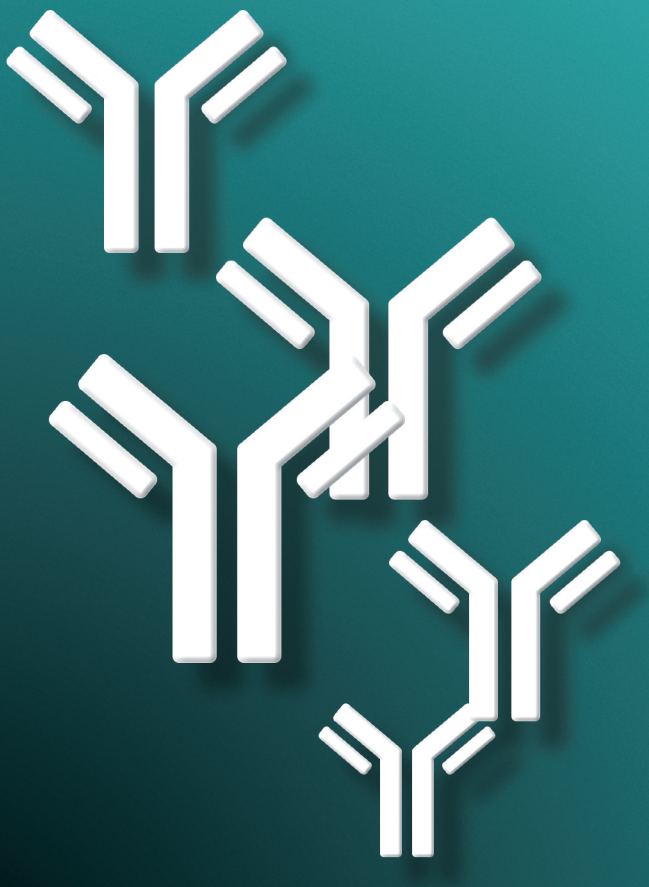
Ingrid Broodman



Early Detection of Lung Cancer

"A Role for Serum Biomarkers?"

Early Detection of Lung Cancer

"A Role for Serum Biomarkers?"

Ingrid Broodman

Financial support for the publication of this thesis was kindly provided by the Erasmus University Rotterdam
ISBN: 978-94-6169-921-3
Cover design: Harro Moens
Layout and printed by: Optima Grafische Communicatie, Rotterdam, the Netherlands

The digital version of this thesis can be downloaded at https://epubs.ogc.nl/?epub=i.broodman

Early Detection of Lung Cancer

"A Role for Serum Biomarkers?"

Vroegtijdige detectie van longkanker

"Een rol voor serum biomarkers?"

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

woensdag 12 oktober 2016 om 13.30 uur

door

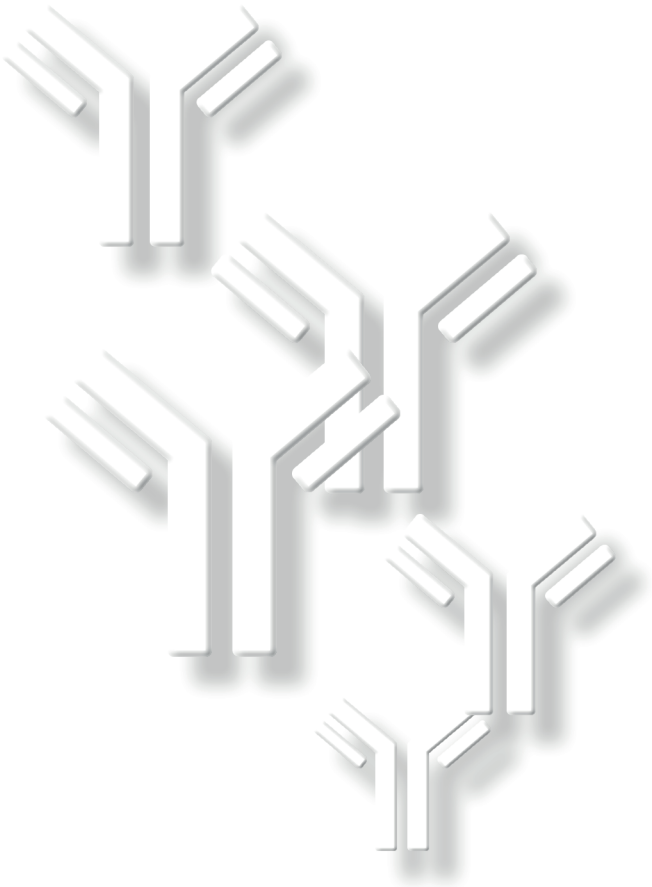
Ingrid Broodman
geboren te Rotterdam

Erasmus University Rotterdam

Promotiecommissie

Promotor:	Prof.dr. J. Lindemans
Overige leden:	Prof.dr. J.G.J.V. Aerts Prof.dr. R.P.H. Bischoff Dr. D. Rizopoulos
Copromotor:	Dr. T.M. Luider

Table of Contents



Introduction

CHAPTER 1

Part of the text has been submitted for a review

LUNG CANCER

Incidence and etiology

Lung cancer is the most common cancer type. Worldwide, more than 1.8 million men and women were diagnosed with lung cancer in 2012.[1, 2] In that year, an estimated 1.6 million died of lung cancer, accounting for one-fifth (19%) of all cancer deaths in the world.[1]

Cigarette smoking is the most important risk factor for lung cancer and accounts for about 80-90% of the lung cancer cases.[3, 4] Besides cigarette smoking, exposure to second-hand smoke, air pollution, asbestos, radon gas and certain chemicals such as arsenic, cadmium, chromium and nickel also increase the risk of developing lung cancer.[3, 5, 6] Genetic inheritance factors may play an important role in the individual risk for lung cancer caused by exposure to these carcinogens.[7, 8]

Lung cancer mainly occurs in older people. Almost 70% of the people diagnosed with lung cancer are 65 or older. Only 2% of all lung cancer cases are found in people younger than 45.[9] The median age at time of diagnosis is about 70 years.[9, 10] Lung cancer is more common in men than women. The male to female age-standardized incidence rates ratio is about 60% higher in men.[4] This high male to female ratio is mainly due to the higher prevalence of cigarette smoking in men than women.[11]

Survival rates of lung cancer patients vary depending on the stage of the cancer when it is diagnosed. The 5-year survival rate for lung cancer is about 15%.[4, 12] However, its 5-year survival rate may increase up to 49%, when lung cancer is diagnosed at an early stage.[9]

Types of lung cancer

Lung cancer, also known as carcinoma of the lung, is a malignant lung tumor formed by uncontrolled cell growth in the tissues of the lung, usually in the bronchi, the airways of the lungs. This growth may spread to a site distant from the lungs and produce metastatic tumors in brain, bone, liver, or adrenal glands.[4] Primary lung cancers are carcinomas that start in the lung and are derived from epithelial cells. These carcinomas are mainly classified by the size and appearance of the malignant cells under a microscope. These histological classifications are necessary for accurate treatment and prognosis of the different types of lung cancer. The two main primary types of lung cancer are non–small cell lung cancer (NSCLC) and small-cell lung cancer (SCLC). About 85% of all lung cancers are NSCLC. The three main subtypes of NSCLC are adenocarcinoma (40%), squamous cell carcinoma (25%-30%), and large cell carcinoma (10%-15%).[9] Adenocarcinoma (ADC) starts in glandular epithelial cells, called adenomatous cells, which secrete fluids such as mucus. It usually arises in the peripheral airways and can spread to other parts of the body. In general, adenocarcinoma grows more slowly than the other types of lung cancer. Although it occurs mainly in current and former smokers, it is the most common type of lung cancer among non-smokers. Women with adenocarcinoma are more frequently non-smokers than men.[9, 13, 14]

Squamous cell carcinoma (SCC) originates mostly from squamous cells in the central airways (trachea, main, lobar and segmental bronchi), but can also arise in the peripheral airways.[4] Squamous cells are large flat cells that often produce keratin.[4] SCC metastasizes later than the other types of NSCLC. It is also more strongly correlated to smoking than the other types of NSCLC, and is more common in men than in women.[13, 14]

Large cell carcinoma (LCC) is composed of poorly differentiated large cells without the specific glandular, squamous or neuroendocrine features of other lung carcinomas.[8, 15] LCC can arise in any part of the lung, grows quickly and tends to metastasize at an early stage to other parts of the body. The prognosis for LCC is unfavorable compared to other types of NSCLC.

About 10-15% of lung cancers are small-cell lung cancer (SCLC), a neuroendocrine carcinoma also called "oat-cell carcinoma". As the name suggests, the neuroendocrine epithelial cells of SCLC are extremely small and look like oat grains. SCLC often starts in the larger airways, the primary (main) and secondary (lobar) bronchi. It is the most aggressive type of lung cancer, grows more quickly than NSCLC and often metastasizes to other parts of the body early in the development of the disease.[9] Most of the SCLC patients have widespread metastasis at the time of diagnosis. SCLC is often associated with paraneoplastic syndromes (PNS).[9] The 5-year survival for SCLC (6%) is lower than that for NSCLC (21%).[9, 10] Nearly all cases of SCLC are due to cigarette smoking.[4, 9, 16] The incidence of SCLC is increasing in women and is now similar to that in men.[16]

Diagnosis and staging

Lung cancer tumors usually grow silent for many years without causing any signs or symptoms in early stages of the disease. About 5-10% of the lung cancer patients are diagnosed when still asymptomatic during a physical examination for an unrelated medical problem or after a routine chest X-ray with lung cancer in an early stage. Unfortunately, most symptomatic lung cancer patients are diagnosed with lung cancer in an advanced stage. The most common symptoms of lung cancer, in order of frequency, are cough, dyspnea, weight loss, chest pain, hemoptysis, bone pain and fatigue.[4] Patients that report such symptoms to a physician often receive a routine chest X-ray as the first step of investigation. This chest X-ray may usually detect large tumors, but may miss smaller or hidden tumors and does not exclude lung cancer.[17, 18] Computerized tomographic (CT) scans are able to detect smaller tumors than X-ray and also the size, shape and location of the tumor because of the three dimensional measurement. For instance, chest X-ray fails to detect almost 80% of CT-detected lung tumors of 20 mm or less in diameter.[19] Magnetic Resonance Imaging (MRI) is more sensitive than CT scanning and may be used when a certain region is difficult to interpret on a CT scan. MRI scans are also useful for diagnosis of bone or central nervous system metastases.[8] Positron emission tomography (PET) scans measure the metabolic activity and function of the tissues. These scans determine the stage and type of the tumor and are very useful for finding metastatic tumors in other parts of the body. The final determination whether a

tumor is malignant, and definitive diagnosis of the type of lung cancer can only be made by examining of a tissue sample by a pathologist. This tissue sample can be obtained by bronchoscopy, sputum cytology, or fine needle aspiration biopsy.[4, 9]

Lung cancer staging is a system that describes the growth and extent of spread of the cancer to other parts of the body. This staging system helps the physician to determine the most effective treatment and prognosis of the disease. The TNM classification system is based on the primary tumor characteristics (T), regional lymph node involvement (N), and distant metastasis status (M) and is established by the International Association for the Study of Lung Cancer (IASLC).[4, 9, 20] This information is combined to classify patients in five stages, 0 (*in situ*), I, II, III or IV. Patients with a higher stage number have a poorer prognosis and lower survival rate.[20]

LUNG CANCER SCREENING

The purpose of lung cancer screening is to detect lung cancer at an early and still curable stage to improve the survival rate of the lung cancer patients. Survival rate improves significantly with early detection of the disease, with a respective 5-year survival rate increasing from 2%, 7%, 19%, 25%, 36% and 43% for stages IV, IIIB, IIIA, IIB, IIA and IB to 50% for stage IA.[20] While the overall survival rate remains poor, patients diagnosed with stage I are potentially curable.[9] Stage I NSCLC patients are usually treated by surgery to remove the cancer, sometimes in combination with chemotherapy or radiation therapy. This surgery offers the best chance to cure early stage NSCLC patients. Because the lung cancer is diagnosed and treated at a localized stage, their 5-year survival rate may increase up to 49%.[9] Since these patients are usually asymptomatic, only 15% of all diagnoses of lung cancer are from stage I.[4] In contrast, CT screening detected 48-85% of lung cancers in stage I.[17, 21] Therefore screening is performed on apparently healthy, asymptomatic people at high risk of lung cancer. People at high risk of lung cancer are current smokers and former smokers.

Randomized screening studies

Randomized screening studies for early detection of lung cancer in high-risk individuals are ongoing. An overview of the main large-scale lung cancer screening studies is presented in Table 1. These studies are comparing low-dose spiral CT screening with chest X-ray or usual care. The aim of these lung cancer screening studies is to reduce the lung cancer mortality with 20-25% by lung cancer detection at an early and still curable stage. Four of the eight randomized screening studies have now published their results. Three trials in Europe, the DANTE (Detection and Screening of Early Lung Cancer by Novel Imaging Technology and Molecular Essays), DLCST (Danish Lung Cancer Screening Trial), and MILD (Multicentric Italian Lung Detection) trials, reported no significant reduction in lung cancer mortality.[22-26]

However, these small randomized trials do not have the statistical power to demonstrate a reliable clinical outcome.

The largest study, the NLST (U.S. National Lung Screening Trial) study reported a significant lung cancer mortality reduction of 20.3% in high-risk individuals who were screened annually with low-dose spiral CT (LDCT) compared to those who were screened annually by chest X-ray.[27, 28] Since this publication, many medical societies have recommended LDCT screening of high-risk individuals in reducing lung cancer mortality.[29-34] In their recommendations, they define high-risk individuals as apparently healthy individuals aged 55-74 years who have a smoking history of at least 30-pack years and currently smoke, or have quit smoking within the past 15 years, or some modification of these inclusion criteria. Effectiveness in terms of survival benefit for LDCT screening of high-risk individuals has been demonstrated by the NLST. However, uncertainty remains about the effectiveness of LDCT screening in other settings or populations screened than in the NLST trial. Nevertheless, no other interventions, besides primary prevention, up to now have shown such reduction in lung cancer mortality. At present, the NELSON, ITALUNG, LUSI and the UKLS screening studies (Table 1) are still ongoing. When data of all randomized screening studies becomes available, a definitive conclusion of the effectiveness of LDCT screening can be made.

The NELSON study

The NELSON trial (Nederlands-Leuvens Longkanker Screeningsonderzoek), -Dutch-Belgian Lung Cancer Screening trial- is world's second-largest randomized lung cancer computer tomography screening trial and differs from the NLST study by screening interval, referral policy, and a control arm wherein individuals receive no screening (usual care).[39, 45] The NELSON trial started in 2003. The main purpose of the trial was to investigate whether LDCT (low dose CT) screening leads to a reduction of lung cancer mortality of at least 25% at 10 years of follow-up in a high risk population. The second purpose was to estimate the cost-effectiveness of lung cancer screening. Participants were recruited between 2003 and 2005 by sending questionnaires to 548,489 individuals between 50–75 years of age from population registries of 7 public health districts in the Netherlands and population registries of 14 municipalities around Leuven in Belgium. Current or former smokers with a smoking history of at least 15 cigarettes per day for at least 25 years or at least 10 cigarettes per day for at least 30 years were included in the trial. Individuals with a bad or moderate self-reported health status, the inability to climb two flights of stairs, or a body weight over 140 kilograms were excluded. Furthermore, individuals with current or past renal cancer, melanoma or breast cancer and lung cancer diagnosed less than 5 years ago or still under treatment were also excluded. A total of 15,822 participants were randomized (1:1) to a screen or a control arm. The screen arm received computed tomography (CT) screening at baseline (first screening round), one year later (second screening round), three years later (third screening round), and five and a half years later (fourth screening round), whereas the control arm received no CT screen-

Table 1. Main large-scale randomized controlled lung cancer screening trials

Trial	Initiation Complete	N	Design	Screens	♂ %	Age yrs	Pack yrs	Quit* yrs
DANTE[23, 24, 35] Italy	2001 2009	2,472	LDCT vs. none	5	100	60-74	≥20	<10
NLST[27, 28] USA	2002 2011	53,454	LDCT vs. CXR	3	59	55-74	≥30	<15
ITALUNG[36] Italy	2004 ongoing	3,206	LDCT vs. none	4	65	55-69	≥20	<10
NELSON[37-39] Netherlands/ Belgium	2004 ongoing	15,822	LDCT vs. none	5	84	50-75	≥15[a]	≤10
DLCST[25, 40] Denmark	2004 2011	4,104	LDCT vs. none	5	55	50-70	≥20	<10
MILD[26] Italy	2005 2012	4,099	LDCT vs. none	5	66	≥49	≥20	<10
LUSI[41, 42] Germany	2007 ongoing	4,052	LDCT vs. none	5	65	50-69	≥15[a]	≤10
UKLS[43, 44] United Kingdom	2011 ongoing	4,055	LDCT vs. none	Pilot study	75	50-75	NA[b]	NA

CXR, chest X-ray; DANTE, Detection and Screening of Early Lung Cancer by Novel Imaging Technology and Molecular Essays; DLCST, Danish Lung Cancer Screening Trial; ITALUNG, Italian lung study; LDCT, low-dose spiral computed tomography; MILD, Multicentric Italian Lung Detection trial; N, patient number; NA, not applicable; NELSON, Dutch-Belgian Lung Cancer Screening Trial (Dutch acronym); NLST, National Lung Screening Trial; LUSI, German Lung Screening and Intervention trial; UKLS, UK Lung Screening trial; yrs, years; ♂, male.
*quit smoking; [a] inclusion criteria ≥ 15 cigarettes per day for 25 years or ≥10 cigarettes per day for 30 years;
[b] inclusion criteria ≥5% risk of lung cancer in 5 yrs.

ing (Figure 1).[37, 39, 109] The difference in lung cancer mortality between the screen arm and the control group will be determined in 2016. Initial CT screening results were based on the lung nodule presence and volume. Screening results were considered positive for (part) solid lung nodules with a volume of >500 mm^3 (>9.8 mm in diameter) and was considered indeterminate for (part) solid lung nodules with a volume of 50 to 500 mm^3 (4.6 to 9.8 mm in diameter). Participants with an initial indeterminate screening result received a follow-up CT scan three months later to classify their final screening test result as negative or positive,

based on nodule volume doubling time (VDT).[38, 46] If the nodule had a VDT<400 days, the final screening result was considered positive. Participants with a positive screening result were referred to a pulmonologist for a diagnostic follow-up. If lung cancer was diagnosed, the participant was offered a treatment protocol and went off screening. Participants with a negative screening result re-entered the protocol and underwent a second-round CT scan 12 months later.

From the 7,915 participants of the screen arm 7,582 (95.8%) participants received at least one screening. In the first three screening rounds were 493 positive screening results found and 200 (40.6%) participants diagnosed with a total of 209 lung cancers.[45] The lung cancers in the NELSON trial were more frequently detected at an early stage (70.8% stage I) and less frequently at an advanced (8.1 % stage IIB-IV) than in other screening trials.[38, 39, 45] The definition of a positive screening result differed considerably between the NLST and the NELSON trial. The NLST defined any solid nodule with a diameter ≥4 mm as a positive screening result.[27, 28] The NELSON trial considered only solid lung nodules with a volume >500 mm3 (>9.8 mm in diameter) or VDT<400 days as a positive screening result. This policy is more stringent than the NLST policy. By using this policy, the positive predictive value was higher in the NELSON trial (40.6%) than in the NLST trial (3.6%). Consequently, the percentage of false-positive results was substantially lower in the NELSON trial (59.4%) than in the NLST

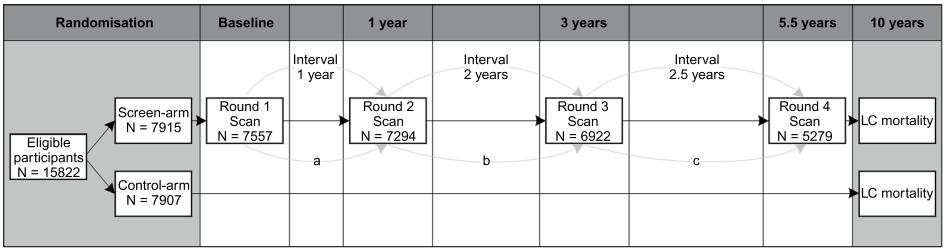


Figure 1. Design of the NELSON Lung Cancer Screening Trial.

Round 1. The first round of screening was conducted from April 2004 through December 2006.[39]

Round 2. The second round of screening was conducted from April 2005 through April 2008, 384.3 days on average (SD 59.2 days) after the baseline scan.[39]

Round 3. The third round of screening was conducted from January 2007 through October 2010, 733.2 days on average (SD=71.9 days) after the second-round scan.[39]

Round 4. The fourth round was conducted from November 2009 through March 2012, 942.3 days on average (SD=127.4 days) after the third-round scan.[109]

a. 7450 participants were invited for screening (7557 of round 1 plus 25 participants who missed screening in round 1, minus 70 participants diagnosed with lung cancer in round 1 and minus 62 deceased participants).

b. 7081 participants were invited for screening (7294 of round 2 minus 55 participants diagnosed with lung cancer in round 2 and minus 158 deceased participants).

c. 6735 participants were invited for screening (6922 of round 3 minus 75 participants diagnosed with lung cancer in round 3, minus 27 participants diagnosed with interval cancer and minus 85 deceased participants).

trial (96.4%).[27, 39, 45] Lung cancer mortality results of the NELSON trial are upcoming. The first results on mortality reduction after 10-year follow-up are expected in 2016.
More than 6,600 serum samples of the NELSON trial were collected at baseline. For the studies described in this thesis we used serum samples from cases and controls of the screen arm.

BIOMARKERS

There is a medical need for additional biomarkers for early detection of lung cancer, as CT screening leaves 15-52% of cases undetected.[17, 21] CT screening has also other limitations. First, CT screening has a high rate of false-positive results due to the high prevalence of benign pulmonary nodules. This results in unnecessary follow-up CT scans, additional tests, biopsies or even surgery. In the NLST, 24% of benign patients underwent an unnecessary surgical procedure (thoracotomy, thoracoscopy, or mediastinoscopy).[27] Secondly, these invasive follow-up procedures are costly and may harm patients.[30, 47] Alternatively, false-negative and indeterminate results may harm patients due to the delayed diagnosis and treatment of lung cancer.
Biomarkers in blood could be a non-invasive, cost-effective tool to stratify individuals at high risk of lung cancer (pre-cancer) who would benefit from CT screening. These biomarkers may be used for early diagnosis, optimal treatment selection, and prognosis of the disease. They may not only reduce the number of unnecessary invasive procedures, but may also lead to the earlier resection of malignant nodules which will substantially improve the prognosis of the patient. Unfortunately, there is still no clinically relevant blood biomarker available for lung cancer although various groups have proposed proteins, mostly in panels of antigens or autoantibodies. In searching for a clinically relevant biomarker for lung cancer it is vital to understand the biological processes of lung cancer. Lung cancer cells have defects in their regulatory processes that maintain normal cell proliferation and homeostasis. Critical changes in their cell physiology lead to cancer growth. Lung cancer cells are insensitive to growth-inhibitory signals and show escape from apoptosis, unlimited replication, sustained angiogenesis, tissue invasion and metastasis.[48] Transformation from a normal to a malignant lung epithelial cell arises after a series of genetic and epigenetic changes, eventually leading to invasive lung cancer by clonal expansion.[49] The molecular composition of lung cancer is complex and heterogeneous, which leads to variable biological, histological and clinical presentations. Various oncogenes, tumor suppressor genes, signaling pathway components, and other cellular processes are involved in the molecular pathogenesis of lung cancer.[4, 50] These cellular processes can lead to the mutation or overexpression of various proteins that may be released into body fluids. Therefore, identification of these lung cancer associated proteins in body fluids as potential biomarkers is a potential way to search for molecules that can detect lung cancer at an early stage, resulting in more optimal treatment and better

prognosis of lung cancer at that early stage. Serum or plasma is considered the most optimal body fluid for this biomarker discovery because of its minimal invasive and easily accessible source.

Lung cancer associated proteins

Lung cancer is often associated with the differential expression of several proteins. These differentially expressed proteins may be potential biomarkers for lung cancer. Table 2 represents a selection of lung cancer associated proteins as potential blood-based biomarkers for lung cancer that have been described in literature.

The well-known and clinically available lung cancer protein biomarkers in serum are carcinoembryonic antigen (CEA), CYFRA 21-1 (cytokeratine 19 fragment), neuron-specific enolase (NSE), progastrin-releasing peptide (ProGRP), and squamous cell carcinoma antigen (SCCA). Although these proteins are elevated in serum of a fraction of lung cancer patients, they are not sensitive or specific enough to detect lung cancer, or to have enough value as biomarker for the diagnosis of asymptomatic patients with lung cancer.[61, 62] Only biomarkers with a sensitivity of at least 95% and specificity of at least 95% are considered to have diagnostic value for early detection of lung cancer. Therefore, these protein biomarkers are not recommended for the diagnosis of lung cancer.

Lung tumor cells may secrete or release small amounts of various tumor-associated proteins in an early stage of lung cancer. Detection of these lung cancer associated proteins in biological samples is proposed for the early diagnosis, prognosis and optimal treatment of lung cancer. However, the large dynamic range of other proteins in a complex biological blood sample , which extends over 10 orders of magnitude, and the high-abundance of albumin (55%) in serum or plasma is a major problem to detect these low-abundant proteins by mass spectrometry.[63] Depletion of high-abundant proteins or targeted enrichment of lung cancer associated proteins are the main strategies to overcome this problem and to enhance the detection of these low-abundant proteins. New DNA-based aptamers have been developed that contain chemically modified nucleotides which bind with high affinity to different low-abundant proteins. Ostroff et al. used an aptamer-based proteomic assay in a multi-center case-control study of 291 NSCLC cases and 1035 non-matched smoking controls.[57] They developed a panel of twelve highly differential proteins (cadherin-1, CD30 ligand, endostatin, HSP90α, LRIG3, MIP-4, pleiotrophin, PRKCI, RGM-C, SCF-sR, sL-selectin, and YES) that was able to distinguish 213 NSCLC cases (62% stage I-II) from 772 controls with 91% sensitivity and 84% specificity. This panel was tested on a validation set consisting of 78 NSCLC cases (63% stage I-II) and 263 controls, including patients with COPD and benign nodules. In this validation set, a similar sensitivity of 89% at a similar specificity of 83%, and a relatively high AUC of 0.90 was found. The reason that this panel is not clincally implemented might be because this sensitivity and specificity is still too low for clinicians. For clinical utility a

Table 2. Characteristics and performance of blood-based proteins as potential biomarkers for lung cancer

Reference	proteins	N	Remarks subjects	Stage (%)					Sensitivity (%)	Specificity (%)	AUC	Method
				I	II	III	IV	Tx				
Li et al.[51]	13 peptides panel	143	D, IPNs[a]	-	-	-	-	100	93	45	0.82	MRM-MS
		104	V, IPNs[a]	-	-	-	-	100	90	27	0.60	
Patz et al.[52]	CEA, AAT, SCCA*	509	D, PNs	41	13	26	16	4	82	84	NA	LCBA
		399	V, PNs	46	18	21	13	2	80	89	NA	
Pecot et al[53]	Clinical data +	100	IPNs[b]	-	-	-	-	100	NA	NA	0.57	
	CT data +								NA	NA	0.67	
	MALDI-MS signature**								NA	NA	0.72	MALDI-MS
Bigbee et al.[54]	10 protein panel	60	NSCLC, SC	-	-	-	-	100	73	93	NA	Luminex
Diamandis et al.[55]	Penatraxin-3	383	LC, SC	14	2	4	2	78	37	90	0.60	ELISA
Takano et al.[56]	Nectin-4	295	NSCLC, HC	27 (I-IIIA)		73 (IIIB-IV)		-	54	98	NA	ELISA
Ostroff et al.[57]	12 proteins panel	985	D, NSCLC, SC	47	15	38	0	-	91	84	0.91	Aptamers
		341	V, NSCLC, SC	49	14	35	2	-	89	83	0.90	
Patz et al[58]	CEA, RBP4, AAT,	100	D, LC, HC	40	4	30	26	-	89	85	NA	ELISA
	SCCA	97	V, LC, HC	33	6	39	22	-	78	75	NA	

Table 2. Characteristics and performance of blood-based proteins as potential biomarkers for lung cancer (continued)

Reference	proteins	N	Remarks subjects	Stage (%)					Sensitivity (%)	Specificity (%)	AUC	Method
				I	II	III	IV	Tx				
Yildiz et al.[59]	MALDI-MS	182	D, NSCLC, SC	39 (ES)		61 (LS)		-	67	89	0.82	MALDI-MS
	signature**	106	V, NSCLC, SC	40 (ES)		60 (LS)		-	58	86	0.82	
Gao et al.[60]	CRP, SAA, MUC1	80	LC, SC	-	-	-	-	100	71	93	NA	protein microarray

Note: Data are listed by most recent publication first (2013-2005).
I, stage I; II, stage II; III, stage III; IV, stage IV; AUC, area under the curve; D, discovery set; ELISA, enzyme-linked immuno sorbent assay; ES, early stage: NSCLC I, II and limited SCLC; HC, healthy controls; IPNs, indeterminate pulmonary nodules; LC, lung cancer (NSCLC and SCLC); LCBA, lung cancer biomarker (Immuno)assay; LS, late stage: NSCLC III, IV and extensive SCLC; MRM-MS, multiple-reaction-monitoring mass spectrometry; MS, mass spectrometry; N, patient number; NA, not applicable; PNs, benign and malignant nodules; SC, smoking controls; Tx, tumor stage unknown (or not described), V, validation set. *logistic regression model based on LBCA data and nodule size; **signature of seven features; [a] nodule size 10-20 mm [b] 5-20 mm.

sensitivity and specificity of at least 95% is mostly acceptable (personal communication with a pulmonologist). Li et al. used immunoaffinity columns for the tandem depletion of high-abundant proteins. They developed and validated a 13-protein blood-based classifier using multiple-reaction-monitoring mass spectrometry (MRM-MS) in a retrospective study consisting of 52 NSCLC and 52 benign controls. Their classifier distinguished benign from early-stage (IA) NSCLC nodules with a relatively high sensitivity of 90% sensitivity, but quite low specificity of 27%.

