;15(12):1342-50.
32. Black WC, Gareen IF, Soneji SS, Sicks JD, Keeler EB, Aberle DR, et al. Cost-effectiveness of CT screening in the National Lung Screening Trial. The New England journal of medicine. 2014;371(19):1793-802.
33. Berg J, Tymoczko J, Stryer L. Biochemistry. 8 ed. New York: W.H. Freeman and Company; 2015.
34. Weinberg F, Chandel NS. Mitochondrial metabolism and cancer. Annals of the New York Academy of Sciences. 2009;1177:66-73.
35. Harvey R, Ferrier D. Biochemistry. 5 ed. Philadelphia: Lipincott Williams & Wilkins; 2011.
36. Lunt SY, Vander Heiden MG. Aerobic glycolysis: meeting the metabolic requirements of cell proliferation. Annual review of cell and developmental biology. 2011;27:441-64.
37. Jozwiak P, Forma E, Brys M, Krzeslak A. O-GlcNAcylation and Metabolic Reprogramming in Cancer. Frontiers in endocrinology. 2014;5:145.

CHAPTER 6

The plasma metabolic phenotype permits to discriminate between lung and breast cancer

Based on:

Evelyne Louis, Peter Adriaensens, Wanda Guedens, Karolien Vanhove, Kurt Vandeurzen, Karen Darquennes, Johan Vansteenkiste, Christophe Dooms, Eric de Jonge, Michiel Thomeer, Liesbet Mesotten. Metabolic phenotyping of human blood plasma: a powerful tool to discriminate between cancer types? Accepted for publication in Annals of Oncology (2015), DOI: 10.1093/annonc/mdv499.

Abstract

Background. Accumulating evidence has shown that cancer cell metabolism differs from that of normal cells. However, up to now it is not clear whether different cancer types are characterized by a specific metabolite profile. Therefore, this study aims to evaluate whether the plasma metabolic phenotype allows to discriminate between lung and breast cancer.

Patients and methods. The proton nuclear magnetic resonance spectrum of plasma is divided into 110 integration regions, representing the metabolic phenotype. These integration regions reflect the relative metabolite concentrations and were used to train a classification model in discriminating between 80 female breast cancer patients and 54 female lung cancer patients, all with an adenocarcinoma. The validity of the model was examined by permutation testing and by classifying an independent validation cohort of 60 female breast cancer patients and 81 male lung cancer patients, all with an adenocarcinoma.

Results. The model allows to classify 99% of the breast cancer patients and 93% of the lung cancer patients correctly with an area under the curve of 0.96 and can be validated in the independent cohort with a sensitivity of 89%, a specificity of 82% and an area under the curve of 0.94. The most discriminating variables indicate that the disturbed biochemical pathways in lung cancer as compared to breast cancer include i) a more pronounced hepatic glycogenolysis, gluconeogenesis and ketogenesis, ii) a more impaired Krebs cycle in hepatocytes and lung cancer cells and iii) an enhanced membrane synthesis.

Conclusion. Metabolic phenotyping of plasma allows to discriminate between lung and breast cancer, indicating that the metabolite profile reflects more than a common cancer marker.

Introduction

The application of metabolomics in the search for cancer biomarkers has increased enormously over the past years (1-3). This discipline comprises the simultaneous and comprehensive analysis of metabolites within a biological system (4). Since metabolites constitute the end products of cellular processes, changes in their concentration reflect alterations in the metabolic phenotype (5). Metabolomics combines analytical characterization tools such as proton nuclear magnetic resonance (^{1}H-NMR) spectroscopy and multivariate statistics to retrieve diagnostic information regarding diseases and to identify disease-related changes in biochemical pathways (6). ^{1}H-NMR spectroscopy allows a fast (< 10 min), non-invasive identification and quantification of complex mixtures of metabolites, as appearing in plasma (4, 7).

Over the past decade, accumulating evidence has shown that cancer cell metabolism differs from that of normal cells (8-10). More specifically, the entire metabolism of cancer cells is reprogrammed to promote cell proliferation and survival and is driven by aberrant signaling pathways induced by the activation of oncogenes or the inactivation of tumor suppressor genes (11). One of the main adaptations of cancer cells is that, even in the presence of normal oxygen levels, they rely on anaerobic energy production through glycolysis, a hallmark known as the Warburg effect (12). The main advantage of predominantly relying on glycolysis for production of energy and essential building blocks is that many glycolytic intermediates can be shunted to anabolic pathways, thereby preserving cancer cell proliferation (10).

Previously, we revealed that metabolic phenotyping of plasma enables to discriminate between breast cancer patients and controls (13). Furthermore, ongoing research of our group is demonstrating, for a large cohort of 450 subjects, that the metabolic phenotype allows differentiate between lung cancer patients and controls (14). Nevertheless, up to now it is not clear whether the metabolic phenotype reflects a common cancer marker or whether different cancer types have characteristic profiles. Therefore, this study aims to 1) investigate whether the plasma metabolic phenotype allows to discriminate between lung and breast cancer, 2) evaluate the predictive accuracy of the trained classification model in

an independent cohort and 3) improve the understanding of the disturbed biochemical pathways.

Patients and methods

Subjects

Lung cancer patients with an adenocarcinoma (n=145; 55 female and 90 male) were included in the Limburg Positron Emission Tomography (PET) center (n=105) (Hasselt, Belgium) and at the Department of Respiratory Medicine of University Hospitals Leuven (n=40) (Leuven, Belgium) from March 2011 to May 2014. The diagnosis of adenocarcinoma was confirmed by a pathological biopsy. Clinical staging of the adenocarcinomas was performed according to the 7th edition of the tumor, nodes and metastases classification of malignant tumors (15). Breast cancer patients with an adenocarcinoma (n=147) were included at the day of primary surgery at the Department of Gynaecology of Ziekenhuis Oost-Limburg (Genk, Belgium) between March 2010 and August 2012. The diagnosis of adenocarcinoma was confirmed by a core needle biopsy. The stage of the adenocarcinomas was defined by the revised staging system for breast cancer (16). For both patients groups, blood sampling and sample preparation was done according to a fixed protocol and by trained staff.

Exclusion criteria were: 1) not fasted for at least 6 h; 2) fasting blood glucose concentration ≥ 200 mg/dl; 3) medication intake on the morning of blood sampling and 4) treatment or history of cancer in the past 5 years. The study was conducted in accordance with the ethical rules of the Helsinki Declaration and Good Clinical Practice and was approved by the ethical committees of Ziekenhuis Oost-Limburg, Hasselt University (Hasselt, Belgium) and University Hospitals Leuven. All study participants provided written informed consent. The study is registered at clinical trials.gov (NCT02362776).

Both patient groups were subdivided into a training cohort and a validation cohort (**Figure 6.1**). More specifically, the 55 female lung cancer patients were assigned to the training cohort, whereas the 90 male lung cancer patients were allocated to the validation cohort in order to examine the confounding effect of gender. Regarding the breast cancer patients, 84 were allocated to the training cohort and 63 to the validation cohort. Five statistical outliers of the training cohort (1 lung cancer patient and 4 breast cancer patients) and 12 of the validation cohort (9 lung cancer patients and 3 breast cancer patients) were excluded. According to

their medical files, they showed abnormal concentrations of glucose, lipids or ketone bodies. The individuals with high glucose levels were diagnosed with diabetes or had an increased fasting glucose while those with high lipid levels suffered from obesity, hyperlipidemia or took cholesterol-lowering medication. Most of the individuals with high ketone body levels had a low BMI. Notice that in order to explain the disturbed biochemical pathways, it was decided to train an OPLS-DA classification model with the 135 lung cancer patients and the 140 breast cancer patients who remained after excluding the outliers instead of dividing both patient groups into a training and a validation cohort (**Figure 6.1**).

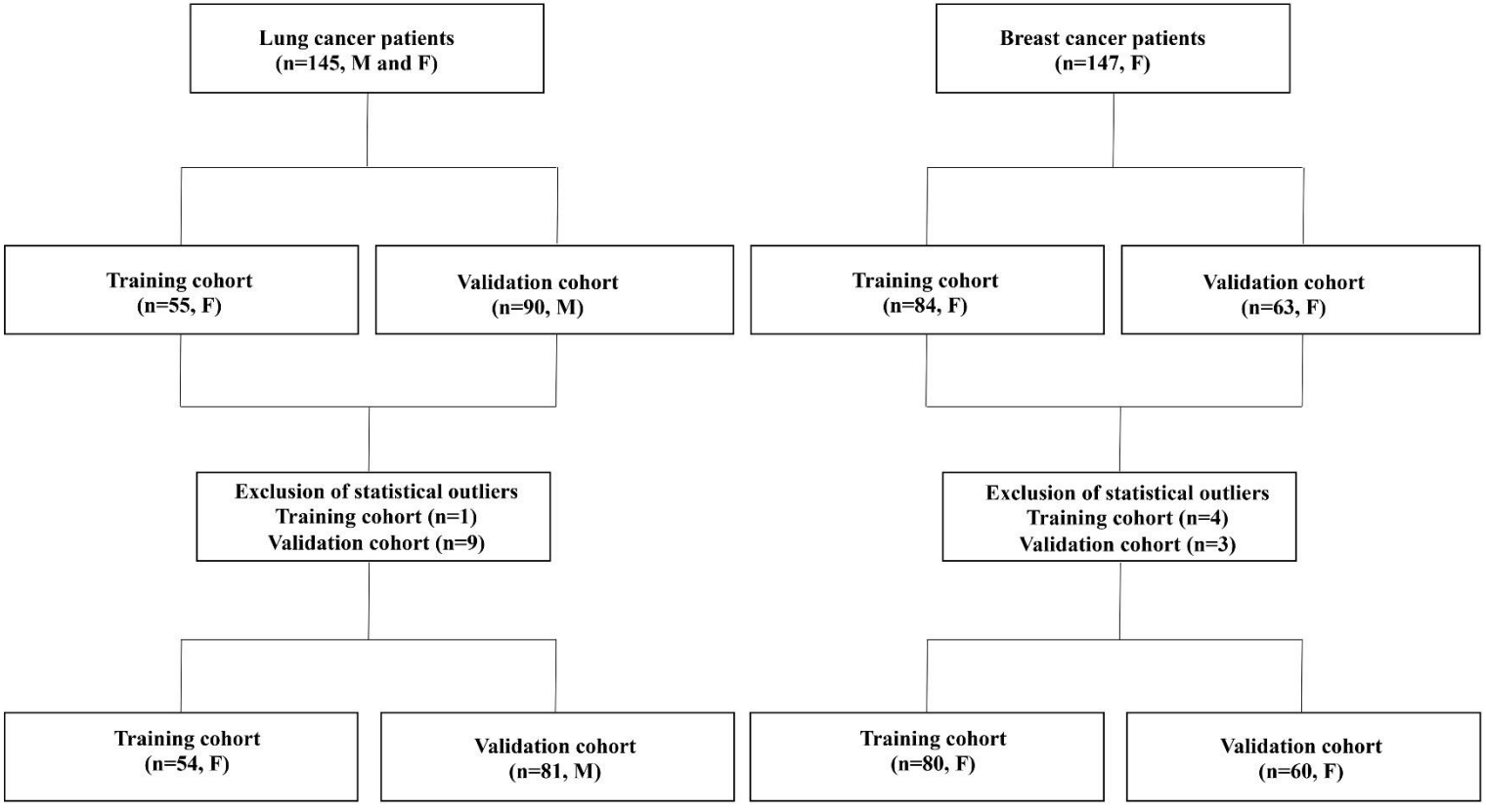


Figure 6.1. CONSORT diagram of the study. F: females; M: males.

Blood sampling and processing

See **Chapter 2**: Blood sampling and processing.

NMR sample preparation and analysis

See **Chapter 2**: NMR sample preparation and analysis.

Spectral processing

See **Chapter 2**: Spectral processing.

Statistical analysis

Multivariate statistics was performed using SIMCA-P+ (Version 14, Umetrics, Umea, Sweden). After mean-centering and Pareto scaling of the variables, unsupervised principal component analysis (PCA) was performed to identify outliers by means of a Hotelling's T2 range test and a distance to model plot. After removing outliers, orthogonal partial least squares discriminant analysis (OPLS-DA) was used to train a classification model (17). The validity of the model was confirmed by i) permutation testing, ii) classifying an independent cohort with a classification cut-off value of 0.5 and iii) receiver operating characteristic curve explorer & tester (18). Via an S-plot, the most discriminating variables and their variable importance for the projection (VIP) values were identified (19). Variables with a VIP exceeding 0.5 were used to describe the biochemical pathways. Additionally, a student t-test with correction for multiple testing by the Benjamini-Hochberg method (20) was applied (IBM SPSS Version 22, Chicago, Illinois, USA). Note that noisy variables, which were identified in **Chapter 2**, were not excluded since none of the noisy variables had a VIP value exceeding 0.5 and therefore, were not selected to explain the biochemical pathways.

Results

Multivariate OPLS-DA statistics was used to train a classification model (classifier) in discriminating between breast and lung cancer based on data input from their metabolic phenotype. The resulting model was validated in an independent patient cohort. **Table 6.1** shows the characteristics of the training and validation cohorts. However, before starting with the OPLS-DA statistics, a PCA analysis was conducted to look for possible confounders. **Figure 6.2** shows PCA score plots stained for disease type (A), smoking habits (B) and tumor stage (C). The plots demonstrate that smoking habits and tumor stage are no confounders.

Table 6.1. Characteristics of the subjects included in the study (without outliers).

	Training cohort		Validation cohort	
	BC	**LC**	**BC**	**LC**
Number of subjects, n	80 F	54 F	60 F	81 M
Age, yrs	58 ± 11	61 ± 10	60 ± 12	66 ± 9
(range)	(24-86)	(43-88)	(40-84)	(46-83)
BMI, kg/m^2	27 ± 5	25 ± 5	26 ± 5	26 ± 4
(range)	(20-44)	(17-42)	(18-43)	(19-39)
Smoking habits				
Non-smoker, n (%)	55 (69)	28 (52)	49 (82)	44 (54)
Smoker, n (%)	18 (22)	26 (48)	9 (15)	37 (46)
Unknown, n (%)	7 (9)	0 (0)	2 (3)	0 (0)
Menopause				
Post-menopausal, n (%)	62 (78)	44 (81)	41 (68)	
Pre-menopausal, n (%)	18 (22)	10 (19)	19 (32)	
Histological subtype				
Adenocarcinoma, n (%)	80 (100)	54 (100)	60 (100)	81 (100)
Tumor stage				
I, n (%)	29 (36)	11 (20)	23 (38)	16 (20)
II, n (%)	44 (55)	6 (11)	33 (55)	8 (10)
III, n (%)	7 (9)	17 (32)	3 (5)	26 (32)
IV, n (%)	0 (0)	20 (37)	1 (2)	31 (38)
Receptor status				
ER				
Positive, n (%)	64 (80)		51 (85)	
Negative, n (%)	16 (20)		9 (15)	
PR				
Positive n (%)	56 (70)		46 (82)	
Negative, n (%)	24 (30)		14 (18)	
HER2 status				
Positive, n (%)	14 (17)		11 (18)	
Negative, n (%)	66 (83)		49 (82)	

Data are presented as mean ± standard deviation and range, unless otherwise indicated. The outliers (5 of the training cohort and 12 of the validation cohort) are not included in this table. BC: breast cancer; BMI: body mass index; ER: estrogen receptor; F: females; HER2: human epidermal growth factor receptor; LC: lung cancer; M: males; PR: progesterone receptor.

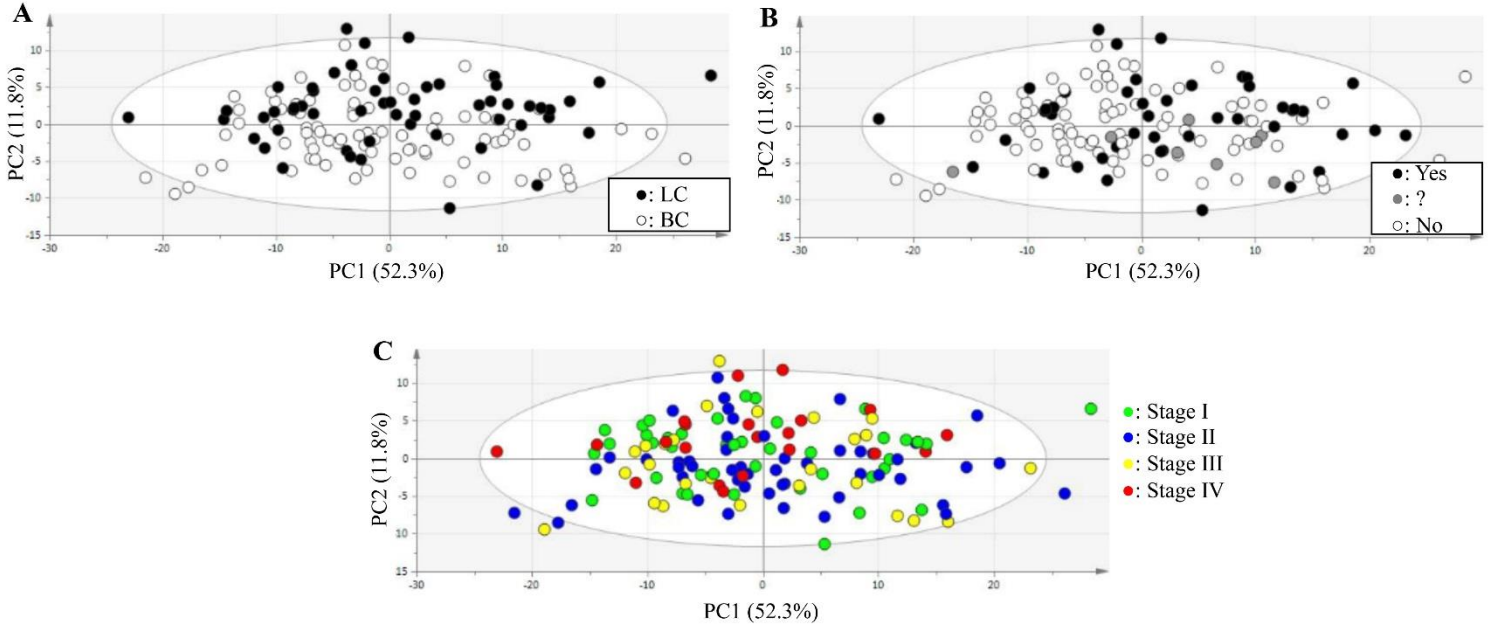


Figure 6.2. PCA score plot of all subjects (145 lung cancer patients and 147 breast cancer patients) stained for (A) disease type, (B) smoking habits and (C) tumor stage. BC: breast cancer; LC: lung cancer; PC: principal component.

Applying OPLS-DA statistics on the training set resulted in a model that allows to classify 93% of the 54 female lung cancer patients and 99% of the 80 female breast cancer patients correctly with an area under the curve (AUC) of 0.96 (**Table 6.2, Figure 6.3A-B**). An independent permutation test confirms that the classification is not due to overfitting (**Figure 6.4**). The predictive accuracy of the model was assessed by applying it to the independent validation cohort consisting of 60 female breast cancer patients and 81 male lung cancer patients (**Table 6.2, Figure 6.3B-C**), resulting in a sensitivity of 89% (89% of the lung cancer patients are correctly classified), a specificity of 82% (82% of the breast cancer patients are correctly classified) and an AUC of 0.94.

Table 6.2. Characteristics of the trained OPLS-DA classification models.