Unfortunately, the methods in Table 2 were not able to offer overall sensitivity and specificity of at least 95% to reliably discriminate lung cancer patients from controls. Sensitivity and specificity were even lower for early stage lung cancer. In addition, most of the proposed lung cancer proteins were not validated between lung cancer cases and controls well-matched for smoking-habit, which is the most relevant group for screening purposes. None of the proteins in Table 2 are currently in use as a clinically relevant biomarker for the early detection of lung cancer.

Immunological biomarkers

The presence of tumor cells can activate the immune system to respond to tumor-specific antigens or to tumor-associated antigens.[64, 65] Tumor-specific antigens (TSA) are only expressed in tumor cells, whereas tumor-associated antigens (TAA) are expressed differently by tumor cells and normal cells. The immune system not only protects the host against the development of primary tumors but may also, strangely enough, promote development of primary tumors. This process, also known as cancer immuno-editing, consists of three phases: elimination, equilibration, and escape. Immunosurveillance occurs during the elimination phase, whereas the immune system recognizes tumor cells as foreign cells and eliminates many of them. Tumor cells that survive this phase enter into the equilibrium phase. In the equilibrium phase variants of the tumor cells are saved or mutated to tumor cell variants with increased resistance to immune attack. This equilibrium phase is assumed to be the longest of the three phases and may continue for several years. The tumor cell variants start to grow in an uncontrolled manner and eventually will be detected in the escape phase.[66, 67] These tumor cells express tumor-associated antigens (TAAs) that distinguish them from normal cells. Most of the TAAs are overexpressed, mutated, misfolded or aberrantly degraded in such a way that they initiate an autoreactive immune response.[64, 68, 69] Post-translational modifications (PTMs) of TAAs, such as acetylation, glycosylation, oxidation, phosphorylation and proteolytic cleavage, could contribute to an immune response by creating a neo-epitope or by improving self-epitope presentation and affinity to the major histocompatibility complex (MHC) or the T-cell receptor.[64, 68, 70] Identification of tumor-associated antigens and autoantibodies to these antigens provide an opportunity for early detection of lung cancer.[71]

Antibodies as biomarker

Autoantibodies to tumor-associated antigens (TAAs) are potential biomarkers for early detection of lung cancer. First, autoantibodies can be detectable in the asymptomatic stage of lung cancer, up to 5 years before radiological detection by CT.[72, 73] Second, in contrast to antigens, autoantibodies are stable and persist in serum for a relatively long period of time at relatively high levels.[64] Tumor-associated antigens may be temporarily present at very low levels due to temporary changes in only a few (pre)neoplastic cells. However, the immune system is very sensitive in detecting these very low levels of TAAs, and may respond by producing very high-affinity T cells and autoantibodies.[74] Such an autoantibody response to a tumor-associated antigen may endure over years. Thus, autoantibodies may be more detectable and at an earlier stage than their corresponding TAAs.

Human IgG antibodies, also known as immunoglobulins, are large molecules (~150 kDa) and composed of four polypeptide chains, two identical heavy chains (50 kDa) and two identical light chains (25 kDa). Each light chain has a variable (V_L) and constant (C_L) region. The heavy chains have three different constant regions (C_H1, C_H2 and C_H3) and one variable region (V_H). The first constant region and variable region of the heavy chain together with the constant and variable part of the light chain form the antigen binding fragment (Fab). The other two constant regions (C_H2 and C_H3) of the heavy chain form the Fc fragment (Figure 2). Three hypervariable complementarity- determining regions (CDR1, CDR2 and CDR3) in the variable regions of the heavy and light chains of an immunoglobulin form the binding surface complementary to the antigen. As such, these CDRs in combination determine the specificity of the immunoglobulin to the antigen. The vast diversity in immunoglobulins initiates during immune response and B-cell development, when CDRs are generated by somatic rearrangements of different V, D and J germline genes, each forming a specific combination of germline genes. These rearranged genes can be further diversified by somatic hypermutations to increase antibody affinity.[75-79] In both light and heavy chains, the diversity of CDR3 is even further enhanced by the insertions and deletions of nucleotides between the genes. The high diversity of CDR3 makes it the key part of antigen recognition, it is the region that most directly interacts with the antigen.[80] The estimated potential diversity in immunoglobulins ranges from 10^{13} to more than 10^{50}.[78, 81] Despite this large range, there is evidence for a repertoire bias, which means that specific germline genes are preferred in the repertoire of immunoglobulins that is elevated during the immune response to a particular antigen.[82, 83] Antigen-specific immunoglobulin sequences may be shared among different lung cancer patients and could serve as biomarkers for lung cancer.

Lung cancer associated autoantibodies

During tumor development lung cancer patients produce specific autoantibodies to tumor-associated antigens (TAAs) that are potential biomarkers for lung cancer. Table 3 represents a

list of autoantibodies to TAAs as potential blood-based biomarkers for lung cancer that have been described in literature.

Although, autoantibodies are an active area of research, this research has not yet led to clinically relevant biomarkers. Table 3 represents a list of autoantibodies to TAAs that have been described in literature. All these studies were able to detect autoantibodies to TAAs, but none of the proposed autoantibodies has found application as a significant biomarker in the clinic. These autoantibodies studies have limitations. First, most of the studies described in Table 3 lack adequate clinical validation. Second, most proposed markers are not specific for lung cancer. For instance, Annexin, CAGE, CEA, HER2, MUC1, NYESO-1 and p53 also arise in other cancers and autoimmune diseases. Third, the studies that were validated were not able to show a clinically relevant sensitivity and specificity of at least 95%. Fourth, some of the methods are time-consuming and therefore not applicable in the clinic. Furthermore, because of the heterogeneity of lung cancer, it is not likely that an autoantibody to any single tumor-associated antigen will detect all types of lung cancer. Various target antigens are involved in the immune response to the different tumors. Therefore, it is more likely that autoantibodies to an antigen panel will detect the different types of lung cancer. The

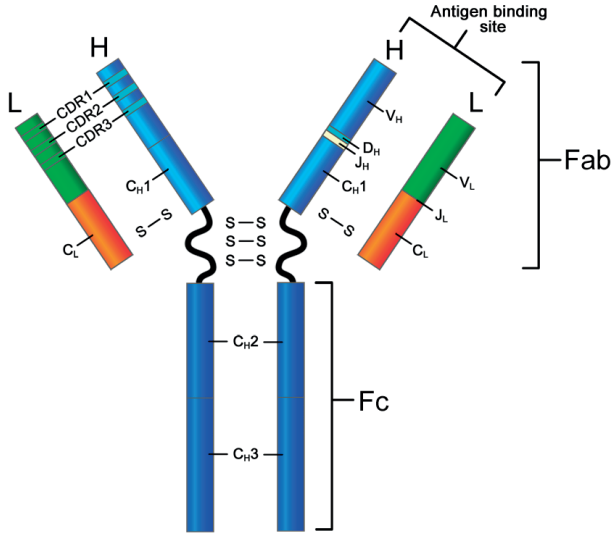


Figure 2. Structure of an immunoglobulin molecule (IgG).
Three hypervariable complementarity- determining regions (CDR1, CDR2 and CDR3) in the variable regions of the heavy (V_H) and light (V_L) chains of an immunoglobulin form the binding surface complementary to the antigen. As such, these CDRs combined determine the specificity of the immunoglobulin to the antigen. Fab, fragment antigen binding; H, heavy chain; L, light chain; J_L, joining region of the light chain; C_L, constant region of the light chain; D_H, diversity region of the heavy chain; J_H, joining region of the heavy chain; C_H1, first constant region of the heavy chain; C_H2, second constant region of the heavy chain; C_H3, third constant region of the heavy chain; Fc, fragment crystallizable.

*Early*CDT-Lung is currently used as an aid to risk assessment and the early detection of lung cancer in high-risk patients. This blood test measures autoantibodies to a panel of seven TAAs (p53, NY-ESO-1, CAGE, GBU4-5, SOX2, HuD and MAGE A4) and was validated in large cohorts including early and late stages NSCLC and SCLC. This autoantibody panel showed overall high specificity of about 91%, but rather low sensitivity of about 37% in NSCLC and 55% in SCLC (Table 3).[73, 85, 91-93, 99]. Another disadvantage of most methods in Table 3 is the limitation that the antigen or antigen panel must be known at the start of the study. Therefore, the development of a sensitive and specific autoantibody detection method without prior knowledge about the antigens involved offer opportunities to explore the complete inventory of tumor-associated antigens and would be of clinical importance.

To reach the highest sensitivity and specificity and to cover the histological heterogeneity of lung cancer, we propose that a panel of peptide sequences in the antigen-binding site of autoantibodies has potential as a relevant screening test for early stage lung cancer. While antibody diversity is huge, the selection pressures during B-cell development restrict the potential antibody diversity that is elevated to a particular tumor antigen. Antibodies are subjected to selection pressures after rearrangement and affinity maturation.[82, 83] During B-cell development and maturation, germline sequences in the hypervariable regions of the antigen-binding site of antibodies are rearranged to form a specific VDJ germline combination. This specific germline combination is further refined by somatic mutations to form an antibody sequence with the highest-affinity to the targeted antigen. As such, these high-affinity sequences are highly specific for the antigen involved. Different studies have demonstrated that it is possible to identify similar or identical autoantibody sequences among different individuals.[104-108] Specific sequences of high-affinity antibodies can be expressed in response to low levels of tumor-associated antigens in early stage lung cancer and could serve as biomarker for early detection of lung cancer.

Table 3. Characteristics and performance of blood-based autoantibodies to TAAs as potential biomarkers for lung cancer.

Reference	Antibodies to TAAs	N	Remarks subjects	Stage (%)					Sensitivity (%)	Specificity (%)	AUC	Method
				I	II	III	IV	Tx				
Doseeva et al.[84]	NY-SO-1 (and	230	D, NSCLC, SC	77	20	3	0	-	74	80	0.81	Luminex
	CEA, CA125, CYFRA 21–1)*	150	V, NSCLC, SC	32	35	20	13	-	77	80	0.85	
Jett et al.[85]	6 AABs panel**	752	LC, SC	39	11	21	15	13	46	83	NA	ELISA
	7 AABs panel***	847	LC, SC	39	11	21	15	13	37	91	NA	
Jia et al.[86]	p53, NY-ESO-1, Livin, Ubiquilin, BIRC, p62, PRDX	98	LC, SC	-	-	-	-	100	80	60	0.82	Luminex
Wang et al.[87]	ANXA1IgG	499	NSCLC, SC	45	19	32	4	-	24	90	0.64	ELISA
	DDX53IgG	499	NSCLC, SC	45	19	32	4	-	14	90	0.52	
Lowe et al.[88]	9 marker panel	200	D, AAH, SC	-	-	-	-	100[a]	92	90	0.87	Phage-display +
		300	V, AAH, SC	-	-	-	-	100[a]	82	70	0.81	protein microarray
	13 marker panel	200	D, SCD, SC	-	-	-	-	100[a]	98	96	0.96	
		300	V, SCD, SC	-	-	-	-	100[a]	86	78	0.88	
Zhang et al.[89]	Anti-p16 IgA	497	NSCLC, SC	44	19	33	4	-	10	90	0.46	ELISA
	Anti-p16 IgG	446	NSCLC, SC	47	17	32	4	-	20	90	0.57	

Table 3. Characteristics and performance of blood-based autoantibodies to TAAs as potential biomarkers for lung cancer. (continued)

Reference	Antibodies to TAAs	N	Remarks subjects	Stage (%)					Sensitivity (%)	Specificity (%)	AUC	Method
				I	II	III	IV	Tx				
Pedchenko et al.[90]	6 scFv panel IgM	43	NSCLC, SC	86	14	0	0	-	80	87	0.88	FMAT
Ye et al.	Anti-CD25	486	NSCLC, SC	45	18	33	4	-	35	90	0.70	ELISA
Chapman et al.[91]	6 AABs panel**	776	LC, SC	-	-	-	-	100	40	82	NA	LLC-LIMS
	7 AABs panel***	836	LC, SC	-	-	-	-	100	47	90	NA	
Lam et al.[92]	p53, NY-ESO1, CAGE, GBU4-5, Annexin I, SOX2**	1,254	LC, SC	60 (ES)		26 (LS)		14	34	87	NA	ELISA
Boyle et al.[93]	p53, NY-ESO-1, CAGE, GBU4-5, Annexin I, SOX2**	481	D, LC, SC	12 (ES)		70 (LS)		18	39	89	0.63	ELISA
		538	V, LC, SC	63 (ES)		10 (LS)		27	37	90	0.64	
Rom et al.[94]	c-myc, Cyclin A, Cyclin B1, Cyclin D1 CDK2, survivin	194	LC, SC	-	-	-	-	100	81	97	0.91	ELISA
Farlow et al.[95]	IMPDH, Ubiquillin, phosphoglycerate mutase, Annexin I, Annexin II, HSP70-9B	196	NSCLC, SC	66	13	18	3	-	95	91	0.97	Luminex Immunoassay

Table 3. Characteristics and performance of blood-based autoantibodies to TAAs as potential biomarkers for lung cancer. (continued)

Reference	Antibodies to TAAs	N	Remarks subjects	Stage (%)					Sensitivity (%)	Specificity (%)	AUC	Method
				I	II	III	IV	Tx				
Yao et al.[96]	DKK1	180	NSCLC, HC	-	-	-	-	100	62	84	NA	ELISA
Wu et al.[97]	OLFM1	180	NSCLC, HC	-	-	-	-	100	92	92	0.96	Phage display + ELISA
Leidinger et al.[98]	1827 peptide clones	127	LC, HC	47	4	38	4	7	98	97	0.81	Phage-display
Murray et al.[99]	p53, NY-ESO-1, CAGE, GBU4-5, Annexin I, SOX2**	481	LC, SC	12 (ES)		70 (LS)		18	34	91	NA	ELISA
Qiu et al.[100]	Annexin I, LAMR1, 14-3-3 theta,	170	NSCLC, SC	-	-	-	-	100[b]	51	82	0.73	Protein micro array
Leidinger et al.[101]	20 peptide clones	79	NSCLC, HC	46	31	15	3	5	93	93	0.98	Phage-display
Chapman et al.[73]	p53, NY-ESO-1, CAGE GBU4-5, c-myc, HER2, MUC1	154	LC, HC	5	4	15	37	39	76	92	NA	ELISA
Pereire-Faca et al.[102]	14-3-3 theta,	37	LC, SC	-	-	-	-	100[b]	55	95	0.84	Western blot +

Table 3. Characteristics and performance of blood-based autoantibodies to TAAs as potential biomarkers for lung cancer. (continued)

Reference	Antibodies to TAAs	N	Remarks subjects	Stage (%)					Sensitivity (%)	Specificity (%)	AUC	Method
				I	II	III	IV	Tx				
	Annexin I, PGP 9.5											protein microarray
Zhong et al.[72]	Paxillin, SEC15L2,	46	D, NSCLC, SC	100	0	0	0	-	91	91	0.99	Phage-display
	BAC cloneRP-11-499F19, XRCC5, MALATI1	102	V, NSCLC, SC	13[c]/68	15[c]	13[c]	4[c]	-	80[c]/83	88	NA	
Yagihashi et al.[103]	Survivin, Livin	38	D, LC, HC	11	5	30	54	-	71	100[a]	NA	ELISA

Note: Data are listed by most recent publication first (2015-2005).

I, stage I; II, stage II; III, stage III; IV, stage IV; AABs, tumor-associated autoantibodies; AAH, atypical adenomatous hyperplasia (pre-neoplastic adenocarcinoma); AUC, area under the curve; D, discovery set; ELISA, enzyme-linked immuno sorbent assay; ES, early stage: NSCLC I, II and limited SCLC; FMAT, fluorometric microvolume assay technology; HC, healthy controls; LC, lung cancer (NSCLC and SCLC); LLC-LIMS, *Early*CDT-Lung® test based on ELISA; LS, late stage: NSCLC III, IV and extensive SCLC; N, patient number; NA, not applicable; SC, smoking controls; SCD, squamous cell carcinoma dysplasia (pre-neoplastic squamous cell carcinoma); scFv, single chain fragment variable antibodies; TAAs, tumor-associated antigens; Tx, tumor stage unknown (or not described), V, validation set.

*measurement of antigen panel: CEA, CA125, CYFRA 21–1; **6 AABs panel corresponds to AAB positivity to any one of the TAA s: p53, NY-ESO-1, CAGE, GBU4-5, Annexin I, and SOX2 (*Early*CDT-Lung); ***7 AABs panel: p53, NY-ESO-1, CAGE, GBU4-5, SOX2, HuD, and MAGE A4 (new *Early*CDT-Lung).

[a] pre-neoplastic samples; [b] preclinical samples within 1 year before diagnosis; [c] preclinical samples 1-5 years before diagnosis.

OUTLINE OF THIS THESIS

High sensitivity and specificity of CT screening can only be realized after follow-up CT examinations to assess nodule growth at different time points, as a consequence it delays lung cancer diagnosis. This delay in diagnosis of lung cancer may even take up to one year after baseline screening. Therefore, additional biomarkers to CT screening are needed to reduce the false-positive and false-negative results at baseline screening. The aim of this thesis was to find lung cancer related proteins, especially sequences of autoantibodies, that can differentiate early stage lung cancer patients from healthy individuals at high-risk in a well-controlled multicenter population study, stratified for smoking (NELSON). We applied immunological and high-performance proteomics techniques to identify and quantify these proteins.

The first two chapters of this thesis describe the development of immunomics methods to identify similar or identical CDR sequences of autoantibodies that can distinguish early stage lung cancer patients from matched controls with high sensitivity and specificity. Detection and identification of CDRs can significantly be improved by reduction of the complexity of the immunoglobulin molecule. In **Chapter 2** we describe molecular dissection of IgG into Fab-κ, Fab-λ, κ and λ fragments to reduce the complexity of this molecule for mass spectrometry measurement. We compared the number of CDRs identified in these immunoglobulin fragments of lung cancer cases and controls with the number of CDRs identified in the Fab fragments. In **Chapter 3** we apply our IgG Fab purification method on a case-control study. The aim of the study was to find sequences in the Fab of immunoglobulins that are shared among early stage lung cancer patients using proteomics techniques without the need of prior knowledge about the antigens involved.
Case-control studies suggested that autoantibodies to survivin protein are potential biomarkers for early diagnosis. In **Chapter 4** we test the hypothesis that sandwich ELISA can detect autoantibodies to survivin before radiologic diagnosis in lung cancer patients from the NELSON trial. **Chapter 5** describes a validation study of a 13-protein and 5-protein blood-based classifier, which has been described in literature as a diagnostic tool to distinguish benign from early-stage malignant nodules in patients with indeterminate lung nodules. In analogy, we used immunodepletion on IgY14-Supermix resin columns and MRM-MS analysis with stable isotope-labeled internal standard (SIS) peptides to analyze the classifier proteins. Finally, the study results of this thesis are summarized and discussed.

REFERENCES

1. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer, 2015.
2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J and Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin*. 2015; 65: 87-108.
3. Alberg AJ, Brock MV, Ford JG, Samet JM and Spivack SD. Epidemiology of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest*. 2013; 143: e1S-29S.
4. Pass HI, Carbone DP, Johnson DH, Minna JD, Scagliotti GV and A.T. T. *Principles and practice of lung cancer. The official reference text of the IASLC*. 4th ed.: Lippincott Williams & Wilkins, 2010.
5. Youlden DR, Cramb SM and Baade PD. The International Epidemiology of Lung Cancer: geographical distribution and secular trends. *J Thorac Oncol*. 2008; 3: 819-31.
6. Hamra GB, Laden F, Cohen AJ, Raaschou-Nielsen O, Brauer M and Loomis D. Lung Cancer and Exposure to Nitrogen Dioxide and Traffic: A Systematic Review and Meta-Analysis. *Environ Health Perspect*. 2015.
7. Alberg AJ and Samet JM. Epidemiology of lung cancer. *Chest*. 2003; 123: 21S-49S.
8. Travis WD, Brambilla E, Müller-Hermelink HK, Harris CC and Biernat W. *Pathology and Genetics of Tumours of the Lung, Pleura,Thymus and Heart*. IARCPress, 2004, p.344.
9. Society AC. Lung Cancer | American Cancer Society. 2015.
10. DeSantis CE, Lin CC, Mariotto AB, Siegel RL, Stein KD, Kramer JL, et al. Cancer treatment and survivorship statistics, 2014. *CA Cancer J Clin*. 2014; 64: 252-71.
11. Alberg AJ and Nonemaker J. Who is at high risk for lung cancer? Population-level and individual-level perspectives. *Semin Respir Crit Care Med*. 2008; 29: 223-32.
12. Davidson MR, Gazdar AF and Clarke BE. The pivotal role of pathology in the management of lung cancer. *Journal of thoracic disease*. 2013; 5 Suppl 5: S463-78.
13. Radzikowska E, Glaz P and Roszkowski K. Lung cancer in women: age, smoking, histology, performance status, stage, initial treatment and survival. Population-based study of 20 561 cases. *Ann Oncol*. 2002; 13: 1087-93.
14. Papadopoulos A, Guida F, Leffondre K, Cenee S, Cyr D, Schmaus A, et al. Heavy smoking and lung cancer: are women at higher risk? Result of the ICARE study. *Br J Cancer*. 2014; 110: 1385-91.
15. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JH, Beasley MB, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J Thorac Oncol*. 2015; 10: 1243-60.
16. Kalemkerian GP, Akerley W, Bogner P, Borghaei H, Chow L, Downey RJ, et al. Small cell lung cancer. *Journal of the National Comprehensive Cancer Network : JNCCN*. 2011; 9: 1086-113.
17. 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: 9-15.

18. Blanchon T, Brechot JM, Grenier PA, Ferretti GR, Lemarie E, Milleron B, et al. Baseline results of the Depiscan study: a French randomized pilot trial of lung cancer screening comparing low dose CT scan (LDCT) and chest X-ray (CXR). *Lung Cancer*. 2007; 58: 50-8.

19. 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: 137-45.

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: 706-14.

21. Henschke CI and International Early Lung Cancer Action Program I. Survival of patients with clinical stage I lung cancer diagnosed by computed tomography screening for lung cancer. *Clin Cancer Res*. 2007; 13: 4949-50.

22. 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. *Ann Intern Med*. 2013; 159: 411-20.

23. 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. *Am J Respir Crit Care Med*. 2009; 180: 445-53.

24. Infante M, Cavuto S, Lutman FR, Passera E, Chiarenza M, Chiesa G, et al. Long-Term Follow-up Results of the DANTE Trial, a Randomized Study of Lung Cancer Screening with Spiral Computed Tomography. *Am J Respir Crit Care Med*. 2015; 191: 1166-75.

25. 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: 296-301.

26. 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. *European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation*. 2012; 21: 308-15.

27. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011; 365: 395-409.

28. Aberle DR, DeMello S, Berg CD, Black WC, Brewer B, Church TR, et al. Results of the two incidence screenings in the National Lung Screening Trial. *N Engl J Med*. 2013; 369: 920-31.

29. Jaklitsch MT, Jacobson FL, Austin JH, Field JK, Jett JR, Keshavjee S, et al. The American Association for Thoracic Surgery guidelines for lung cancer screening using low-dose computed tomography scans for lung cancer survivors and other high-risk groups. *J Thorac Cardiovasc Surg*. 2012; 144: 33-8.

30. 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: 2418-29.

31. Wender R, Fontham ET, Barrera E, Jr., Colditz GA, Church TR, Ettinger DS, et al. American Cancer Society lung cancer screening guidelines. *CA Cancer J Clin*. 2013; 63: 107-17.

32. Roberts H, Walker-Dilks C, Sivjee K, Ung Y, Yasufuku K, Hey A, et al. Screening high-risk populations for lung cancer: guideline recommendations. *J Thorac Oncol*. 2013; 8: 1232-7.

33. Couraud S, Cortot AB, Greillier L, Gounant V, Mennecier B, Girard N, et al. From randomized trials to the clinic: is it time to implement individual lung-cancer screening in clinical practice? A multidisciplinary statement from French experts on behalf of the French intergroup (IFCT) and the groupe d'Oncologie de langue francaise (GOLF). *Ann Oncol*. 2013; 24: 586-97.

34. Moyer VA and Force USPST. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2014; 160: 330-8.

35. Infante M, Lutman FR, Cavuto S, Brambilla G, Chiesa G, Passera E, et al. Lung cancer screening with spiral CT: baseline results of the randomized DANTE trial. *Lung Cancer*. 2008; 59: 355-63.

36. Lopes Pegna A, Picozzi G, Mascalchi M, Maria Carozzi F, Carrozzi L, Comin C, et al. Design, recruitment and baseline results of the ITALUNG trial for lung cancer screening with low-dose CT. *Lung Cancer*. 2009; 64: 34-40.

37. 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). *Int J Cancer*. 2007; 120: 868-74.

38. van Klaveren RJ, Oudkerk M, Prokop M, Scholten ET, Nackaerts K, Vernhout R, et al. Management of lung nodules detected by volume CT scanning. *N Engl J Med*. 2009; 361: 2221-9.

39. Horeweg N, van der Aalst CM, Vliegenthart R, Zhao Y, Xie X, Scholten ET, et al. Volumetric computed tomography screening for lung cancer: three rounds of the NELSON trial. *Eur Respir J*. 2013; 42: 1659-67.