	LV (P+O)	R^2X (cum)	R^2Y (cum)	Q^2 (cum)	Sens (%)	Spec (%)	AUC
			Training cohort				
80 LC vs. 54 BC	5 (1+4)	0.82	0.73	0.63	93	99	0.96
135 LC vs. 140 BC	7 (1+6)	0.85	0.74	0.68	96	94	/
Stage I LC vs. Stage I BC	8 (1+7)	0.89	0.80	0.52	93	100	/
LC vs. TN BC	6 (1+5)	0.85	0.72	0.42	98	95	/
			Validation cohort				
LC vs. BC	/	/	/	/	89	82	0.94

AUC: area under the curve; BC: breast cancer; LC: lung cancer; LV: latent variable; O: number of orthogonal components; P: number of predictive components; R^2X(cum): total explained variation in X; R^2Y(cum): total explained variation in Y; TN: triple negative; Sens: sensitivity; Spec: specificity; Q^2(cum): predicted variation as determined by 7-fold cross-validation.

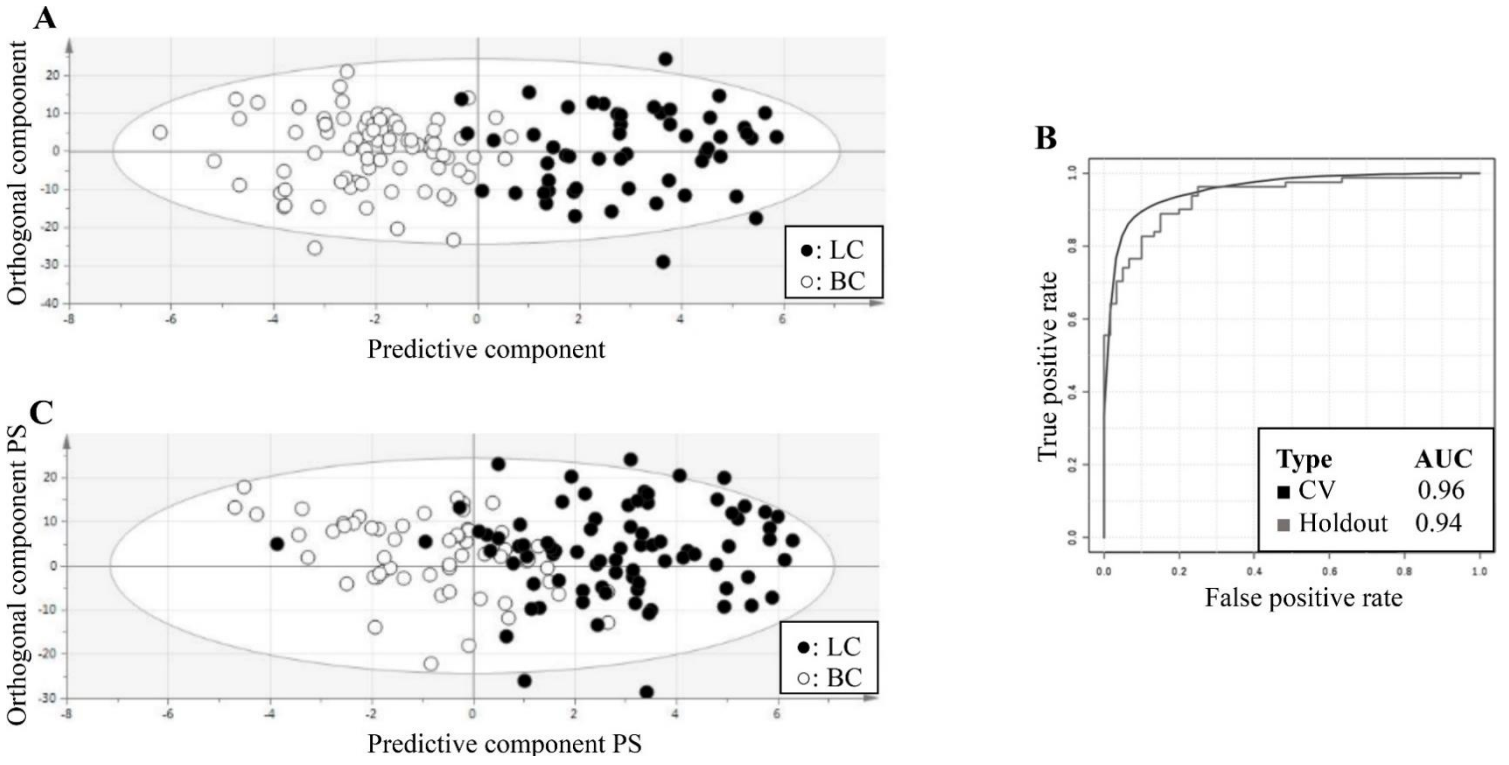


Figure 6.3. (A) OPLS-DA score plot derived from the training cohort of 80 breast cancer patients and 54 lung cancer patients, (B) Receiver operating characteristic curves showing for the cross-validation (CV) of the training cohort as well as for the independent validation the high predictive accuracy of the OPLS-DA model. Regarding the CV, sub-sampling CV was used, i.e. 2/3 of the observations were used for model training and the remaining 1/3 of the observations were used for validation. This procedure was repeated 50 times. (C) OPLS-DA score plot for the classification of the independent cohort of 81 lung cancer patients and 60 breast cancer patients by means of the trained classifier. AUC: area under the curve; BC: breast cancer; CV: cross-validation; LC: lung cancer; PS: predicted scores.

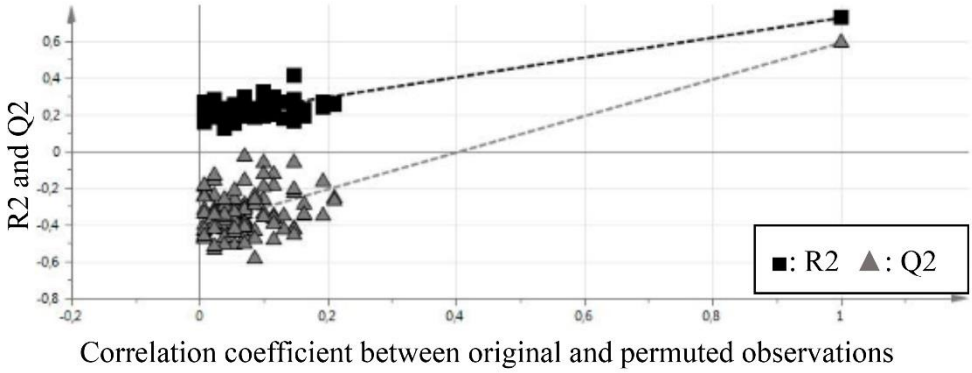


Figure 6.4. Result of the permutation test demonstrating that the obtained classification model is not the result of overfitting. All resulting R^2 and Q^2 values (at the left) are lower than these of the model (at the right), indicative for a valid model. R^2: explained variation; Q^2: predicted variation as determined by 7-fold cross-validation.

Although PCA showed that tumor stage is no confounder, an OPLS-DA model was trained in discriminating between stage I lung cancer patients (n=27) and stage I breast cancer patients (n=52) to confirm this finding. The resulting model classifies 93% of the stage I lung cancer patients and 100% of the stage I breast cancers patients correctly (**Figure 6.5, Table 6.2**). Hence, it can be definitively concluded that tumor stage is no confounder.

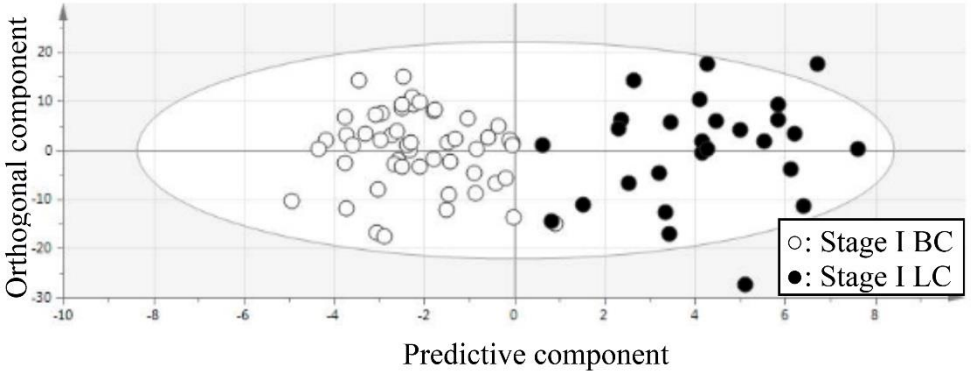


Figure 6.5. OPLS-DA score plot of stage I lung cancer patients (n=27) and stage I breast cancer patients (n=52). BC: breast cancer; LC: lung cancer.

Besides tumor stage, it could be argued that the discrimination is attributable to the fact that most of the breast tumors are estrogen receptor positive (ER^+) and progesterone receptor positive (PR^+, **Table 6.1**). To investigate whether group discrimination is not due to the predominantly positive hormone receptor status

of the breast tumors, the PCA score plot of breast cancer patients was stained for ER and PR status **(Figure 6.6)**. Since ER^+ and PR^+ breast tumors do not cluster, the contribution of hormone receptor status to group discrimination can be excluded. To further affirm this finding, an OPLS-DA model was trained in discriminating between lung cancer patients (n=54) and patients with triple negative breast cancer (n=19, ER^-, PR^- and human epidermal growth factor $receptor^-$ ($HER2^-$)). The obtained model classifies 98% of the lung cancer patients and 95% of the patients with triple negative breast cancer correctly (**Figure 6.7, Table 6.2**). Therefore, it can be concluded that hormone receptor status is no confounder.

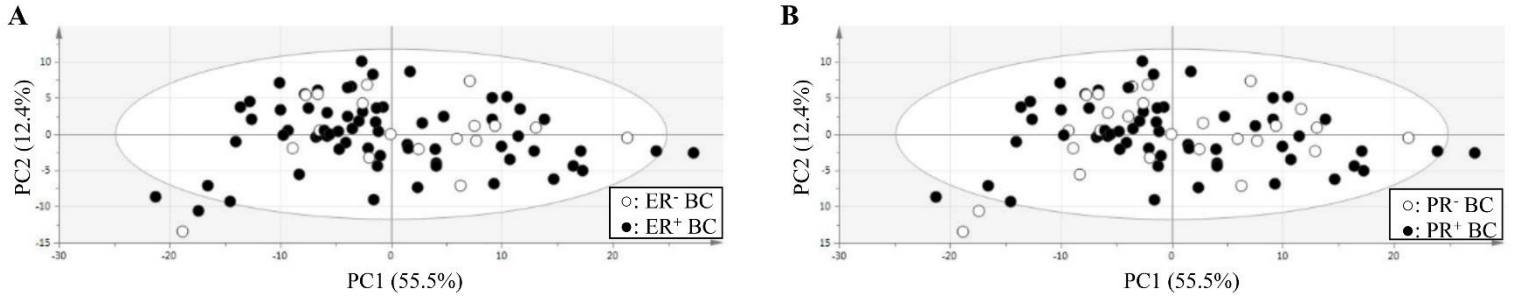


Figure 6.6. (A) PCA class score plot of breast cancer patients stained for estrogen receptor status, (B) PCA class score plot of breast cancer patients stained for progesterone receptor status. BC: breast cancer; ER: estrogen receptor; PC: principal component; PR: progesterone receptor.

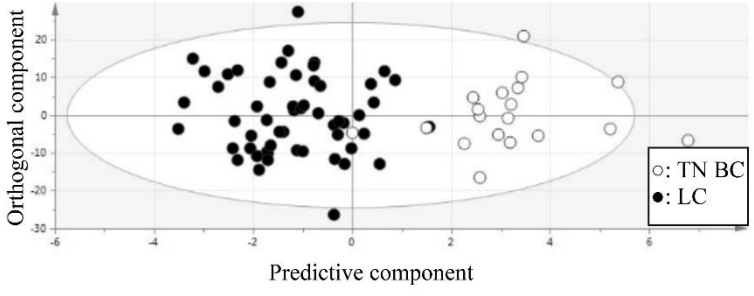


Figure 6.7. OPLS-DA score plot of lung cancer patients (n=54) and patients with triple negative breast cancer (n=19). BC: breast cancer; LC: lung cancer; TN: triple negative.

In an attempt to explain the disturbed biochemical pathways, a classification model was trained in discriminating between the 135 lung cancer patients and the 140 breast cancer patients who remained after the exclusion of outliers rather than dividing both patient groups into a training and a validation cohort (**Figure 6.8**, **Table 6.2**). Only the 50 variables having a VIP value exceeding 0.5 were selected as shown in **Figure 6.9**. **Tables 6.3 and 6.4** show the variables (and contributing metabolites) which are respectively increased or decreased for lung cancer patients together with their univariate p-value and corresponding fold change (FC). A positive/negative FC denotes a relatively higher/lower concentration for lung cancer.

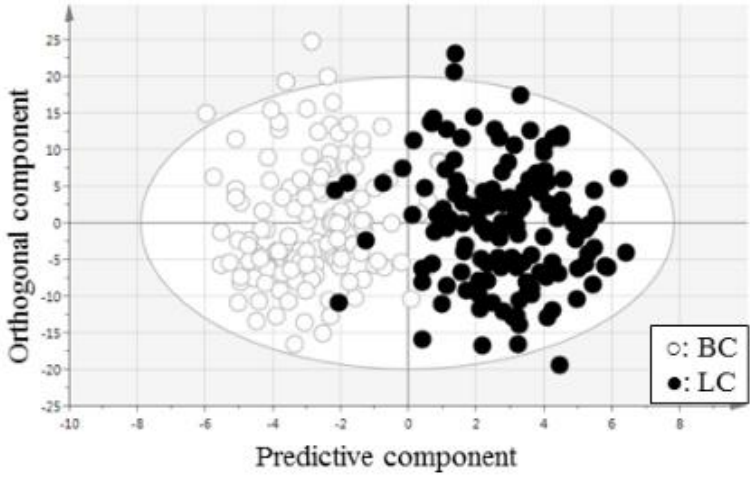


Figure 6.8. OPLS-DA score plot derived from the training cohort of 140 breast cancer patients and 135 lung cancer patients. BC: breast cancer; LC: lung cancer.

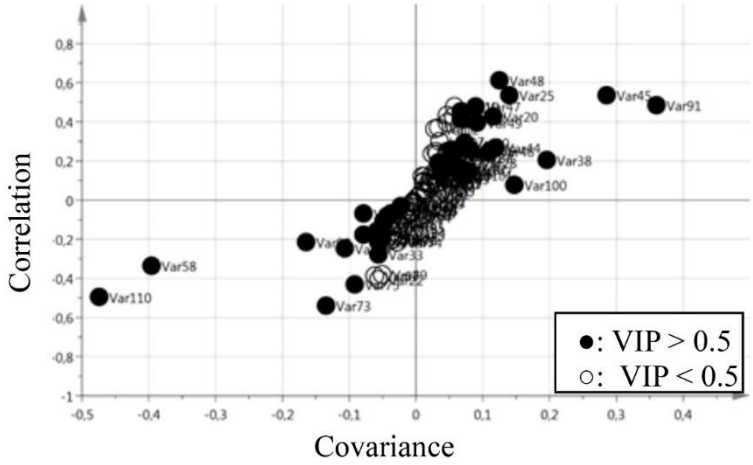


Figure 6.9. S-plot of the OPLS-DA model showing the variables contributing most to group discrimination. Variables situated at the right end are increased in the plasma of lung cancer patients, whereas those situated at the left end are increased in the plasma of breast cancer patients. Variables used to explain the disturbed biochemical pathways (VIP > 0.5) are marked (●). Var: variable; VIP: variable importance for the projection.

The metabolites of which the concentration is increased for lung cancer are glucose, glycerol, tyrosine, aspartate, threonine, leucine, valine, histidine, methionine, N-acetylated glycoproteins, β-hydroxybutyrate and phospholipids with long, saturated fatty acid chains. Furthermore, the metabolites of which the concentration is decreased for lung cancer include alanine, glutamine, creatinine, citrate, sphingomyelin and phosphatidylcholine (and other cholinated phospholipids), lactate and phospholipids with short, unsaturated fatty acid chains.

Table 6.3. **Integration regions (variables) with a VIP value > 0.5 which are increased in the plasma spectra of lung cancer patients.**

VAR	Integration region (ppm)	VIP	p-value	FC	Contributing metabolites	Assignment
100*	1.3450 – 1.2458	3.41	$2.14x10^{-1}$	1.03	Thr, Ile, lipids: CH_3-(CH_2)$_n$- in FAC	Lipids: CH_3-(CH_2)$_n$- in FAC
91*	2.1230 – 1.9720	2.98	$5.53x10^{-13}$	1.07	Glu, Met, Pro, Ile, CH_3 of N-acetylated glycoproteins, lipids: -CH_2-CH=CH- in FAC	CH_3 of NAG
38°	3.9590 – 3.8330	2.34	$4.31x10^{-3}$	1.06	Asp, Met, Ser, Tyr, glucose	Glucose
45§	3.7141 – 3.6680	2.32	$8.35x10^{-15}$	1.24	Ile, glycerol, CH_2-$N^+(CH_3)_3$ of SM & PC	Glycerol
44	3.7390 – 3.7141	1.22	$1.05x10^{-4}$	1.08	Glucose	Glucose
42°	3.7820 – 3.7550	1.15	$6.28x10^{-2}$	1.04	Ala, Glu, Gln, Leu, Lys, glucose	Glucose
25*	4.3159 – 4.2332	1.10	$8.22x10^{-18}$	1.28	Thr, O-CH_2-CH_2-$N^+(CH_3)_3$ of SM & PC	Thr
48	3.6376 – 3.6240	1.00	$1.17x10^{-19}$	1.29	Val	Val
20	5.2526 – 5.2030	0.97	$3.16x10^{-8}$	1.38	C_2H in glycerol backbone of TG & PL	C_2H in glycerol backbone of TG & PL
28§	4.1750 – 4.1260	0.95	$6.57x10^{-3}$	1.09	Lactate, C_1H & C_3H in glycerol backbone of TG & PL	C_1H & C_3H in glycerol backbone of TG & PL
46	3.6680 – 3.6500	0.92	$7.80x10^{-5}$	1.34	Glycerol	Glycerol
21	4.6940 – 4.6620	0.87	$5.90x10^{-1}$	1.01	Glucose	Glucose
43°	3.7550 – 3.7390	0.84	$1.33x10^{-1}$	1.03	Ala, Leu, glucose	Glucose
59	3.2186 – 3.1930	0.81	$1.90x10^{-3}$	1.23	Tyr	Tyr
49	3.6240 – 3.6097	0.78	$2.94x10^{-8}$	1.16	Thr	Thr
101*	1.2458 – 1.2180	0.78	$2.38x10^{-1}$	1.04	Ile, β-hydroxybutyrate	β-hydroxybutyrate
19	5.2752 – 5.2526	0.74	$6.17x10^{-2}$	1.03	Glucose	Glucose
94*	1.8060 – 1.6860	0.74	$4.96x10^{-1}$	1.02	Leu, Lys	Leu
47	3.6500 – 3.6376	0.73	$2.70x10^{-12}$	1.21	Glycerol	Glycerol
51§	3.5914 – 3.5649	0.73	$2.38x10^{-1}$	1.02	Glycerol, Gly, Thr, glucose	Glycerol
37	3.9810 – 3.9590	0.69	$1.17x10^{-3}$	1.11	Asn, His, Ser, Tyr, creatine	?
50	3.6097 – 3.5914	0.69	$1.52x10^{-5}$	1.11	Thr	Thr
24	4.4100 – 4.3159	0.65	$1.14x10^{-3}$	1.09	C_1H & C_3H in glycerol backbone of TG	C_1H & C_3H in glycerol backbone of TG
71*	2.7360 – 2.6600	0.65	$1.42x10^{-2}$	1.06	Asp, Met, citrate	Asp
39°	3.8330 – 3.8100	0.62	$1.52x10^{-1}$	1.03	Ala, Ser, glucose	Glucose
61*	3.1760 – 3.1462	0.56	$1.03x10^{-2}$	1.19	NI	NI
53°	3.5510 – 3.5360	0.52	$9.06x10^{-1}$	1.01	Glucose, acetoacetate, Pro	Glucose

Table 6.3. continued.