40. Pedersen JH, Ashraf H, Dirksen A, Bach K, Hansen H, Toennesen P, et al. The Danish randomized lung cancer CT screening trial--overall design and results of the prevalence round. *J Thorac Oncol*. 2009; 4: 608-14.

41. Becker N, Motsch E, Gross ML, Eigentopf A, Heussel CP, Dienemann H, et al. Randomized study on early detection of lung cancer with MSCT in Germany: study design and results of the first screening round. *J Cancer Res Clin Oncol*. 2012; 138: 1475-86.

42. Becker N, Motsch E, Gross ML, Eigentopf A, Heussel CP, Dienemann H, et al. Randomized Study on Early Detection of Lung Cancer with MSCT in Germany: Results of the First 3 Years of Follow-up After Randomization. *J Thorac Oncol*. 2015; 10: 890-6.

43. Baldwin DR, Duffy SW, Wald NJ, Page R, Hansell DM and Field JK. UK Lung Screen (UKLS) nodule management protocol: modelling of a single screen randomised controlled trial of low-dose CT screening for lung cancer. *Thorax*. 2011; 66: 308-13.

44. Field JK, Duffy SW, Baldwin DR, Whynes DK, Devaraj A, Brain KE, et al. UK Lung Cancer RCT Pilot Screening Trial: baseline findings from the screening arm provide evidence for the potential implementation of lung cancer screening. *Thorax*. 2016; 71: 161-70.

45. Horeweg N, van der Aalst CM, Thunnissen E, Nackaerts K, Weenink C, Groen HJ, et al. Characteristics of lung cancers detected by computer tomography screening in the randomized NELSON trial. *Am J Respir Crit Care Med*. 2013; 187: 848-54.

46. Xu DM, Gietema H, de Koning H, Vernhout R, Nackaerts K, Prokop M, et al. Nodule management protocol of the NELSON randomised lung cancer screening trial. *Lung Cancer*. 2006; 54: 177-84.

47. de Koning HJ, Meza R, Plevritis SK, ten Haaf K, Munshi VN, Jeon J, et al. Benefits and harms of computed tomography lung cancer screening strategies: a comparative modeling study for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2014; 160: 311-20.

48. Hanahan D and Weinberg RA. The hallmarks of cancer. *Cell*. 2000; 100: 57-70.

49. Wistuba, II and Gazdar AF. Lung cancer preneoplasia. *Annu Rev Pathol*. 2006; 1: 331-48.

50. Larsen JE and Minna JD. Molecular biology of lung cancer: clinical implications. *Clinics in chest medicine*. 2011; 32: 703-40.

51. Li XJ, Hayward C, Fong PY, Dominguez M, Hunsucker SW, Lee LW, et al. A blood-based proteomic classifier for the molecular characterization of pulmonary nodules. *Sci Transl Med*. 2013; 5: 207ra142.

52. Patz EF, Jr., Campa MJ, Gottlin EB, Trotter PR, Herndon JE, 2nd, Kafader D, et al. Biomarkers to help guide management of patients with pulmonary nodules. *Am J Respir Crit Care Med*. 2013; 188: 461-5.

53. Pecot CV, Li M, Zhang XJ, Rajanbabu R, Calitri C, Bungum A, et al. Added value of a serum proteomic signature in the diagnostic evaluation of lung nodules. *Cancer Epidemiol Biomarkers Prev*. 2012; 21: 786-92.

54. Bigbee WL, Gopalakrishnan V, Weissfeld JL, Wilson DO, Dacic S, Lokshin AE, et al. A multiplexed serum biomarker immunoassay panel discriminates clinical lung cancer patients from high-risk individuals found to be cancer-free by CT screening. *J Thorac Oncol*. 2012; 7: 698-708.

55. Diamandis EP, Goodglick L, Planque C and Thornquist MD. Pentraxin-3 is a novel biomarker of lung carcinoma. *Clin Cancer Res*. 2011; 17: 2395-9.

56. Takano A, Ishikawa N, Nishino R, Masuda K, Yasui W, Inai K, et al. Identification of nectin-4 oncoprotein as a diagnostic and therapeutic target for lung cancer. *Cancer Res*. 2009; 69: 6694-703.

57. Ostroff RM, Bigbee WL, Franklin W, Gold L, Mehan M, Miller YE, et al. Unlocking biomarker discovery: large scale application of aptamer proteomic technology for early detection of lung cancer. *PLoS One*. 2010; 5: e15003.

58. Patz EF, Jr., Campa MJ, Gottlin EB, Kusmartseva I, Guan XR and Herndon JE, 2nd. Panel of serum biomarkers for the diagnosis of lung cancer. *J Clin Oncol*. 2007; 25: 5578-83.

59. Yildiz PB, Shyr Y, Rahman JS, Wardwell NR, Zimmerman LJ, Shakhtour B, et al. Diagnostic accuracy of MALDI mass spectrometric analysis of unfractionated serum in lung cancer. *J Thorac Oncol*. 2007; 2: 893-901.

60. Gao WM, Kuick R, Orchekowski RP, Misek DE, Qiu J, Greenberg AK, et al. Distinctive serum protein profiles involving abundant proteins in lung cancer patients based upon antibody microarray analysis. *BMC Cancer*. 2005; 5: 110.

61. Molina R, Filella X, Auge JM, Fuentes R, Bover I, Rifa J, et al. Tumor markers (CEA, CA 125, CYFRA 21-1, SCC and NSE) in patients with non-small cell lung cancer as an aid in histological diagnosis and prognosis. Comparison with the main clinical and pathological prognostic factors. *Tumour Biol*. 2003; 24: 209-18.

62. Barlesi F, Gimenez C, Torre JP, Doddoli C, Mancini J, Greillier L, et al. Prognostic value of combination of Cyfra 21-1, CEA and NSE in patients with advanced non-small cell lung cancer. *Respir Med*. 2004; 98: 357-62.

63. Anderson N and Anderson N. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics*. 2002; 1: 845 - 67.

64. Anderson KS and LaBaer J. The sentinel within: exploiting the immune system for cancer biomarkers. *J Proteome Res*. 2005; 4: 1123-33.

65. Qiu J and Hanash S. Autoantibody profiling for cancer detection. *Clin Lab Med*. 2009; 29: 31-46.

66. Dunn GP, Old LJ and Schreiber RD. The immunobiology of cancer immunosurveillance and immunoediting. *Immunity*. 2004; 21: 137-48.

67. Finn OJ. Cancer immunology. *N Engl J Med*. 2008; 358: 2704-15.

68. Caron M, Choquet-Kastylevsky G and Joubert-Caron R. Cancer immunomics using autoantibody signatures for biomarker discovery. *Mol Cell Proteomics*. 2007; 6: 1115-22.

69. Backes C, Ludwig N, Leidinger P, Harz C, Hoffmann J, Keller A, et al. Immunogenicity of autoantigens. *BMC Genomics*. 2011; 12: 340.

70. Hanash S. Harnessing immunity for cancer marker discovery. *Nat Biotechnol*. 2003; 21: 37-8.

71. Tan HT, Low J, Lim SG and Chung MC. Serum autoantibodies as biomarkers for early cancer detection. *The FEBS journal*. 2009; 276: 6880-904.

72. Zhong L, Coe SP, Stromberg AJ, Khattar NH, Jett JR and Hirschowitz EA. Profiling tumor-associated antibodies for early detection of non-small cell lung cancer. *J Thorac Oncol*. 2006; 1: 513-9.

73. Chapman CJ, Murray A, McElveen JE, Sahin U, Luxemburger U, Tureci O, et al. Autoantibodies in lung cancer: possibilities for early detection and subsequent cure. *Thorax*. 2008; 63: 228-33.

74. Finn OJ. Immune response as a biomarker for cancer detection and a lot more. *N Engl J Med*. 2005; 353: 1288-90.

75. Tonegawa S. Reiteration frequency of immunoglobulin light chain genes: further evidence for somatic generation of antibody diversity. *Proc Natl Acad Sci U S A*. 1976; 73: 203-7.

76. de Wildt RM, van Venrooij WJ, Winter G, Hoet RM and Tomlinson IM. Somatic insertions and deletions shape the human antibody repertoire. *J Mol Biol*. 1999; 294: 701-10.

77. Meffre E, Catalan N, Seltz F, Fischer A, Nussenzweig MC and Durandy A. Somatic hypermutation shapes the antibody repertoire of memory B cells in humans. *J Exp Med*. 2001; 194: 375-8.

78. Murphy K. TP, Walport M. *Janeway's immunobiology*. 7th ed.: Garland Science, 2008.

79. Schroeder HW, Jr. and Cavacini L. Structure and function of immunoglobulins. *J Allergy Clin Immunol*. 2010; 125: S41-52.

80. Xu JL and Davis MM. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity*. 2000; 13: 37-45.

81. Saada R, Weinberger M, Shahaf G and Mehr R. Models for antigen receptor gene rearrangement: CDR3 length. *Immunol Cell Biol*. 2007; 85: 323-32.

82. Baranzini SE, Jeong MC, Butunoi C, Murray RS, Bernard CC and Oksenberg JR. B cell repertoire diversity and clonal expansion in multiple sclerosis brain lesions. *J Immunol*. 1999; 163: 5133-44.

83. Andersen PS, Haahr-Hansen M, Coljee VW, Hinnerfeldt FR, Varming K, Bregenholt S, et al. Extensive restrictions in the VH sequence usage of the human antibody response against the Rhesus D antigen. *Mol Immunol*. 2007; 44: 412-22.

84. Doseeva V, Colpitts T, Gao G, Woodcock J and Knezevic V. Performance of a multiplexed dual analyte immunoassay for the early detection of non-small cell lung cancer. *J Transl Med*. 2015; 13: 55.

85. Jett JR, Peek LJ, Fredericks L, Jewell W, Pingleton WW and Robertson JF. Audit of the autoantibody test, EarlyCDT(R)-lung, in 1600 patients: an evaluation of its performance in routine clinical practice. *Lung Cancer*. 2014; 83: 51-5.

86. Jia J, Wang W, Meng W, Ding M, Ma S and Wang X. Development of a multiplex autoantibody test for detection of lung cancer. *PLoS One*. 2014; 9: e95444.

87. Wang W, Guan S, Sun S, Jin Y, Lee KH, Chen Y, et al. Detection of circulating antibodies to linear peptide antigens derived from ANXA1 and DDX53 in lung cancer. *Tumour Biol*. 2014; 35: 4901-5.

88. Lowe FJ, Shen W, Zu J, Li J, Wang H, Zhang X, et al. A novel autoantibody test for the detection of pre-neoplastic lung lesions. *Mol Cancer*. 2014; 13: 78.

89. Zhang C, Ye L, Guan S, Jin S, Wang W, Sun S, et al. Autoantibodies against p16 protein-derived peptides may be a potential biomarker for non-small cell lung cancer. *Tumour Biol*. 2014; 35: 2047-51.

90. Pedchenko T, Mernaugh R, Parekh D, Li M and Massion PP. Early detection of NSCLC with scFv selected against IgM autoantibody. *PLoS One*. 2013; 8: e60934.

91. Chapman CJ, Healey GF, Murray A, Boyle P, Robertson C, Peek LJ, et al. EarlyCDT(R)-Lung test: improved clinical utility through additional autoantibody assays. *Tumour Biol*. 2012; 33: 1319-26.

92. Lam S, Boyle P, Healey GF, Maddison P, Peek L, Murray A, et al. EarlyCDT-Lung: an immunobiomarker test as an aid to early detection of lung cancer. *Cancer Prev Res (Phila)*. 2011; 4: 1126-34.

93. Boyle P, Chapman CJ, Holdenrieder S, Murray A, Robertson C, Wood WC, et al. Clinical validation of an autoantibody test for lung cancer. *Ann Oncol*. 2011; 22: 383-9.

94. Rom WN, Goldberg JD, Addrizzo-Harris D, Watson HN, Khilkin M, Greenberg AK, et al. Identification of an autoantibody panel to separate lung cancer from smokers and nonsmokers. *BMC Cancer*. 2010; 10: 234.

95. Farlow EC, Patel K, Basu S, Lee BS, Kim AW, Coon JS, et al. Development of a multiplexed tumor-associated autoantibody-based blood test for the detection of non-small cell lung cancer. *Clin Cancer Res*. 2010; 16: 3452-62.

96. Yao X, Jiang H, Zhang C, Wang H, Yang L, Yu Y, et al. Dickkopf-1 autoantibody is a novel serological biomarker for non-small cell lung cancer. *Biomarkers*. 2010; 15: 128-34.

97. Wu L, Chang W, Zhao J, Yu Y, Tan X, Su T, et al. Development of autoantibody signatures as novel diagnostic biomarkers of non-small cell lung cancer. *Clin Cancer Res*. 2010; 16: 3760-8.

98. Leidinger P, Keller A, Heisel S, Ludwig N, Rheinheimer S, Klein V, et al. Identification of lung cancer with high sensitivity and specificity by blood testing. *Respir Res*. 2010; 11: 18.

99. Murray A, Chapman CJ, Healey G, Peek LJ, Parsons G, Baldwin D, et al. Technical validation of an autoantibody test for lung cancer. *Ann Oncol*. 2010; 21: 1687-93.

100. Qiu J, Choi G, Li L, Wang H, Pitteri SJ, Pereira-Faca SR, et al. Occurrence of autoantibodies to annexin I, 14-3-3 theta and LAMR1 in prediagnostic lung cancer sera. *J Clin Oncol*. 2008; 26: 5060-6.

101. Leidinger P, Keller A, Ludwig N, Rheinheimer S, Hamacher J, Huwer H, et al. Toward an early diagnosis of lung cancer: an autoantibody signature for squamous cell lung carcinoma. *Int J Cancer*. 2008; 123: 1631-6.

102. Pereira-Faca SR, Kuick R, Puravs E, Zhang Q, Krasnoselsky AL, Phanstiel D, et al. Identification of 14-3-3 theta as an antigen that induces a humoral response in lung cancer. *Cancer Res*. 2007; 67: 12000-6.

103. Yagihashi A, Asanuma K, Kobayashi D, Tsuji N, Shijubo Y, Abe S, et al. Detection of autoantibodies to livin and survivin in Sera from lung cancer patients. *Lung Cancer*. 2005; 48: 217-21.

104. Foreman AL, Lemercier B, Lim A, Kourlisky P, Kenny T, Gershwin ME, et al. VH gene usage and CDR3 analysis of B cell receptor in the peripheral blood of patients with PBC. *Autoimmunity*. 2008; 41: 80-6.

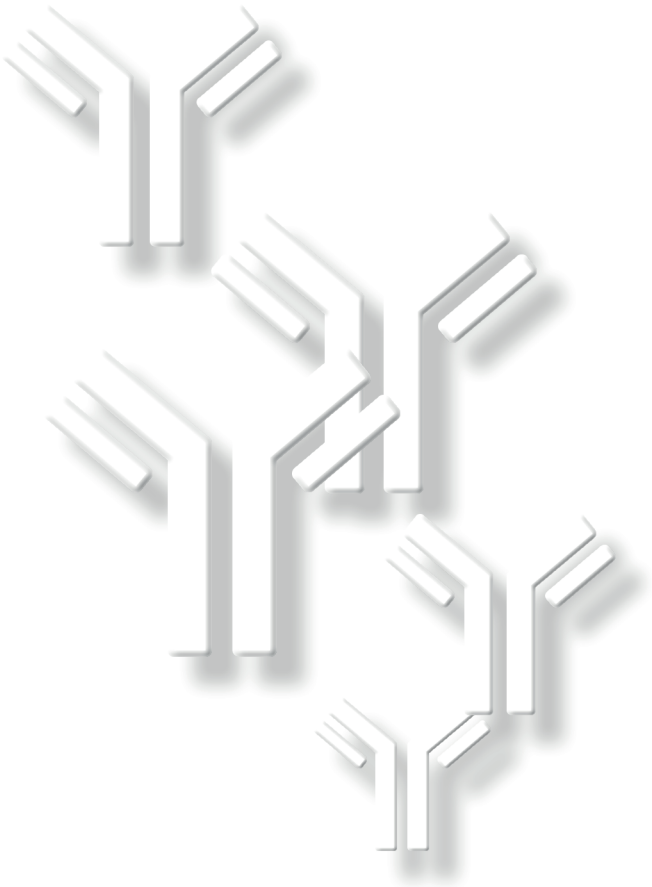
105. Weinstein JA, Jiang N, White RA, 3rd, Fisher DS and Quake SR. High-throughput sequencing of the zebrafish antibody repertoire. *Science*. 2009; 324: 807-10.

106. VanDuijn MM, Dekker LJ, Zeneyedpour L, Smitt PA and Luider TM. Immune responses are characterized by specific shared immunoglobulin peptides that can be detected by proteomic techniques. *J Biol Chem*. 2010; 285: 29247-53.

107. Scheid JF, Mouquet H, Ueberheide B, Diskin R, Klein F, Oliveira TY, et al. Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science*. 2011; 333: 1633-7.

108. Maat P, Vanduijn M, Brouwer E, Dekker L, Zeneyedpour L, Luider T, et al. Mass spectrometric detection of antigen-specific immunoglobulin peptides in paraneoplastic patient sera. *J Autoimmun*. 2012.

109. Yousaf-Khan U, van der Aalst C, de Jong PA, Heuvelmans M, Scholten E, Lammers JW, et al. Final screening round of the NELSON lung cancer screening trial: the effect of a 2.5-year screening interval. *Thorax*. 2016.



Ingrid Broodman, Dominique de Costa, Christoph Stingl,
Lennard J.M. Dekker, Martijn M. VanDuijn, Jan Lindemans,
Rob J. van Klaveren and Theo M. Luider

Mass spectrometry analyses of kappa and lambda fractions result in increased number of complementarity-determining regions identifications

CHAPTER 2

Proteomics 2012; 12: 183–191

ABSTRACT

Sera from lung cancer patients contain antibodies against tumor-associated antigens. Specific amino acid sequences of the complementarity determining regions (CDRs) in the antigen-binding fragment (Fab) of these antibodies have potential as lung cancer biomarkers. Detection and identification of CDRs by mass spectrometry can significantly be improved by reduction of the complexity of the immunoglobulin molecule. Our aim was to molecular dissect IgG into kappa and lambda fragments to reduce the complexity and thereby identify substantially more CDRs than by just total Fab isolation. We purified Fab, Fab-κ, Fab-λ, κ and λ light chains from serum from 10 stage I lung adenocarcinoma patients and 10 matched controls from current and former smokers. After purification, the immunoglobulin fragments were enzymatically digested and measured by high-resolution mass spectrometry. Finally, we compared the number of CDRs identified in these immunoglobulin fragments with that in the Fab fragments. Twice as many CDRs were identified when Fab-κ, Fab-λ, κ and λ (3330) were combined than in the Fab fraction (1663) alone. The number of CDRs and κ:λ ratio was statistically similar in both cases and controls. Molecular dissection of IgG identifies significantly more CDRs, which increases the likelihood of finding lung cancer-related CDR sequences.

INTRODUCTION

Only 15-20% of all lung cancers are detected at an early and potential curable stage today.[1] An early detection and treatment of lung cancer can reduce the high lung cancer mortality rate. This is currently investigated in several randomized lung cancer CT screening trials.[2-5] At the moment, there is no early detection biomarker available for lung cancer. Biomarkers could be used to stratify people according to their risk to develop lung cancer. The different strata could, dependent on their cancer risk, be invited for baseline CT screening and for subsequent screening rounds. A biomarker for early detection of lung cancer could be used as a complement to CT screening in order to reduce the rate of false-positive test results and the number of unnecessary biopsies, surgical interventions or serial CT scans.[6]

There is increasing evidence that during tumor development a humoral immune response evolves into various tumor types, including lung cancer.[7-9] Immunoglobulins against different tumor-associated antigens (TAAs) in lung cancer have been identified by different strategies[10-15] up to 5 years before the tumor was detectable by a CT scan.[16, 17] These strategies use immunoglobulins to identify the targeted tumor antigens as potential biomarkers, rather than using the reactive immunoglobulins as potential biomarkers. In contrast to antigens, immunoglobulins are excreted and circulate in the blood at relatively high levels, which support their detection.

We previously described a new approach in which tryptic fragments of the immunoglobulins themselves are used as potential biomarkers.[18] Three hypervariable complementarity determining regions (CDR1, CDR2 and CDR3) in the variable regions of the light and heavy chains of an immunoglobulin form the binding surface complementary to the antigen. As such, these CDRs determine the specificity of the immunoglobulin to the antigen. During immune response and B-cell development, CDRs are generated by somatic rearrangements of different (V, or V, D and J) germline genes to form a specific combination. In both light and heavy chains, the diversity of CDR3 is even further enhanced by the insertions and deletions of nucleotides between the genes. The estimated potential immunoglobulin diversity varies from 10^{13} to more than 10^{50}.[19, 20] Despite this large range there is evidence for repertoire bias, which means that certain germline genes are preferentially used in response to a particular antigen.[21, 22] Moreover, similar and identical CDR3 sequences have been found in humans and in zebrafish, respectively.[23, 24] Our hypothesis is that a specific molecular profile of CDRs may distinguish lung cancer patients from controls and can thus be used as lung cancer biomarker.

The ability to find differences in CDRs between lung cancer cases and controls depends on the number of CDRs identified, which in turn depends on the accuracy, resolution, sensitivity and reproducibility of the mass spectrometry (MS) to identify these very low-abundant CDR peptides. However, ion suppression in the mass spectrometer especially for complex peptide mixtures can reduce the sensitivity.[25] Reduction of this complexity reduces ion suppression

and leads to a significantly higher sensitivity to detect CDR peptides. In our previous paper, we presented our method to sequence Fab fragments by using mass spectrometry.[26] To identify as many CDRs as possible, the complexity of the immunoglobulin molecule can be reduced by separating Fab into Fab-κ and Fab-λ, and even further by purifying only the kappa (κ) or lambda (λ) light chain. The normal overall κ: λ ratio in human immunoglobulins is approximately 2 (κ:λ of: IgG 2.34 ± 0.80; IgA 1.59 ± 0.40; IgM 1.86 ± 0.76) with most of the immunoglobulins consisting of IgG.[27]

Our aim was to use molecular dissection of IgG into kappa and lambda fragments to identify substantially more CDRs than obtained by the Fab method. To determine if we would be able to identify more CDRs by molecular dissection of IgG in kappa and lambda fragments than of Fab, we designed a pilot study. In this study, we purified Fab, Fab-κ, Fab-λ, κ and λ light chains from serum from 10 stage I lung adenocarcinoma patients and 10 matched controls from current and former smokers of the NELSON study.[4] After purification, the immunoglobulin fragments were enzymatically digested by trypsin and measured by high-resolution mass spectrometry. Finally, we compared the number of CDRs identified in these immunoglobulin fragments with the number of CDRs identified in the Fab fragments.

MATERIALS AND METHODS

Cases and controls from the NELSON trial

Sera from 20 current and former smokers were obtained from the Dutch-Belgian randomized lung cancer screening trial (NELSON), as described previously,[4] and collected under uniform conditions. The subjects were between 53 and 73 years of age (50% males and 50% females, median age 61 years). Ten serum samples of stage I lung adenocarcinoma patients without history of other cancer were collected. As non-cancer controls, 10 matched serum samples were taken from participants in the same trial. The controls were matched for gender, smoking status, COPD status, absence of previous cancer and asbestos history. All participants gave written informed consent as approved by the Dutch Minister of Health and the ethics board at each participating center. Samples were blinded and analyzed in random order.

Serum-collection protocol

During the participants' visits to the center, one serum gel tube was collected per participant. The venous blood was allowed to clot, and was centrifuged for 10 minutes at 1400 x g and 4°C within 2 hours after collection. After centrifugation, the serum was stored immediately in aliquots at -80°C. All samples were blinded and analyzed in random order.

Reference sample

One reference donor sample (male: 59 years), with a normal serum IgG of 9.75 g/L, was used as a quality control for each analysis step.[26] In accordance with the general guidelines of the Sanquin Blood Bank Rotterdam (the Netherlands), the healthy donor gave written consent for the serum to be used for scientific research.

Purification of IgG

Serum IgG (80 µL) was purified using the Melon Gel IgG purification kit (Pierce, Rockford, IL), according to the manufacturer's instructions. The concentration of the purified IgG protein (800 µL) was determined by means of the mass extinction coefficient of 1.37 (mg/mL) cm^{-1} at 280 nm on a NanoDrop Spectrophotometer (ND-1000, Nanodrop Technologies, Wilmington, DE).

Purification of Fab

After purification, purified IgG (400 µL) was digested in Fab and Fc fragments overnight by immobilized papain on agarose beads according to the manufacturer's instructions (Pierce, Rockford, IL). Then this digest (2.800 mL) was concentrated approximately ten times by an Amicon Ultra 3K centrifugal filter device (Millipore, Amsterdam, the Netherlands).
Finally, the Fab fragments were separated from Fc fragments and undigested IgG by SDS-PAGE under reducing conditions.[26] The proteins were fixed and visualized with the Colloidal Blue staining kit (Invitrogen, Breda, the Netherlands) and gels were washed for a minimum of 4 h in deionized water (Figure 1).

Purification of IgG Fab-κ and IgG Fab-λ

For purification of IgG Fab-κ 100 µL λ-specific-anti-human IgG and for purification of IgG Fab-λ 200 µL κ-specific-anti-human IgG was immobilized onto the MicroLink Protein support of two AminoLink Plus coupling gel spin column (Pierce, Rockford, IL). Concentrated papain-digested IgG (100 µL) was loaded onto the columns and incubated at 4°C overnight with gentle end-over-end mixing. Finally, the IgG Fab-κ and IgG Fab-λ were individually collected in the flow-throughs and separated from the Fc proteins by SDS-PAGE under reducing conditions and visualized as described above (Figure 1).

Purification of κ and λ

The Melon Gel purified IgG (100 µL) was loaded onto the λ-specific-anti-human IgG (100 µL) and κ-specific-anti-human IgG (200 µL) columns and incubated at 4°C overnight with gentle end-over-end mixing. Finally, the IgG-κ and IgG-λ were individually collected in the flow-throughs and the heavy chain (H) and light chain, κ or λ, were separated by SDS-PAGE under reducing conditions and visualized as described above (Figure 1).

Sample preparation for LTQ Orbitrap MS

Recovery and reproducibility of the purifications were determined by densitometry. Intensities of the protein bands of the reference sample on the stained SDS-PAGE gel were quantified by scanning on a Molecular Imager GS-800 Calibrated densitometer (Bio-Rad, Veenendaal, the Netherlands) with Quantity One® 1-D analysis software (version 4.6.5: Bio-Rad). After imaging and analysis of the SDS-PAGE gels, the selected protein bands were excised from the gels and cut into plugs. The in-gel trypsin digestion was performed in Rapigest detergent solution according to the manufacturer's instructions (Waters, Milford, MA, USA).

Nano-LC Orbitrap MS analyses

LCMS measurements of the tryptic peptides were performed on an Ultimate 3000 nano-LC system (Dionex, Amsterdam, the Netherlands) online coupled to a hybrid linear ion trap/Orbitrap MS (LTQ Orbitrap XL: Thermo Fisher Scientific, Bremen, Germany).
For identification of the IgG peptides we used collisional activated dissociation (CAD) fragmentation. High-resolution full-scan MS was obtained in the Orbitrap (resolution 30,000; AGC 1,000,000) and CAD fragmentation was performed on the five most abundant masses in the full-scan spectra.