VAR	Integration region (ppm)	VIP	p-value	FC	Contributing metabolites	Assignment
4	7.7780 – 7.7480	0.52	$6.23x10^{-8}$	1.47	His	His
72	2.6600 – 2.6300	0.52	$1.22x10^{-7}$	1.55	Met	Met
106*	1.0220 – 1.0020	0.52	$1.28x10^{-1}$	1.03	Val, Ile, Leu	Val

°It can be deduced that these composite regions (variables) emanate from glucose because variable 19, 21 and 44 which represent a unique signal of glucose, are also increased in blood plasma of lung cancer patients. §It can be deduced that these composite regions (variables) emanate from glycerol because variable 20, 24, 46 and 47 which represent a unique signal of glycerol, are also increased in blood plasma of lung cancer patients. *The assignment of metabolites in crowded regions with signal overlap was improved by spiking plasma samples with known metabolites in relevant concentrations on a 900 MHz NMR spectrometer (Lille, France), having a higher spectral resolution and signal-to-noise ratio. Amino acids are presented by their 3-letter code. FAC: fatty acid chain; FC: fold change; NAG: N-acetylated glycoproteins; NI: non-identified; PC: phosphatidylcholine; ppm: parts per million; PL: phospholipids; SM: sphingomyelin; TG: triglycerides; VAR: variable; VIP: variable importance for the projection.

Table 6.4. Integration regions (variables) with a VIP value > 0.5 which are decreased in the plasma spectra of lung cancer patients.

VAR	Integration region (ppm)	VIP	*P* value	FC	Contributing metabolites	Assignment
110	0.9660 – 0.8000	3.92	$4.56x10^{-13}$	-1.08	Lipids: C**H**$_3$-(CH$_2$)$_n$- in FAC	Lipids: C**H**$_3$-(CH$_2$)$_n$- in FAC
58*	3.3230 – 3.2186	3.64	$1.21x10^{-6}$	-1.13	His, Phe, Tyr, N^+(C**H**$_3$)$_3$ of SM & PC, glucose	N^+(C**H**$_3$)$_3$ of SM & PC
99*	1.3740 – 1.3450	2.05	$6.08x10^{-1}$	-1.01	Lactate, Thr	Lactate
54	3.5460 – 3.3980	2.03	$3.17x10^{-1}$	-1.02	Pro, glucose, acetoacetate	?
98	1.4200 – 1.374	1.93	$4.32x10^{-3}$	-1.13	Lactate	Lactate
18	5.4300 – 5.2752	1.22	$9.12x10^{-1}$	-1.01	Lipids: -CH=CH- in FAC	Lipids: -CH=CH- in FAC
95#	1.6860 – 1.5600	1.21	$4.53x10^{-1}$	-1.03	Lys, lipids: -C**H**$_2$-CH$_2$-C=O and -C**H**$_2$-CH$_2$-CH=CH in FAC	Lipids: -C**H**$_2$-CH$_2$-C=O and -C**H**$_2$-CH$_2$-CH=CH in FAC
73	2.5960 – 2.5340	1.06	$5.26x10^{-16}$	-1.41	Citrate	Citrate
96*	1.5400 – 1.4900	1.05	$1.87x10^{-4}$	-1.12	Ala, Ile, Lys	Ala
90*	2.1970 – 2.1230	0.93	$1.00x10^{-2}$	-1.05	Gln, Glu, Pro, Met	Gln
66	3.0640 – 2.9950	0.85	$1.22x10^{-1}$	-1.03	Lys, α-ketoglutarate	?
67	2.9950 – 2.8860	0.81	$5.28x10^{-2}$	-1.05	Asn, lipids: =CH-C**H**$_2$-CH= in FAC	?
75	2.5150 – 2.4920	0.75	$3.83x10^{-10}$	-1.21	Gln	Gln
102	1.2180 – 1.1300	0.71	$9.11x10^{-1}$	-1.01	NI	NI
76*	2.4920 – 2.4500	0.66	$9.46x10^{-3}$	-1.07	Gln, α-ketoglutarate, β-hydroxybutyrate	Gln
69#	2.8550 – 2.7500	0.64	$1.51x10^{-1}$	-1.02	Asn, Asp, lipids: =CH-C**H**$_2$-CH= in FAC	Lipids: =CH-C**H**$_2$-CH= in FAC
40	3.8100 – 3.7956	0.56	$9.32x10^{-2}$	-1.04	Ala, Glu, glucose, Gln	?
64*	3.0860 – 3.0716	0.55	$7.98x10^{-3}$	-1.07	Tyr, Lys, creatinine	Creatinine
41	3.7956 – 3.7820	0.54	$1.45x10^{-1}$	-1.03	Ala, Glu, glucose, Gln, Leu, Lys	?
92	1.9720 – 1.9240	0.53	$7.50x10^{-4}$	-1.08	Ile, Lys, acetate	?

#It can be deduced that these composite regions (variables) emanate from lipids because variable 18 and 110 which represent a unique signal of lipids are also increased in blood plasma of breast cancer patients. *The assignment of metabolites in crowded regions with signal overlap was improved by spiking plasma samples with known metabolites in relevant concentrations on a 900 MHz NMR spectrometer (Lille, France), having a higher spectral resolution and signal-to-noise ratio. Amino acids are presented by their 3-letter code. FAC: fatty acid chain; FC: fold change; PC: phosphatidylcholine; ppm: parts per million; SM: sphingomyelin; VAR: variable; VIP: variable importance for the projection.

Discussion

The rapidly expanding field of metabolomics has been making progress in the area of oncology (21-23) and studies have explored the disturbed metabolism of lung or breast cancer by means of ^{1}H-NMR spectroscopy (13, 24-27). However, to our knowledge, this is the first report which examines whether cancer types have specific profiles. Hereto, a homogenous study population of cancer patients suffering with an adenocarcinoma was selected. Furthermore, since all breast cancer patients were females, only females were included in the training cohort in order to exclude gender as a confounder. This study demonstrates that the metabolic phenotype allows to classify 93% of the female lung cancer patients and 99% of the female breast cancer patients correctly with an AUC of 0.96. Additionally, male lung cancer patients and female breast cancer patients of an independent cohort were classified by the trained classification model with a sensitivity of 89%, a specificity of 82% and an AUC of 0.94, confirming that gender is no confounder.

Investigation of the underlying disturbed metabolism demonstrates that plasma of lung cancer patients is characterized by increased levels of glucose, glycerol, tyrosine, aspartate, threonine, leucine, valine, histidine, methionine, N-acetylated glycoproteins, β-hydroxybutyrate and phospholipids with long, saturated fatty acid chains and by decreased levels of alanine, glutamine, creatinine, citrate, sphingomyelin and phosphatidylcholine, lactate and phospholipids with short, unsaturated fatty acid chains. Note that the increase/decrease in concentrations of plasma metabolites is always expressed for lung cancer as compared to breast cancer. The increase of phospholipids with long, saturated fatty acid chains can be derived from variable 100, i.e. the integration value of the region between 1.25-1.35 ppm in the ^{1}H-NMR spectrum representing CH_2 groups that are separated by at least two carbons from a double bond, and so mainly reflects the relative concentration of phospholipids with long, saturated fatty acid chains. In contrast, variable 110 represents the CH_3 groups and so rather reflects the relative concentration of phospholipids with short, unsaturated fatty acid chains.

Because cancer cells often outgrow the surrounding vasculature, they have adopted an oxygen-independent metabolism to produce sufficient energy for proliferation. Even in the presence of normal oxygen levels, cancer cells prefer to

convert glucose into lactate rather than oxidizing it via oxidative phosphorylation, a hallmark known as the Warburg effect or "aerobic glycolysis" (12). This latter term is somewhat confusing, because oxygen is not involved in the fermentation of pyruvate to lactate. Concerning the elevated glycolytic rate in cancer cells, literature reports that the degree of uptake of the radioactive fluorodeoxyglucose (^{18}F-FDG) tracer as measured by PET/CT is correlated with metabolic activity and long-term prognosis (28-30). Lung tumors display a higher ^{18}F-FDG uptake than breast tumors, indicating that lung tumors are metabolically more active and by implication more aggressive.

An ongoing study of our research group in a large cohort of 233 lung cancer patients and 226 controls reveals increased plasma levels of glucose and decreased levels of lactate for lung cancer (14). These metabolic alterations demonstrate that measurements of the metabolite profile in blood plasma are reflecting the response of the body to the Warburg effect, i.e. hepatic glycogenolysis and gluconeogenesis, rather than the Warburg effect itself (as occurring in the cancer cells) (14). Since lung tumors are metabolically more active than breast tumors, it is expected that the body response to the Warburg effect will be more prominent. Note that this study compares metabolite concentrations between different cancer types and that a decrease of a specific metabolite in lung cancer patients as compared to breast cancer patients might still represent an increase relative to controls and vice versa.

In more detail, this study reveals that lung cancer patients have increased plasma levels of glucose and decreased levels of alanine and lactate, in line with the fact that lung tumors are metabolically more active as compared to breast cancer. More specifically, the response of the body to compensate for the lack of glucose as energy and carbon source for normal cells will be more prominent in the lung cancer patients. As a result, less oxaloacetate will be available to condense with acetyl-CoA and the Krebs cycle and subsequent oxidative phosphorylation in the liver will be more hampered. Therefore, the conversion of acetyl-CoA to ketone bodies will be more pronounced. The increased hepatic synthesis of ketone bodies is confirmed by the increased levels of the ketone body β-hydroxybutyrate and the ketogenic amino acid leucine.

Besides hepatocytes, which display an impaired Krebs cycle due to the compensating gluconeogenesis in response to the Warburg effect, also cancer cells are characterized by a disturbed Krebs cycle (31). Since lung tumors are metabolically more active, the Krebs cycle is expected to be more hampered, which is affirmed by decreased levels of citrate and further supported by increased levels of amino acids which normally would be consumed for the production of Krebs cycle intermediates (aspartate, threonine, valine, histidine, methionine, tyrosine) (32). The plasma concentration of citrate is decreased since, once it is shuttled to the cytoplasm, it becomes converted to acetyl-CoA, being a critical precursor of fatty acids and cholesterol, i.e. fatty acids and cholesterol are synthesized from the carbons provided by acetyl-CoA (33). After conversion to phospholipids, the fatty acids are incorporated in the plasma membrane of new daughter cells together with cholesterol. In order to maintain fatty acid and cholesterol synthesis, citrate is replenished via the glutaminolysis, i.e. a salvage pathway in which glutamine is converted to α-ketoglutarate (31). This α-ketoglutarate is subsequently reduced to isocitrate which can be converted to citrate (33). Since lung tumors are metabolically more active and consequently need more anabolic precursors for their increased proliferation, lipogenesis and subsequent membrane synthesis will be more pronounced. All the above is affirmed by decreased levels of glutamine, citrate and sphingomyelin and phosphatidylcholine. Since long, saturated fatty acid chains have very strong Van der Waals interactions, the increased plasma levels of phospholipids with long, saturated fatty chains and the decreased levels of phospholipids with short, unsaturated fatty acid chains indicate that the cell membrane of lung tumors is more rigid and less sensitive to lipid peroxidation (34). Notice that although the increased plasma levels of glycerol seem to be in contrast with the enhanced lipogenesis and subsequent membrane synthesis in lung cancer cells, this discrepancy might be explained by the fact that cancer cells do not use glycerol obtained from plasma to create their membranes but rather glycerol derived from the glycolytic pathway. Finally, plasma of lung cancer patients is characterized by increased levels of N-acetylated glycoproteins. Since the enzyme catalyzed addition of N-acetylglucosamine to serine or threonine residues of proteins plays a role in the metabolic reprogramming of cancer cells (35) and since lung tumors seem to be metabolically more active, this is rather expected.

Conclusions

This study reveals that the plasma metabolic phenotype allows to discriminate between lung and breast cancer, indicating that the plasma metabolite profile reflects more than a common cancer marker, i.e. the cancer types studied are characterized by their own, specific metabolite profile. To the best of our knowledge, this has never been published before.

References

1. Bag S, Banerjee DR, Basak A, Das AK, Pal M, Banerjee R, et al. NMR ((1)H and (13)C) based signatures of abnormal choline metabolism in oral squamous cell carcinoma with no prominent Warburg effect. Biochemical and biophysical research communications. 2015;459(4):574-8.
2. Kumar D, Gupta A, Mandhani A, Sankhwar SN. Metabolomics-derived prostate cancer biomarkers: fact or fiction? Journal of proteome research. 2015;14(3):1455-64.
3. Zhang L, Jin H, Guo X, Yang Z, Zhao L, Tang S, et al. Distinguishing pancreatic cancer from chronic pancreatitis and healthy individuals by (1)H nuclear magnetic resonance-based metabonomic profiles. Clinical biochemistry. 2012;45(13-14):1064-9.
4. Lindon JC, Nicholson JK. Spectroscopic and statistical techniques for information recovery in metabonomics and metabolomics. Annu Rev Anal Chem 2008;1:45-69.
5. Fiehn O. Metabolomics--the link between genotypes and phenotypes. Plant molecular biology. 2002;48(1-2):155-71.
6. Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica; the fate of foreign compounds in biological systems. 1999;29(11):1181-9.
7. Bervoets L, Louis E, Reekmans G, Mesotten L, Thomeer M, Adriaensens P, et al. Influence of preanalytical sampling conditions on the ^{1}H NMR metabolic profile of human blood plasma and introduction of the standard preanalytical code used in biobanking. Metabolomics. 2015;11:1197-1207.
8. Amoedo ND, Valencia JP, Rodrigues MF, Galina A, Rumjanek FD. How does the metabolism of tumour cells differ from that of normal cells. Bioscience reports. 2013;33(6).
9. Munoz-Pinedo C, El Mjiyad N, Ricci JE. Cancer metabolism: current perspectives and future directions. Cell death & disease. 2012;3:e248.
10. Sciacovelli M, Gaude E, Hilvo M, Frezza C. The metabolic alterations of cancer cells. Methods in enzymology. 2014;542:1-23.
11. Iurlaro R, Leon-Annicchiarico CL, Munoz-Pinedo C. Regulation of cancer metabolism by oncogenes and tumor suppressors. Methods in enzymology. 2014;542:59-80.
12. Chen X, Qian Y, Wu S. The Warburg effect: evolving interpretations of an established concept. Free radical biology & medicine. 2015;79:253-63.
13. Louis E, Bervoets L, Reekmans G, De Jonge E, Mesotten L, Thomeer M, et al. Phenotyping human blood plasma by ^{1}H-NMR: a robust protocol based on metabolite spiking and its evaluation in breast cancer. Metabolomics. 2015;11:225-36.
14. Louis E, Vanhove K, Stinkens K, Mesotten L, Reekmans G, Guedens W, et al., editors. Validation of ^{1}H-NMR-based metabolomics as a new, complementary tool for the detection of lung cancer via human blood plasma. EACR - Precision Medicine for Cancer; 2015; Luxembourg, Luxembourg.
15. Goldstraw P, Crowley J, Chansky K, Giroux DJ, Groome PA, Rami-Porta R, et al. The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours. Journal of thoracic oncology. 2007;2(8):706-14.
16. Singletary SE, Allred C, Ashley P, Bassett LW, Berry D, Bland KI, et al. Revision of the American Joint Committee on Cancer staging system for breast cancer. Journal of clinical oncology. 2002;20(17):3628-36.
17. Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). J Chemometr. 2002;16(3):119-28.
18. Xia J, Broadhurst DI, Wilson M, Wishart DS. Translational biomarker discovery in clinical metabolomics: an introductory tutorial. Metabolomics. 2013;9(2):280-99.
19. Wiklund S, Johansson E, Sjostrom L, Mellerowicz EJ, Edlund U, Shockcor JP, et al. Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. Analytical chemistry. 2008;80(1):115-22.
20. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met. 1995;57(1):289-300.

21. Armitage EG, Barbas C. Metabolomics in cancer biomarker discovery: current trends and future perspectives. Journal of pharmaceutical and biomedical analysis. 2014;87:1-11.
22. Duarte IF, Gil AM. Metabolic signatures of cancer unveiled by NMR spectroscopy of human biofluids. Progress in nuclear magnetic resonance spectroscopy. 2012;62:51-74.
23. Spratlin JL, Serkova NJ, Eckhardt SG. Clinical applications of metabolomics in oncology: a review. Clinical cancer research. 2009;15(2):431-40.
24. Carrola J, Rocha CM, Barros AS, Gil AM, Goodfellow BJ, Carreira IM, et al. Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of urine. Journal of proteome research. 2011;10(1):221-30.
25. Gu H, Pan Z, Xi B, Asiago V, Musselman B, Raftery D. Principal component directed partial least squares analysis for combining nuclear magnetic resonance and mass spectrometry data in metabolomics: application to the detection of breast cancer. Analytica chimica acta. 2011;686(1-2):57-63.
26. Oakman C, Tenori L, Biganzoli L, Santarpia L, Cappadona S, Luchinat C, et al. Uncovering the metabolomic fingerprint of breast cancer. The international journal of biochemistry & cell biology. 2011;43(7):1010-20.
27. Rocha CM, Carrola J, Barros AS, Gil AM, Goodfellow BJ, Carreira IM, et al. Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of blood plasma. Journal of proteome research. 2011;10(9):4314-24.
28. Sachs S, Bilfinger TV, Komaroff E, Franceschi D. Increased standardized uptake value in the primary lesion predicts nodal or distant metastases at presentation in lung cancer. Clinical lung cancer. 2005;6(5):310-3.
29. Higashi K, Ueda Y, Arisaka Y, Sakuma T, Nambu Y, Oguchi M, et al. 18F-FDG uptake as a biologic prognostic factor for recurrence in patients with surgically resected non-small cell lung cancer. Journal of nuclear medicine. 2002;43(1):39-45.
30. Vesselle H, Schmidt RA, Pugsley JM, Li M, Kohlmyer SG, Vallires E, et al. Lung cancer proliferation correlates with [F-18]fluorodeoxyglucose uptake by positron emission tomography. Clinical cancer research. 2000;6(10):3837-44.
31. Weinberg F, Chandel NS. Mitochondrial metabolism and cancer. Annals of the New York Academy of Sciences. 2009;1177:66-73.
32. Harvey R, Ferrier D. Biochemistry. 5 ed. Philadelphia: Lipincott Williams & Wilkins; 2011.
33. Lunt SY, Vander Heiden MG. Aerobic glycolysis: meeting the metabolic requirements of cell proliferation. Annual review of cell and developmental biology. 2011;27:441-64.
34. Rysman E, Brusselmans K, Scheys K, Timmermans L, Derua R, Munck S, et al. De novo lipogenesis protects cancer cells from free radicals and chemotherapeutics by promoting membrane lipid saturation. Cancer research. 2010;70(20):8117-26.
35. Jozwiak P, Forma E, Brys M, Krzeslak A. O-GlcNAcylation and Metabolic Reprogramming in Cancer. Front Endocrinol. 2014;5:145.

CHAPTER 7

The ^{1}H-NMR-derived metabolic phenotype of plasma enables to differentiate between lung, breast and colorectal cancer

Based on:

Robby Louis, Evelyne Louis, Kirsten Stinkens, Liesbet Mesotten, Eric de Jonge, Michiel Thomeer, Philip Caenepeel, Peter Adriaensens. The metabolic phenotype of blood plasma allows to discrimininate between colorectal cancer, breast cancer and lung cancer. Under review.