Data analysis

Progenesis software (Version 2.5: Nonlinear Dynamics, Newcastle, UK) was used for the label-free quantification of MS data. In total five Progenesis analyses were performed, one for each individual dissected IgG fraction. The raw data files were aligned by their retention time, features were selected and intensities were normalized (Nonlinear Dynamics http://www.nonlinear.com/support/progenesis/lc-ms/faq/how-alignment-works.aspx, http://www.nonlinear.com/support/progenesis/lc-ms/faq/how-normalisation-works.aspx).[26] Data matrices containing the feature intensities (area under the peak) were exported for further calculations.
Database searches were performed with Mascot (version 2.2.06: Matrix Science Inc., London, UK) against the NCBInr human database (version nrHuman_database_20090311; *Homo sapiens* species restriction; 2220660 sequences). Parameters used for the database search were as follows: a maximum of two miss cleavages; carbamidomethylation of cysteine as a fixed modification and oxidation of methionine as a variable modification; trypsin as enzyme; a peptide mass tolerance of 10 ppm; a fragment mass tolerance of 0.5 Da; an ion score of 25 as a cut-off.
De novo sequencing was used for features not identified by a Mascot search against the database (NCBInr). Therefore, raw data files were processed by the Peaks Studio 5.1 software package (Bioinformatics Solutions, Waterloo, ON, Canada). The Average Local Confidence score (ALC%) was assigned on the basis of the positional confidence for each amino acid in the peptide sequence divided by the total number of amino acids.

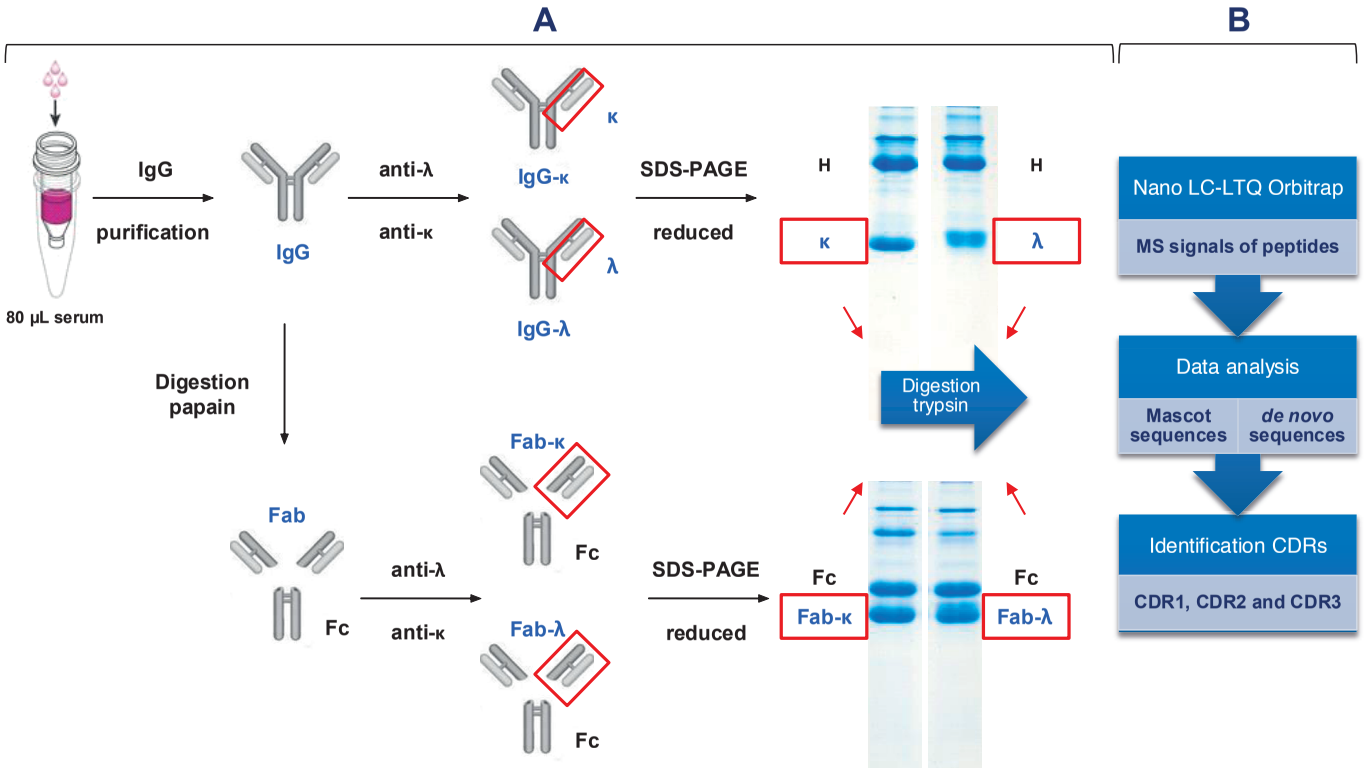


Figure 1. Flow-chart of the different steps in molecular dissection and data analysis. A) Purified IgG and concentrated papain-digested IgG were loaded onto anti-λ IgG and anti-κ IgG columns. Subsequently, IgG-κ, IgG-λ, Fab-κ and Fab-λ were individually collected in the flow-throughs. The heavy chain (H) and light chain (κ or λ) of IgG-κ and IgG-λ, and Fab and Fc of IgG Fab-κ and IgG Fab-λ were separated by SDS-PAGE under reducing conditions. B) After in-gel-tryptic digestion, the peptides were measured by nano-LC-LTQ Orbitrap (MS). MS signals were quantified by Progenesis and analyzed by Mascot and *de novo* sequencing. Mascot and *de novo* peptide sequences were used for the identification of CDR sequences.

Peptide identifications from both Mascot and Peaks were imported into Progenesis, which keeps the best scoring sequence for each MS signal. To this end, Peaks data were manually converted to Mascot XML format for import into Progenesis, and the ALC% scores were divided by a factor of 100. By doing so Mascot scores obtained from the data-dependent search always overruled the ALC% scores. Finally, all intensities and sequences from Mascot and Peaks were combined in a single Progenesis file per individual fraction for further analysis. Mascot and *de novo* peptide sequences were used for the identification of CDR sequences. Irrespective of the protein identification, the BLAST algorithm was subsequently used to align all peptides to databases containing human V, D, J or C-region germline sequences obtained from the IMGT database (IMGT®, the international ImMunoGeneTics information system® http://www.imgt.org). All peptides with a bit score of at least 12.5 were assigned to these germline sequences and selected for further analysis. Peptides aligned to a V-region germline sequence were also aligned using the IMGT/DomainGapAlign tool. This tool positions the peptide to the germline sequence in the IMGT unique residue numbering system and helps to identify the peptide as a framework or CDR in the immunoglobulin molecule. Only peptides with an identity score of at least 70% were assigned to a CDR sequence. Total numbers of CDRs were calculated based on the CDRs found by Mascot and *de novo* sequencing.

Statistical Analysis

Coefficient of variation was used to measure the reproducibility of three replicate purifications of the reference sample for each individual IgG fraction. For each individual IgG fraction and each combination of IgG fractions, descriptive summary statistics (number of measurements (*N*), mean, standard deviation (SD) and Confidence Interval (95% CI)) were provided for the number of CDRs identified in the cases and controls.

The two sample t-test (two-sided) was used to compare differences in the κ:λ ratio in Fab molecules between cases and controls. We used Microsoft Excel 2007 for the descriptive summary statistics and the t-tests. Pearson chi-square (X^2) tests were performed to establish the existence of significant differences between cases and controls in the number of CDRs identified in each specific molecular dissected IgG fraction and each combination of IgG fractions compared with that of Fab. These tests, odds ratios and 95% confidence intervals were calculated by the application of Vassarstats software (http://faculty.vassar.edu/lowry/VassarStats.html). The non-parametric Kruskal-Wallis test (two-sided) was performed to compare the CDR3 ratio in the three Fab fractions (Fab, Fab-κ, Fab-λ) with the CDR3 ratio in the light chain fractions (κ and λ). Analyses were done using STATA, version 11 (Stata, TX, USA).

To determine whether the number of significant different CDRs increases by molecular dissection of IgG, we used the Anova available in the Progenesis program and the two sample t-test (two-sided). By a permutation test that was repeated 20 times we determined the random chance on such an event. Two standard deviations of this permutation test were

used as a threshold to determine whether the number of CDRs identified was significantly different. For all statistical tests a *p*-value <0.01 was considered statistically significant.

RESULTS

Fab, Fab-κ, Fab-λ, κ and λ purification

To calculate the recovery and reproducibility of protein band intensities, triplicate purifications of the reference sample were quantified by densitometry. A total recovery of 91% for IgG of total IgG-κ (heavy + light chain) and total IgG-λ (heavy + light chain) combined and, ≥95% for Fab of Fab-κ and Fab-λ combined was calculated. Coefficients of variation of the densitometry intensity of triplicate purifications of the reference sample were 2.1% for κ, 4.8% for λ, 3.1% for Fab-κ and 4.4% for Fab-λ.

κ-to-λ ratio

Calculated serum IgG concentration of controls and cases were on average 9.5 g/L (95% CI 7.4-11.5 g/L) and therefore within the normal range (7.0-16.0 g/L). To determine the κ:λ ratio of the κ and λ purification, Fab, Fab-κ, Fab-λ, κ and λ fractions from the reference donor sample were purified in duplicate and each fraction was measured twice on an LTQ Orbitrap XL. Figure 2 represent the distribution of all Mascot peptide sequences corresponding to the V,D, J, and C region of the κ, λ and heavy chain (BLAST identity score ≥ 70%) and their normalized intensities in the Fab and the total IgG light chain (κ and λ). The κ:λ ratio was calculated by counting all V, D, J and C spectra, with a normalized intensity >0. We found a normal κ:λ ratio of 2.0 in the Fab and 2.1 in the total light chain of IgG (κ and λ) fraction in this healthy donor sample. In addition, we calculated the κ:λ ratio in the Fab of all controls and cases and observed a normal mean ratio of 1.9 (SD: 0.1) and 2.0 (SD: 0.0), respectively. The unpaired two sample t-test revealed no statistically significant difference (*p*=0.10) between the κ:λ ratio in cases and controls.

Enrichment of κ and λ peptides.

Kappa and lambda enrichment of the cases and controls was calculated by the κ:λ ratio in the Fab-κ, Fab-λ, κ and λ fractions, as described above. After purification we obtained a 7-fold enrichment of κ in the Fab-κ (κ:λ ratio 14:1) and a 3-fold enrichment of λ in the Fab-λ fraction (κ:λ ratio 2:3). An 8-fold increase in enrichment factor was observed in the κ and a 4-fold increase in the λ fractions of IgG. In addition, <9% peptides of the heavy chain were found in the light chain fractions of IgG.

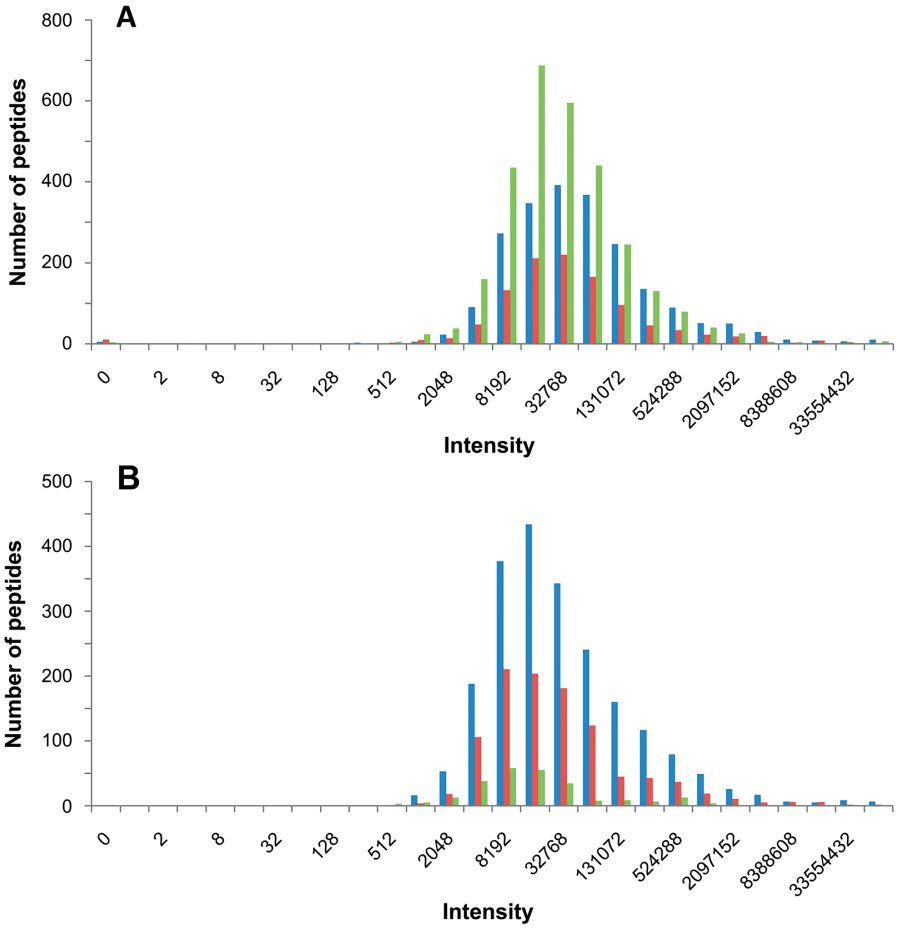


Figure 2. Distribution of VDJC peptides in Fab and IgG light chain of reference sample. All peptide sequences corresponding to the V region, D region, J region and C region of the κ (Blue), λ (Red) and heavy chain (Green) found by Mascot in the IMGT database and their normalized intensities in the Fab (A) and the IgG light chain (B).

Replicate measurements of reference sample

In mass spectrometry, replicate measurements can increase the number of peptides identified. To determine whether the total number of CDRs identified in the multiple IgG fractions might increase due to the multiple MS measurements of the sample, we compared the number of CDRs identified in four replicate measurements of the Fab, Fab-κ, Fab-λ, κ, λ fractions of the reference sample. Although, the number of CDRs reached a maximum at each third or fourth replicate measurement, the combined fractions revealed more CDRs than the individual fractions (Figure 3). We found 617 CDRs in the Fab and 1238 CDRs in the combined fractions.

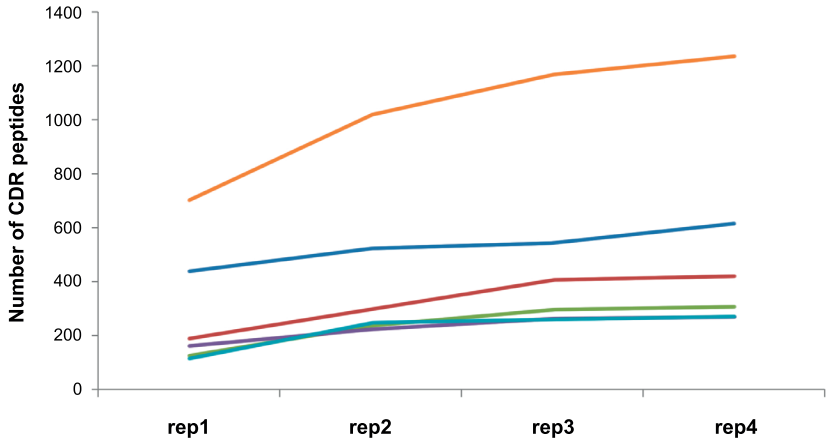


Figure 3. Number of CDRs peptides identified in replicate MS measurements of the fractions individually and combined of the reference sample. Fab (Blue), Fab-κ (Red), Fab-λ (Green), kappa (κ) (Purple), lambda (λ) (Light Blue) fraction and all fractions combined (Orange).

Number of CDRs identified in cases and controls

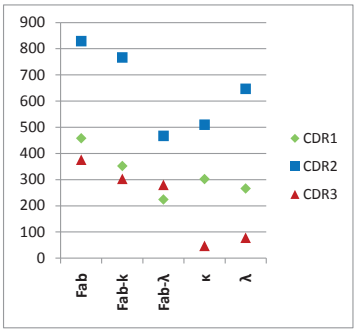
Nano-LC-LTQ Orbitrap MS measurements of the respective digested Fab, Fab-κ, Fab-λ, κ and λ of the cases and controls yielded a combined peak list of 13061, 12441, 8246, 10294 and 11853 features reported by Progenesis. Automatic alignment in Progenesis was not possible for one Fab, one Fab-λ and one κ sample of the controls and one Fab-λ sample of the cases and therefore were excluded from data analysis. Figure 4 shows the number of features (MS signals) and the number of CDR peptide sequences identified by Mascot and *de novo* sequencing according to IMGT for each individual IgG fraction. In addition, we have listed the V, D, J and C regions (VDJC) and the V-region-related peptides.

To compare the number of CDRs identified between the fractions, only samples analyzed by Progenesis for each fraction were selected. Redundant peptide sequences of these 16 samples were counted once in order to reveal the number of unique CDRs per fraction. Alignment to the IMGT database showed that 1663 peptide sequences of Fab, 1422 of Fab-κ, 971 of Fab-λ, 859 of κ and 991 of λ corresponded to CDRs. The numbers of the three types of CDR are shown in Figure 5.

In all three types of Fab (Fab, Fab-κ, Fab-λ) we observed a mean CDR1:CDR2:CDR3 ratio of approximately 1.0:2.0:1.0 and in the light chains different CDR ratios of approximately 1.4:2.4:0.2 for κ and 1.0:2.7:0.3 for λ were seen. The CDR3 ratio in the three Fab fractions (Fab, Fab-κ, Fab-λ) was significantly (p<0.001) higher than the CDR3 ratio in the light chain fractions (κ and λ).

The mean number of CDRs identified in the individual and combined IgG fractions compared with that of the Fab fraction of lung cancer cases and controls are listed in Table 1. We found 1.73 times more CDRs in the combination of Fab-κ, Fab-λ, κ and λ (Comb 6, Table 1),

	Fab		
	13061 MS signals of controls and cases (9/10)		
Peptides	**Total**	**Mascot**	***de novo***
VDJC-region	8013	3087	4926
V-region	5318	2567	2751
V-region unique	3191	1890	1301
CDRs unique	**1663**	1017	646
CDR1	**458**	228	230
CDR2	**829**	517	312
CDR3	**376**	272	104



	Fab-k		
	12441 MS signals of controls and cases (10/10)		
Peptides	Total	Mascot	*de novo*
VDJC-region	7552	2466	5086
V-region	4966	1986	2980
V-region unique	2881	1375	1506
CDRs unique	**1422**	594	828
CDR1	**352**	120	232
CDR2	**767**	421	346
CDR3	**303**	53	250

	Fab-λ		
	8246 MS signals of controls and cases (9/9)		
	Total	Mascot	*de novo*
	4950	1845	3105
	3193	1522	1671
	1885	1134	751
	971	618	353
	224	110	114
	467	295	172
	280	213	67

	kappa (κ)		
	10294 MS signals of controls and cases (9/10)		
Peptides	Total	Mascot	*de novo*
VDJC-region	5492	1893	3599
V-region	3274	1355	1919
V-region unique	1462	710	752
CDRs unique	**859**	422	437
CDR1	**302**	118	184
CDR2	**510**	290	220
CDR3	**47**	14	33

	lambda (λ)		
	11853 MS signals of controls and cases (10/10)		
	Total	Mascot	*de novo*
	6387	1651	4736
	3804	1295	2509
	1774	789	985
	991	437	554
	266	74	192
	647	359	288
	78	4	74

Figure 4. Number of MS signals and number of peptide sequences identified by Mascot and *de novo* sequencing for each individual IgG fraction. Redundant peptides corresponding to the VDJC-region and the V-region, and non-redundant (unique) peptides corresponding to the V-region and CDR (CDR1, 2, 3) region germline sequences from the IMGT database are shown. The graph illustrates the total number of CDR1, CDR2 and CDR3 identified for each individual IgG fraction. (*Published as Supporting Information Figure 4.*)

than in Fab for both cases and controls. Pearson chi-square tests with odds ratios were performed to measure the association between cases and controls in the number of CDRs identified in the individual and combined IgG fractions compared with that of Fab (Table 1). We found no statistically significant difference (p>0.50) between cases and controls for these numbers.

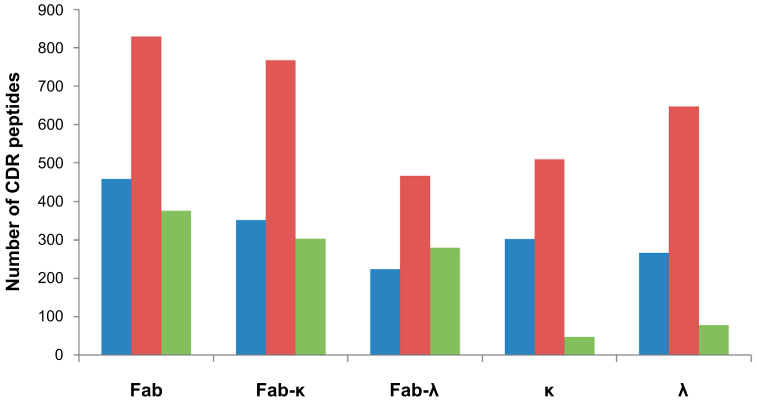


Figure 5. Numbers of CDR1, CDR2 and CDR3 identified in the individual IgG fractions. The total number of CDR1 (Blue), CDR2 (Red) and CDR3 (Green) identified in each individual IgG fraction.

Table 1. Number of CDRs identified in individual and combined IgG fractions compared with Fab: Lung cancer cases versus controls.

Fractions	Number of CDRs						OR	95% CI	*p*
	Lung cancer cases			Controls					
	N	mean (±SD)	Odds	N	mean (±SD)	Odds			
Fab	9	1412 (± 72)	ND	7	1370 (± 80)	ND	ND	ND	ND
Fab-κ	9	1119 (± 115)	0.79	7	1072 (± 97)	0.78	1.01	0.91-1.13	0.82
Fab-λ	9	715 (± 90)	0.51	7	706 (± 113)	0.52	0.98	0.86-1.12	0.79
κ	9	585 (± 82)	0.41	7	582 (± 83)	0.42	0.98	0.85-1.12	0.72
λ	9	748 (± 55)	0.53	7	700 (± 57)	0.51	1.04	0.91-1.18	0.58
Comb 1	32	1211 (± 100)	0.86	32	1172 (± 91)	0.86	1.00	0.90-1.12	1.00
Comb 2	32	1652 (± 127)	1.17	32	1609 (± 121)	1.17	1.00	0.90-1.10	0.92
Comb 3	32	1724 (± 85)	1.22	32	1690 (± 83)	1.23	0.99	0.90-1.09	0.84
Comb 4	32	2105 (± 144)	1.49	32	2032 (± 125)	1.48	1.01	0.91-1.11	0.92
Comb 5	48	2400 (± 146)	1.70	48	2335 (± 117)	1.70	1.00	0.91-1.10	1.00
Comb 6	64	2449 (± 159)	1.73	64	2376 (± 170)	1.73	1.00	0.91-1.10	1.00
Comb 7	80	3119 (± 165)	2.21	80	3024 (± 167)	2.21	1.00	0.92-1.09	1.00

Odds ratios between lung cancer cases and controls of the number of CDRs identified in individual and combined IgG fractions compared with Fab. ND, the value was not determined; OR, odds ratio; CI, confidence interval; *p, p*-value of Pearson chi-square (X^2) test; Comb 1, κ+λ; Comb 2, Fab-κ+Fab-λ; Comb 3, Fab+Fab-λ; Comb 4, Fab+Fab-κ; Comb 5, Fab+Fab-κ+Fab-λ; Comb 6, Fab-κ+Fab-λ+κ+λ; Comb 7, Fab+Fab-κ+Fab-λ+κ+λ.

We calculated the mean additional number of unique CDRs (mean %, CI: 95%) found in the different individual and combined IgG fractions to the mean number of CDRs of the Fab fraction. Kappa (κ) gave 320 CDRs (23.1%, 20.6-25.5%), lambda (λ) 501 CDRs (36.0%, 34.2-37.8%), Fab-κ 679 CDRs (48.7%, 46.2-51.2%) and Fab-λ 315 CDRs (22.7%, 20.1-25.4%) additional to Fab. Combined κ and λ fractions resulted in 804 additional CDRs

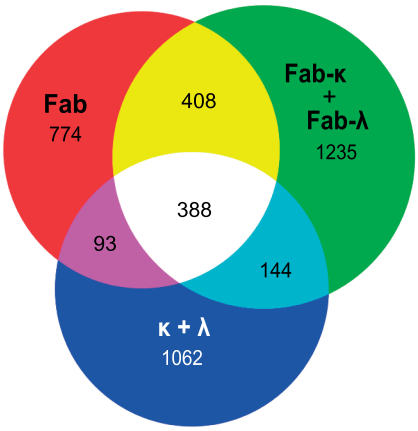


Figure 6. Venn diagram of all fractions of the cases and controls and the total number of CDRs found by Mascot and de novo sequencing.

(57.8%, 54.4-61.3%), and combined Fab-κ and Fab-λ fractions in 978 additional CDRs (70.3, 66.8-73.7%). In addition, these four fractions combined showed an additional 1683 unique CDRs (121.0%, 115.5-126.5%) compared with the original Fab.
Figure 6 shows a Venn diagram of all the fractions of the cases and controls and the total number of CDRs found by Mascot and *de novo* sequencing. We found a total of 1663 CDRs in the Fab and an additional 2441 unique CDRs (146.8%) in all the other fractions combined.

Number of significantly different CDRs between cases and controls

We analyzed the CDR-identified sequences of cases and controls obtained via database-dependent and *de novo* sequencing with Anova and the two sample t-test. We observed that in Fab-κ the number of significantly identified CDR ($p<0.01$) was significantly increased compared to random chance as determined by a permutation test.

DISCUSSION

In this study, we demonstrated that molecular dissection of IgG into kappa and lambda fragments (Fab-κ, Fab-λ, κ and λ) identifies approximately twice as many CDRs than the Fab method.

In all fractions CDRs were identified exclusive to the specific fraction resulting in a total of 4104 CDR sequences. The results from all fractions and of combinations of them were evaluated. Although multiple MS measurements of an individual fraction also increase the number of CDRs identified, the combination of the various molecular fractions exceeds this significantly. In addition, the method we used to isolate Fab-κ, Fab-λ, κ and λ chains of IgG

is reproducible and shows high recovery rates. The technical variation in MS measurements and variation in individual IgG molecules was described previously.[26] More specifically, the normal κ:λ ratio in the Fab and light chain of the healthy donor sample and the Fab of the controls and cases demonstrated a sufficient purification with preserved κ:λ ratio, and was validated by MS measurement. By means of the κ:λ ratio we were able to determine the κ or λ enrichment for the different IgG fractions.

As the whole Fab fraction includes the heavy and light chains (both κ and λ), this fraction yielded the most features and as a result more CDRs were identified in the Fab fraction than in the other individual IgG fractions. The MS sample of the Fab-κ fraction yielded more features than the Fab-λ fraction, and therefore the Fab-κ fraction revealed more CDRs than the Fab-λ fraction. The lower number of CDRs identified in κ and λ light chain fractions is very likely caused by the fact that CDRs specific to the heavy chain are missing, including the highly diverse CDR3 of the heavy chains.[28] This is supported by the observation of a different ratio for CDR1:CDR2:CDR3 in the light chains, which shows an approximately 4-fold lower ratio of CDR3 in the light chains than in the three types of Fab (Fab, Fab-κ, and Fab-λ). In general, in all fractions the CDR3 peptides were relatively difficult to assign. These peptides are highly diverse and their N-terminal side often contains a cleavage site for trypsin (lysine or arginine). As a result, their tryptic-digested peptides often contain a mostly conserved V-region fragment or a highly diverse fragment, which makes it difficult to align to the germline sequence. Interestingly, more significantly different CDRs were observed in the Fab-κ than in the other fractions. This points to the possibility of finding lung cancer-related CDRs and in general of finding tumor-related CDRs.

Peptides from the constant regions gave the highest peak intensities. By choosing the maximum injection volume based on the highest peak intensity in the UV chromatogram we were able to maximize the loading of the CDRs on the C18 trap column. Nano-LC-LTQ Orbitrap MS measurements of additional fractions cause an increase in measurement time. Even though, measuring both Fab and Fab-κ fractions instead of only the Fab fraction requires twice as much measurement time, it makes the effort worthwhile because of the additional 50% CDRs identified. In addition, when measuring twice this is the best combination of all fractions because the immunoglobulin molecules that are occasionally expressed by cancer cells have been reported to consist predominantly of the heavy chains and κ chains.[29] This can be explained by the fact that during B-cell differentiation first the heavy chain genes rearrange followed by the κ chain genes. Only if none of the κ chain gene rearrangements leads to a functional κ chain the λ chain genes start to rearrange.[19] Another explanation is that the heavy chain contains the highly diverse CDR3, which plays a prominent role in antigen binding.[28] Both heavy chains and κ chains are present in the Fab and Fab-κ fractions. Recent studies have shown that antibody specificity is determined by a limited number of amino acid residues of the CDRs. Synthesized small peptides based on these CDRs retained the antigen-binding properties and functions of the intact immunoglobulin.[30, 31] Administra-

tion of synthetic CDR peptides inhibits tumor cell growth in mice and thereby increases their survival time.[32] These reports support the hypothesis that a specific molecular profile of CDRs may distinguish lung cancer patients from controls. In agreement, we found an increase in the number of significantly different CDRs in the Fab-κ fraction.

In our study, the lung cancer cases did not differ in their normal κ:λ ratio of Fab and in the number of detected CDRs in all the different IgG fractions from the controls. These findings show that our method is technical suitable to compare CDRs in IgG fractions between lung cancer patients and controls. Our approach revealed more CDRs than the original Fab method, which may enhance the possibility to identify a biomarker model for the early detection of lung cancer. However, there is most probably a larger sample set required to identify such statistically and physiologically relevant model. Sample size calculations[33] based on unpublished data estimate that a sample set of approximately 30 lung cancer cases and 30 controls is required to acquire this.

Improvements in sequence coverage and annotation may help to further increase the number of CDRs that is possible to identify. Alternative proteases could be used to obtain larger sequences coverage for a better alignment to the germline sequence. Other potential improvements are using ultra high pressure chromatography techniques to improve resolution for a better identification of sequences and depletion of constant regions by partial digestion of immunoglobulins to enrich CDR regions. In addition, complementing fragmentation spectra by higher energy collision induced dissociation (HCD) and electron transfer dissociation (ETD) can improve *de novo* peptide sequencing compared to CID fragmentation.[34-36]

The ability to detect specific tumor-related CDR peptides by mass spectrometry depends on the proportion of total IgG that has affinity to the tumor antigen. Affinity purification of rat sera revealed that 1-3% of IgG had affinity for the antigen used for the immunizations.[18] Such a polyclonal antibody response to an antigen has been estimated to derive from approximately 100 B-cell clones.[23] In another study, an upper limit of 0.1-0.3% of the human B-cell population was found to have originated from a particular clone.[37] Based on these data, we estimate that 0.01-0.3% of the total IgG may present a particular immunoglobulin, depending on the degree of the immune response against the antigen and the diversity of the B-cell clones. In previously published papers we showed that it is possible to detect CDRs of specific immunoglobulins at these levels[18, 26] and in particular, by our recent paper.[36] In this paper, we showed that specific CDR peptides of a spiked antibody could be detected at attomole levels which were 5 orders of magnitude lower than the total IgG serum.

In conclusion, we have demonstrated that molecular dissection of IgG into kappa and lambda fragments is a valuable addition to Fab purification. Molecular dissection of IgG into kappa and lambda fragments identifies significantly more CDRs than Fab purification alone. This approach will increase the likelihood of finding lung cancer-related CDR sequences.

ACKNOWLEDGEMENT

The authors acknowledge financial support from an NWO (Nederlandse organisatie voor Wetenschappelijk Onderzoek), Zenith grant 93511034.

REFERENCES

1. Society AC. Cancer Facts & Figures 2010. 2010.
2. Lopes Pegna A, Picozzi G, Mascalchi M, Maria Carozzi F, Carrozzi L, Comin C, et al. Design, recruitment and baseline results of the ITALUNG trial for lung cancer screening with low-dose CT. *Lung Cancer*. 2009; 64: 34-40.
3. Infante M, Lutman FR, Cavuto S, Brambilla G, Chiesa G, Passera E, et al. Lung cancer screening with spiral CT: baseline results of the randomized DANTE trial. *Lung Cancer*. 2008; 59: 355-63.
4. 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). *Int J Cancer*. 2007; 120: 868-74.
5. van Klaveren RJ, Oudkerk M, Prokop M, Scholten ET, Nackaerts K, Vernhout R, et al. Management of lung nodules detected by volume CT scanning. *N Engl J Med*. 2009; 361: 2221-9.
6. Reich JM. A critical appraisal of overdiagnosis: estimates of its magnitude and implications for lung cancer screening. *Thorax*. 2008; 63: 377-83.
7. Dunn GP, Bruce AT, Ikeda H, Old LJ and Schreiber RD. Cancer immunoediting: from immunosurveillance to tumor escape. *Nat Immunol*. 2002; 3: 991-8.
8. Dunn GP, Old LJ and Schreiber RD. The immunobiology of cancer immunosurveillance and immunoediting. *Immunity*. 2004; 21: 137-48.
9. Qiu J and Hanash S. Autoantibody profiling for cancer detection. *Clin Lab Med*. 2009; 29: 31-46.
10. Chen G, Wang X, Yu J, Varambally S, Thomas DG, Lin MY, et al. Autoantibody profiles reveal ubiquilin 1 as a humoral immune response target in lung adenocarcinoma. *Cancer Res*. 2007; 67: 3461-7.
11. Qiu J, Choi G, Li L, Wang H, Pitteri SJ, Pereira-Faca SR, et al. Occurrence of autoantibodies to annexin I, 14-3-3 theta and LAMR1 in prediagnostic lung cancer sera. *J Clin Oncol*. 2008; 26: 5060-6.
12. Leidinger P, Keller A, Heisel S, Ludwig N, Rheinheimer S, Klein V, et al. Identification of lung cancer with high sensitivity and specificity by blood testing. *Respir Res*. 2010; 11: 18.
13. Rom WN, Goldberg JD, Addrizzo-Harris D, Watson HN, Khilkin M, Greenberg AK, et al. Identification of an autoantibody panel to separate lung cancer from smokers and nonsmokers. *BMC Cancer*. 2010; 10: 234.
14. Farlow EC, Patel K, Basu S, Lee BS, Kim AW, Coon JS, et al. Development of a multiplexed tumor-associated autoantibody-based blood test for the detection of non-small cell lung cancer. *Clin Cancer Res*. 2010; 16: 3452-62.
15. Wu L, Chang W, Zhao J, Yu Y, Tan X, Su T, et al. Development of autoantibody signatures as novel diagnostic biomarkers of non-small cell lung cancer. *Clin Cancer Res*. 2010; 16: 3760-8.
16. Zhong L, Coe SP, Stromberg AJ, Khattar NH, Jett JR and Hirschowitz EA. Profiling tumor-associated antibodies for early detection of non-small cell lung cancer. *J Thorac Oncol*. 2006; 1: 513-9.
17. Chapman CJ, Murray A, McElveen JE, Sahin U, Luxemburger U, Tureci O, et al. Autoantibodies in lung cancer: possibilities for early detection and subsequent cure. *Thorax*. 2008; 63: 228-33.

18. VanDuijn MM, Dekker LJ, Zeneyedpour L, Smitt PA and Luider TM. Immune responses are characterized by specific shared immunoglobulin peptides that can be detected by proteomic techniques. *J Biol Chem*. 2010; 285: 29247-53.

19. Murphy KP, Travers P and Walport M. Janeway's Immunobiology. New York: Garland Science, 2008.

20. Saada R, Weinberger M, Shahaf G and Mehr R. Models for antigen receptor gene rearrangement: CDR3 length. *Immunol Cell Biol*. 2007; 85: 323-32.

21. Andersen PS, Haahr-Hansen M, Coljee VW, Hinnerfeldt FR, Varming K, Bregenholt S, et al. Extensive restrictions in the VH sequence usage of the human antibody response against the Rhesus D antigen. *Mol Immunol*. 2007; 44: 412-22.

22. Baranzini SE, Jeong MC, Butunoi C, Murray RS, Bernard CC and Oksenberg JR. B cell repertoire diversity and clonal expansion in multiple sclerosis brain lesions. *J Immunol*. 1999; 163: 5133-44.

23. Poulsen TR, Meijer PJ, Jensen A, Nielsen LS and Andersen PS. Kinetic, affinity, and diversity limits of human polyclonal antibody responses against tetanus toxoid. *J Immunol*. 2007; 179: 3841-50.

24. Weinstein JA, Jiang N, White RA, 3rd, Fisher DS and Quake SR. High-throughput sequencing of the zebrafish antibody repertoire. *Science*. 2009; 324: 807-10.

25. Annesley TM. Ion suppression in mass spectrometry. *Clin Chem*. 2003; 49: 1041-4.

26. de Costa D, Broodman I, Vanduijn MM, Stingl C, Dekker LJ, Burgers PC, et al. Sequencing and quantifying IgG fragments and antigen-binding regions by mass spectrometry. *J Proteome Res*. 2010; 9: 2937-45.

27. Chui SH, Lam CW and Lai KN. Light-chain ratios of immunoglobulins G, A, and M determined by enzyme immunoassay. *Clin Chem*. 1990; 36: 501-2.

28. Xu JL and Davis MM. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity*. 2000; 13: 37-45.

29. Chen Z, Qiu X and Gu J. Immunoglobulin expression in non-lymphoid lineage and neoplastic cells. *Am J Pathol*. 2009; 174: 1139-48.

30. Eisenhardt SU, Schwarz M, Schallner N, Soosairajah J, Bassler N, Huang D, et al. Generation of activation-specific human anti-alphaMbeta2 single-chain antibodies as potential diagnostic tools and therapeutic agents. *Blood*. 2007; 109: 3521-8.

31. Padlan EA, Abergel C and Tipper JP. Identification of specificity-determining residues in antibodies. *FASEB J*. 1995; 9: 133-9.

32. Polonelli L, Ponton J, Elguezabal N, Moragues MD, Casoli C, Pilotti E, et al. Antibody complementarity-determining regions (CDRs) can display differential antimicrobial, antiviral and antitumor activities. *PLoS One*. 2008; 3: e2371.

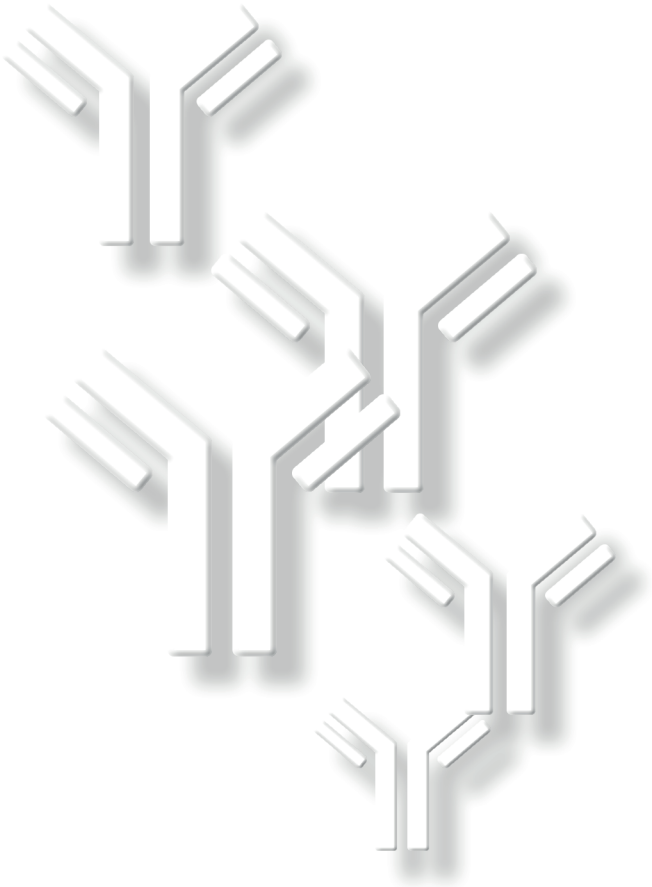
33. Geller NL and Pocock SJ. Interim analyses in randomized clinical trials: ramifications and guidelines for practitioners. *Biometrics*. 1987; 43: 213-23.

34. Olsen JV, Macek B, Lange O, Makarov A, Horning S and Mann M. Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods*. 2007; 4: 709-12.

35. Shen Y, Tolic N, Xie F, Zhao R, Purvine SO, Schepmoes AA, et al. Effectiveness of CID, HCD, and ETD with FT MS/MS for degradomic-peptidomic analysis: comparison of peptide identification methods. *J Proteome Res*. 2011.

36. Dekker LJ, Zeneyedpour L, Brouwer E, van Duijn MM, Sillevis Smitt PA and Luider TM. An antibody-based biomarker discovery method by mass spectrometry sequencing of complementarity determining regions. *Anal Bioanal Chem*. 2011; 399: 1081-91.

37. Boyd SD, Marshall EL, Merker JD, Maniar JM, Zhang LN, Sahaf B, et al. Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing. *Sci Transl Med*. 2009; 1: 12ra23.



Dominique de Costa D, **Ingrid Broodman**, Wim Calame, Christoph Stingl,
Lennard J.M. Dekker, René M. Vernhout, Harry J. de Koning,
Henk C. Hoogsteden, Peter A. E. Sillevis Smitt, Rob J. van Klaveren,
Theo M. Luider, Martijn M. VanDuijn

Peptides from the variable region of specific antibodies are shared among lung cancer patients

CHAPTER 3

PLoS One 2014; 9: e96029

ABSTRACT

Late diagnosis of lung cancer is still the main reason for high mortality rates in lung cancer. Lung cancer is a heterogeneous disease which induces an immune response to different tumor antigens. Several methods for searching autoantibodies have been described that are based on known purified antigen panels. The aim of our study is to find evidence that parts of the antigen-binding-domain of antibodies are shared among lung cancer patients. This was investigated by a novel approach based on sequencing antigen-binding-fragments (Fab) of immunoglobulins using proteomic techniques without the need of previously known antigen panels. From serum of 93 participants of the NELSON trial IgG was isolated and subsequently digested into Fab and Fc. Fab was purified from the digested mixture by SDS-PAGE. The Fab containing gel-bands were excised, tryptic digested and measured on a nano-LC-Orbitrap-Mass-spectrometry system. Multivariate analysis of the mass spectrometry data by linear canonical discriminant analysis combined with stepwise logistic regression resulted in a 12-antibody-peptide model which was able to distinguish lung cancer patients from controls in a high risk population with a sensitivity of 84% and specificity of 90%. With our Fab-purification combined Orbitrap-mass-spectrometry approach, we found peptides from the variable-parts of antibodies which are shared among lung cancer patients.

INTRODUCTION

Lung cancer is currently the most common cancer with the highest mortality rate (28%) in the World due to diagnosis at an advanced stage.[1, 2] However, with the demonstration of a 20% lung cancer mortality reduction by the NLST trial (National Cancer Screening Trial) low dose CT screening for lung cancer is receiving increasing interest.[3] The NELSON trial (Dutch-Belgian lung cancer screening trial) showed that after three screening rounds 3.6% of all participants of this study had a false-positive screen result.[4] Although, still approximately 27% of the participants were subjected to invasive procedures that revealed benign lung diseases at baseline screening (first round NELSON trial).[5] A good biomarker (panel) will reduce this number of unnecessary invasive procedures. At the moment selection of high risk individuals for screening is done by age and smoking history. A biomarker or biomarker panel would be helpful in selecting high risk individuals for CT screening as this may detect lung cancer at an earlier stage than CT.

Antibodies can be interesting as markers for distinguishing lung cancer patients from lung cancer-free individuals. These antibodies are produced by the immune response that target specific tumor-associated antigens (TAAs) during cancer development, probably at an early stage.[6-12] Recently Liu et al. showed that the concentration of circulating IgG autoantibodies against ABCC3 transporter was significantly higher in female adenocarcinoma patients than in female controls.[13]

Human antibodies consist of four chains, two identical heavy chains and two identical light chains. Each light chain has a variable (V_L) and constant (C_L) domain. The heavy chains have three different constant domains (C_H1, C_H2 and C_H3) and a variable domain (V_H). The first constant and variable parts form the antigen binding fragment (Fab). The remaining two constant parts of the heavy chain form the Fc region. Within the Fab six complementarity determining regions (CDR1, CDR2 and CDR3) are located between frameworks. These CDRs determine the antigen specificity and form a surface complementary to a shape that is part of the antigen. CDRs are hypervariable regions of the antibody.[14] Antibodies, or immunoglobulins, are highly complex molecules with large variation in their amino acid sequence. The possible diversity in immunoglobulins is estimated between 10^{13} and 10^{50} and therefore the finding of similar or even identical sequences in different individuals by chance is in theory, highly unlikely.[14, 15] However, studies of different research groups have recently demonstrated that despite this theoretical small chance to have identical antibodies among individuals, it is possible to identify similar or identical sequences.[16-19] A study performed by us showed that in PNS (paraneoplastic neurological syndrome) patients identical mutated primary amino acid sequences of complementarity determining regions (CDRs) exist. These CDRs are specific for known onconeural antigens, such as HuD and Yo in PNS patients, and most interestingly were shared between different PNS patients.[20]

The aim of this study is to find evidence that specific antibody peptides are shared between lung cancer patients in contrast to lung cancer-free individuals. As lung cancer is a heterogeneous disease and with the variability of an antibody it might be a challenge to detect identical tumor-related antibodies in serum. We experimentally test the hypothesis that specific highly variable regions of an antibody including complementarity determining regions (CDRs) can be shared between lung cancer patients. Our experimental approach to verify this hypothesis is based on sequencing antibody peptides by mass spectrometry. Measurement of serum by a mass spectrometer might be too complex due to the high variability as mentioned above. Purifying IgG Fab from serum will reduce the complexity of the sample from a lung cancer patient and will give the possibility to focus on pure antibody fractions.

MATERIALS AND METHODS

Ethics and legal approval

The NELSON trial was approved by the Dutch Health Council, the Minister of Health and by the Medical Ethical Committees of all participating centers (clinical trial number ISRCTN63545820). All participants for this study provided written informed consent for the use of their serum samples. The donor of the reference sample used throughout this study provided written consent for the use of his/her serum for scientific purposes according to the guidelines of the Blood Bank Sanquin, Rotterdam, the Netherlands.

NELSON trial

The NELSON (Dutch-Belgian Lung Cancer Screening trial) trial has started recruitment in 2003 by sending questionnaires to 548,489 males and females between 50–75 years of age. Participants had to be current or former smokers for at least 25 years, smoking at least 15 cigarettes per day or smoking at least 30 years, smoking at least 10 cigarettes per day. From the 548,489 males and females 15,822 participants were included in the trial. These participants were randomized to a screen or control arm. The screening arm received CT screening in years 1,2 and 4. The control arm received no screening (usual care). Participants with a positive test result were referred to a pulmonologist. If the diagnosis lung cancer was established the patient was treated and went off screening. Participants with an indeterminate test result underwent a follow-up scan three months later. If a negative test result was obtained the second-round CT scan was scheduled for 12 months later.[5, 21]

Study population

For this study, we selected 44 lung cancer cases and 49 controls (Supplementary Figure S1) from the NELSON lung cancer screening trial.[5, 21] For the cases of the discovery set, NELSON 1, only early stage (I and II) squamous cell (n=4) or adenocarcinomas (n=21) were selected.

They were carefully matched to the controls by age, gender, smoking status, duration and number of cigarettes smoked per day, chronic obstructive pulmonary disease (COPD) status, asbestos exposure and site of blood sampling (Supplementary Table S1). The selection criteria for the cases of the NELSON 2 (validation) set (n=19) were similar, except that all non-small cell histology's and disease stages were allowed (Supplementary Table S1) in order to challenge the results of the discovery phase. On purpose the clinical characteristics of the control patients are dissimilar with the NELSON 1 set in respect to smoking and COPD. Therefore, this NELSON 2 set is not matched with the NELSON 1 set. By using a validation sample set (NELSON 2) chosen in this way, the robustness of the method can be determined. Serum samples were collected for both NELSON 1 and NELSON 2 obtained from baseline CT screening (first round).

IgG Fab purification and nano-LC Orbitrap MS analyses

Prior to all sample preparation procedures, all samples were blinded and the key for unblinding was put at the database coordinator of the NELSON trial. IgG Fab purification and nano-LC Orbitrap MS analyses were performed according to the method described before.[22] In brief, IgG was isolated from serum and digested into Fab and Fc (Figure 1). The Fab part was isolated from the digested mixture by SDS-PAGE. The Fab containing gel bands were excised and tryptic digested. A blank piece of gel that was not loaded with protein was excised and treated like the excised Fab bands for background assessment.

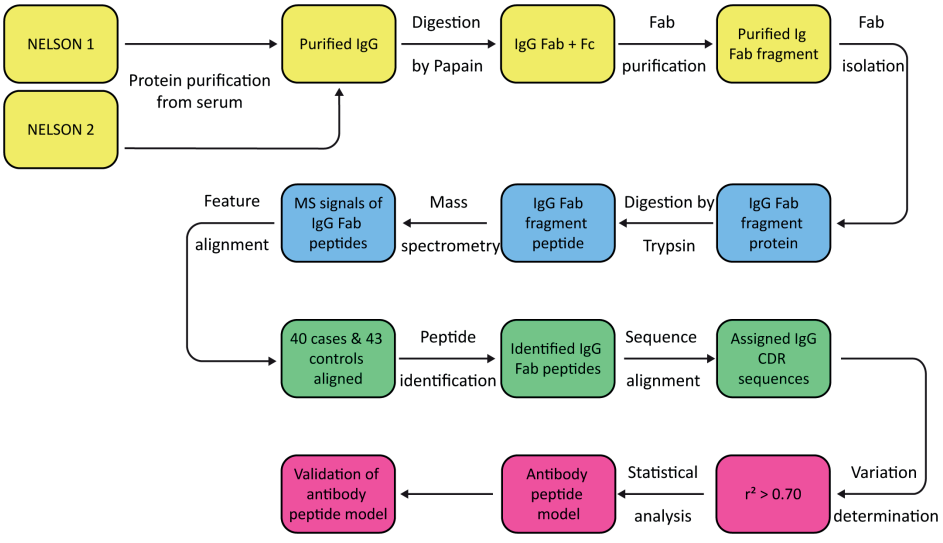


Figure 1. Flow-chart of the method and analysis used. In this flow-chart the different steps in Fab purification, Fab measurement and data analysis are illustrated. In yellow the Fab purification is shown, in blue the mass spectrometry measurement, in green the data analysis and in pink the statistical analysis.

LCMS measurements were performed on an Ultimate 3000 nano LC system (Thermo Fisher Scientific/Dionex, Amsterdam, the Netherlands) online coupled to a hybrid linear ion trap/Orbitrap MS (LTQ Orbitrap XL; Thermo Fisher Scientific, Bremen, Germany). Four µL of the digested Fab was loaded onto the system. For further settings and solutions we refer to previous published work.[22] All samples were randomized before measurement and were measured in batches of 11 samples including a reference sample. A reference sample was used as a quality control for each measurement and analysis step. A blank sample was run at the start and end of the measurement to determine background and the existence of carry-over during chromatography.

Data Analyses

Raw data files were loaded into the software Progenesis (Figure 1) (Version 3.1; Nonlineair Dynamics Ltd, New Castle, UK) and processes as described previously.[22] In addition, we performed a Progenesis analysis where instead of detecting features (peptide masses (m/z)) in all the samples at the same time by the software program, feature detection was performed individually per sample. Features picked thereby were matched to the Progenesis result table containing all samples with a mass tolerance of 5 ppm. This was of advantage, since often features occur with low intensities in one sample and are subsequently matched by Progenesis in all other samples. This result in errors related to background if one takes the respective mass spectra into account. With this relative small adjustment it ensures that a feature is detected more accurately throughout the samples. The data acquired by this approach was filtered using the same default settings.[22] A separate data matrix for every case and control was generated consisting of all features with corresponding raw abundance and retention time. To generate one large data matrix that includes all cases and controls from these separate data matrices, we searched masses from the separate data matrices per case or control in the complete data matrix generated from the standard Progenesis analyses. Every mass had to meet three criteria: 1) m/z (±5 ppm), 2) retention time (±1 min) and 3) identical charge. If a mass met these three criteria the raw abundance from the complete matrix (generated by a general procedure[22] recommended by the manufacturer) was used. If a mass did not meet these criteria a zero was generated for the raw abundance.

MS/MS spectra were extracted from raw data files and converted into Mascot compatible files using extract-msn (part of Xcalibur version 2.0.7, Thermo Fisher Scientific Inc.). Mascot (version 2.3.01; Matrix Science Inc., London, UK) was used to perform database searches against the human subset NCBInr database (version March 11th, 2009; Homo sapiens species restriction; 222,066 sequences) of the extracted MS/MS data (Figure 1). Database (NCBInr) dependent peptide identification and *de novo* sequencing results (software PEAKS; Version 5.2; Bioinformatics Solutions Inc., Waterloo, Canada) were also included in the Progenesis provided matrix. For settings used for the database search and *de novo* sequencing we refer to previous published work and methods S1.[22] For *de novo* sequences so far not known from

a database, the Peaks software identifies a leucine for the isobaric amino acids leucine and isoleucine. Database dependent peptide identification results or *de novo* sequencing results were included in the matrix based on the highest peptide identity score (Data S1, Data S2 and Data S3). All peptide sequences from the cases and controls identified by Mascot or PEAKS were subsequently aligned to databases containing V, D, J or C-region germline sequences derived from IMGT database (IMGT, the international ImMunoGeneTics information system http://www.imgt.org) using the BLAST algorithm (Figure 1).[23] Peptides with sufficient match (bitscore ≥12.5 and alignment score ≥70%) to the V-region database were assigned to a position on the immunoglobulin molecule with varying CDR lengths.

Raw data files of the reference samples of each data set were separately loaded into the software Progenesis and followed the standard procedures as mentioned above. To determine the proportion of variation between the reference sample measurements performed on different time points, median r-squares were calculated for each sample. Each sample was compared to all the other reference samples measured in that dataset and a median r-square was calculated for each sample. The comparison was based on the raw abundance of each feature. This was performed separately for both independent datasets, NELSON 1 and NELSON 2.

To determine the proportion of variation (Figure 1) between the samples (cases and controls) of the two separate datasets, the same calculations were performed as described above for each case and control sample. This analysis was performed separately for the two datasets. Based on the distribution of the median r-squares of each sample, we decided to set a cut-off at r-square >0.70. The cases and controls that obtained a median r-square below 0.70 were excluded from the dataset and further analyses. Calculations were conducted using Microsoft Excel 2007.

Statistical Analysis

Two independent data sets have been used, NELSON 1 and NELSON 2. The initial step in the statistical analysis consisted of testing for normality using skewness and kurtosis distribution characteristics on the intensity of the raw abundance of the features.[24]

Subsequently, univariate analysis was performed, applying either an unpaired t-test (parametric) or a Mann-Whitney U-test (non-parametric) to detect significant differences in raw abundance between cases and controls in the NELSON 1 set.[25] The significance limit was set at 0.05 (two-sided). All identified features that were found significantly different were used for the selection of features to distinguish lung cancer patients from controls.

Secondly, we used for multivariate analysis only the significantly identified features that had ≥2 triggered MS spectra. We applied a multivariate analysis on features fulfilling these criteria with a (logistic) stepwise regression model ($y=a_1x_1 + a_2x_2 + a_3x_3 \ldots a_nx_n + c$) in combination with canonical linear discriminant analysis.[26, 27] This resulted in a combination of features with high sensitivity and specificity in the NELSON 1 dataset. This combination of features was

then tested in the NELSON 2 dataset using the same methodology as described above.[26, 27] Note that for the NELSON 2 dataset it was necessary to optimize the coefficients in the model equation in order to optimize the sensitivity and specificity in the NELSON 2 dataset.