Abstract

Although many studies have demonstrated that plasma metabolic phenotyping allows to discriminate between cancer patients and controls, it has not yet been thoroughly investigated whether different cancer types elicit distinguishable metabolic signatures. Therefore, the present study was designed to examine whether metabolic phenotyping of blood plasma by proton nuclear magnetic resonance (^{1}H-NMR) spectroscopy allows to discriminate between 37 colorectal cancer, 37 breast cancer and 37 lung cancer patients. Hereto, plasma ^{1}H-NMR spectra were rationally divided into 110 integration regions defined on the basis of spiking experiments with known metabolites. The normalized integration values of these 110 regions, which represent the metabolic phenotype, were used as statistical variables to construct a classification model which enables to discriminate between the three aforementioned cancer types. The resulting model allows to classify 78% of the colorectal cancer patients, 95% of the breast cancer patients and 84% of the lung cancer patients correctly. Although the number of cancer patients in each group has to be increased and an independent validation study has to be performed in order to confirm these findings, the present study provides preliminary evidence that the plasma metabolic phenotype has potential to become a complementary diagnostic tool to differentiate between cancer types in addition to known common cancer biomarkers.

Introduction

It is widely accepted that cancer cells exhibit a major reprogramming of their energy metabolism in order to fulfill the high metabolic demands that are associated with increased cell proliferation and survival (1). Because the metabolic alterations in cancer cells provoke changes in the metabolic phenotype of the patient, metabolites might serve as attractive biomarkers for facilitating the diagnosis of cancer. Complex mixtures of metabolites in biofluids, such as plasma, serum or urine, can be mined for diagnostic biomarkers by means of the metabolomics approach. This discipline represents a relatively new '-omics' science downstream of genomics, transcriptomics and proteomics that uses an analytical platform in conjunction with multivariate pattern recognition approaches in order to discover and monitor metabolic changes in patient biospecimens related to disease status or in response to a medical or external intervention (2). Since metabolites are the end products of all cellular regulatory processes, their levels can be regarded as the ultimate response of biological systems to genetic, biological and/or environmental perturbations (3). More specifically, alterations in the concentrations of metabolites are the net result of epigenetic changes, genetic variation and changes in the activity and/or levels of enzymes (4). The metabolome, i.e. the complete set of metabolites present in the human body, is therefore the most closely related to the observed phenotype and provides the most accurate representation of the functional status of the patient when compared with the genome, transcriptome and proteome (5). Consequently, metabolomics holds great promise for early cancer detection as the concentrations of metabolites are sensitive to subtle changes in the pathological status of the patient, such as the early onset of tumor growth (6).

One of the main analytical tools that is commonly used in metabolomics to identify and quantify a wide range of metabolites in biological samples is proton nuclear magnetic resonance (^{1}H-NMR) spectroscopy. This technique only requires minimal sample preparation, needs no chemical derivatization, can be easily automated and is fast, non-destructive, highly reproducible and relatively cheap on a per sample basis, making it a promising platform for performing high-throughput diagnostic analyses on a large scale (5, 7). In the past decade, a growing interest in the diagnostic utility of ^{1}H-NMR-based metabolomics of blood plasma has

emerged in the field of oncology, and various studies have already established the potential of this methodology to detect various cancer types (8, 9). However, since the majority of these studies only compared the plasma metabolic phenotype of cancer patients with that of controls, it remains unclear whether different tumors share the same metabolic perturbations or whether certain metabolic alterations are specific for certain cancer types. Recently, our research group has demonstrated that metabolic phenotyping of plasma by ^{1}H-NMR spectroscopy allows to discriminate between breast and lung cancer, classifying 99% of 80 female breast cancer patients and 93% of 54 female lung cancer patients correctly. These results were successfully validated in an independent cohort in which 82% of 60 female breast cancer patients and 89% of 81 male lung cancer patients were correctly classified (10). In order to further explore the ability of the plasma metabolic phenotype to differentiate between cancer types, the present study aims to investigate whether the ^{1}H-NMR-derived metabolic phenotype of blood plasma allows to discriminate between colorectal cancer patients, breast cancer patients and lung cancer patients.

Materials and methods

Subjects

Thirty-seven colorectal cancer patients and an equal number of breast cancer and lung cancer patients, all with an adenocarcinoma, were randomly selected in order to construct a classification model which permits to discriminate between the different cancer types based on the metabolic phenotype of blood plasma. The colorectal cancer patients were included in the Limburg Positron Emission Tomography (PET) Center (Hasselt, Belgium), at the Gastroenterology Department of Ziekenhuis Oost-Limburg (Campus Sint-Jan, Genk, Belgium) and in the Center for Specialized Medicine (Reumacentrum, Genk, Belgium). The diagnosis of colorectal adenocarcinoma was confirmed by histopathological examination of biopsies taken during endoscopy. The lung cancer patients were recruited in the Limburg PET center and at the Department of Respiratory Medicine of University Hospital Leuven (Campus Gasthuisberg, Leuven, Belgium). The diagnosis of lung adenocarcinoma was confirmed by means of a pathological biopsy or by a medical doctor specialized in the interpretation of radiological and clinical data regarding lung cancer. The stage of both the colorectal and lung tumors was defined according to the 7th edition of the tumor, nodes and metastases classification of malignant tumors. The breast cancer patients were included at the day of primary surgery at the Gynaecology Department of Ziekenhuis Oost-Limburg (Campus Sint-Jan). The diagnosis of breast adenocarcinoma was confirmed by a core needle biopsy and the stage of the tumors was defined by the revised staging system for breast cancer (11). For each group of cancer patients, blood collection and sample preparation were performed according to a fixed protocol and by trained staff. Furthermore, for each group, the following exclusion criteria were defined: 1) poorly controlled diabetes (fasting blood glucose ≥ 200 mg/dl), 2) any inflammatory condition, 3) any other cancer type besides colorectal, breast or lung cancer, 4) history or treatment of cancer during the past five years, 5) medication intake on the morning of blood sampling and 6) no fasting for at least six hours prior to blood sampling. All procedures were conducted in accordance with the ethical rules of the Helsinki declaration and Good Clinical Practice and the study protocols were approved by the Medical Ethics Committees of Ziekenhuis Oost-Limburg, Hasselt University (Campus

Diepenbeek, Hasselt, Belgium) and University Hospital Leuven. All participants provided written informed consent prior to their inclusion. The studies in which the patients were enrolled are registered at clinicaltrials.gov (NCT02364154 and NCT02362776).

Blood sampling and processing

See **Chapter 2**: Blood sampling and processing.

NMR sample preparation and analysis

See **Chapter 2**: NMR sample preparation and analysis.

Spectral processing

See **Chapter 2**: Spectral processing.

Statistical analysis

Multivariate statistics were performed using SIMCA-P$^+$ (version 14.0, Umetrics, Umea, Sweden). After the variables had been subjected to mean centering and Pareto scaling, orthogonal partial least squares discriminant analysis (OPLS-DA) was performed to construct a classification model which discriminates between the colorectal cancer, breast cancer and lung cancer patients based on their metabolic phenotype (12). The validity of the established model was evaluated based on 1) the total amount of variation between and within the three groups explained by the model (denoted as $R^2Y(cum)$ and $R^2X(cum)$, respectively) and 2) the predictive ability of the model as determined by sevenfold cross-validation (denoted as $Q^2(cum)$). Furthermore, permutation testing was performed to ensure that the discrimination between the three cancer types was not due to overfitting of the data. Additionally, principal component analysis (PCA) was carried out to identify possible confounding effects of age, gender, BMI or tumor stage on group discrimination.

Results and discussion

Subject characteristics

The demographical and clinical characteristics of the colorectal cancer, breast cancer and lung cancer patients are displayed in **Table 7.1**. All breast cancer patients were female, while the majority of colorectal cancer patients (70%) and lung cancer patients (57%) were male. However, PCA revealed that there was no apparent clustering with respect to gender in the entire patient cohort, indicating that the discrepancy in gender among the three different groups of cancer patients is not a confounding factor (**Figure 7.1A**). Additionally, there were no clusters observed when PCA score plots of the entire patient cohort were stained according to age, BMI or tumor stage, thereby confirming that none of these factors have a confounding effect on the discrimination between colorectal cancer, breast cancer and lung cancer patients (**Figure 7.1B-D**).

Table 7.1. Characteristics of the subjects included in the study.

	CRC	BC	LC
Number of subjects, n	37	37	37
Gender, n (%)			
Male	26 (70)	0 (0)	21 (57)
Female	11 (30)	37 (100)	16 (43)
Age, yrs	65 ± 10	57 ± 10	65 ± 10
(range)	(45-86)	(42-78)	(43-88)
BMI, kg/m^2	26.9 ± 4.5	25.7 ± 3.1	25.6 ± 4.8
(range)	(21.4-40.8)	(20.2-35.4)	(17.0-35.8)
Smoking habits			
Smokers, n (%)	4 (11)	8 (22)	18 (49)
Ex-smoker, n (%)	20 (54)	0 (0)	16 (43)
Non-smoker, n (%)	10 (27)	0 (0)	3 (8)
Unknown, n (%)	3 (8)	29 (78)	0 (0)
Amount of tumors, n	40	39	37
Histological subtype			
Adenocarcinoma, n (%)	40 (100)	39 (100)	37 (100)
Tumor stage			
0, n (%)	1 (2)	0 (0)	0 (0)
I, n (%)	10 (25)	22 (56)	8 (21)
II, n (%)	8 (20)	16 (41)	4 (11)
III, n (%)	13 (33)	1 (3)	11 (30)
IV, n (%)	6 (15)	0 (0)	14 (38)
Unknown, n (%)	2 (5)	0 (0)	0 (0)

Data are presented as mean ± standard deviation and range, unless otherwise indicated. BC: breast cancer patients; BMI: body mass index; CRC: colorectal cancer patients; LC: lung cancer patients.

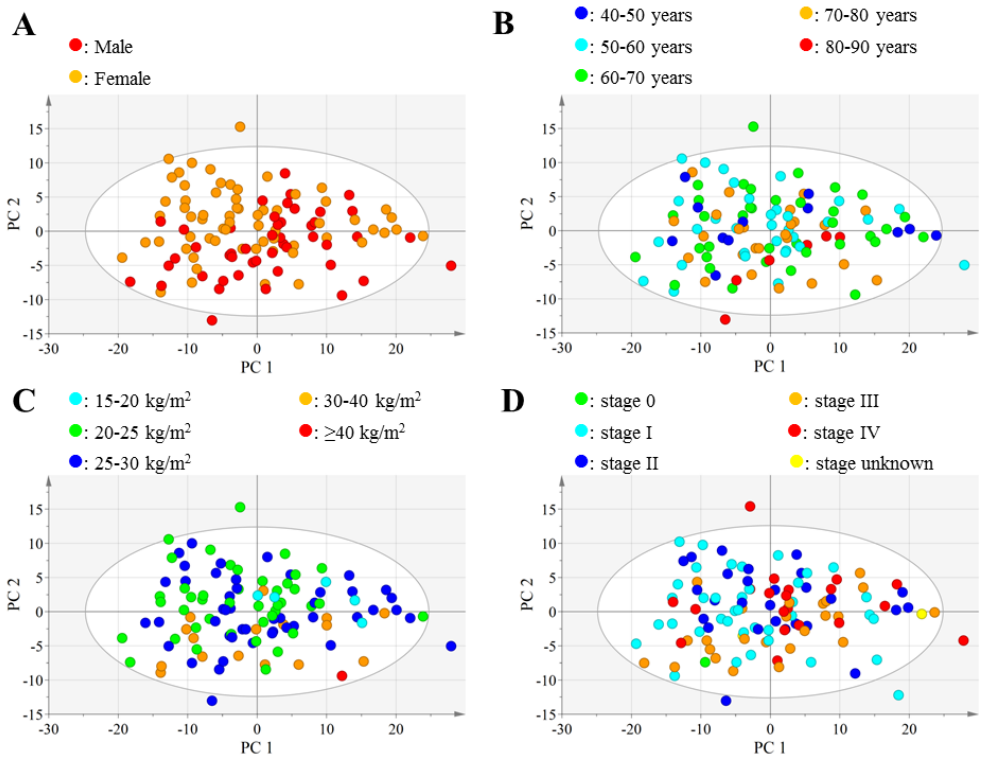


Figure 7.1. Search for possible confounding factors in the discrimination between colorectal, breast and lung cancer based on the metabolic phenotype of blood plasma determined by ^{1}H-NMR spectroscopy. PCA score plots of the entire patient cohort stained according to A) gender, B) age, C) body mass index and D) tumor stage. PC: principal component.

Discrimination between colorectal, breast and lung cancer based on the plasma metabolic phenotype

The OPLS-DA model that was established to discriminate between the three groups of cancer patients allows to classify 29 out of 37 colorectal cancer patients (78%), 35 out of 37 breast cancer patients (95%) and 31 out of 37 lung cancer patients (84%) correctly (R^2X(cum): 0.875 and R^2Y(cum): 0.561) **(Figure 7.2A)**. The predictive ability of the model is relatively high (Q^2(cum) = 0.387), indicating that the discrimination between the three cancer types based on the metabolic phenotype is valid. The validity of the OPLS-DA model was also confirmed by permutation testing, which demonstrated that the R^2(cum) and Q^2(cum) values of the actual model are higher than those of twenty permuted models **(Figure 7.2B)**. Although the number of subjects included in this study is still limited, these

results strongly suggest that colorectal cancer, breast cancer and lung cancer patients display different metabolic alterations in blood plasma, independent of tumor stage. Future perspectives aim to increase the number of cancer patients in each group in order to confirm these findings and unravel the underlying metabolic perturbations responsible for the discrimination between colorectal, breast and lung cancer. In addition, a validation study should be performed in an independent patient cohort to ensure that the ability to differentiate between the aforementioned cancer types by ^{1}H-NMR-based metabolomics is reproducible and therefore valid.

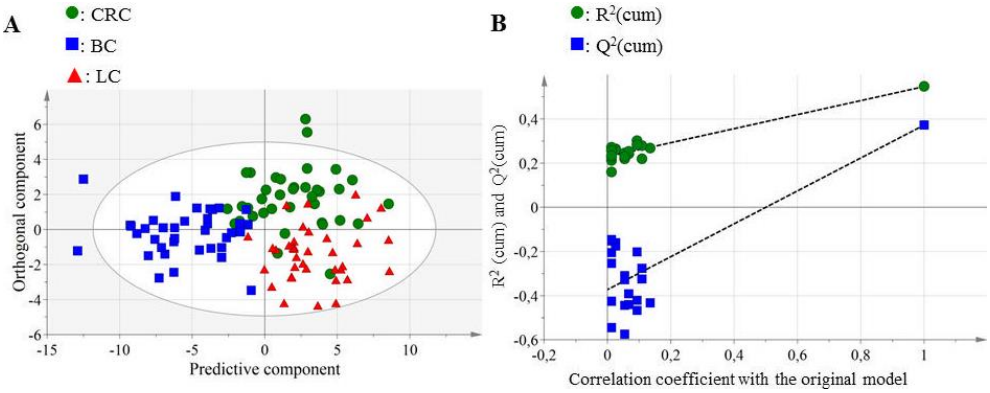


Figure 7.2. Discrimination between colorectal, breast, and lung cancer based on the ^{1}H-NMR-derived metabolic phenotype of blood plasma. A) OPLS-DA score plot discriminating between 37 colorectal cancer patients and an equal number of breast cancer patients and lung cancer patients based on their metabolic phenotype. The model has two predictive components, which describe the variation between the three groups, and five orthogonal components, which display the variation within the different groups, B) Permutation plot comparing the goodness of fit of the constructed OPLS-DA model with that of twenty permuted models. The R^2(cum) and Q^2(cum) values of the actual model are located at the right, whereas those of the permuted models are located at the left. R^2(cum) denotes the variation explained by the model, while Q^2(cum) is the predictive ability of the model as determined by 7-fold cross-validation. BC: breast cancer patients; CRC: colorectal cancer patients; LC: lung cancer patients.

Conclusions

The present study demonstrates that metabolic phenotyping of plasma by ^{1}H-NMR spectroscopy allows to discriminate between the three most common cancer types worldwide, i.e. lung cancer, breast cancer and colorectal cancer, respectively. Although it has already been demonstrated for several types of malignancies that cancer patients can be differentiated from controls on the basis of their plasma metabolic phenotype (8, 9), the present study indicates that, in addition, different cancer types display distinct metabolic characteristics rather than a common cancer metabolome. Keeping in mind that the sample size in this study is still limited, the number of patients in each group should be increased and an independent validation study should be performed in order to confirm the validity of the presented findings. Nevertheless, this study provides preliminary evidence that metabolic phenotyping of plasma has potential to become a diagnostic tool to differentiate between cancer types rather than a common cancer detection method. This, in turn, offers additional incentives to further investigate the diagnostic utility of ^{1}H-NMR-based metabolomics of plasma for many different cancer types in large-scale clinical trials.

References

1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74.
2. Beger RD. A review of applications of metabolomics in cancer. Metabolites. 2013;3(3):552-74.
3. Ma Y, Zhang P, Yang Y, Wang F, Qin H. Metabolomics in the fields of oncology: a review of recent research. Molecular biology reports. 2012;39(7):7505-11.
4. Griffin JL, Shockcor JP. Metabolic profiles of cancer cells. Nature reviews Cancer. 2004;4(7):551-61.
5. O'Connell TM. Recent advances in metabolomics in oncology. Bioanalysis. 2012;4(4):431-51.
6. Gu H, Gowda GA, Raftery D. Metabolic profiling: are we en route to better diagnostic tests for cancer? Future Oncol. 2012;8(10):1207-10.
7. Mamas M, Dunn WB, Neyses L, Goodacre R. The role of metabolites and metabolomics in clinically applicable biomarkers of disease. Archives of toxicology. 2011;85(1):5-17.
8. Duarte IF, Gil AM. Metabolic signatures of cancer unveiled by NMR spectroscopy of human biofluids. Progress in nuclear magnetic resonance spectroscopy. 2012;62:51-74.
9. Louis E, Bervoets L, Reekmans G, De Jonge E, Mesotten L, Thomeer M, et al. Phenotyping human blood plasma by ^{1}H-NMR: a robust protocol based on metabolite spiking and its evaluation in breast cancer. Metabolomics. 2015;11:225-36.
10. Louis E, Adriaensens P, Guedens W, Vanhove K, Vandeurzen K, Darquennes K, et al. Metabolic phenotyping of human blood plasma: a powerful tool to discriminate between cancer types? Annals of Oncology. 2015; DOI:10.1093/annonc/mdv499.
11. Singletary SE, Allred C, Ashley P, Bassett LW, Berry D, Bland KI, et al. Revision of the American Joint Committee on Cancer staging system for breast cancer. Journal of clinical oncology. 2002;20(17):3628-36.
12. Trygg JW, S. Orthogonal projections to latent structures (O-PLS). J Chemometr. 2002;16(3):119-28.

CHAPTER 8

General discussion, conclusions and future perspectives

Lung cancer constitutes a major public health problem, which accounts for approximately 20% of all cancer-related deaths worldwide (1). In order to achieve a significant decline in lung cancer mortality, this disease needs to be detected at the time that treatment is the most effective, namely before patients experience clinical symptoms (2, 3). In the past decade, low-dose computed tomography (LDCT) has renewed interest in lung cancer screening (4, 5). This technique allows to detect lung cancer with a high sensitivity and to reduce lung cancer mortality (6-9). However, the main challenge of LDCT screening is the high rate of false positive results accompanied by unnecessary and possibly harmful follow-up diagnostic tests, which put a massive financial burden on the health care system (10). Therefore, there is an increasing interest in improving the accuracy of current risk models by including lung cancer risk-related biomarkers in order to better select high-risk individuals eligible for LDCT screening, thereby decreasing the false positive rate and the corresponding financial burden (11, 12).

A promising biomarker for risk prediction should be low-cost, highly sensitive and specific and accessible in a non-invasive manner (13, 14). A blood-based diagnostic biomarker signifies an attractive possibility to complement current risk models for LDCT screening as blood samples can be obtained in a non-invasive way and with minimal risk for the patient (15-17). Cellular elements which are currently examined as blood-based biomarkers for facilitating the diagnosis of cancer include DNA and RNA, proteins and metabolites (18). Although it has been demonstrated that the concentration of serum proteins is linked to the presence of lung cancer, their sensitivity and specificity were insufficient to warrant their use as diagnostic tools in clinical practice (19). Because of the disappointing performance of single biomarkers, several studies focused on the development of biomarker panels in an attempt to attain superior levels of sensitivity and specificity (19, 20). Recent technological advances in genomics, epigenomics, transcriptomics and proteomics have facilitated the search for signatures that allow to differentiate between cancer patients and controls (21, 22). In particular, Nolen et al. (19) discovered a three-biomarker protein panel consisting of macrophage migration inhibition factor, prolactin and thrombospondin which classifies 74% of 62 lung cancer patients and 90% of 142 controls in the training cohort correctly. Moreover, the predictive accuracy of the protein panel was validated in an independent cohort of 30 lung cancer patients and 30 controls with

a sensitivity of 70% and a specificity of 90%. In addition, the research group of Boeri et al. (20) found a signature of 15 microRNA's present in blood which allows to detect lung cancer with a sensitivity of 80% and a specificity of 90%.

Besides genomics and proteomics, new –omics sciences, like transciptomics and metabolomics, are gaining increased attention in the search for cancer diagnostic biomarkers (23). Although all –omics sciences are considered complementary, metabolomics has some particular benefits: 1) small alterations in gene expression (the genome and transcriptome) and individual enzyme concentrations and activities (the proteome) have little impact on metabolic fluxes, but do have a significant effect on metabolite concentrations and 2) the metabolome is further down the line from gene to function and is the most closely related to the phenotype of the patient (24-26). Furthermore, metabolomics platforms are high-throughput and relatively low-cost compared to other -omics (27). Obviously, a biomarker which detects lung cancer in an early stage is more valuable than one which only identifies metastatic lung cancer (28). Metabolomics holds great promise for early cancer detection as metabolic changes occur before clinical symptoms manifest (29). ^{1}H-NMR spectroscopy, one of the main analytical platforms used in metabolomics studies, is a very reproducible tool which permits a fast and non-invasive identification and quantification of complex mixtures of metabolites, as in blood plasma, with minimal sample preparation and relatively low costs on a per sample basis (30, 31). Hence, ^{1}H-NMR-based metabolomics of plasma represents an attractive tool in the search for blood-based diagnostic biomarkers to complement current risk models for LDCT screening.

Optimization of a ^{1}H-NMR analysis protocol to quantitatively determine the plasma metabolites

In order to permit a correct implementation of ^{1}H-NMR-based metabolomics of plasma in this doctoral thesis, a ^{1}H-NMR analysis protocol needed to be optimized. The assignment of metabolite signals in ^{1}H-NMR spectra is often based on chemical shift values reported in literature (32). However, reported chemical shift values for metabolites are dependent of the biofluid under study and the applied experimental measuring conditions such as temperature and pH (33, 34). In order to accurately assign metabolite signals present in ^{1}H-NMR spectra of human blood plasma, it was spiked with 37 known metabolites in relevant concentrations and

analyzed on a 400 MHz NMR spectrometer (**Chapter 2**). Based on the metabolite spiking, the 400 MHz plasma spectrum could be rationally divided into 110 well-defined variable-sized integration regions (variables), representing the metabolic phenotype. The main advantage of using spiking experiments compared to fixed spectral binning is that splitting of peaks between adjacent bins is avoided, resulting in a more accurate data interpretation (35).

An additional concern regarding accurate data interpretation is the occurrence of spectral noise. Noisy integration regions (or variables) in the ^{1}H-NMR spectrum can be defined as signals showing unpredictable variation in intensity from moment to moment and under identical experimental conditions. In order to not complicate subsequent multivariate statistics, noisy variables, defined according to a specified threshold for the coefficient of variation, needed to be removed (36). Hence, we identified noisy variables in the plasma ^{1}H-NMR spectrum, classified those according predefined thresholds and studied the impact of their removal on multivariate group classification of a small group of 53 breast cancer patients and 52 controls. For our fixed hardware NMR set-up, a threshold of 15% for the coefficient of variation appeared to be the most optimal, i.e. removing variables with a coefficient of variation above 15% had a beneficial effect on multivariate classification of this small study population. However, in meantime, we observed that when a larger study population was recruited, exclusion of noisy variables had no impact anymore on the discriminative power of the classification model owing to the fact that all noisy variables had a low variable importance for the projection (VIP) value and therefore were not important for group discrimination. Moreover, the low VIP value of the noisy variables (<0.5) explains why they were not selected to explain the disturbed biochemical pathways.

A ^{1}H-NMR analysis protocol was optimized in order to accurately assign the signals present in ^{1}H-NMR spectra of human blood plasma to the corresponding metabolites. On the basis of metabolite spiking experiments, the 400 MHz ^{1}H-NMR spectrum of plasma could be rationally divided into 110 well-defined integration regions (variables), representing the metabolic phenotype.

Further sophistication of the ^{1}H-NMR analysis protocol was required to further improve the assignment of human plasma metabolites in crowded regions of the

400 MHz spectra with severe signal overlap, thereby facilitating the interpretation of the underlying, disease disturbed, biochemical pathways. Hereto, human blood plasma was spiked with the same metabolites and analyzed on a 900 MHz spectrometer (**Chapter 3**). The 900 MHz spectra have less signal overlap, an enhanced spectral resolution and enhanced signal-to noise ratio (S/N) compared to the 400 MHz spectra, enabling to assign the signals better and to define the integration regions more accurately. This results in a larger number of integration regions that represent a single metabolite (68% in the 900 MHz spectrum versus 24% in the 400 MHz spectrum), thereby facilitating the interpretation of the underlying, disease disturbed, biochemical pathways. On the basis of the spiking information, the 900 MHz plasma spectrum could be rationally divided into 105 well-defined variable-sized integration regions, representing the metabolic phenotype.

The ^{1}H-NMR analysis protocol was optimized further via spiking experiments on a high-field 900 MHz spectrometer in order to improve the assignment of plasma metabolite signals in crowded NMR regions with severe signal overlap. 900 MHz spectra have an enhanced spectral resolution and S/N, allowing to assign the signals better and to delineate the integration regions more accurately. This results in a larger number of integration regions (variables) that represent a single metabolite, thereby facilitating the interpretation of the underlying, disease disturbed, biochemical pathways.

In order to further evaluate advantages/disadvantages of high (900 MHz) versus medium (400 MHz) magnetic field strengths, the integration data collected at 400 and 900 MHz were applied to classify 69 lung cancer patients and 74 controls. It was demonstrated that the discriminative power of the models obtained on the basis of the 400 MHz and 900 MHz data is quasi comparable, i.e. a sensitivity of 94% and a specificity of 97% using the 400 MHz data versus 90% and 100% using the 900 MHz data (**Chapter 3**). In line with our findings, Bertram et al. have shown that although the discriminative power of the urinary metabolic phenotype significantly improved when augmenting the magnetic field strength from 250 to 500 MHz, the discriminative power remained quasi constant upon further increasing the magnetic field strength from 500 to 800 MHz (37). In conclusion,

high-field 900 MHz NMR spectra are indispensable at the start of a study to accurately assign the signals of the plasma metabolites in regions with several signal overlap in 400 MHz spectra and this specially if the unraveling of the disturbed metabolism is aimed for. Nevertheless, the increase in spectral resolution and S/N of the 900 MHz spectra does not outweigh the exponential rise in cost and housing facilities and medium-field spectrometers provide sufficient information to discriminate between groups. Hence, we suggest that medium-field 400-600 MHz spectrometers are most suitable for clinical metabolomics studies.

Although the enhanced spectral resolution and S/N of 900 MHz spectra is indispensable to unravel disease disturbed biochemical pathways, medium-field spectrometers (400-600 MHz) provide sufficient information to discriminate between groups, suggesting that they are most suitable for clinical metabolomics studies from a point of view of general cost.

Development of a standardized protocol regarding sample handling to enable implementation of ^{1}H-NMR-based plasma metabolomics in clinical practice

In order to allow the implementation of ^{1}H-NMR-based metabolomics in clinical practice, the impact of preanalytical variation in collection, processing, freezing and storage procedures on the plasma metabolome was investigated (**Chapter 4**). Preanalytical changes can have a major influence on the quality of plasma samples, hindering accurate interpretation of results and reducing the reliability of study findings (38, 39). From all preanalytical conditions studied, only increasing the processing delay between blood collection and centrifugation from 30 min to 3 h and 8 h at 4°C had a significant impact on the plasma metabolome. More specifically, the concentration of glucose and pyruvate decreased whereas the concentration of lactate increased. These alterations might be attributed to a continued anaerobic cell metabolism (i.e. glycolysis) in the collected blood sample owing to the ongoing contact of the metabolites with erythrocytes (40). In agreement with our findings, Bernini et al. reported decreased plasma levels of glucose and pyruvate as well as increased levels of lactate when whole blood is preserved during 4 h at 4°C (41). However, the inter-individual variation for these

variables was much larger than the preanalytical variation due to the increased processing delay. Accordingly, the impact of this preanalytical variation on multivariate cluster analysis will be minimal. Nevertheless, we recommend to keep the time gap between blood collection and centrifugation similar for all samples within a study. Note that in the studies reported in this thesis, blood samples were centrifuged within 8 h after collection (preservation at 4°C). As a result, since blood samples of both cancer patients and controls were processed in the same way (i.e. within 8 h storage at 4°C), the impact of the processing delay on multivariate group classification will be negligible.

The impact of relevant preanalytical conditions on the plasma metabolome was evaluated in order to explore the potential of implementing ^{1}H-NMR-based metabolomics in the clinic. From all preanalytical conditions studied, only increasing the processing delay from 30 min to 3 h and 8 h at 4°C had a significant impact on the plasma metabolome. However, the inter-individual variation was much larger than the preanalytical variation induced by increasing the processing delay. Nevertheless, we recommend to keep the time gap between blood collection and centrifugation comparable for all samples within a study.

Although efforts are currently made within the field of metabolomics to move towards defining standard operation procedures for preanalytical handling (41), complete standardization of preanalytical processing is yet not feasible between and within clinical settings. However, the introduction of a preanalytical sample code that traces preanalytical variations would already allow some useful sample management. To this end, the biobank community recently has developed the Standard PREanalytical Code (SPREC) in order to encode preanalytical conditions (42, 43). In an effort to validate SPREC for ^{1}H-NMR-based plasma metabolomics, we encoded the preanalytical conditions examined in our study according to SPREC and evaluated its potential to identify preanalytical conditions which significantly affect the plasma metabolome (**Chapter 4**). Our results demonstrate that SPREC enables to identify such plasma metabolome-affecting conditions. Consequently, it can be concluded that the implementation of SPREC could contribute to the validation of plasma ^{1}H-NMR-based metabolomics in clinical, biobank and multicenter research settings.

The recently developed SPREC allows to identify the plasma metabolome-affecting conditions. Therefore, its implementation is of utmost importance for a straightforward interpretation and validation of ^{1}H-NMR-based metabolomics in clinical, biobank and multicenter research settings.

Can ^{1}H-NMR-based plasma metabolomics be used as a tool to detect lung cancer?

In **Chapter 5**, the optimized ^{1}H-NMR analysis protocol (see **Chapter 2 and 3**) was implemented in a prospective study including lung cancer patients and controls in order to evaluate whether ^{1}H-NMR-based metabolomics of plasma allows to detect lung cancer. To the best of our knowledge, only the study of Rocha et al. utilized ^{1}H-NMR-based metabolomics to examine whether the metabolic phenotype of plasma permits to detect lung cancer, showing that lung cancer patients and controls can be differentiated with a sensitivity and specificity of nearly 90% (44, 45). Important drawbacks of their study were the relatively small sample size (i.e. 85 lung cancer patients and 78 controls) and the fact that the findings were not validated in an independent cohort. In contrast, our study shows in a large cohort consisting of 233 lung cancer patients and 226 controls that the multivariate statistical model (classifier) constructed by means of the plasma metabolic phenotype allows to classify 78% of the lung cancer patients and 92% of the controls correctly. Moreover, the predictive accuracy of the metabolic phenotype was confirmed in an independent validation cohort, demonstrating that 98 lung cancer patients can be discriminated from 89 controls with a sensitivity of 71% and a specificity of 81%.

In a next step, the underlying metabolic changes were identified in an attempt to improve the understanding of the disturbed metabolism of lung cancer. The variables responsible for group discrimination indicate that plasma of lung cancer patients is characterized by increased levels of glucose, glycerol, N-acetylated glycoproteins, β-hydroxybutyrate, leucine, lysine, tyrosine, threonine, glutamine, valine and aspartate and by decreased levels of alanine, lactate, sphingomyelin and phosphatidylcholine, citrate and other phospholipids. The decreased levels of citrate, of sphingomyelin and phosphatidylcholine and of other phospholipids point

to an increased lipogenesis and subsequent synthesis of membranes in lung cancer cells, in line with the metabolic transformation of cancer cells in order to support the creation of new daughter cells (46, 47). Surprisingly, the other metabolic alterations do not seem to reflect the Warburg effect in cancer cells (see **Chapter 1**) (48). Instead, the plasma metabolite profile seems to express the counteraction of the body to the Warburg effect. This seems to be a plausible theorem, since investigations of plasma attempt to elucidate the complex interaction between cancer cells and host. However, our findings and accompanying theorem should be independently confirmed by more dedicated biochemical studies. In contrast to our results, Rocha and co-workers reported decreased glucose levels and increased lactate levels in plasma of lung cancer patients, findings which are in line with the Warburg effect (44). A highly possible explanation for these apparently discrepant results might be that most patients in their study had early (stage I and stage II) lung cancer (81%) and no metastatic (stage IV) lung cancer, while our study population included lung cancer patients with disease stages ranging from stage I to stage IV. This most probably demonstrates that the counteraction of the body to the Warburg effect is more prominent at more advanced stages of lung cancer.

^{1}H-NMR-based metabolomics of plasma allows to classify 78% of the 233 lung cancer patients and 92% of the 226 controls correctly. Moreover, the predictive accuracy of the metabolic phenotype-based classifier was validated in an independent cohort, showing that 98 lung cancer patients and 89 controls can be discriminated with a sensitivity of 71% and a specificity of 81%. The underlying metabolic changes point to an increased lipogenesis and subsequent synthesis of phospholipid membranes in cancer cells, in line with the metabolic transformation of cancer cells. Remarkably, the other metabolic changes seem to reflect the counteraction of the body to the Warburg effect.

Does the plasma metabolic phenotype allow to detect lung cancer in an early stage?

Besides investigating whether ^{1}H-NMR-based metabolomics of plasma allows to detect lung cancer, it is also indispensable to evaluate whether it permits to detect

lung cancer before patients experience clinical symptoms and when curative treatment is still possible, thereby increasing life expectancy and quality of life (3, 49). Although the number of early stage lung cancer patients in our cohort was still limited, preliminary data showed that the plasma metabolic phenotype allows to classify 74% of the 76 early stage (stage I) lung cancer patients and 78% of the 76 controls correctly (**Chapter 5**). These findings seem to indicate that the metabolic changes present in the initial phase of cancer development can already be detected by ^{1}H-NMR-based metabolomics. Although these results look promising, the number of early stage lung cancer patients needs to be expanded to further investigate the potential of ^{1}H-NMR-based metabolomics to detect lung cancer in an early stage.

^{1}H-NMR-based metabolomics of plasma allows to classify 74% of the 76 early stage (stage I) lung cancer patients and 78% of the 76 controls correctly. These preliminary results look very promising. They indicate that the metabolic alterations present in the initial phase of cancer development can be detected. However, the number of early stage lung cancer patients need to be expanded to confirm our findings in a larger cohort.

Does ^{1}H-NMR-based metabolomics of plasma permit to discriminate between breast and lung cancer?

While the prospective study, described in **Chapter 5**, has established the potential of ^{1}H-NMR-based metabolomics of plasma to differentiate between lung cancer patients and controls, the question arose whether the plasma metabolic phenotype reflects a common cancer marker or whether different cancer types are characterized by their own, specific metabolite profile. For that reason, the plasma metabolic phenotype of lung cancer was compared with that of breast cancer, the most common cancers in men and in women worldwide, respectively (1) (**Chapter 6**). Although several studies, including ours, already have explored the disturbed metabolism of lung or breast cancer by means of ^{1}H-NMR-based metabolomics of body fluids (44, 50-54), to the best of our knowledge, we are the first to apply this methodology to examine whether lung and breast cancer are characterized by a specific metabolite profile and so can be differentiated from

each other. To this end, a homogenous study population of cancer patients with the histological subtype adenocarcinoma was chosen and, as all breast cancer patients were females, only female patients were included in the training cohort in order to exclude gender as a confounding factor in the discrimination between lung and breast cancer. As a result, it was shown that the plasma metabolic phenotype allows to classify 93% of the 54 female lung cancer patients and 99% of the 80 female breast cancer patients correctly. Furthermore, the validity of the metabolic phenotype-based statistical classifier was affirmed in an independent cohort, showing that 60 female breast cancer patients and 81 male lung cancer patients can be differentiated with a sensitivity of 89% (signifying that 89% of the lung cancer patients were correctly classified) and a specificity of 82% (meaning that 82% of the breast cancer patients were correctly classified). This indicates that the plasma metabolite profile reflects more than a common cancer marker, i.e. that breast and lung cancer are characterized by their own, specific metabolite profile.

Subsequently, the metabolic changes responsible for group discrimination were identified in an attempt to explain the differences in the disturbed biochemical pathways between both cancer types. The corresponding metabolites of which the concentration is increased for lung cancer encompass glucose, tyrosine, aspartate, threonine, leucine, valine, histidine, methionine, N-acetylated glycoproteins, β-hydroxybutyrate and phospholipids with long, saturated fatty acid chains. Likewise, the metabolites of which the concentration is decreased for lung cancer comprise alanine, glutamine, α-ketoglutarate, creatinine, citrate, sphingomyelin and phosphatidylcholine, lactate and phospholipids with short, unsaturated fatty acid chains. The decreased concentration of glutamine, α-ketoglutarate, citrate and of sphingomyelin and phosphatidylcholine indicate that the lipogenesis and the following membrane synthesis is more pronounced in lung cancer cells as compared to breast cancer cells, in line with the fact that lung tumors are metabolically more active than breast tumors. Since long, saturated fatty acid chains have very strong Van der Waals interactions, the increased plasma levels of phospholipids with long, saturated fatty chains and the decreased levels of phospholipids with short, unsaturated fatty acid chains indicate that the cell membrane of lung tumors is more rigid and less sensitive to lipid peroxidation (46). In **Chapter 5**, describing the study of lung cancer patients and controls, it

was stated that the plasma metabolite profile reflects the counteraction of the body to the Warburg effect rather than the Warburg effect, as present in cancer cells, itself. Similarly, all metabolic alterations in this study seem to point to a more pronounced counteraction of the body to the Warburg effect for lung as compared to breast cancer. Nevertheless, our findings should be independently affirmed by more dedicated biochemical studies.

^{1}H-NMR-based metabolomics of plasma allows to classify 93% of the 54 female lung cancer patients and 99% of the 80 female breast cancer patients correctly. Additionally, the validity of the metabolic phenotype-derived classifier was confirmed in an independent cohort, showing that 60 female breast cancer patients and 81 male lung cancer patients could be differentiated with a sensitivity of 89% and a specificity of 82%. The underlying metabolic changes point to an increased lipogenesis and following membrane synthesis in lung cancer cells, in line with the higher metabolic activity of lung tumors as compared to breast tumors. The other metabolic alterations seem to reflect a more pronounced counteraction of the body to the Warburg effect in lung as compared to breast cancer.