To avoid a random-error effect in modeling, we verified the statistical background of the combination of features in a permutated dataset. The background evaluation consisted of the same workflow as used for the model building, except that at the beginning the assignment of cases and controls of NELSON 1 were permutated (Figure S2). This permutation was performed twelve times and the results obtained were tested for significance against the model outcome by z-test (one-sided; $p<0.05$). Since model building was based on the data as provided in NELSON 1 after which validation of this model was done using the data in NELSON 2, the same approach was taken after each individual permutation. Also here, note that for NELSON 2 dataset the coefficients in the model equation were optimized.

All analyses on model building, validation and background evaluation were done using STATA, version 12 (StataCorp, Texas, US). Throughout the study, using two-sided testing (except for one-sided testing for Z-values), p-values of 0.05 or lower were considered to be statistically significant. Statistical analyses of the data shown in Table S1 were generated by SPSS (IBM SPSS Statistics 20). The time to cancer was generated by calculating the interval between blood sampling and diagnosis for each case.

RESULTS

Clinical characteristics of the study population

There was no significant difference in the clinical characteristics between the cases and controls in the NELSON 1 set (Table S1). In the NELSON 2 set, current or former smoker and COPD status differed significantly between cases and controls (Table S1). In 72% and 84% of the cases of the NELSON 1 set, and NELSON 2 set, respectively, the time interval between blood sampling and lung cancer diagnosis was between 0–1.5 years. The median follow-up duration after blood sampling was for the control population 1925 days (range 1075–2086 days) and 1861 days (range 347–2135) in the NELSON 1 set and NELSON 2 set, respectively. None of the controls developed lung cancer during the follow-up period.

Technical Variation

During the mass spectrometry measurements of the biological samples we measured a reference sample at different time points. R-square values were calculated from the abundances of identified proteins in each reference measurement to show technical reproducibility. The lowest r-square value observed in the different measurements ranged between 0.84 and 0.93 (Figure 2).

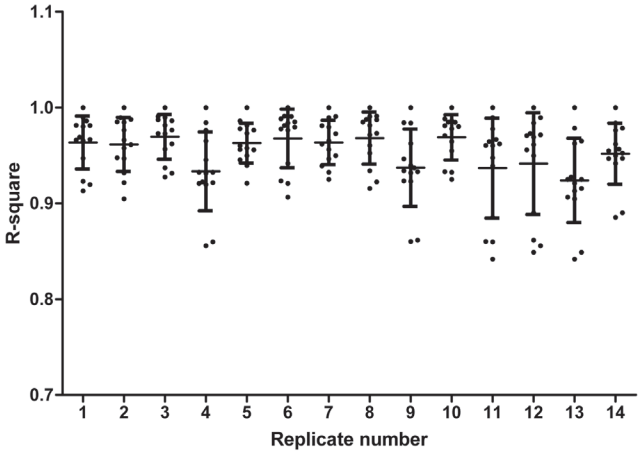


Figure 2. Technical reproducibility of replicate measurements of the reference sample. Reference sample measured at different time points during measurement of the NELSON 1 sample set. A replicate of the reference sample (x-axis) was compared to each other replicate sample based on the raw abundance of each feature. An r-square value was calculated. Each dot represents an r-square (y-axis) value for the comparison of that specific replicate with another replicate. For each replicate the average r-square and standard deviation (SD) is shown.

We performed the same r-square calculation for 5 random biological samples taken from the NELSON 1 set that were measured on two different LC-columns (same batch) at different time points. The technical reproducibility within each column resulted in lowest r-square values ranging from 0.75–0.93, but the technical reproducibility of the five biological samples measured on two independent similar columns was lower. For the two independent similar columns a median r-square of 0.52 was observed. In Figure 3 the correlation between each sample and between columns are shown.

In Figure 4A the retention times are shown for peptides identified with high confidence (Mascot score >60) in the Reference samples measured concurrently with both NELSON 1 and NELSON 2. This Figure shows that column performance was comparable between the two different LC columns for these abundant peptides (r-square 0.996). In addition, the abundances observed for these peptide also correlated well (Figure 4B; r-square 0.995). This suggests that both chromatography and mass spectrometry performed nominally, at least for peptides identified with high confidence at relatively high abundance. Thus, the technical variation we see primarily stems from peptides at lower abundances, closer to the detection limits (Figure S3).

An estimation of the biological variation was performed and resulted in a median r-square of 0.43. This result was much lower than the lowest r-square (0.84) observed for the technical variation. Therefore, the biological variation is higher compared to the technical variation.

These results show that technical variation should be taken into account and adjustment is needed for comparison of independently measured sample sets since the NELSON 1 and

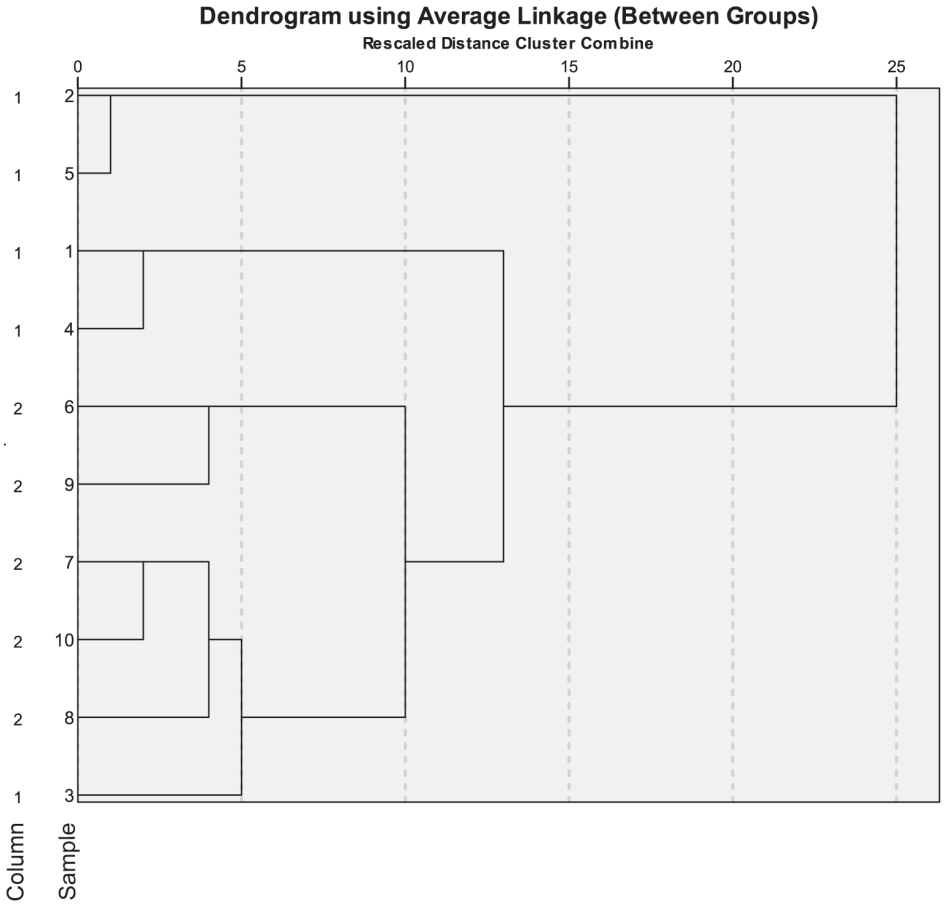


Figure 3. Technical reproducibility of five biological samples measured on two different columns at different time points. This dendrogram shows the correlation between five different biological samples measured on two different columns from same batch, column 1 and column 2 (y-axis). On the y-axis the five different samples are shown. Sample 1–5 are measured on column 1 and 6–10 are measured on column 2. Sample 1 and 6 are from the same individual. This also applies for sample 2 and 7, 3 and 8, 4 and 9 and 5 and 10. On the x-axis the Euclidian distance between each sample is shown. A strong correlation per column is found.

NELSON 2 dataset were measured on two different columns at different time points. To overcome this technical variation, we applied a number of filters on the data before we could start a data analysis as described in the Material and Methods section.

With this data we performed separate univariate analysis on all peptides found in cases and controls from the separate NELSON 1 and NELSON 2 data set. We were able to observe 49 peptides that were significantly different between cases and controls in the NELSON 1 dataset. However, these peptides, with one exception, did not show this difference in the NELSON 2 dataset. There was no trend observed (r-square 0.004) in p-values for the two datasets. Therefore, testing univariately in this manner was either not the right analysis strat-

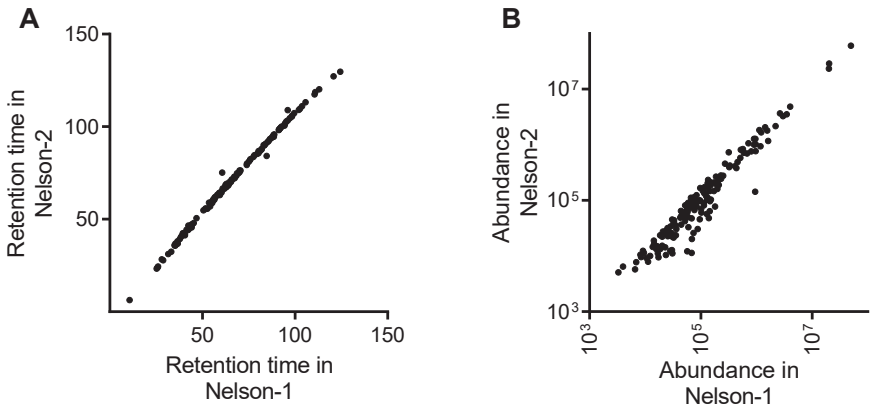


Figure 4. LC-MS performance for high abundant peptides in NELSON 1 and NELSON 2. For Reference samples that were measured during both NELSON 1 and NELSON 2, we compared peptides that were identified with high confidence by a Mascot search with a score of more than 60 in both sets. For this subset of peptides, we compared the retention times observed in NELSON 1 and NELSON 2 (A) and also their abundance (B). For these parameters we observed r-square values of 0.996 and 0.995, respectively.

egy or the process generated randomly selected features (chance). Therefore, the significant peptides from NELSON 1 were analyzed as a next step in a multivariate way.

Antibody Peptide Model

An optimal combination of 12 peptides was identified by the multivariate statistics used on the NELSON 1 set (discovery set). This combination of peptides could distinguish lung cancer patients from controls with sensitivity and specificity of 96% and 100%, respectively. This antibody peptide model was able to detect lung cancer 373 days on average (range 39–1193 days) before the diagnosis was determined. In Figure 5 we show that the combination of the 12 peptides was able to distinguish cases from controls. The 12 peptides corresponded to 1 sequence overlapping with the CDR2 region, 1 sequence overlapping CDR3 region, 7 sequences overlapping the Framework 1 region and 3 sequences overlapping with the Framework 3 region according to the IMGT database (Table 1).

We performed an external validation in the NELSON 2 (validation) set. When we applied the same 12 peptide model to this set, cases and controls could no longer be distinguished. However, with the same peptides but after re-optimization of the model coefficients, we observed a sensitivity and specificity of 84% and 90%, respectively. As the coefficients of the equation are adjusted we had to check for the chance of overfitting of the data. Therefore, a background evaluation was performed which will be described later. Within the NELSON 2 validation set the combination of peptides was able to detect lung cancer 281 days on average (range 54–777 days) before the diagnosis of lung cancer.

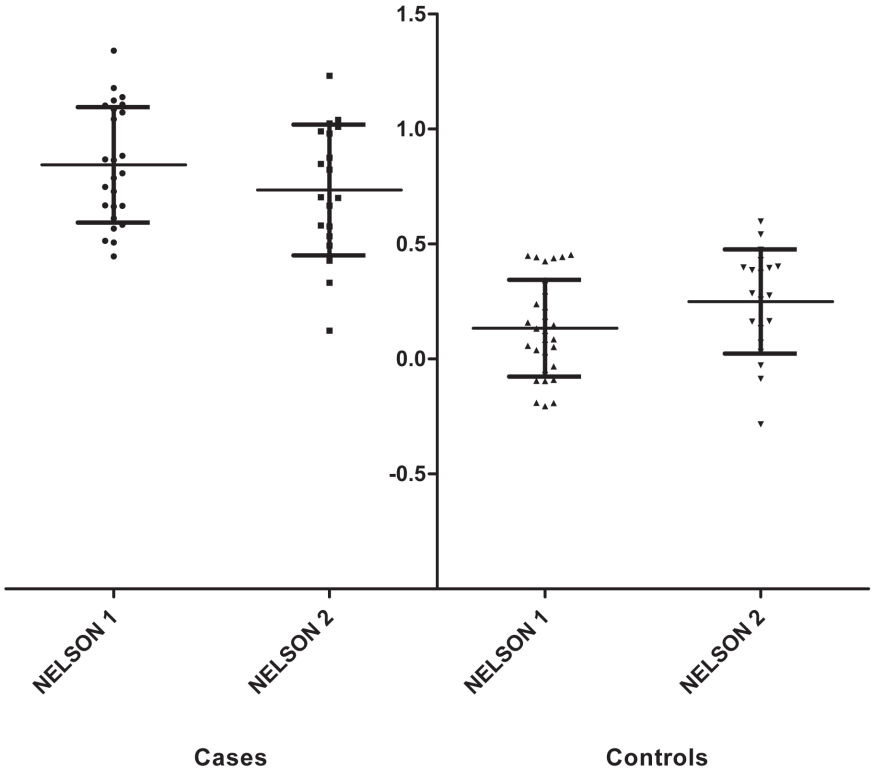


Figure 5. Distribution of the antibody peptide model outcome of the NELSON 1 and NELSON 2 sets. The raw abundances are filled-in in the model equation ($y= a_1x_1 + a_2x_2 + a_3x_3 \ldots.a_nx_n + c$) of the relevant sample set. On the y-axis (in arbitrary units) the figures generated by the equation are shown.

We compared the raw abundance of the 12 peptides between the two NELSON datasets. We observed that the average raw abundance of five peptides was higher in the cases compared to the average abundance of the controls from the NELSON 1 dataset. These data were consistent with the findings from the NELSON 2 dataset The other seven peptides had a higher average raw abundance in the controls of the NELSON 1 dataset compared to the abundance in the cases of this dataset. For only one of these seven peptides, this difference could be confirmed in the NELSON 2 dataset.

Background evaluation of antibody peptide model

In addition to the finding of the optimal combination of peptides which significantly distinguished cases from controls, a background analysis was performed. As the coefficients of the equation of the model were adjusted for each dataset we verified the results for a contribution of random selection of the data and thereby the chance of finding a comparable model by chance. The same workflow was applied for the model building except that at the beginning of the workflow the cases and controls of NELSON 1 were permutated at random

Table 1. Information of the 12 peptides of the antibody peptide model.

Peptide	CDR/Fr	Sequence	m/z	Charge	Protein description	BLAST bit score	IMGT % identified	*p*-value
1	Fr1	GITLSVRP	421.758	2	gi|553734: T-cell receptor	16.8	83.3	0.045
2	Fr3	LMAWLDLK	503.274	2	De Novo Peptide	15.9	75.0	0.009
3	CDR2	IYWDDDKR	555.763	2	gi|39938054: Immunoglobulin heavy chain variable region	32.9	100.0	0.039
4	Fr3	SYPLTFGGGTK	564.288	2	gi|4378188: Immunoglobulin kappa variable region	26.5	100.0	0.014
5	CDR3	LLLYTGGDQR	568.301	2	De Novo Peptide	18.0	75.0	0.030
6	Fr1	EVLLVESGGGLVKPGGSLR	623.025	3	gi|2072264: Immunoglobulin heavy chain	53.2	94.7	0.017
7	Fr3	NTVFLEMNSLR	670.336	2	gi|112699425: Immunoglobulin heavy chain variable region	32.9	81.8	0.040
8	Fr1	HVQLQESGPGLVK	696.386	2	De Novo Peptide	39.7	92.3	0.031
9	Fr1	SYSCQVTHEGSTVEK	827.872	2	gi|16554039: Immunoglobulin heavy chain	50.7	75.0	0.036
10	Fr1	SELTQDPAVSVALGQTVR	936.000	2	gi|87901: Immunoglobulin lambda variable region	57.1	100.0	0.030
11	Fr1	VSSVRCTSGGGLVQPGGSLR	959.501	2	gi|112702369: Immunoglobulin heavy chain variable region	40.1	100.0	0.042
12	Fr1	REMTKPPSVSVSETSHR	964.487	2	De Novo Peptide	29.1	81.8	0.013

CDR, Complementarity Determining Region; Fr, Framework.

(Figure S2). Discovery was performed in the 12 times permutated NELSON 1 datasets, each time with 12 different peptides showing the lowest p-value ($p<0.05$) in the NELSON 1 set for that particular permutation. Validation of these models was performed in NELSON 2. The performance of the multivariate model of the permutated discovery sets (NELSON 1) is shown in Figure 6A (black dots) where the sensitivity is plotted against the specificity. The corresponding power in the validation sets (NELSON 2) is shown in Figure 6B (black dots). Thus, each point in Figure 6A (black dot) corresponds with a point (black dot) in Figure 6B. Also, the performance found for the actual datasets in which the antibody peptide model was found is plotted (red dot). It can be observed that the multivariate fitting from the permutated datasets produces reasonable models even for permutated data in the discovery set.

However, especially in the validation datasets, the real data (antibody peptide model) performed significantly better ($p<0.05$) than the permutated datasets, suggesting that the immunoglobulin peptides harbor information related to the disease state of the patient. Thus, the results we obtained do not stem from an artifact in the data processing.

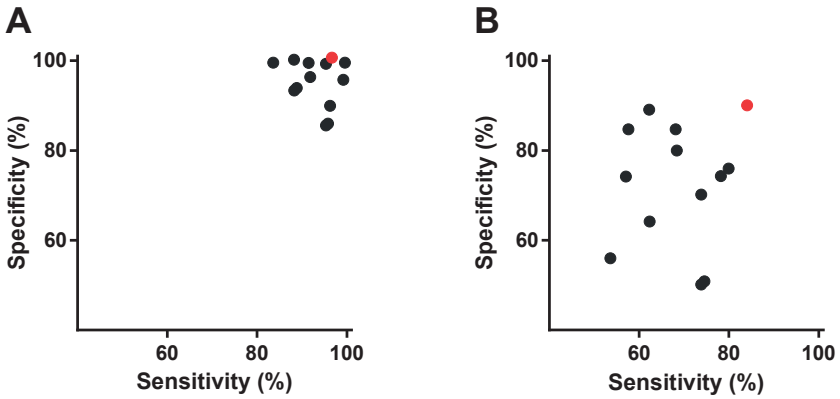


Figure 6. Background determination in NELSON 1 and NELSON 2 datasets. Twelve times a permutation (Background) was performed on the NELSON 1 and NELSON 2 dataset. The sensitivity and specificity of the antibody peptide model are shown in red. Background assessment: A) Twelve permutation runs are shown with the corresponding sensitivity and specificity of the NELSON 1 dataset (black). The same 12 peptides found in the background evaluation of NELSON 1 were tested in NELSON 2. B) The 12 runs are shown with the corresponding sensitivity and specificity of NELSON 2 dataset (black). Note, as some results of the background analysis occurred more than once, a random number between -1 and 1 were added to each sensitivity and specificity number to make sure each analysis (black dot) can be seen in the figure.

CT Screening Result in NELSON 1 and NELSON 2 Dataset

In Figure 7A and 7B the screening results of the baseline CT scans are shown for the NELSON 1 and NELSON 2 set, respectively. According to the screening protocol of the NELSON trial, a repeat CT scan was performed following an indeterminate screening result, approximately 3 months later.

We observed that 68% of the cases had a positive screening result in both the NELSON 1 and NELSON 2 set during the first 3 months of the screening program, the other lung cancers were diagnosed following another repeat CT scan after 3 months or during the second screening round. After on average 367 days (range 39–1193 days) for NELSON 1 and 269 days (range 54–777 days) for NELSON 2, the screening result was positive, i.e. suspect for lung cancer and resulting in clinical work-up by the pulmonologist and eventually finally diagnosis of lung cancer.

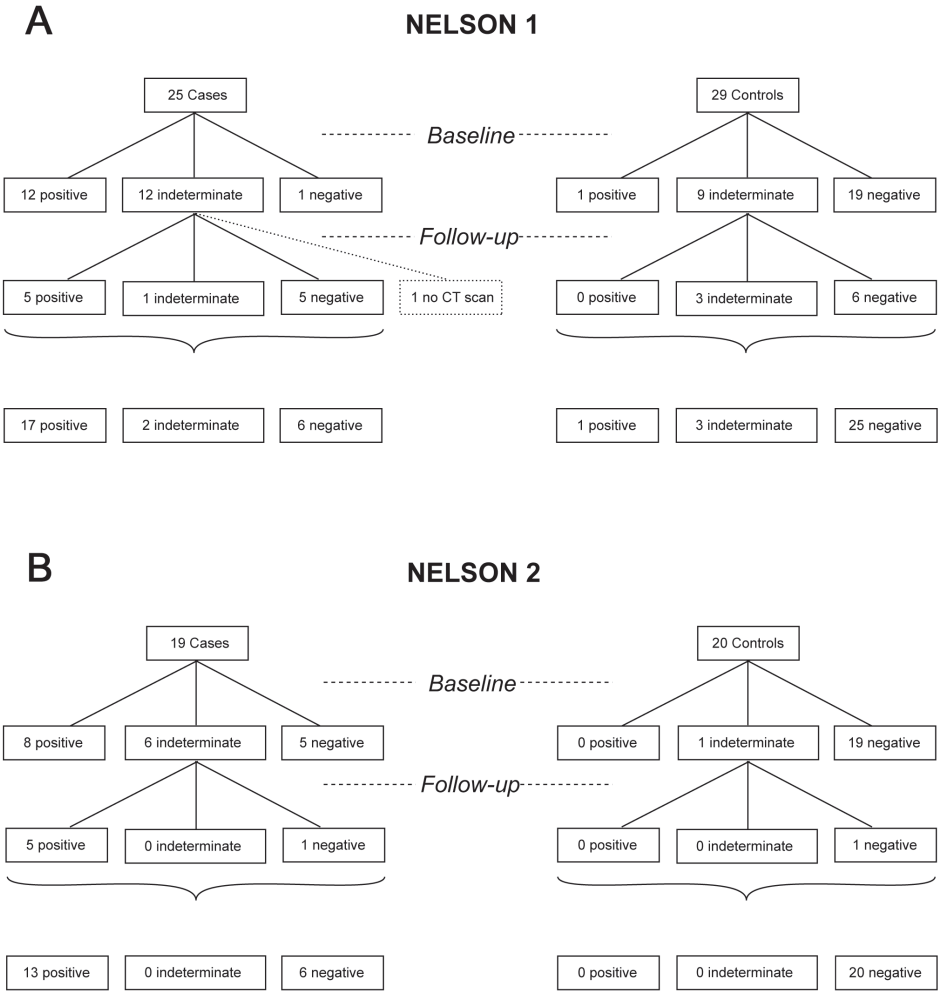


Figure 7. CT scan results of the NELSON 1 and NELSON 2 sample set. CT scan results of the A) NELSON 1 and B) NELSON 2 sample sets are shown at time of blood sampling (Baseline). Also, CT results are shown of the follow-up CT scan after approximately three months (Follow-up). For one case from the NELSON 1 set no Follow-up CT scan result was available. The last row represents the numbers of positive, indeterminate and negative CT scan results of baseline including follow-up results.

DISCUSSION

By mass spectrometry we found evidence that a proportion of peptides of the variable part of antibodies differ between lung cancer patients and controls. A combination of 12 different peptides was able to distinguish lung cancer patients from controls in a high risk population. A sensitivity of 96% and a specificity of 100% were observed in the discovery set. An external validation in an independent case–control set was performed and generated a sensitivity of 84% and a specificity of 90%. The background evaluation showed that the 12 antibody peptide model performed significantly better than a model generated based on permutated data.

Recently, Arentz et al. published that uniquely mutated V regions peptides could be used as a proxy for the detection of anti-Ro52 autoantibodies in sera from primary Sjögren's syndrome patients by mass spectrometry.[28] Why these and other studies were able to identify similar or identical sequences could be explained by repertoire bias and the convergent evolution of antibodies during somatic mutation and selection.[19, 20] This selection favors specific alleles and sequences of antibodies with the optimal affinity towards the specific antigens during immune response[18, 29, 30]

We were able to identify peptide sequences which were distributed differently between lung cancer patients and controls. The antibody peptide model consisted not only of peptide sequences positioned at the CDR regions of an immunoglobulin but also at the framework region surrounding the CDRs. It may appear surprising that most of the peptides that are represented in the antibody peptide model derive from framework regions of the immunoglobulin, rather than from the hypervariable CDRs. This may be explained by their abundance in the immunoglobulin pool. Peptides carrying only few mutations relative to the germline are more likely to occur in several antibody clones, and thus have a higher abundance. This favors their detection by the mass spectrometer, especially in samples of high complexity. While technological advances may enable the reliable quantitation of also lower abundant peptides, it may even be that hypermutated CDRs are not as likely to be common among patients sharing an immune response. But moderately mutated peptides strike the best balance between specificity, abundance and sharing for the purposes of a diagnostic marker. The large heterogeneity of lung cancer could also contribute to the presence of fewer CDR peptides shared between lung cancer patients.

We observed that the average raw abundance of 6 from the 12 peptides was distributed differently in the cases versus controls between the two datasets. The average raw abundance of these six peptides was higher in the controls in the NELSON 1 set but in the NELSON 2 set the average raw abundance was higher in the cases. This may be due to the increased technical variation we observed for lower abundance peptides between the sets that were measured some time apart on different LC columns. While the system operated nominally

for abundant peptides, possibly the performance close to the detection limit cannot be held constant over time, affecting reliable detection and quantification of such peptides.

For our validation set, NELSON 2, we used all disease stages in contrast to NELSON 1. In NELSON 1 we only used early stage I and II. Using different stages of lung cancer could also contribute to the average raw abundance discrepancies between NELSON 1 and NELSON 2. It could be that tumor-specific antibodies are more abundant in sera from early stage lung cancer patients compared to late stage lung cancer patients. We repeated our data analysis for cohorts that were a mixture of NELSON 1 and -2 data. While this reduced the clinical differences between the Discovery and Validation sets, advantages from this improvement were outweighed by the technical differences between the samples. While similar trends were observed, they were not as strong as those shown in Figure 6 (Figure S4).

We also have to cope with the high variability of immunoglobulins, which make the samples probably too complex for the mass spectrometer. A solution to this problem could be reduction of the complexity of the sample before it is measured on the mass spectrometer. This reduction could be established by fractionation into smaller protein fragments such as Fab-κ and Fab-λ, or by producing immunoglobulin fragments containing just the variable domains of the IgG molecule.

It was our aim to offset biological variation by including a relatively large number of patients in this study, but unfortunately large sample numbers translate to extended measurement times of up to 8 weeks for a dataset. These measurement times introduce technical variation that counteracts the advantage gained from the number of included patients.

We were not able to distinguish lung cancer cases from controls univariately by one peptide. Instead we needed a panel of different peptides to discriminate significantly between cases and controls. Lung cancer is a very heterogeneous disease which results in high variability between patients and cancer types. This might induce various immune responses to different tumor antigens. Therefore, finding only one antibody that is shared between all lung cancer patients is highly unlikely. Brichory et al. for instance showed for PGP 9.5, annexin I and II a sensitivity of only 14%, 30% and 33%, respectively.[31, 32] Chapman et al. tested a panel of seven TAAs and found a sensitivity of 41% and a specificity of 93%. Validation of this panel in an independent sample set showed a sensitivity and specificity of 47% and 90%, respectively.[33] Koziol et al. were able to distinguish lung cancer patients from normal individuals with a panel of seven TAAs. A sensitivity of 80% and a specificity of 90% were observed, but no validation was performed.[34] Moreover, Khattar et al. and Zhong et al. were able to identify validated autoantibody peptide panels for lung cancer screening with sensitivity and specificity ranging from 84%–91% and 73%–91%, respectively.[35, 36] It is therefore not surprising that no single peptide could be found in the current data set that distinguishes cases from controls.