Does the plasma metabolome represent a specific diagnostic tool or rather a common cancer biomarker?

Although the findings of **Chapter 6** already demonstrate the potential of ^{1}H-NMR-based plasma metabolomics to discriminate between lung and breast cancer, additional research was accomplished to further examine whether the three most frequently occurring cancer types worldwide (1), i.e. lung, breast and colorectal cancer, can be differentiated (**Chapter 7**). Preliminary results are showing that the plasma metabolic phenotype allows to classify 78% of the 37 colorectal cancer patients, 95% of the 37 breast cancer patients and 84% of the 37 lung cancer patients correctly. Although the number of subjects included is still somewhat limited, the study provides preliminary indications that ^{1}H-NMR-based metabolomics of plasma has potential to become a tool to diagnose specific cancer types rather than to detect only a common cancer marker. Future studies should focus on increasing the number of colorectal cancer patients in order to gain more

statistically reliable insights regarding whether ^{1}H-NMR-based metabolomics of plasma enables to differentiate between lung, breast and colorectal cancer and to unravel the differentiating metabolic perturbations between lung, breast and colorectal cancer. Moreover, an independent validation study should be performed in order to ensure the validity of our preliminary study findings.

^{1}H-NMR-based metabolomics of plasma allows to classify 78% of the 37 colorectal cancer patients, 95% of the 37 breast cancer patients and 84% of the 37 lung cancer patients correctly. These preliminary results indicate that ^{1}H-NMR-based metabolomics of plasma has potential to become a tool to diagnose specific cancer types rather than a common cancer marker. Future studies should focus on increasing the number of colorectal cancer patients in order to strengthen the reliability of these insights, and to unravel the differentiating metabolic perturbations between lung, breast and colorectal cancer as well as to validate our study findings in an independent patient cohort.

In this doctoral thesis, it is demonstrated that the proposed methodology of ^{1}H-NMR-based metabolomics enables to detect lung cancer and to differentiate between lung and breast cancer. Moreover, preliminary evidence is provided regarding the ability of the methodology to discriminate between lung, breast and colorectal cancer.

Implications for clinical practice

The main challenge of low-dose computed tomography (LDCT), which is currently the most studied tool for lung cancer screening, is the high rate of false positive results. This leads to excessive and possibly harmful follow-up investigations, which put an enormous financial burden on our health care system (10, 55). Hence, complementary tools, which reinforce current risk models and thereby improve lung cancer diagnosis, are urgently needed. ^{1}H-NMR-based metabolomics represents an appealing methodology to complement risk models for LDCT screening by addition of complementary data and thereby improve lung cancer diagnosis. It is a highly reproducible tool which enables a fast and non-invasive identification and quantification of complex mixtures of metabolites, as in blood plasma, with minimal sample preparation and relatively low costs on a per sample basis. We have chosen to examine the metabolic composition of blood plasma since a blood-based screening test is relatively non-invasive and inexpensive. Furthermore, the collection of blood samples takes little time from the patient and is already well-accepted in routine clinical practice, making it ideally suitable for attaining high participation rates in the general population (29).

Our research group has joined forces with the Interuniversity Institute for Biostatistics and statistical Bioinformatics in order to investigate whether the addition of parameters which reflect the plasma metabolic phenotype to current risk models, which only take epidemiological and clinical data into account, has potential to improve the identification of high-risk individuals eligible for LDCT screening. Preliminary results from this collaboration have revealed that the addition of NMR metabolic phenotype data (i.e. 102 normalized NMR integration values) to a risk model, which only contains the clinical parameters age, body mass index, presence of chronic obstructive pulmonary disease (COPD), number of smoking pack years, smoking habits and intake of anti-arrhytmetic and anti-coagulants medication, improves its misclassification error from 24% to 19%. In this analysis only 102 out of the 110 normalized NMR integration values were used, since 8 integration values had more than 10% missing values. Furthermore, the variable importance for the projection plot of this risk model shows that 26 out of 30 of the most discriminating variables constitute normalized NMR integration values. Besides the number of smoking pack years, the presence of

COPD and smoking habits, the concentration of threonine (VAR49 and 50), glycerol (VAR45 and 46) and valine (VAR48) contribute the most to the discriminative power of the risk model (**Figure 8.1**).

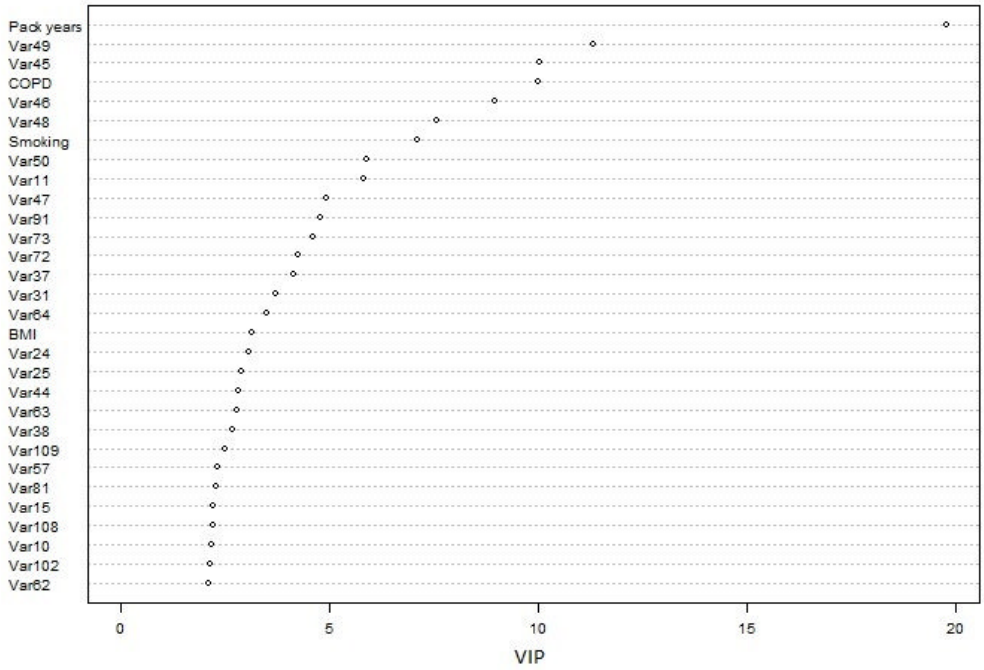


Figure 8.1. Variable importance for the projection plot demonstrating the contribution of clinical, epidemiological and metabolic phenotype data to the discriminative power of the risk model. BMI: body mass index; COPD: chronic obstructive pulmonary disease; VIP: variable importance for the projection.

These findings are comparable to those of earlier studies where the addition of genetic risk markers/DNA repair capacity and mutagen sensitivity data improved the performance of certain risk models which only contain epidemiological and clinical data (11, 12). Our findings indicate that the plasma metabolic phenotype has potential to improve the identification of high-risk individuals eligible for LDCT screening, thereby reducing the false positive rate of LDCT screening and the corresponding financial burden. Thus, the plasma metabolic phenotype can be used as an adjunct to LDCT screening, thereby improving its cost-effectiveness. Note that our findings were achieved in a cohort consisting of patients with a known diagnosis of lung cancer and controls and that they need to be confirmed in asymptomatic individuals who are eligible for LDCT screening.

Besides diagnosis and screening, ^{1}H-NMR-based metabolomics of plasma can be applied in other segments of cancer patient care. In particular, it has great potential in the emerging area of personalized medicine, i.e. the development of an individually tailored approach, by stratifying patients based on their plasma metabolic phenotype (29, 56). More specifically, the plasma metabolic phenotype enables to identify individuals that will most likely respond to a particular drug and to recognize patients that are prone to develop resistance or to suffer from drug toxicities (pharmacometabolomics) (17). Selecting the right drug for the right patient on the basis of the plasma metabolic phenotype should result in a more effective treatment (reduced occurrence and intensity of adverse events), reduced pain and a decreased financial burden for the health care system (29, 57). Furthermore, ^{1}H-NMR-based plasma metabolomics might assist in the identification of biochemical pathways that could serve as new drug targets (57, 58). Moreover, the plasma metabolic phenotype can be measured before and periodically during treatment to evaluate whether the tumor is responding to treatment. Likewise, the plasma metabolic phenotype can be measured after the end of treatment to check whether the cancer has returned (29).

Although results of current studies are encouraging, some obstacles have to be overcome before ^{1}H-NMR-based metabolomics of plasma can gain widespread clinical use in the field of oncology, i.e. educating medical doctors and medical technologists in the acquisition, analysis and interpretation of ^{1}H-NMR-based metabolomics data, and manufacture user-friendly NMR instruments with all the features required for optimal use (59). Moreover, our study findings need to be independently validated in large-scale prospective screening studies with asymptomatic, high-risk individuals who are eligible for LDCT screening before ^{1}H-NMR-based plasma metabolomics can be applied in routine clinical practice (18, 60).

Avenues for future research

The findings obtained during this doctoral project reveal that ^{1}H-NMR-based metabolomics of plasma allows to i) detect lung cancer and ii) discriminate between lung and breast cancer. Nonetheless, large-scale validation studies should be conducted to confirm these findings and to further explore the potential of ^{1}H-NMR-based plasma metabolomics as a complementary screening tool for lung cancer.

Due to the heterogeneity of lung tumors, the number of lung cancer patients should be increased in order to investigate whether the plasma metabolic phenotype can further discriminate between histological subtypes and disease stages. Future studies also need to address whether the plasma metabolic phenotype enables to differentiate between lung cancer patients and individuals with benign pulmonary lesions such as tuberculosis and sarcoidosis. Moreover, although our findings have already shown that ^{1}H-NMR-based metabolomics of plasma permits to discriminate between lung and breast cancer, additional research is required in order to further examine whether the plasma metabolic phenotype represents a specific diagnostic tool rather than a common cancer biomarker. Preliminary evidence has shown that ^{1}H-NMR-based metabolomics of plasma enables to discriminate between lung, breast and colorectal cancer. However, the number of colorectal cancer patients should be expanded in order to gain statistically reliable insights and to enable validation of our study findings in an independent patient cohort. The final step in evaluating the potential of ^{1}H-NMR-based metabolomics of plasma as a complementary screening tool for lung cancer will be to examine whether it allows to identify lung cancer patients among asymptomatic individuals who are eligible for LDCT screening. To this end, large-scale prospective screening studies with lung cancer mortality as a primary endpoint have to be performed in the target population.

Future studies might combine different analytical platforms (i.e. ^{1}H-NMR spectroscopy and mass spectrometry) in order to expand the number of metabolites which can be identified in plasma (61, 62). Last but not least, study findings regarding the disturbed metabolism should be confirmed by alternative approaches, e.g. enzyme catalytic studies or NMR studies in which ^{13}C-labeled glucose is used.

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. International journal of cancer. 2015;136(5):E359-86.
2. Manser R, Lethaby A, Irving LB, Stone C, Byrnes G, Abramson MJ, et al. Screening for lung cancer. Cochrane Database of Systematic Reviews 2004. 2013(1):CD001991.
3. Wood DE, Eapen GA, Ettinger DS, Hou L, Jackman D, Kazerooni E, et al. Lung cancer screening. Journal of the National Comprehensive Cancer Network. 2012;10(2):240-65.
4. Aberle DR, Abtin F, Brown K. Computed tomography screening for lung cancer: has it finally arrived? Implications of the national lung screening trial. Journal of clinical oncology. 2013;31(8):1002-8.
5. Bach PB, Mirkin JN, Oliver TK, Azzoli CG, Berry DA, Brawley OW, et al. Benefits and harms of CT screening for lung cancer: a systematic review. Jama. 2012;307(22):2418-29.
6. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. The New England journal of medicine. 2011;365(5):395-409.
7. Horeweg N, Scholten ET, de Jong PA, van der Aalst CM, Weenink C, Lammers JW, et al. Detection of lung cancer through low-dose CT screening (NELSON): a prespecified analysis of screening test performance and interval cancers. The Lancet Oncology. 2014;15(12):1342-50.
8. Humphrey LL, Deffebach M, Pappas M, Baumann C, Artis K, Mitchell JP, et al. Screening for lung cancer with low-dose computed tomography: a systematic review to update the US Preventive services task force recommendation. Annals of internal medicine. 2013;159(6):411-20.
9. van Klaveren RJ, Habbema JDF, Pedersen JH, de Koning HJ, Oudkerk M, Hoogsteden HC. Lung cancer screening by low-dose spiral computed tomography. The European respiratory journal. 2001;18(5):857-66.
10. Tammemagi MC, Lam S. Screening for lung cancer using low dose computed tomography. BMJ. 2014;348:g2253.
11. Raji OY, Agbaje OF, Duffy SW, Cassidy A, Field JK. Incorporation of a genetic factor into an epidemiologic model for prediction of individual risk of lung cancer: the Liverpool Lung Project. Cancer Prev Res. 2010;3(5):664-9.
12. Spitz MR, Etzel CJ, Dong Q, Amos CI, Wei Q, Wu X, et al. An expanded risk prediction model for lung cancer. Cancer Prev Res. 2008;1(4):250-4.
13. Broadhurst DI, Kell DB. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics. 2006;2(4):171-96.
14. Knottnerus JA, van Weel C, Muris JW. Evaluation of diagnostic procedures. BMJ. 2002;324(7335):477-80.
15. Tsay JC, DeCotiis C, Greenberg AK, Rom WN. Current readings: blood-based biomarkers for lung cancer. Seminars in thoracic and cardiovascular surgery. 2013;25(4):328-34.
16. Mamas M, Dunn WB, Neyses L, Goodacre R. The role of metabolites and metabolomics in clinically applicable biomarkers of disease. Archives of toxicology. 2011;85(1):5-17.
17. Vermeersch KA, Styczynski MP. Applications of metabolomics in cancer research. Journal of carcinogenesis. 2013;12:9.
18. Hassanein M, Callison JC, Callaway-Lane C, Aldrich MC, Grogan EL, Massion PP. The state of molecular biomarkers for the early detection of lung cancer. Cancer Prev Res. 2012;5(8):992-1006.
19. Nolen BM, Langmead CJ, Choi S, Lomakin A, Marrangoni A, Bigbee WL, et al. Serum biomarker profiles as diagnostic tools in lung cancer. Cancer biomarkers : section A of Disease markers. 2011;10(1):3-12.
20. Boeri M, Verri C, Conte D, Roz L, Modena P, Facchinetti F, et al. MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(9):3713-8.

21. Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. Nature reviews Cancer. 2004;4(4):309-14.
22. Sawyers CL. The cancer biomarker problem. Nature. 2008;452(7187):548-52.
23. Ransohoff DF. Bias as a threat to the validity of cancer molecular-marker research. Nature reviews Cancer. 2005;5(2):142-9.
24. Goodacre R. Metabolomics of a superorganism. The Journal of nutrition. 2007;137(1 Suppl):259S-66S.
25. Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB. Metabolomics by numbers: acquiring and understanding global metabolite data. Trends in biotechnology. 2004;22(5):245-52.
26. Hollywood K, Brison DR, Goodacre R. Metabolomics: current technologies and future trends. Proteomics. 2006;6(17):4716-23.
27. Griffin JL, Shockcor JP. Metabolic profiles of cancer cells. Nature reviews Cancer. 2004;4(7):551-61.
28. Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, et al. Phases of biomarker development for early detection of cancer. Journal of the National Cancer Institute. 2001;93(14):1054-61.
29. Palmnas MS, Vogel HJ. The future of NMR metabolomics in cancer therapy: towards personalizing treatment and developing targeted drugs? Metabolites. 2013;3(2):373-96.
30. Emwas AH, Salek RM, Griffin JL, Merzaban J. NMR-based metabolomics in human disease diagnosis: applications, limitations, and recommendations. Metabolomics. 2013;9:1048-72.
31. Lindon JC, Nicholson JK. Spectroscopic and statistical techniques for information recovery in metabonomics and metabolomics. Annu Rev Anal Chem. 2008;1:45-69.
32. Salek RM, Maguire ML, Bentley E, Rubtsov DV, Hough T, Cheeseman M, et al. A metabolomic comparison of urinary changes in type 2 diabetes in mouse, rat, and human. Physiological genomics. 2007;29(2):99-108.
33. Kohl SM, Klein MS, Hochrein J, Oefner PJ, Spang R, Gronwald W. State-of-the art data normalization methods improve NMR-based metabolomic analysis. Metabolomics. 2012;8(Suppl 1):146-60.
34. Staab JM, O'Connell TM, Gomez SM. Enhancing metabolomic data analysis with Progressive Consensus Alignment of NMR Spectra. BMC bioinformatics. 2010;11:123.
35. Emwas AH. The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. Methods Mol Biol. 2015;1277:161-93.
36. Halouska S, Powers R. Negative impact of noise on the principal component analysis of NMR data. Journal of magnetic resonance. 2006;178(1):88-95.
37. Bertram HC, Malmendal A, Petersen BO, Madsen JC, Pedersen H, Nielsen NC, et al. Effect of magnetic field strength on NMR-based metabonomic human urine data. Comparative study of 250, 400, 500, and 800 MHz. Analytical chemistry. 2007;79(18):7110-5.
38. Kamlage B, Maldonado SG, Bethan B, Peter E, Schmitz O, Liebenberg V, et al. Quality markers addressing preanalytical variations of blood and plasma processing identified by broad and targeted metabolite profiling. Clinical chemistry. 2014;60(2):399-412.
39. Yin P, Peter A, Franken H, Zhao X, Neukamm SS, Rosenbaum L, et al. Preanalytical aspects and sample quality assessment in metabolomics studies of human blood. Clinical chemistry. 2013;59(5):833-45.
40. Baynes JW, Dominiczak MH. Medical Biochemistry. Amsterdam: Elsevier; 2010.
41. Bernini P, Bertini I, Luchinat C, Nincheri P, Staderini S, Turano P. Standard operating procedures for pre-analytical handling of blood and urine for metabolomic studies and biobanks. Journal of biomolecular NMR. 2011;49(3-4):231-43.
42. Betsou F, Lehmann S, Ashton G, Barnes M, Benson EE, Coppola D, et al. Standard preanalytical coding for biospecimens: defining the sample PREanalytical code. Cancer epidemiology, biomarkers & prevention. 2010;19(4):1004-11.

43. Lehmann S, Guadagni F, Moore H, Ashton G, Barnes M, Benson E, et al. Standard preanalytical coding for biospecimens: review and implementation of the Sample PREanalytical Code (SPREC). Biopreservation and biobanking. 2012;10(4):366-74.
44. Rocha CM, Carrola J, Barros AS, Gil AM, Goodfellow BJ, Carreira IM, et al. Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of blood plasma. Journal of proteome research. 2011;10(9):4314-24.
45. Duarte IF, Rocha CM, Gil AM. Metabolic profiling of biofluids: potential in lung cancer screening and diagnosis. Expert review of molecular diagnostics. 2013;13(7):737-48.
46. Rysman E, Brusselmans K, Scheys K, Timmermans L, Derua R, Munck S, et al. De novo lipogenesis protects cancer cells from free radicals and chemotherapeutics by promoting membrane lipid saturation. Cancer research. 2010;70(20):8117-26.
47. Lunt SY, Vander Heiden MG. Aerobic glycolysis: meeting the metabolic requirements of cell proliferation. Annual review of cell and developmental biology. 2011;27:441-64.
48. Upadhyay M, Samal J, Kandpal M, Singh OV, Vivekanandan P. The Warburg effect: insights from the past decade. Pharmacology & therapeutics. 2013;137(3):318-30.
49. Obuchowski NA, Graham RJ, Baker ME, Powell KA. Ten criteria for effective screening: their application to multislice CT screening for pulmonary and colorectal cancers. American journal of roentgenology. 2001;176(6):1357-62.
50. Carrola J, Rocha CM, Barros AS, Gil AM, Goodfellow BJ, Carreira IM, et al. Metabolic signatures of lung cancer in biofluids: NMR-based metabonomics of urine. Journal of proteome research. 2011;10(1):221-30.
51. Deja S, Porebska I, Kowal A, Zabek A, Barg W, Pawelczyk K, et al. Metabolomics provide new insights on lung cancer staging and discrimination from chronic obstructive pulmonary disease. Journal of pharmaceutical and biomedical analysis. 2014;100:369-80.
52. Gu H, Pan Z, Xi B, Asiago V, Musselman B, Raftery D. Principal component directed partial least squares analysis for combining nuclear magnetic resonance and mass spectrometry data in metabolomics: application to the detection of breast cancer. Analytica chimica acta. 2011;686(1-2):57-63.
53. Louis E, Bervoets L, Reekmans G, De Jonge E, Mesotten L, Thomeer M, et al. Phenotyping human blood plasma by ^{1}H-NMR: a robust protocol based on metabolite spiking and its evaluation in breast cancer. Metabolomics. 2015;11:225-36.
54. Oakman C, Tenori L, Biganzoli L, Santarpia L, Cappadona S, Luchinat C, et al. Uncovering the metabolomic fingerprint of breast cancer. The international journal of biochemistry & cell biology. 2011;43(7):1010-20.
55. Black WC, Gareen IF, Soneji SS, Sicks JD, Keeler EB, Aberle DR, et al. Cost-effectiveness of CT screening in the National Lung Screening Trial. The New England journal of medicine. 2014;371(19):1793-802.
56. Aboud OA, Weiss RH. New opportunities from the cancer metabolome. Clinical chemistry. 2013;59(1):138-46.
57. Beger RD. A review of applications of metabolomics in cancer. Metabolites. 2013;3(3):552-74.
58. Patel S, Ahmed S. Emerging field of metabolomics: big promise for cancer biomarker identification and drug discovery. Journal of pharmaceutical and biomedical analysis. 2015;107:63-74.
59. Bezabeh T, Ijare OB, Nikulin AE, Somorjai RL, Smith IC. MRS-based Metabolomics in Cancer Research. Magnetic resonance insights. 2014;7:1-14.
60. Cuperlovic-Culf M. NMR metabolomics in cancer research. 1 ed. Cambridge: Woodhead Publishing; 2013.
61. Pan Z, Raftery D. Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. Analytical and bioanalytical chemistry. 2007;387(2):525-7.
62. Nagrath D, Caneba C, Karedath T, Bellance N. Metabolomics for mitochondrial and cancer studies. Biochimica et biophysica acta. 2011;1807(6):650-63.

Lung cancer is a major public health concern, accounting for approximately 20% of all cancer-related deaths worldwide. The main challenge of low-dose computed tomography (LDCT) screening, the tool which is currently studied for lung cancer screening, is the high rate of false positive results. Hence, there is an increasing interest in improving the accuracy of current risk models by incorporating lung cancer risk-related biomarkers in order to better select high-risk individuals eligible for LDCT screening. Metabolomics holds great potential for cancer diagnosis as the metabolome reflects alterations in the observed metabolic phenotype. ^{1}H-NMR spectroscopy enables a fast and non-invasive identification of complex mixtures of metabolites with minimal sample preparation and relatively low costs. Therefore, ^{1}H-NMR-based metabolomics represents an appealing option in the search for blood-based diagnostic biomarkers in order to complement current risk models for LDCT screening, thereby decreasing the false positive rate and the corresponding financial burden.

First, a ^{1}H-NMR analysis protocol was optimized via metabolite spiking experiments on a medium-field 400 and a high-field 900 MHz spectrometer. Next, a standardized protocol regarding sample handling, which enables the implementation of ^{1}H-NMR-based metabolomics in clinical practice, was developed. Subsequently, it was evaluated whether ^{1}H-NMR-based metabolomics of plasma can be used as a tool to detect lung cancer. It was shown that the plasma metabolic phenotype allows to classify 78% of 233 lung cancer patients and 92% of 226 controls correctly. Moreover, the predictive accuracy of the plasma metabolic phenotype-based classifier was validated in an independent cohort, showing that 98 lung cancer patients and 89 controls can be differentiated with a sensitivity of 71% and a specificity of 81%. The underlying metabolic alterations were identified in an attempt to improve the understanding of the disturbed metabolism of lung cancer. The decreased levels of phospholipids indicate an increased lipogenesis and following membrane synthesis in lung cancer cells, in agreement with the metabolic transformation of cancer cells in order to support the creation of new daughter cells. Surprisingly, the other metabolic changes do not seem to reflect the Warburg effect in cancer cells, but rather the counteraction of the body to the Warburg effect. Although this appears to be a plausible theorem, it should be independently confirmed by more dedicated biochemical studies.

In a next step, it was examined whether the plasma metabolic phenotype enables to differentiate between lung and breast cancer. It was demonstrated that the plasma metabolic phenotype enables to classify 93% of 54 female lung cancer patients and 99% of 80 female breast cancer patients correctly. Additionally, the validity of the plasma metabolic phenotype-derived statistical classifier was confirmed in an independent cohort, demonstrating that 60 female breast cancer patients and 81 male lung cancer patients can be discriminated with a sensitivity of 89% (89% of the lung cancer patients were correctly classified) and a specificity of 82% (82% of the breast cancer patients were correctly classified), indicating that breast and lung cancer are characterized by their own, specific metabolite profile. The metabolic alterations responsible for group discrimination were identified in an attempt to explain the differences in the disturbed biochemical pathways between both cancer types. The decreased concentration of sphingomyelin and phosphatidylcholine indicate that the lipogenesis and subsequent membrane synthesis is more pronounced in lung cancer cells as compared to breast cancer cells. These findings are in line with the fact that lung tumors display a higher uptake of the radioactive fluorodeoxyglucose tracer as measured by positron emission tomography, indicating that lung tumors are metabolically more active than breast tumors. Since long, saturated fatty acid chains have very strong Van der Waals interactions, the increased plasma levels of phospholipids with long, saturated fatty acid chains and the decreased plasma levels of phospholipids with short, unsaturated fatty acid chains indicate that the cell membrane of lung tumors is more rigid (less fluid) and less sensitive to lipid peroxidation than that of breast tumors.

Afterwards, it was examined whether the plasma metabolome permits to differentiate between lung, breast and colorectal cancer. Preliminary results have shown that the plasma metabolic phenotype-based classifier allows to classify 78% of 37 colorectal cancer patients, 95% of 37 breast cancer patients and 84% of 37 lung cancer patients correctly. Although the number of patients included in this study is still limited, this study provides preliminary evidence that ^{1}H-NMR-based metabolomics of plasma has potential to become a complementary tool to diagnose specific cancer types rather than a common cancer marker

Finally, preliminary results of a collaboration between our research group and the group of Biostatistics have shown that the incorporation of plasma metabolic phenotype data in risk models, which only take epidemiological and clinical data into account, reduces the misclassification error of the risk model from 24% to 19%. These findings suggest that the plasma metabolic phenotype has potential to enhance the identification of high-risk individuals eligible for LDCT screening, thereby lowering the false positive rate of LDCT screening and the corresponding financial burden. However, before ^{1}H-NMR-based plasma metabolomics can be applied in routine clinical practice, the study findings need to be independently validated in large-scale studies with asymptomatic, high-risk individuals who are eligible for LDCT screening.

Longkanker is een belangrijke aandoening die verantwoordelijk is voor ongeveer 20% van alle sterfgevallen ten gevolge van kanker wereldwijd. Het voornaamste nadeel van lage dosis computertomografie (LDCT), de methode die op dit moment bestudeerd wordt voor longkankerscreening, is het hoge aantal vals positieve resultaten. Daarom is er steeds meer interesse om de nauwkeurigheid van huidige risicomodellen te verbeteren door biomerkers toe te voegen die verband houden met het risico op longkanker. Op deze manier kunnen individuen die een verhoogd risico op longkanker hebben en aldus in aanmerking komen voor LDCT screening beter geselecteerd worden. Metabolomics heeft veel potentieel om kanker te diagnosticeren omdat het metaboloom een weerspiegeling is van de wijzigingen die gebeuren in het geobserveerde metabole fenotype. Met behulp van ^{1}H-NMR spectroscopie kunnen complexe mengsels van metabolieten op een snelle en niet-invasieve wijze geïdentificeerd worden zonder een uitgebreide staalvoorbereiding en hoge kosten. Daarom is ^{1}H-NMR metabolomics een aantrekkelijke keuze om op zoek te gaan naar diagnostische biomerkers in het bloed die nodig zijn om huidige risicomodellen voor LDCT screening te versterken. Op deze manier zouden het aantal vals positieve resultaten en de overeenkomstige kosten voor de maatschappij gereduceerd kunnen worden.

Eerst werd een protocol geoptimaliseerd om bloedplasma te analyseren met ^{1}H-NMR spectroscopie. Hiervoor werden spike experimenten uitgevoerd met gekende metabolieten op een spectrometer met een gemiddelde veldsterkte (400MHz) enerzijds en een spectrometer met een hoge veldsterkte anderzijds (900 MHz). Daarnaast werd een gestandaardiseerd protocol met betrekking tot staalverwerking ontwikkeld om de implementatie van ^{1}H-NMR metabolomics in de klinische praktijk toe te laten. Vervolgens werd nagegaan of ^{1}H-NMR plasma metabolomics gebruikt kan worden om longkanker op te sporen. Er werd aangetoond dat het metabole plasma fenotype toelaat om 78% van 233 longkankerpatiënten en 92% van 226 controles correct te classificeren. Bovendien werd de predictieve nauwkeurigheid van het metabole plasma fenotype gevalideerd in een onafhankelijk cohort. Er werd aangetoond dat 98 longkankerpatiënten en 89 controles onderscheiden konden worden met een sensitiviteit van 71% en een specificiteit van 81%. De onderliggende metabole wijzigingen werden geïdentificeerd om het verstoorde longkankermetabolisme beter te kunnen begrijpen. De verlaagde fosfolipidenconcentratie wijst op een

verhoogde lipogenese en daaropvolgende membraansynthese in longtumoren. Deze bevindingen kunnen gelinkt worden aan de metabole transformatie die plaatsvindt in tumorcellen om de aanmaak van nieuwe dochtercellen te ondersteunen. Tegen de verwachting in lijken de andere metabole wijzigingen niet overeen te komen met het Warburg effect dat plaatsvindt in tumorcellen. In de plaats daarvan lijken ze de tegenreactie van het lichaam op het Warburg effect te weerspiegelen. Hoewel deze verklaring aannemelijk lijkt, moet ze onafhankelijk bevestigd worden door meer gerichte biochemische studies.

In een volgende stap werd onderzocht of het metabole plasma fenotype toelaat om een onderscheid te maken tussen long- en borstkanker. Hierbij werd aangetoond dat het metabole plasma fenotype toelaat om 93% van 54 vrouwelijke longkankerpatiënten en 99% van 80 vrouwelijke borstkankerpatiënten correct te classificeren. Daarnaast werd de validiteit van het metabole plasma fenotype bevestigd in een onafhankelijk cohort. Er werd aangetoond dat er een onderscheid gemaakt kan worden tussen 60 vrouwelijke borstkankerpatiënten en 81 mannelijke longkankerpatiënten met een sensitiviteit van 89% (89% van de longkankerpatiënten worden correct geclassificeerd) en een specificiteit van 82% (82% van de borstkankerpatiënten worden correct geclassificeerd). Deze resultaten tonen aan dat borst- en longkanker gekenmerkt worden door hun eigen, specifiek metaboliet profiel. Vervolgens werden de metabole wijzigingen die verantwoordelijk zijn voor het onderscheid tussen beide groepen geïdentificeerd om de verschillen in de verstoorde biochemische paden tussen beide kankertypes te verklaren. De verlaagde concentraties van sfingomyeline en fosfatidylcholine tonen aan dat de lipogenese en daaropvolgende membraansynthese meer uitgesproken is bij longtumoren dan bij borsttumoren. Deze bevindingen komen overeen met het feit dat longtumoren een hogere opname vertonen van de radioactieve tracer fluorodeoxyglucose op een positron emissie tomografie-scan en bevestigen aldus dat longtumoren metabool meer actief zijn dan borsttumoren. Aangezien lange, verzadigde vetzuurketens zeer sterke Van der Waals interacties hebben, tonen de verhoogde plasmaconcentratie van fosfolipiden met lange, verzadigde vetzuurketens en de verlaagde plasmaconcentratie van fosfolipiden met korte, onverzadigde vetzuurketens aan

dat het celmembraan van longtumoren steviger is (lagere fluiditeit) en zodus minder gevoelig is aan lipidenperoxidatie dan het celmembraan van borsttumoren.

Bovendien werd bestudeerd of het plasma metaboloom toelaat om long-, borst-, en dikkedarmkanker van elkaar te onderscheiden. Preliminaire resultaten hebben aangetoond dat het metabole plasma fenotype toelaat om 78% van 37 dikkedarmkankerpatiënten, 95% van 37 borstkankerpatiënten en 84% van 37 longkankerpatiënten correct te classificeren. Hoewel het aantal patiënten in deze studie nog beperkt is, tonen deze preliminaire resultaten aan dat ^{1}H-NMR plasma metabolomics potentieel heeft om een complementaire tool te worden die het mogelijk maakt om specifieke kankertypes te diagnosticeren.

Ten slotte tonen preliminaire resultaten van een samenwerking met de onderzoeksgroep Biostatistiek aan dat het toevoegen van plasma metaboliet data aan risicomodellen, die voordien enkel rekening hielden met epidemiologische en klinische gegevens, de misclassificatiefout van het risicomodel verlaagt van 24% naar 19%. Deze bevindingen tonen aan dat het metabole fenotype van bloedplasma potentieel heeft om de identificatie van individuen met een verhoogd risico op longkanker, die aldus in aanmerking komen voor LDCT screening, te verbeteren. Op deze manier zou het aantal vals positieve resultaten van LDCT screening en de overeenkomstige kosten voor de maatschappij verminderd kunnen worden. Alvorens ^{1}H-NMR plasma metabolomics toegepast kan worden in de routine klinische praktijk moeten bovenstaande studieresultaten echter nog onafhankelijk gevalideerd worden in grootschalige studies met asymptomatische individuen die een verhoogd risico hebben op longkanker en bijgevolg in aanmerking komen voor LDCT screening.

Curriculum Vitae

Evelyne Louis was born on 10 September 1988 in Hasselt (Belgium). She graduated in Latin-Sciences at the Heilig-Graf Instituut in Bilzen in 2006. In the same year, she started her bachelor studies in Biomedical Sciences at Hasselt University. In June 2011, she obtained her master degree in Biomedical Sciences (clinical molecular sciences) at Hasselt University. In October 2011, she accepted a new challenge: performing a PhD project at Hasselt University within the Limburg Clinical Research Program (LCRP), a collaboration between the Faculty of Medicine and Life Sciences of Hasselt University, Jessa Hospital (Hasselt, Belgium) and Ziekenhuis Oost-Limburg (Genk, Belgium). As a member of the Doctoral School for Medicine and Life Sciences, she followed several courses in order to improve her scientific and transferrable skills and she participated in the writing of a grant proposal. Furthermore, she had teaching responsibilities in several courses, supervised various bachelor and master students and participated in the organization of a PhD symposium.

Publications

Louis E[*], Bervoets L[*], Reekmans G, De Jonge E, Mesotten L, Thomeer M, et al. Phenotyping human blood plasma by ^{1}H-NMR: a robust protocol based on metabolite spiking and its evaluation in breast cancer. Metabolomics. 2015;11:225-36 (ISI IF_{2015}: 3.855)

[*]These authors contributed equally to this work

Bervoets L[*], **Louis E**[*], Reekmans G, Mesotten L, Thomeer M, Adriaensens P, et al. Influence of preanalytical sampling conditions on the ^{1}H NMR metabolic profile of human blood plasma and introduction of the standard preanalytical code used in biobanking. Metabolomics. 2015;11:1197-1207 (ISI IF_{2015}: 3.855)

[*]These authors contributed equally to this work

Louis E, Adriaensens P, Guedens W, Vanhove K, Vandeurzen K, Darquennes K, Vansteenkiste J, Dooms C, de Jonge E, Thomeer M, Mesotten L. Metabolic phenotyping of human blood plasma: a powerful tool to discriminate between cancer types? Annals of Oncology. 2015, DOI: 10.1093/annonc/mdv499 (ISI IF_{2015}: 7.040)

Publications in preparation

Louis E, Adriaensens P, Guedens W, Bigirumurame T, Baeten K, Vanhove K, Vandeurzen K, Darquennes K, Vansteenkiste J, Dooms C, Shkedy Z, Mesotten L, Thomeer M. Detection of lung cancer via metabolic changes in blood plasma. Under revision for resubmission to Journal of Thoracic Oncology (ISI IF_{2015}: 5.282)

Louis E, Cantrelle F, Mesotten L, Reekmans G, Vanhove K, Thomeer M, Lippens G, Adriansens P. Metabolic phenotyping of human plasma by NMR at higher magnetic field strengths: advantages and disadvantages of 900 versus 400 MHz. Submitted to metabolomics (ISI IF_{2015}: 3.855)

Louis R, **Louis E**, Stinkens K, Mesotten L, Reekmans G, de Jonge E, Thomeer M, Caenepeel P, Adriaensens P. The ^{1}H-NMR-derived metabolic phenotype of blood plasma discriminates between colorectal cancer, breast cancer and lung cancer. Submitted to Metabolomics (ISI IF_{2015}: 3.855)

Louis E*, Bigirumurame T*, Adriaensens P, Mesotten L, Vanhove K, Shkedy Z, Thomeer M. Risk models to select for lung cancer screening with low-dose computed tomography: the added value of including NMR metabolic phenotype data to reduce the misclassification error. In preparation for submission to Cancer Prevention Research (ISI IF_{2015}: 4.444)

*These authors contributed equally to this work

Awards

Winner of the Glasgow Polyomics & University of Strathclyde Young Scientist Award, *9th Annual Conference of the Metabolomics Society*, 1-4 July 2013, Glasgow, United Kingdom

Oral presentations

Louis E. Klinisch onderzoek binnen de cluster oncologie. *Dag van de Zorg*, 18 March 2012, Genk, Belgium.

Louis E. Bepalen van het metabool fenotype bij longkanker: klinisch toepasbaar of irrelevant? *Phd-symposium: Patiëntgericht wetenschappelijk onderzoek in de Limburgse ziekenhuizen*, 24 November 2012, Genk, Belgium.

Louis E. Metabolic phenotyping by ^{1}H-NMR spectroscopy detects lung cancer via a simple blood sample. *Belgian-Netherlands Joint symposium on Metabolomics*, 13-14 May 2013, Spa, Belgium.

Louis E. Metabolic phenotyping by ^{1}H-NMR spectroscopy to detect lung cancer via a simple blood sample. *9th Annual Conference of the Metabolomics Society*, 1-4 July 2013, Glasgow, United Kingdom.

Louis E. Metabolic phenotyping of blood plasma by ^{1}H-NMR spectroscopy to detect lung cancer? *Young Belgium Magnetic Resonance Scientist*, 2-3 December 2013, Blankenberge, Belgium.

Louis E. Opsporen van longkanker met behulp van een ^{1}H-NMR spectroscopische analyse van bloedplasma, *GlaxoSmithKline Awards in Pneumology*, 11 June 2014, Brussel, Belgium.

Louis E. Validation of ^{1}H-NMR-based metabolomics as a tool to detect lung cancer in human blood plasma. *Young Belgium Magnetic Resonance Scientist*, 24-25 November 2014, Spa, Belgium.