Using a multivariate model, we were able to distinguish lung cancer patients from controls. However, due to the experimental and biological variation, it was necessary that we

recalibrated our model for each group of patients. This limits the current applicability of the method in the clinical practice, at least until significant technical advances enable a more robust quantification and identification of peptides in such complex samples. Still, we conclude from our data that differences exist between the immunoglobulin-derived peptides from early lung cancer patients and controls. This is corroborated by data from earlier studies in our own group as well as in other research groups that showed conservation and sharing of rearranged immunoglobulin sequences in immunoglobulins against a particular antigen.[19, 20, 28]

So far, only age and smoking history have been used as selection criteria for enrolment in screening trials, but it is well known that even though over 80% of all lung cancer cases are directly related to smoking, only 11% of female smokers and 17% of male smokers will be diagnosed with lung cancer during their lifetimes.[37, 38] Therefore, additional diagnostic tests might select high risk individuals more precise when combined with the selection criteria age and smoking history in screening trials. The cases and controls we used for this study were selected based on their diagnosis of lung cancer within three years (range 39–1193 days) after the baseline CT scan. Therefore, calculation of sensitivity and specificity of CT screening in our subset of cases and controls from the NELSON trial are not applicable in this retrospective study. However, in this study we have demonstrated that 68% of the cases were detectable for lung cancer by CT screening. At the same time point the CT scan was performed, the antibody peptide model was able to detect lung cancer in 96% and 84% of the cases in the NELSON 1 and NELSON 2 set, respectively. Eventually after approximately 1 year the screening result of all cases were positive by CT screening.

In the high risk population of the NELSON trial still approximately 27% of the participants are subjected to invasive and expensive follow-up studies that revealed in benign disease at baseline CT screening.[5] The performance of CT improves after follow-up scans, but only after an amount of time has passed, on average a year for the sets in this study. Thus, there is need for additional diagnostic capabilities that can improve the performance of the current testing at baseline. For example, the group of Massion recently published their results on a combination of a serum proteomic biomarker panel with clinical and CT data.[39] In the current study, we were able to detect lung cancer with an antibody peptide model in the NELSON 1 and NELSON 2 set with sensitivities of 96% and 84% and specificities of 100% and 90%, respectively at an early stage. This indicates that specific antibodies are present at an early disease stage and that such a panel of antibodies is able to detect lung cancer at an earlier stage than CT. Auto-antibody profiling has the potential to be a tool for early detection when incorporated into a comprehensive screening strategy if technical challenges described in this study can be overcome.

In conclusion, a panel of antibody peptides is identified that discriminates samples of lung cancer patients from controls. This is a first indication that peptides generated from the variable part of antibodies are shared between lung cancer patients and can be used to

discriminate lung cancer patients and control groups. More quantitative work is still needed to assess the use of these peptides in clinical settings.

ACKNOWLEDGMENTS

The authors would like to thank Frank Santegoets and Roel Faber from the department Public Health, Erasmus Medical Center for providing us the CT scan data of the used cases and controls from the NELSON trial in this study.

REFERENCES

1. Ferlay J SI, Ervik M, Dikshit R, Eser S, et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer, 2012.
2. Jemal A, Siegel R, Xu J and Ward E. Cancer statistics, 2010. *CA Cancer J Clin*. 2010; 60: 277-300.
3. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011; 365: 395-409.
4. Horeweg N, van der Aalst CM, Vliegenthart R, Zhao Y, Xie X, Scholten ET, et al. Volumetric computed tomography screening for lung cancer: three rounds of the NELSON trial. *Eur Respir J*. 2013; 42: 1659-67.
5. van Klaveren RJ, Oudkerk M, Prokop M, Scholten ET, Nackaerts K, Vernhout R, et al. Management of lung nodules detected by volume CT scanning. *N Engl J Med*. 2009; 361: 2221-9.
6. Anderson KS and LaBaer J. The sentinel within: exploiting the immune system for cancer biomarkers. *J Proteome Res*. 2005; 4: 1123-33.
7. Baldwin RW. Immunity to transplanted tumour: the effect of tumour extracts on the growth of homologous tumours in rats. *Br J Cancer*. 1955; 9: 646-51.
8. Caron M, Choquet-Kastylevsky G and Joubert-Caron R. Cancer immunomics using autoantibody signatures for biomarker discovery. *Mol Cell Proteomics*. 2007; 6: 1115-22.
9. Gure AO, Altorki NK, Stockert E, Scanlan MJ, Old LJ and Chen YT. Human lung cancer antigens recognized by autologous antibodies: definition of a novel cDNA derived from the tumor suppressor gene locus on chromosome 3p21.3. *Cancer Res*. 1998; 58: 1034-41.
10. Hanash S. Harnessing immunity for cancer marker discovery. *Nat Biotechnol*. 2003; 21: 37-8.
11. Mintz PJ, Kim J, Do KA, Wang X, Zinner RG, Cristofanilli M, et al. Fingerprinting the circulating repertoire of antibodies from cancer patients. *Nat Biotechnol*. 2003; 21: 57-63.
12. Stockert E, Jager E, Chen YT, Scanlan MJ, Gout I, Karbach J, et al. A survey of the humoral immune response of cancer patients to a panel of human tumor antigens. *J Exp Med*. 1998; 187: 1349-54.
13. Liu L, Liu N, Liu B, Yang Y, Zhang Q, Zhang W, et al. Are circulating autoantibodies to ABCC3 transporter a potential biomarker for lung cancer? *J Cancer Res Clin Oncol*. 2012.
14. Murphy K. TP, Walport M. *Janeway's immunobiology*. 7th ed.: Garland Science, 2008.
15. Saada R, Weinberger M, Shahaf G and Mehr R. Models for antigen receptor gene rearrangement: CDR3 length. *Immunol Cell Biol*. 2007; 85: 323-32.
16. Foreman AL, Lemercier B, Lim A, Kourlisky P, Kenny T, Gershwin ME, et al. VH gene usage and CDR3 analysis of B cell receptor in the peripheral blood of patients with PBC. *Autoimmunity*. 2008; 41: 80-6.
17. Scheid JF, Mouquet H, Ueberheide B, Diskin R, Klein F, Oliveira TY, et al. Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science*. 2011; 333: 1633-7.
18. VanDuijn MM, Dekker LJ, Zeneyedpour L, Smitt PA and Luider TM. Immune responses are characterized by specific shared immunoglobulin peptides that can be detected by proteomic techniques. *J Biol Chem*. 2010; 285: 29247-53.

19. Weinstein JA, Jiang N, White RA, 3rd, Fisher DS and Quake SR. High-throughput sequencing of the zebrafish antibody repertoire. *Science*. 2009; 324: 807-10.

20. Maat P, VanDuijn M, Brouwer E, Dekker L, Zeneyedpour L, Luider T, et al. Mass spectrometric detection of antigen-specific immunoglobulin peptides in paraneoplastic patient sera. *J Autoimmun*. 2012; 38: 354-60.

21. 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). *Int J Cancer*. 2007; 120: 868-74.

22. de Costa D, Broodman I, Vanduijn MM, Stingl C, Dekker LJ, Burgers PC, et al. Sequencing and quantifying IgG fragments and antigen-binding regions by mass spectrometry. *J Proteome Res*. 2010; 9: 2937-45.

23. Lefranc MP, Giudicelli V, Ginestoux C, Jabado-Michaloud J, Folch G, Bellahcene F, et al. IMGT, the international ImMunoGeneTics information system. *Nucleic Acids Res*. 2009; 37: D1006-12.

24. Snedecor GW and W.G. C. Statistical Methods. 7 ed. Iowa: The Iowa State University Press, 1980.

25. Armitage P, Berry G and S. MJN. Statistical Methods in Medical Research. 4 ed. Oxford: Blackwell Scientific., 2002.

26. Afifi A, Clark V and May S. Computer-aided multivariate analysis. 4 ed. Florida: Chapman & Hall, 2004.

27. Kleinbaum DG, Kupper LL, Muller KE and Nizam A. Applied regression analysis and other multivariable methods. 3 ed. California: Duxbury Press, 1998.

28. Arentz G, Thurgood LA, Lindop R, Chataway TK and Gordon TP. Secreted human Ro52 autoantibody proteomes express a restricted set of public clonotypes. *J Autoimmun*. 2012; 39: 466-70.

29. Andersen PS, Haahr-Hansen M, Coljee VW, Hinnerfeldt FR, Varming K, Bregenholt S, et al. Extensive restrictions in the VH sequence usage of the human antibody response against the Rhesus D antigen. *Mol Immunol*. 2007; 44: 412-22.

30. Baranzini SE, Jeong MC, Butunoi C, Murray RS, Bernard CC and Oksenberg JR. B cell repertoire diversity and clonal expansion in multiple sclerosis brain lesions. *J Immunol*. 1999; 163: 5133-44.

31. Brichory F, Beer D, Le Naour F, Giordano T and Hanash S. Proteomics-based identification of protein gene product 9.5 as a tumor antigen that induces a humoral immune response in lung cancer. *Cancer Res*. 2001; 61: 7908-12.

32. Brichory FM, Misek DE, Yim AM, Krause MC, Giordano TJ, Beer DG, et al. An immune response manifested by the common occurrence of annexins I and II autoantibodies and high circulating levels of IL-6 in lung cancer. *Proc Natl Acad Sci U S A*. 2001; 98: 9824-9.

33. Chapman CJ, Healey GF, Murray A, Boyle P, Robertson C, Peek LJ, et al. EarlyCDT-Lung test: improved clinical utility through additional autoantibody assays. *Tumour Biol*. 2012.

34. Koziol JA, Zhang JY, Casiano CA, Peng XX, Shi FD, Feng AC, et al. Recursive partitioning as an approach to selection of immune markers for tumor diagnosis. *Clin Cancer Res*. 2003; 9: 5120-6.

35. Khattar NH, Coe-Atkinson SP, Stromberg AJ, Jett JR and Hirschowitz EA. Lung cancer-associated auto-antibodies measured using seven amino acid peptides in a diagnostic blood test for lung cancer. *Cancer Biol Ther*. 2010; 10: 267-72.

36. Zhong L, Coe SP, Stromberg AJ, Khattar NH, Jett JR and Hirschowitz EA. Profiling tumor-associated antibodies for early detection of non-small cell lung cancer. *J Thorac Oncol*. 2006; 1: 513-9.

37. Pass H.I. CDP, Johnson D.H., Minna J.D., Scagliotti G.V., Turrisi A.T. *Principles and practice of lung cancer. The official reference text of the IASLC*. 4th ed.: Lippincott Williams & Wilkins, 2010.

38. Villeneuve PJ and Mao Y. Lifetime probability of developing lung cancer, by smoking status, Canada. *Can J Public Health*. 1994; 85: 385-8.

39. Pecot CV, Li M, Zhang XJ, Rajanbabu R, Calitri C, Bungum A, et al. Added value of a serum proteomic signature in the diagnostic evaluation of lung nodules. *Cancer Epidemiol Biomarkers Prev*. 2012; 21: 786-92.

SUPPLEMENTARY DATA

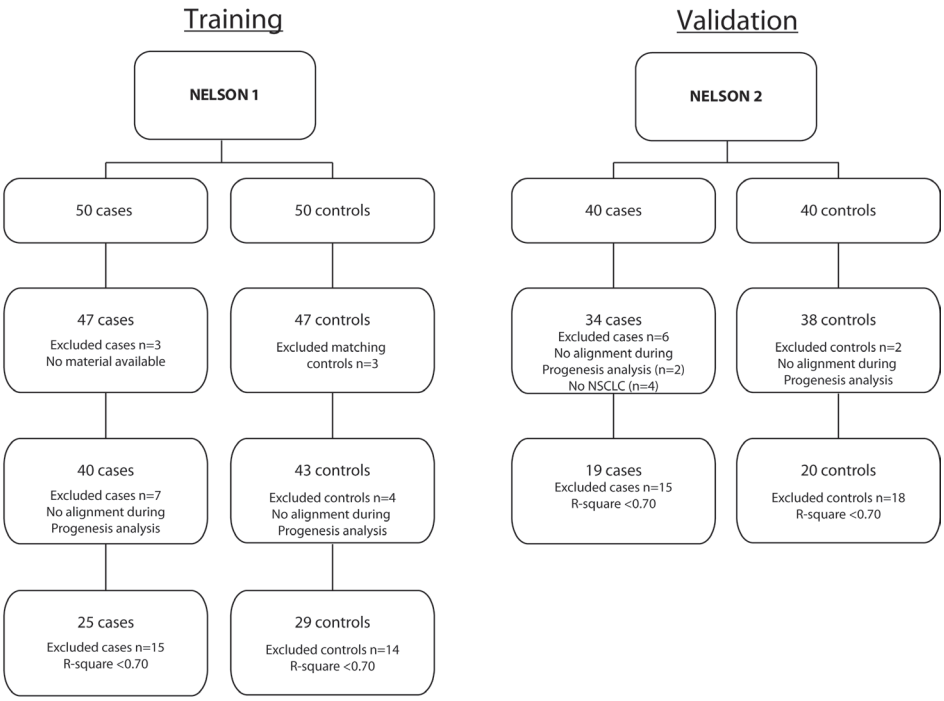


Figure S1. Study Flow-chart. A flow-chart diagram of the samples used in this study. NSCLC: Non-small cell lung carcinoma.

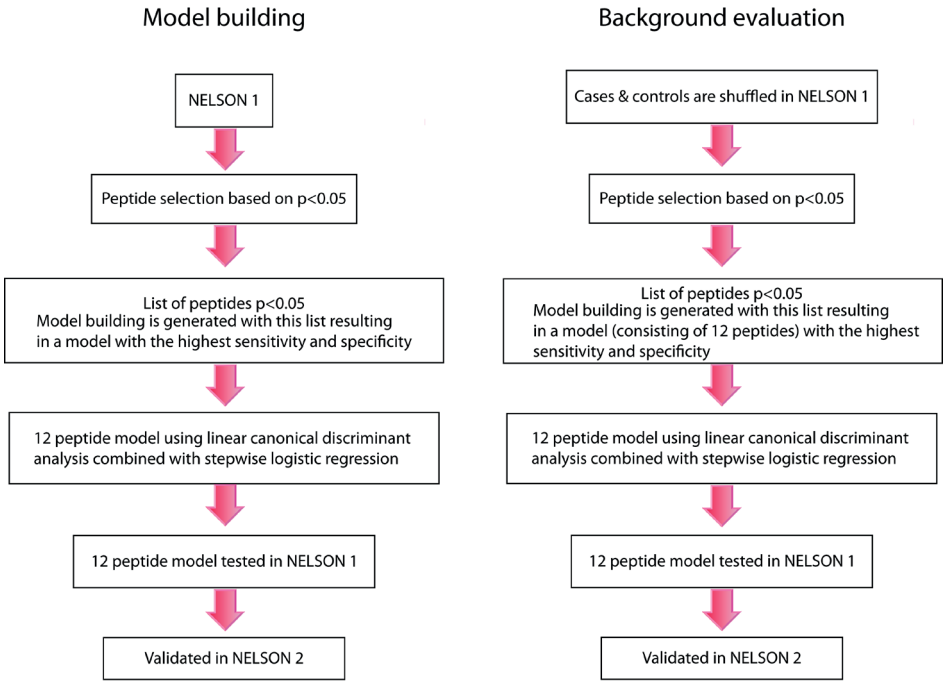


Figure S2. Statistical analysis flow-chart. Before background analysis is performed, cases and controls of the NELSON 1 dataset are shuffled randomly.

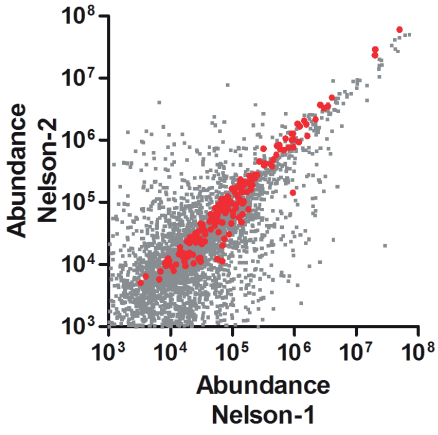


Figure S3. Variation at different abundances. The abundances of all peptides in the reference sample compared in data from the Nelson-1 and Nelson-2 datasets. Superimposed, the subset of peptides that was identified with high confidence, as plotted in Figure 4B, has been superimposed in red.

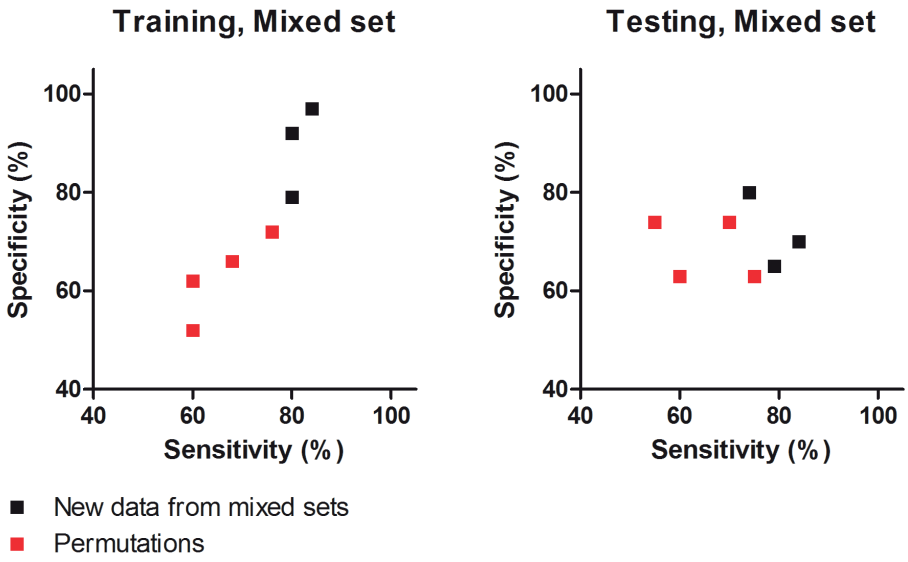


Figure S4. The performance of the prediction model was tested in Training and Testing sets, for both real data, and data in which the assignment of cases and controls had been randomized. This approach is the same as in Figure 6, except that each set was composed of samples drawn from a combination of both the Nelson-1 and Nelson-2 sets. We assessed three such combinations, and four permutations

Table S1. Clinical characteristics of the NELSON 1 and NELSON 2 sample sets.

	NELSON 1					NELSON 2				
	Cases		Controls		*p*-value	Cases		Controls		*p*-value
	n=25		n=29			n=19		n=20		
Gender					0.847					0.589
Male	22	88.0%	26	89.7%		16	84.2%	18	90.0%	
Female	3	12.0%	3	10.3%		3	15.8%	2	10.0%	
Age (years)	62.0	(±6.5)	62.2	(±4.8)	0.754	62.2	(±7.3)	62.2	(±6.8)	0.944
Men	61.8	(±6.6)	62.5	(±5.0)	0.611	63.3	(±7.4)	63.3	(±6.2)	1.000
Women	63.3	(±6.8)	60.0	(±1.7)	0.637	56.7	(±3.5)	52.5	(±0.7)	0.139
Smoking status[a]					0.389					<0.001
Current	15	60.0%	14	48.3%		10	52.6%	0	0.0%	
Former	10	40.0%	15	51.7%		9	47.4%	20	100.0%	
Smoking duration (years)					0.108					0.308
26-40	14	56.0%	11	37.9%		12	63.2%	11	55.0%	
41-45	6	24.0%	15	51.7%		4	21.1%	8	40.0%	
> 45	5	20.0%	3	10.3%		3	15.8%	1	5.0%	
Cigarettes/day					0.891					0.355
0-15	5	20.0%	6	20.7%		3	15.8%	8	40.0%	
16-20	8	32.0%	7	24.1%		2	10.5%	2	10.0%	
21-25	5	20.0%	8	27.6%		4	21.1%	4	20.0%	
> 25	7	28.0%	8	27.6%		10	52.6%	6	30.0%	
COPD[a]					0.641					0.024
Yes	10	40.0%	12	41.4%		5	26.3%	0	0.0%	
No	14	56.0%	14	48.3%		13	68.4%	20	100.0%	
Unknown	1	4.0%	3	10.3%		1	5.3%	0	0.0%	

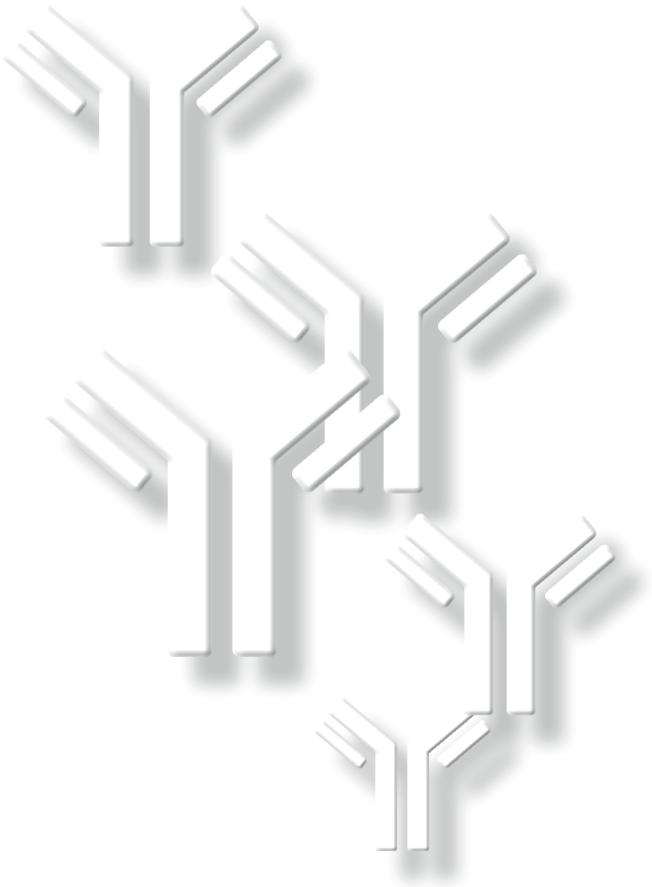
Table S1. Clinical characteristics of the NELSON 1 and NELSON 2 sample sets. (continued)

	NELSON 1					NELSON 2				
	Cases		Controls		*p*-value	Cases		Controls		*p*-value
	n=25		n=29			n=19		n=20		
Asbestos exposure					0.542					0.676
Yes	5	20.0%	4	13.8%		2	10.5%	3	15.0%	
No	20	80.0%	25	86.2%		17	89.5%	17	85.0%	
Center					0.275					0.158
Groningen	4	16.0%	9	31.0%		4	21.1%	7	35.0%	
Utrecht	9	36.0%	12	41.4%		4	21.1%	6	30.0%	
Haarlem	11	44.0%	6	20.7%		7	36.8%	7	35.0%	
Leuven	1	4.0%	2	6.9%		4	21.1%	0	0.0%	
Time to cancer (years)[b]										
0-0.5	11	44.0%	-	-		12	63.2%	-	-	
0.5-1.5	7	28.0%	-	-		4	21.0%	-	-	
1.5-2.5	6	24.0%	-	-		3	15.8%	-	-	
2.5-3.5	1	4.0%	-	-		0	0.0%	-	-	
Histology										
Adenocarcinoma	21	84.0%	-	-		8	42.1%	-	-	
Squamous cell carcinoma	4	16.0%	-	-		2	10.5%	-	-	
Other[c]	0	0.0%	-	-		9	47.4%	-	-	
Pathological stage										
IA	19	76.0%	-	-		6	31.6%	-	-	
IB	3	12.0%	-	-		0	0.0%	-	-	
II	3	12.0%	-	-		2	10.5%	-	-	
III	0	0.0%	-	-		8	42.1%	-	-	
IV	0	0.0%	-	-		3	15.8%	-	-	

Table S1. Clinical characteristics of the NELSON 1 and NELSON 2 sample sets. (continued)

	NELSON 1					NELSON 2				
	Cases		Controls		*p*-value	Cases		Controls		*p*-value
	n=25		n=29			n=19		n=20		
VDT (Days)										
> 600	0	0.0%	-	-		0	0.0%	-	-	
400-600	4	16.0%	-	-		1	5.3%	-	-	
<400	7	28.0%	-	-		6	31.6%	-	-	
Not applicable[d]	14	56.0%	-	-		13	68.4%	-	-	
Volume (mm^3)										
< 50	0	0.0%	-	-		0	15.8%	-	-	
50-500	6	24.0%	-	-		3	78.9%	-	-	
> 500	19	76.0%	-	-		15	79.4%	-	-	
Unknown	0	0.0%	-	-		1	5.3%	-	-	
Consistency										
Solid	21	84.0%	-	-		17	89.4%	-	-	
Partial solid	3	12.0%	-	-		1	5.3%	-	-	
Non-solid	0	0.0%	-	-		0	0.0%	-	-	
Unknown	1	4.0%	-	-		1	5.3%	-	-	
Benign nodules					0.940					0.861
0	14	56.0%	16	55.2%		10	52.6%	12	60.0%	
1	5	20.0%	5	17.2%		4	21.1%	3	15.0%	
>1	6	24.0%	8	27.6%		5	26.3%	5	25.0%	

COPD: Chronic Obstructive Pulmonary Disease; VDT: Volume Doubling Time; significance of all characteristics were tested by Pearson chi-square except for age (Mann-Whitney U test);[a] Significantly different in NELSON 2 sample set; [b] Interval between blood sampling and diagnosis; [c] NSCLC histology other than adenocarcinoma or squamous cell carcinoma; [d] No repeat scan, no growth (decrease in size or resolution).



Ingrid Broodman, Martijn M. VanDuijn, Christoph Stingl,
Lennard J.M. Dekker, Anastasios E. Germenis, Harry J. de Koning,
Rob J. van Klaveren, Joachim G. Aerts, Jan Lindemans, Theo M. Luider

Survivin autoantibodies are not elevated in lung cancer when assayed controlling for specificity and smoking status

CHAPTER 4

Cancer Immunology Research 2015; 4(2): 1–8

ABSTRACT

The high mortality rate in lung cancer is largely attributable to late diagnosis. Case-control studies suggest that autoantibodies to the survivin protein are potential biomarkers for early diagnosis. We tested the hypothesis that sandwich ELISA can detect autoantibodies to survivin before radiological diagnosis in early-stage non-small-cell lung cancer (NSCLC) patients. Because previous studies assayed survivin autoantibodies with the direct antigen-coating ELISA (DAC-ELISA), we first compared that assay with the sandwich ELISA. Based on the more robust results from the sandwich ELISA, we used it to measure survivin autoantibodies in the serum of 100 individuals from a well-controlled population study (the Dutch-Belgian lung cancer screening trial (NELSON) trial) composed of current and former smokers (50 patients with NSCLC, both before and after diagnosis and 50 matched, smoking-habit controls), and another 50 healthy non-smoking controls. We found no difference in specific autoantibodies to survivin in NSCLC patients, although non-specific median optical densities were 24% higher ($p<0.001$) in both NSCLC patients and smokers, than in healthy non-smokers. Finally, we confirmed the ELISA results with western blot analysis of recombinant and endogenous survivin (HEK-293), which showed no anti-survivin reactivity in patient sera. We conclude that specific anti-survivin autoantibody reactivity is most likely not present in sera before or after diagnosis. Autoantibody studies benefit from a comparison to a well-controlled population, stratified for smoking habit.