Louis E. Metabolic phenotyping of blood plasma as a tool to detect lung cancer, LCRP @ Jessa & ZOL, 6 November 2015, Hasselt, Belgium

Louis E. ^{1}H-NMR-based metabolomics to detect and to differentiate cancer types and its added value for lung cancer risk models, *Young Belgium Magnetic Resonance Scientist,* 30 November -1 December 2015, Blankenberge, Belgium.

Poster presentations

Louis E, Baeten K, Vandeurzen K, Vanhove K, Darquennes K, Mesotten L, Thomeer M, de Jonge E, Van der Speeten K, Mebis J, Bulens P, Adriaensens P. Metabolic phenotype by NMR spectroscopy: A biomarker for lung cancer. *Bringing cancer research into business*, 24 November 2011, Genk, Belgium.

Louis E, Baeten K, Vandeurzen K, Vanhove K, Darquennes K, Mesotten L, Thomeer M, de Jonge E, Van der Speeten K, Mebis J, Adriaensens P. Metabolic phenotype by NMR spectroscopy: A biomarker for lung cancer. *Knowledge for Growth*, 24 May 2012, Gent, Belgium.

Louis E, Reekmans G, Baeten K, Mesotten L, Thomeer M, Vanhove K, Mebis J, Vandeurzen K, Darquennes K, Adriaensens P. Metabolic phenotype by NMR spectroscopy: A biomarker for lung cancer. *Belgian-German (Macro) Molecular Meeting, Advanced Materials by Modular Strategies: From Synthesis to Industrial Applications*, 3-4 December 2012, Houffalize, Belgium.

Louis E, Vandeurzen K, Darquennes K, Vanhove K, Mesotten L, Thomeer M, Reekmans G, Adriaensens P. Metabolic phenotyping by ^{1}H-NMR spectroscopy detects lung cancer via a simple blood sample. *Belgian-Netherlands Joint symposium on Metabolomics*, 13-14 May 2013, Spa, Belgium.

Louis E, Mesotten L, Thomeer M, Vanhove K, Vandeurzen K, Sadowska A, Wauters E, Vansteenkiste J, Dooms C, Reekmans G, Adriaensens P. Validation of ^{1}H-NMR spectroscopy based metabolomics as a tool to detect lung cancer via a simple blood sample. *Markers in Cancer*, 7-9 November 2013, Brussel, Belgium.

Louis E, Thomeer M, Mesotten L, Vanhove K, Vandeurzen K, Sadowska A, Reekmans G, Adriaensens P. Validation of ^{1}H-NMR-based metabolomics as a tool to detect lung cancer in human blood plasma. *European Society for Medical Oncology 2014*, 26-30 September 2014, Madrid, Spain.

Louis E, Vanhove K, Stinkens K, Mesotten L, Reekmans G, Guedens W, Vandeurzen K, Sadowska A, Vansteenkiste J, Dooms C, Thomeer M, Adriaensens P. Validation of ^{1}H-NMR-based metabolomics as a new, complementary tool for the detection of lung cancer via human blood plasma. *Precision Medicine for Cancer*, 1-4 March 2015, Luxembourg, Luxembourg.

Louis E, Vanhove K, Stinkens K, Mesotten L, Reekmans G, Guedens W, Vandeurzen K, Sadowska A, Vansteenkiste J, Dooms C, de Jonge E, Thomeer M, Adriaensens P. Metabolic phenotyping of human blood plasma: a powerful biomarker to dicriminate between cancer types? *Precision Medicine for Cancer*, 1-4 March 2015, Luxembourg, Luxembourg.

Louis E, Vanhove K, Stinkens K, Mesotten L, Reekmans G, Guedens W, Vandeurzen K, Sadowska A, Vansteenkiste J, Dooms C, Thomeer M, Adriaensens P. Validation of ^{1}H-NMR spectroscopy as a tool to detect lung cancer in human blood plasma. *Knowledge for Growth*, 21 May, Gent, Belgium.

Louis E, Vanhove K, Stinkens K, Mesotten L, Reekmans G, Guedens W, Vandeurzen K, Sadowska A, Vansteenkiste J, Dooms C, Thomeer M, Adriaensens P. Lung Cancer Can Be Detected By Metabolic Changes In Blood Plasma. *Biomedica*, 2-3 June 2015, Genk, Belgium

Dankwoord

Ik zou graag iedereen die mij tijdens dit vierjarige avontuur geholpen en gesteund heeft bedanken. Iedereen bij naam noemen is onmogelijk (en ik zal er waarschijnlijk ook wel enkelen vergeten), maar toch wil ik de volgende personen in het bijzonder bedanken:

Allereerst wil ik graag mijn promotor, **Prof. dr. Michiel Thomeer** en mijn copromotoren, **Prof. dr. Peter Adriaensens** en **Prof. dr. Liesbet Mesotten**, bedanken.

Michiel, ook al liep onze samenwerking niet altijd van een leien dakje, ik denk dat we fier mogen zijn op het eindresultaat dat we bereikt hebben. Wat me altijd zal bijblijven is uw (over)enthousiasme en uw streven naar "sexy" manuscripten, zoals u dit ludiek verwoordde :o). Ook bent u zeer gepassioneerd in het uitoefenen van uw functie als longarts. Het grote aantal longkankerpatiënten dat u dagelijks op de raadpleging zag passeren liet u duidelijk niet onberoerd. U benutte dan ook elke kans om me er aan te herinneren dat we dit onderzoek uitvoeren om de overlevingskansen en de levenskwaliteit van de longkankerpatiënten te verbeteren. Bedankt om mij de kans te geven om dit project uit te voeren en om mij de klinische inzichten bij te brengen die nodig waren om mijn doctoraat succesvol te beëindigen.

Peter, u stond altijd voor mij klaar. Zelfs als u het zeer druk had, probeerde u toch een gaatje in uw agenda te vinden (net zoals voor al uw andere doctoraatsstudenten trouwens). Ik denk met veel plezier terug aan onze vele meetings in de koffiekamer (terwijl de anderen koffiepauze hadden of kwamen eten, vergaderden wij rustig verder :o)). Daarnaast heeft u vele avonden en soms zelfs uw slaap opgeofferd om posters/presentaties/projecten/manuscripten van mij na te lezen en te verbeteren. Mailtjes om één uur s 'nachts waren dan ook geen uitzondering :o). Het is vooral dankzij uw wetenschappelijke ervaring en het lezen-herlezen tot u helemaal tevreden was dat onze manuscripten (volgens de reviewer commentaren toch) zo vlot leesbaar zijn. Ook bedankt voor uw deskundige uitleg over de NMR techniek. Peter, nog eens bedankt voor alles! U was een grote hulp en steun tijdens mijn doctoraat!

Liesbet, eerst en vooral heel erg bedankt dat ik binnen uw afdeling controles en longkankerpatiënten mocht verzamelen. Zonder uw hulp (u staat goed uw "mannetje" tussen al dat mannelijk geweld op de afdeling :o)) was het mij nooit gelukt om op korte tijd zo een grote groep longkankerpatiënten en controles te includeren. U was ook diegene die me wegwijs maakte in de kliniek en u leerde me hoe ik gemakkelijk alle benodigde patienteninformatie kon terugvinden in Mediweb. Bedankt daarvoor! Ook leerde u me de PET beelden van de longkankerpatiënten interpreteren en gingen we aan de slag met de tool van Prof. Boellaard. Omdat Karolien het PET gedeelte van mijn doctoraat heeft overgenomen is ons contact doorheen de jaren wat afgezwakt. Toch wil ik u heel graag bedanken voor de ondersteuning tijdens het begin van mijn doctoraat en de vele momenten die u gedurende deze vier jaar vrijmaakte in uw drukke werkschema. Uiteraard ook bedankt aan uw echtgenoot, **Pol Degryse**, voor zijn waardevolle bijdrage aan de statistische discussies.

Vervolgens wil ik ook de leden van de doctoraatscommissie en de jury bedanken voor hun constructieve feedback op mijn thesis. **Prof. dr. Veerle Somers**, ik wil u graag bedanken voor de waardevolle input tijdens de jaarlijkse samenkomsten van de doctoraatscommissie. **Prof. dr. Jean Paul Noben**, bedankt voor uw nuttige inzichten betreffende het gewijzigde kankermetabolisme en om mij de kans te geven om enkele werkzittingen te geven tijdens het vak metabolisme. Omdat ik deze lessen had gegeven, had ik al de nodige achtergrond om het gewijzigde kankermetabolisme beter te begrijpen. **Prof. dr. Ziv Shkedy**, bedankt voor de statistische input en de samenwerking betreffende het risicomodel manuscript.

Prof. dr. Wanda Guedens, samen met Peter stond u klaar om mij te helpen het gewijzigde kankermetabolisme te ontrafelen. Bedankt voor de waardevolle inzichten betreffende het metabolisme, uw hulp bij de desbetreffende manuscripten en zeker ook voor de tijd die u vrijmaakte om de draft versies keer op keer na te lezen en voor de vele meetings in de koffiekamer. Ondanks het feit dat u geen promotor was en niet in mijn doctoraatscommissie zetelde, maakte u toch de nodige tijd voor me vrij! Hoedje af! Ik heb heel graag met u samengewerkt, Wanda, dankjewel!

Prof. dr. Eric de Jonge, bedankt voor de nodige hulp bij het opzoeken van de ontbrekende informatie van de borstkankerpatiënten die Kurt in het verleden op uw afdeling verzameld had. Ook bedankt voor uw nuttige feedback op de manuscripten. **Prof. dr. Christophe Dooms**, **Prof. dr. Johan Vansteenkiste**, **Prof. dr. Karen Darquennes**, en **Prof. dr. Kurt Vandeurzen**, bedankt voor jullie bijdrage aan de inclusie van de longkankerpatiënten en jullie constructieve feedback op de manuscripten.

Uiteraard ook bedankt aan het personeel van de afdeling Nucleaire Geneeskunde, **Ilka**, **Karolien**, **Katrien**, **Lambert**, **Lieve**, **Rina**, **Pascal**, **Sonia**, **Sonja** en de **stagiaires**, voor het verzamelen van de bloedstalen. Een extra dikke merci aan **Ilka**, **Karolien**, **Katrien** en **Sonja**, jullie namen zowel stalen bij de longkankerpatiënten op de PET als bij de controles op de afdeling Nucleaire Geneeskunde. Hoe druk het ook was, jullie namen met plezier een bloedstaaltje af voor mijn onderzoek. Dankjewel ook aan **Greet**, als secretaresse ontving je mij altijd met een glimlach en zorgde je ervoor dat de patiënten die voor een hartonderzoek kwamen, eerst langs mij passeerden. Ook bedankt voor je oprechte interesse in mijn onderzoek! Ik zou ook graag **Kristien de Bent** willen bedanken voor de collectie, verwerking en opslag van de stalen van de longkankerpatiënten in Leuven. Uiteraard ook bedankt aan de patiënten en controles die bereid waren om deel te nemen aan mijn onderzoek.

Ik zou ook graag de **medische secretaresses van de dienst pneumologie** van het Algemeen Ziekenhuis Vesalius in Tongeren (o.a. Linda) en Mariaziekenhuis Noord-Limburg in Overpelt (o.a. Carina) bedanken voor de toffe ontvangst. Jullie hielpen me graag op weg met het uitpluizen van de medische dossiers (want om het moeilijk te maken heeft ieder ziekenhuis een ander programma hiervoor :o)) en schoten ter hulp als ik een vraag had. Dank jullie wel! **Christel Oyen** van het Universitair Ziekenhuis Leuven, ook u bedankt voor de hulp bij de administratie die nodig was om mij toegang te verlenen tot de klinische gegevens van de longkankerpatiënten die geïncludeerd werden in het kader van dit doctoraatsproject. Ook bedankt aan het team van de Universitaire Biobank Limburg, **Caroline**, **Igna**, **Kim**, **Loes** en **Tine** om de kwaliteitsvolle opslag van de stalen te garanderen. **Kim**, bedankt voor het labelen van de stalen en voor je hulp bij de staalregistratie. **Loes**, bedankt voor je hulp bij het in orde maken van

de aanvraag voor de opstart van een nieuwe biobankcollectie en voor de waardevolle input bij het pre-analytische testen artikel.

Gunter en **Koen**, jullie waren altijd bereid om de "400" af te staan voor mijn metabolomics onderzoek, waardoor de routine stalen vaak wat langer moesten wachten. Gelukkig kwam er toen de multisampler, waardoor de verloren tijd ook s' nachts ingehaald kon worden ;o). Ook stonden jullie altijd klaar om mij te helpen als de temperatuur weer eens niet meewerkte of als het tunen wat moeilijker ging. Dikke merci! **Gunter**, nog eens bedankt voor je hulp bij het spike artikel. **Prof. dr. Guy Lippens**, bedankt om mij de kans te geven om NMR metingen uit te voeren op een 900 MHz spectrometer. Ook hartelijk dank aan **Francois-Xavier Cantrelle** om enkele dagen tijd vrij te maken om mij te helpen met de staalvoorbereiding en de NMR metingen. Eén dag was het zelfs bijna 24u en kwam de security kijken wie er nog zo laat in het labo aanwezig was :o)! Merci beaucoup!

Bedankt **Tatsiana** en **Theophile** voor de nodige statistische input! Hartelijk dank ook aan **Dave Bosmans** voor de hulp met de lay-out. U heeft me vaak geholpen om ervoor te zorgen dat de figuren in mijn manuscripten van voldoende hoge kwaliteit waren en niet te groot waren om up te loaden. Ik kan u niet genoeg bedanken! **Hilde**, **Jean**, **Jessica** en **Laura** van de financiële dienst, ik kon altijd met mijn vragen bij jullie terecht! Bedankt. **Astrid**, **Helene**, **Ilse**, **Kathleen** en **Marleen**, bedankt voor jullie hulp bij de administratie. **Veronique**, bedankt voor de hulp bij de voorbereiding van mijn doctoraatsverdediging.

Bedankt ook aan de studenten die ik begeleid heb, **Robby** en **Vincent**, jullie hebben heel mooi werk afgeleverd! Ik heb jullie met veel plezier begeleid en wens jullie veel succes in de toekomst!

Liene, aangezien we voor onze PhD projecten allebei gebruik moesten maken van de NMR, hebben we samen heel wat beleefd. We hebben samen zo goed als heel België doorkruist voor cursussen, meetings en congressen en soms gingen we ook al eens de landsgrenzen over. Ik kijk met veel plezier terug op deze tijd, want met twee is toch altijd veel leuker dan alleen. Ik zal ook nooit vergeten hoe we bijna te laat waren in Zaventem voor onze vlucht naar Glasgow! We waren al helemaal aan het uitdenken hoe we dat moesten uitleggen op de unief :o)! Liene,

een heel dikke merci voor de gezellige babbels op onze bureau, voor je hulp als ik vragen had over PhD kwesties en om een luisterend oor te bieden als ik dit nodig had. Je was een zeer toffe collega en ik hoop dat we af en toe nog eens gaan afspreken! Veel plezier in Australië en heel veel succes met je Amerikaanse avontuur!

Bojoura, **Kristel** en **Laura**, jullie ook bedankt voor de gezellige babbels (al dan niet tijdens de middaglunch ;o)). Deze waren een welkome afleiding tijdens het drukke werkschema en deden veel deugd. Hopelijk praten we regelmatig nog eens bij! **Laura**, veel succes met je carrière bij LifeTechLimburg! **Kristel**, doe dat goed in het Jessa ziekenhuis en **Bojoura**, veel succes met het laatste anderhalf jaar van je doctoraat!

Karolien, één jaar na de start van mijn doctoraat kwam je ons team vervoegen. Bedankt voor het nakijken van de TNM classificaties van de longtumoren. Ook nog eens merci voor op zaterdagnamiddag te komen helpen stalen uitzoeken in de -80°C vriezer! Pff dat was toch een werkje he! Gelukkig was er dan altijd een warme choco in de buurt :o)! Ook als ik klinische vragen had, kon ik altijd bij jou terecht! Bedankt voor alles, je bent de meest toegankelijke dokter die ik ken :o). Succes nog met uw doctoraat!

Kirsten, je maakte maar kort deel uit van ons team, maar toch wil ik je bedanken. Dankjewel voor Robby te begeleiden, voor de klinische inzichten en voor de leuke babbels! Ik ben blij dat je je draai in Leuven gevonden hebt! Succes met je verdere carrière!

Ook dankjewel aan de PhD collega's van het ZOL, **Amber**, **Anneleen**, **Annelies**, **Christophe**, **Cornelia**, **Dorien**, **Frederik**, **Ingrid**, **Joren**, **Kathleen**, **Lars**, **Lieselotte**, **Petra**, **Philippe**, **Sharona** en **Ward**. In de periode dat ik nog proefpersonen verzamelde, was ik dagelijks op de bureau in het ZOL. Er was altijd een gezellige sfeer, soms zelfs iets te gezellig om goed door te kunnen werken :o) (voel u zeker niet aangesproken Christophe ;o)). Merci voor deze toffe werksfeer en voor diegene die hun doctoraat nog moeten afleggen: veel succes!

Ook op privé-vlak kreeg ik de nodige mentale steun. Bedankt aan ons kliekje **David en Daisy**, **Davy en Sanne**, **Dirk en Sarah**, **Jurgen en Stéphanie** en **Michael en Sophie** voor de gezellige weekend-uitjes (café, cinema, stapje

zetten) waardoor ik even de stress van het werk kon vergeten. **Caroline**, **Dorien**, **Elke**, **Emmy**, **Katrien** en **Sylvia** dankjewel voor de girlsnights! Deze waren altijd zeer gezellig waardoor mijn batterijen terug helemaal terug opgeladen waren voor de volgende werkdag ;o).

Benny, lieve schat, voor jou was het ook niet altijd gemakkelijk. Zo vond je het niet zo leuk als ik weer maar eens 's avonds een belangrijke mail van het werk binnenkreeg en deze dadelijk wou beantwoorden. Dan zei je altijd: "Ik koop mij nooit een smartphone! Ze moeten maar wachten tot morgenvroeg :o)". Dankjewel om mij af en toe te "verplichten" om in het weekend samen iets leuks te doen en om naar mij te luisteren als ik eens stoom moest aflaten na een zware werkdag. Love you!

Robby, broer, 5 jaar geleden begon je aan de opleiding Biomedische wetenschappen. Bijgevolg zat je gedurende de periode van mijn doctoraat in hetzelfde vakgebied en kon ik al eens iets aan jou vragen (de omgekeerde wereld ;o)). Aangezien je kennis van het Engels beter is dan die van mij, heb ik je regelmatig gevraagd om manuscripten en zelfs deze thesis na te lezen op de Engelse taal. Merci voor al je hulp! Dankzij de verbeteringen van jou en Peter lezen de manuscripten een stuk vlotter! Ik wil dan ook nog zeggen: Veel succes met je zoektocht naar een nieuwe job, geloof in je eigen kunnen, je hebt veel meer in je mars dan je zelf denkt! En laat je niet te snel ontmoedigen ;o)!

Mama en **papa**, zonder jullie steun had ik deze "mijlpaal" in mijn leven nooit kunnen verwezenlijken. Allereerst bedankt om mij de kans te geven om verder te studeren aan de universiteit. Ook wil ik jullie bedanken voor het feit dat ik altijd bij jullie terecht kon als ik een moeilijke dag achter de rug had. Jullie snapten dan wel niet altijd waar ik het precies over had (in tegenstelling tot Robby :o)), maar waren wel altijd bereid om te luisteren en om mij met raad en daad bij te staan. Hoewel ik niet altijd in mijn eigen kunnen heb geloofd, hebben jullie wel altijd in mij geloofd. Bedankt voor jullie grenzeloze vertrouwen in mij! Ik zie jullie graag! Uiteraard ook een dikke merci aan onze hondjes **Sisqo** en **Remco** voor de lieve knuffels wanneer ik die nodig had :o).

Bedankt,

Evelyne, 8 januari 2016

"Don't cry because it's over, smile because it happened."

Dr. Seuss