INTRODUCTION

Long-term survival in lung cancer depends mainly on early detection and immediate start of treatment of the tumor.[1] During tumor development, tumor-antigen expression elicits cellular and humoral immune responses.[2-4] Identifying autoantibodies to tumor-associated antigens (TAAs) is thought to be a promising method of early lung cancer diagnosis,[5-8] especially because these autoantibodies have been found up to 5 years before CT detection.[9, 10]

Antibodies to survivin are one of the autoantibodies described most frequently in lung cancer.[5, 11-18] Survivin, also known as baculoviral IAP repeat-containing protein 5 (BIRC5), is a member of the inhibitor apoptosis proteins (IAP) family. It promotes cell proliferation and inhibits apoptosis, thereby favoring the growth and progression of transformed cells and tumors. Although survivin is abundantly expressed in fetal cells, transformed cell lines and various tumors, it is undetectable in most normal, differentiated adult tissues.[19] If over-expressed in lung cancer, it may lead to antibody responses to this protein. Various studies have compared amounts of survivin autoantibodies in lung cancer patients and healthy blood-donor control subjecst; the presence of antibodies against survivin in lung cancer sera collected after diagnosis was reported to range between 8%[16] and 58%.[15]

In our institute a well-controlled multi-center population study was conducted aiming at early detection of lung cancer with specific emphasis on smoking habit, the NELSON trial.[20, 21] We wanted to test the hypothesis that,in this population, anti-survivin autoantibodies are detectable before the radiological diagnosis of NSCLC and establish how these autoantibodies may emerge over time in the disease. For that purpose, we collected NSCLC-cases from the NELSON trial with well-matched controls to the NSCLC cases, subdivided into individuals who currently smoking or were former smokers. In addition, we included late-stage NSCLC patients from the Dutch Association of Pulmonologists for Lung Diseases and Tuberculosis (NVALT)-12 study[22] and a control group of generally healthy blood donors, similar to that used in the publications on survivin autoantibodies.[5, 11-18]

All published studies on survivin autoantibodies in lung cancer have used the direct antigen-coating (DAC) form of an ELISA with a recombinant survivin.[5, 11-18] However, the DAC-ELISA assay can give false-positive results if antibodies bind to impurities in the antigen preparation. In our study we therefore chose to assess the presence of autoantibodies to survivin with the sandwich-ELISA. In this assay, a highly specific capture antibody is absorbed to the solid phase and incubated with an antigen solution. Only autoantibodies to survivin are then detected in the serum samples. At the cost of less signal, this greatly increases the specificity of the assay.

MATERIALS AND METHODS

Subjects

We obtained serum from 50 NSCLC cases (82% adenocarcinoma; 90% stage I/II) before and after diagnosis and from 50 smoking-habit controls drawn from current and former smokers in the NELSON trial, as described previously[20, 21] Smoking-habit controls were matched to the NSCLC cases for age, gender, smoking status, smoking duration, number of cigarettes smoked per day, chronic obstructive pulmonary disease (COPD) status, asbestos exposure, and blood collection center (Supplementary Table S1). For the reference population, serum was derived from 50 healthy non-smoking blood donors of the Sanquin Blood Supply Rotterdam. These non-smoking controls were matched to cases and smoking-habit controls for age and gender (Supplementary Table S1). From the NVALT-12 study we obtained baseline serum from 20 patients (10 males and 10 females; 6 current smokers; 12 former smokers; 2 non-smokers; mean age 63.2 ± 9.1 years) with stage IV NSCLC (80% adenocarcinoma; 20% large cell carcinoma). From Sanquin Blood Supply Rotterdam we obtained seven independent serum samples from healthy non-smoking controls with a normal serum IgG (mean 10.5 ± 1.78 g/L). Professor Anastasios E. Germenis[16] made available nine serum samples labeled positive for anti-survivin antibodies by the DAC-assay from NSCLC patients (8 males and 1 female, mean age 57.6 ± 8.0 years) in his laboratory. Upon diagnosis, these patients—part of a cohort of 117 NSCLC patients (108 males and 9 females, mean age 64.2 ± 9.3 years)—had been shown by DAC-ELISA to have higher anti-survivin antibody levels than 100 healthy, age- and gender-matched individuals (99 males and 1 female, mean age 68.2 ± 5.8 years).
For all human samples, a medical ethical statement according to the Declaration of Helsinki was given by the METC Erasmus MC Rotterdam. Written informed consent was obtained from all participants for the use of their samples.

Serum-collection protocol

During the participants' visits to the center, one serum gel tube was collected per participant. The venous blood was allowed to clot, and was centrifuged for 10 min at 1400 x g and 4°C within 2 hours after collection. After centrifigation, the serum was stored immediately in aliquots at -80°C. All samples were blinded and analyzed in random order.

Detection of anti-survivin antibodies by ELISA

Antibody reactivity against recombinant full-length human survivin fused to *calmodulin* (CaM)-tag (BIRC5, Abcam) was measured by DAC-ELISA as described previously[16] and by sandwich ELISA. To validate that both assays detected survivin autoantibodies and to assess their dynamic ranges, we used a rabbit monoclonal antibody (mAb) to survivin (0.06-1000 ng/mL; Abcam, Cambridge, UK) to establish standard curves (Supplementary Figure S1). The standard curve of rabbit anti-survivin showed a dynamic range from 0.06 to 40 ng/mL in the

DAC-ELISA and from 0.06 to 1000 ng/mL in the sandwich ELISA (Supplementary Figure S1). To determine if these assays could distinguish human autoantibody–positive from –negative sera, we analyzed serum from seven NSCLC patients that had been previously reported[16] to be positive for survivin autoantibodies and from seven healthy non-smoking controls, the type of control group used by Karanikas and colleagues.[16] The replicability of our sandwich ELISA was established by repeatedly testing a single sample for intra-assay variation, as well as testing inter-assay variation. The intra-assay coefficient of variation (CV) for the the optical densities (ODs) with survivin of 20 replicates of 1 serum from a non-smoking control sample was 6.9 %, the inter-assay (plate-to-plate) CV for 6 serum samples from non-smoking control controls ranged from 5.3% to 11.0%.

DAC-ELISA for anti-survivin antibodies. One half of a microtiter plate (Nunc-Immuno Maxisorp flat bottom, Thermo Scientific, IL, USA) was incubated with 100 µL of the 2 µg/mL recombinant CaM-tagged survivin in 0.05 mol/L carbonate buffer (pH 9.6) and the other half only with 0.05 mol/L carbonate buffer for 20 hours at 4°C. After washing five times with phosphate-buffered saline containing 0.05% Tween-20 (PBST), plates were blocked with 200 µL 5% bovine serum albumin (BSA, Sigma, USA) in PBST for 20 hours at 4°C. Plates were washed five times with PBST, and wells with and without survivin on the same plate were incubated for 1 hour at room temperature with 100 µL of serum sample either diluted 1:40 as described previously,[16] diluted 1:100, or with serial dilutions of rabbit anti- survivin in PBST. After washing as before, plates were incubated for 1 hour at room temperature with 100 µL of horseradish peroxidase (HRP)-conjugated goat anti-human IgG (H+L) (100 ng/mL; Vector, CA, USA) or HRP-conjugated goat anti-rabbit IgG (Fc) (80 ng/mL; Jackson, PA, USA) in 1% BSA-PBST. Plates were washed six times, and enzyme activity was visualized by adding 100 µL of 3,3',5,5'-tetramethylbenzidine substrate (TMB; Sigma-Aldrich, MO, USA). Absorbance at 450 nm was measured after 10 min. Absorbance values were corrected by blank subtraction.

Sandwich ELISA for anti-survivin antibodies. Microtiter plates were coated overnight at 4°C with 100 µL of 0.50 µg/mL Cam-tag specific mAb (Santa Cruz Biotechnology, TX, USA) in 0.05 mol/L carbonate buffer. Plates were washed five times with PBST and blocked with 200 µL 5% BSA-PBST for 1.5 hours at room temperature. After washing five times with PBST, one half of the plate was incubated with 100 µL of 400 ng/mL recombinant CaM-tagged survivin in 5% BSA-PBST, and the other half with only 5% BSA-PBST for 3.5 hours at room temperature. Plates were washed as before, and 100 µL of serum sample diluted 1:100 or serial dilutions of rabbit anti-survivin in PBST were added to wells with and without survivin on the same plate. After 1-hour incubation at room temperature followed by washing, wells were incubated with 100 µL of (HRP)-conjugated goat anti-human IgG (H+L) (100 ng/mL) or (HRP)-conjugated goat anti-rabbit IgG (Fc) (80 ng/mL) in 1% BSA-PBST for 1 hour at room temperature. After washing six times, enzyme activity was visualized by adding 100 µL of

TMB. After 10 min incubation, absorbance at 450 nm was measured. Absorbance values were corrected by blank subtraction.

Sensitivity and specificity of reagents. The linearity of the sandwich ELISA was determined by diluting the rabbit anti-survivin positive control into six non-smoking control sera and ten NSCLC patient sera, to create a standard curve. ODs of serial dilutions of rabbit anti-survivin in human sera were identical to those spiked in diluent (Supplementary Figures S3 and S4). To validate the utility and specificity of our secondary antibody (goat anti-human IgG (H+L)), we designed a similar sandwich ELISA using as a positive control a different human protein and sera that were known to contain autoantibodies to it. Patients with paraneoplastic neurological syndromes (PNS) can generate autoantibodies against the onconeural antigen HuD.[23] Recombinant HuD antigen (ELAV4) was applied to serial dilutions of a patient serum, known to be positive for anti-HuD antibodies. For this test, one healthy donor sample was used as negative control.
To determine non-specific reactions, samples were analyzed in wells with and without survivin on the same plate. After blank subtraction, the net absorbance at 450 nm for each sample was calculated using the following equation: Net OD = OD with survivin – OD without survivin.

Western blot analysis

With CaM-survivin antigen. To determine the specificity of the DAC-ELISA and sandwich ELISA, selected samples diluted 1:200 were analyzed by Western blotting of CaM-tagged survivin (50 ng/lane). Rabbit anti-survivin serially diluted (1.6 to 200 ng/mL) was analyzed to check the sensitivity of the western blot. Blots were incubated with a 1:10,000 dilution of IRDye© 800CW goat anti-human IgG (H+L) or IRDye© 680RD donkey anti-rabbit IgG (H+L), and scanned with an Odyssey Infrared Imager (LI-COR, NE, USA) to simultaneously visualize proteins at anti-human and anti-rabbit dual wavelengths.

With HEK-293 cell lysate. Blots were performed using HEK-293 cell lysate as a source of endogenous survivin. For equal loading of proteins, 100 µL HEK-293 cell lysate was loaded into a single-well comb slot for western blot analysis. After protein transfer, the blot was cut into strips containing equal amounts of protein per centimeter of strip. Antibody response to HEK-293 antigens was analyzed with either serum samples diluted at 1:200 or rabbit anti-survivin diluted 1:2,500. Proteins were detected by fluorescence as described previously or by chemiluminescence using the ECL western blotting substrate (Thermo Scientific, IL, USA) according to the protocol provided.

Preparation of HEK-293 cell lysate as source of endogenous survivin

Potentially, posttranslational modification of survivin may be necessary for immune recognition and may be a reason that a positive survivin antibody serum is missed. We prepared human embryonic kidney (HEK)-293 cell lysate that endogenously produces survivin.[24] HEK-293 cells were cultured in DMEM with GlutaMax (Life Technologies) supplemented with 10% fetal bovine serum, 1% penicillin and 1% streptomycin at 37°C in 5% CO_2. The identity and homogeneity of the HEK-293 culture was verified by light-microscopic inspection of cell morphology and growth patterns. No additional authentication was performed. Cells were washed with PBS and lysed in 1 mL SDS sample buffer (± 10^6 cells/mL). The cell lysate was centrifuged and stored in aliquots at -20°C until analysis.

NanoLC Orbitrap mass spectrometry measurement of endogenous survivin in HEK-293 cell lysate

Proteins in the HEK-293 cell lysate were separated by SDS-PAGE. Protein bands (band 1-5) in the 15-16 kDa range were tryptic digested for LC-MS measurement as described previously.[25] Tryptic peptides were measured by LC-MS on an Ultimate 3000 nano-RSLC system (Dionex, Amsterdam, Netherlands) coupled online to a hybrid linear ion trap/Orbitrap MS (Orbitrap Fusion, Thermo Fisher Scientific, San Jose, CA, USA). The trap column was then switched online with the analytical column (PepMap C18, 75 µm ID x 500 mm, 3 µm particle and 100 Å pore size; Dionex), and peptides were eluted by a 90-minute gradient from 4 to 38% acetonitrile at a flow rate of 250 nL/minute. For electro-spray ionization (ESI), metal-coated nano ESI emitters (New Objective, Woburn, MA, USA) were used at a spray voltage of 1.7 kV. Two different methods for MS detection were used. First, all samples were run according to a data-dependent shotgun method whereby MS1 survey scans of 400-1600 m/z were acquired in the Orbitrap at 120,000 resolution. Subsequently, CID MS/MS spectra were acquired in the linear iontrap using the top speed mode of the instrument. Second, in a further measurement of protein band 2, the method described above was adjusted to exclusively trigger MS/MS spectra of precursor masses matched with the predicted tryptic peptides of survivin within a 12 ppm m/z window.

Statistical analysis

Before samples were included in this study, the clinical characteristics of non-smoking controls, NSCLC cases and smoking-habit controls were statistically analyzed. To detect significant differences in absorbances (OD) among the data sets, the Mann-Whitney U-test (two-tailed) was performed. Pearson correlation and linear regression analysis were performed to assess the relationship between the OD of samples measured in wells with survivin and those measured without survivin. All data were analyzed by SPSS (IBM SPSS Statistics 21). A *p*-value of <0.05 was considered statistically significant.

RESULTS

Before we started our studies on the development of survivin autoantibodies in NSCLC, we tested both the DAC-ELISA and the sandwich ELISA for robustness and specificity.

DAC-ELISA for survivin antibodies

The antibody response to human survivin was measured in seven NSCLC serum samples previously found positive for survivin antibodies in the provider's laboratory[16] and in seven serum samples from healthy non-smoking controls. Tests were carried out at a 1:40 (Figure 1A) and a 1:100 dilution (Supplementary Figure S2). Comparisons between patients and healthy non-smokers were made. The Mann-Whitney U-test showed no significant difference between the NSCLC cases and the healthy non-smoking controls at two different dilutions (1:40 dilution, $p = 1.000$; 1:100 dilution, $p = 0.620$). In addition, no significant differences were found in specific binding, between OD readings in assays that contained the antigen, survivin, and those without antigen, either at a 1:40 ($p = 0.535$) or 1:100 ($p = 0.535$) dilution.

Sandwich ELISA for survivin antibodies

We first determined the replicability of our sandwich ELISA by repeatedly testing a single sample for intra-assay variation, as well as testing inter-assay variation (see Materials and Methods), and found no sample matrix interference in the sandwich ELISA (Supplementary Figures S3 and S4).

To determine whether our sandwich ELISA technique can detect known human autoantibodies, we tested the assay using the onconeural antigen HuD (ELAV-like protein 4) and serial dilutions of an anti-HuD-positive serum from a patient with PNS. Serum from a healthy donor was used as negative control. The patient serum gave a linear standard curve at serum dilutions of 1:10 to 1:640 and a relatively high OD (2.378) at the 1:100 dilution that is commonly used for immunohistology (Supplementary Figure S5). The negative control serum had a low OD (0.147) at a dilution of 1:100. When antigen was omitted from the sandwich ELISA and serum used at a1:100 dilution, the patient serum had a much lower OD (0.102), but the OD of the negative donor serum was virtually unchanged (0.146). Thus, the anti-human secondary antibody reaction works correctly in demonstrating the presence of immobilized antigen-bound human auto-antibody in our sandwich assay.

Comparison of DAC-ELISA to sandwich ELISA

Having validated the sandwich ELISA as capable of detecting human autoantibodies, we retested the same seven NSCLC patient and healthy control samples that we had assayed with the DAC-ELISA, at a serum dilution of 1:100. We found that specific binding was essentially zero, ranging from 0.000 to 0.008. We observed that ODs with and without antigen were significantly higher for patients (median 0.108; interquartile range (IQR) 0.095-0.121)

than for the non-smoking control subjects (median 0.075; IQR 0.065-0.092; Mann-Whitney U-test, p = 0.011) and were strongly correlated (Figure 1B).

Survivin sandwich ELISAs of NELSON samples

We tested sera for the presence of autoantibodies to survivin from 50 NSCLC cases,before and after diagnosis, 50 matched, smoking-habit controls from the NELSON trial, and another 50 non-smoking controls (n=200). All samples were randomized before measurement. Specific reactions (subtracting background values obtained in the absence of survivin in the sandwich ELISA) were almost zero, ranging from –0.004 to –0.001 (Figure 2). However, as

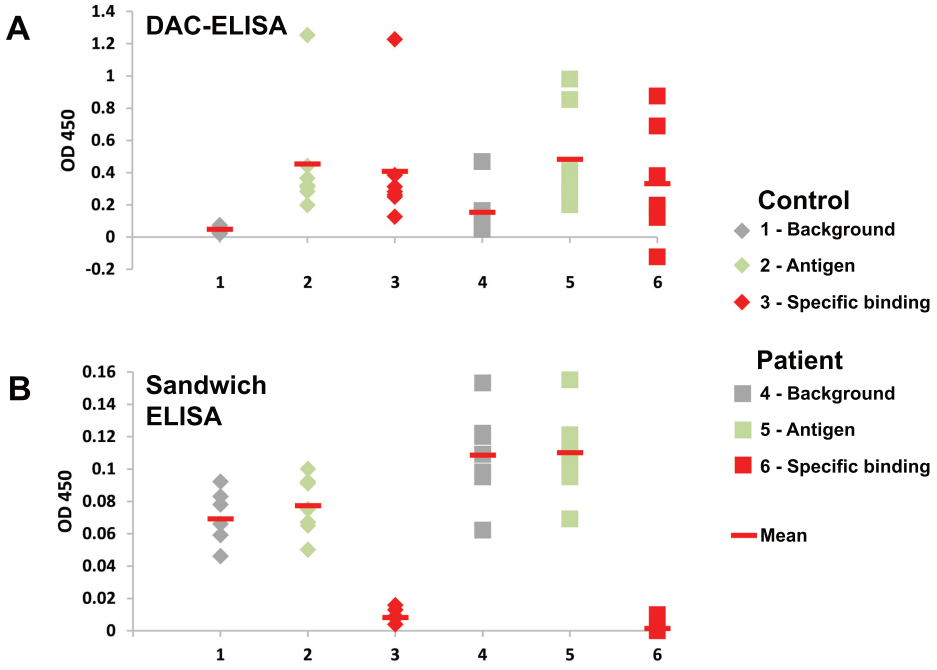


Figure 1. Antibody response to recombinant survivin was measured in seven NSCLC patient serum samples previously reported to be positive for survivin antibodies[16] and in serum samples from seven healthy non-smoking control subjects. DAC-ELISA (A) was carried out at a 1:40 dilution and sandwich ELISA at a 1:100 dilution (B). ODs were measured in assays that contained the antigen, survivin (green), and those without antigen (grey). Specific binding (red) was obtained by subtracting background ODs (grey) obtained in the absence of survivin. A, No significant difference was found between ODs with antigen from the NSCLC patients (square; median 0.345; IQR 0.271-0.852) and those of the healthy non-smoking control subjects (rhombus; median 0.320; IQR 0.284-0.442; Mann-Whitney U-test (MW), p = 1.000). In addition, no significant differences were found in specific binding (MW, p = 0.535). B, ODs with and without antigen were significantly higher for NSCLC patients (median 0.108; IQR 0.095-0.121) than for the healthy non-smoking control subjects (median 0.075; IQR 0.065-0.092; MW, p = 0.011) and were strongly correlated. Pearson correlation and linear regression between OD with survivin and OD without survivin of these samples resulted in an R^2 value of 0.96 (p<0.001) with an intercept of 0.014 (95% CI, 0.003-0.025) and a slope of 0.896 (95% CI, 0.781-1.011). Specific binding was essentially zero, median Net ODs (ODs with survivin – ODs without survivin) ranged from 0.000 to 0.008.

noted with the sera from the study by Karanikas and colleagues,[16] the background median OD was significantly higher in NSCLC patients ($p< 0.001$) and smoking-habit controls than in non-smoking controls (Figure 2 and Table 1). Thus, the choice of control may be critical for these studies if background is not removed before reporting data.

We also tested for autoantibodies in sera of 20 NSCLC patients from the late-stage NVALT-12 study (median 0.051; IQR 0.027), 13 smoking-habit controls (median 0.059; IQR 0.045, $p = 0.253$), and 13 non-smoking controls (median 0.062; IQR 0.035, $p = 0.113$). None of these late-stage NSCLC cases was positive for antibody binding to survivin.

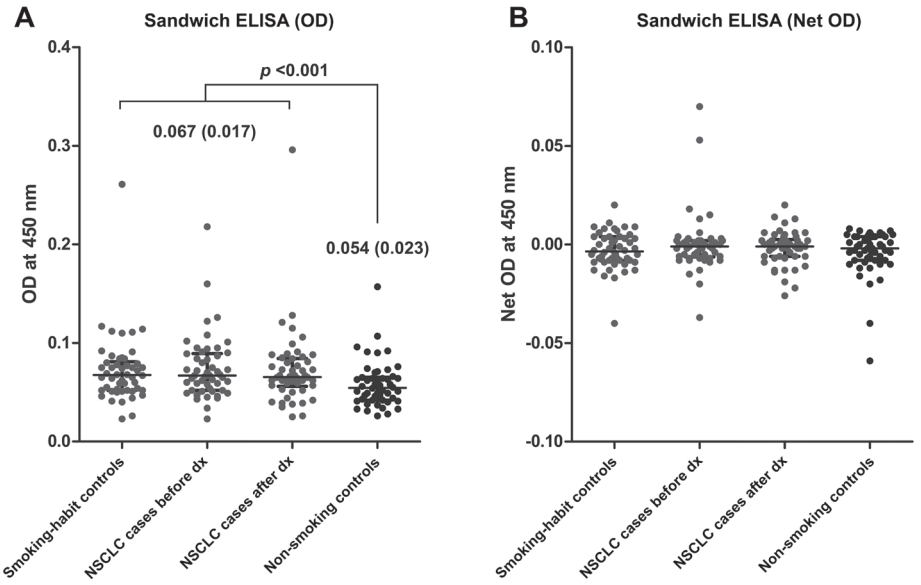


Figure 2. Results of sandwich ELISA for antibody response to recombinant human survivin in sera from 50 smoking-habit controls, 50 cases measured before and after diagnosis (dx) of NSCLC, and 50 healthy non-smoking controls. Data, OD at 450 nm (A) and Net OD at 450 nm (Net OD = OD with survivin – OD without survivin) (B) with bars representing median OD (IQR) for each group. A, no difference in specific autoantibodies to survivin was found in NSCLC patients, although nonspecific median ODs were 24% higher ($p < 0.001$) in both NSCLC patients and smokers, than in healthy non-smokers. B, specific median ODs (subtracting background values obtained in the absence of survivin in the sandwich ELISA) were almost zero, ranging from –0.004 to –0.001. For the 200 samples, ODs with survivin and ODs without survivin were strongly correlated. Pearson correlation and linear regression analysis of log-transformed OD values resulted in an R2 value of 0.83 ($p < 0.001$) with an intercept of –0.117 (95% CI, –0.280-0.053) and a slope of 0.971 (95% CI, 0.781-1.011).

Sandwich ELISA results confirmed by western blot analysis

CaM-tagged survivin was used in western blots for further testing of serum samples from nine NSCLC patients previously found positive by DAC-ELISA[16], five patients before diagnosis of NSCLC, five patients after diagnosis of NSCLC, two smoking-habit controls, and two non-smoking controls (Supplementary Figure S6). To check the sensitivity of the western blot, we

Table 1. Survivin antibody absorbances of NSCLC cases, smoking-habit controls, and non-smoking controls

Sample set	*N*	OD at 450 nm Median (IQR)	Sample set	*p*[a]
			vs.	
Non-smoking controls	50	0.054 (0.023)		
			Smoking-habit controls	0.003
			NSCLC cases before dx	0.001
			NSCLC cases after dx	0.006
			Smoking-habit controls and NSCLC cases[b]	<0.001
Smoking-habit controls	50	0.068 (0.029)		
			NSCLC cases before dx	0.751
			NSCLC cases after dx	0.978
NSCLC cases before dx	50	0.067 (0.037)		
			NSCLC cases after dx	0.652
NSCLC cases after dx	50	0.066 (0.028)		
Smoking-habit controls and NSCLC cases[b]	150	0.067 (0.017)		

dx, diagnosis of lung cancer; IQR, interquartile range; N, number of sandwich ELISA measurements; OD, absorbance. [a]Differences in absorbance values were statistically determined by the Mann–Whitney U test (two-tailed).[b]NSCLC cases before and after diagnosis.

analyzed rabbit anti-survivin serially diluted from 1:5,000 to 1:625,000 (1.6 to 200 ng/mL), finding a sensitivity of at least 1.6 ng/mL for rabbit anti-survivin antibodies. The 20 NVALT-12 study samples were also analyzed as described above. Sandwich ELISA and western blot analysis both showed all selected human serum samples to be negative.

Western blot analysis with HEK-293 cell lysate

To rule out the possibility that autoantibodies are not reactive to recombinant survivin but only to endogenous survivin, blots were also performed with HEK-293 cell lysate as the source of endogenous survivin. Proteins in the HEK-293 cell lysate were separated by SDS-PAGE and bands of interest (band 1-5) were excised from the gel for MS.

MS/MS database search identified survivin (predicted: 16 kDa) in 16-kDa protein bands 2 and 3 (Figure 3). The presence of survivin was verified by a targeted MS analysis on protein band 2, which identified four peptides of BIRC5: KKEFEETAK, HSSGCAFLSVK, AIEQLAAMD, and RAIEQLAAMD. Western blot analysis showed that rabbit anti-survivin recognized a band of 15- to 16-kDa protein in the HEK-293 lysate (Figure 3) that was in agreement with band 3 of the mass spectrometry (MS) analysis. Thus, western blot and MS both confirmed the presence of survivin in HEK-293 cells.

Antibody response to antigens in HEK-293 cell lysate was analyzed in serum of seven randomly selected non-smoking controls and nine longitudinally collected sample sets, each set

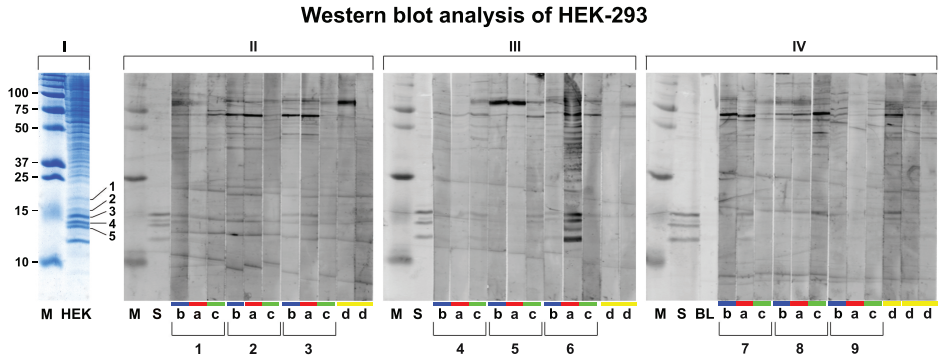


Figure 3. Western blot analysis of endogenous survivin in HEK-293 cell lysate. I, SDS-PAGE of HEK-293 cell lysate. Molecular weight control (M), HEK-293 cell lysate (HEK). Protein bands of interest (1–5) in the 16-kDa region were excised from the gel for MS analysis. MS/MS database search identified survivin in bands 2 and 3. Proteins that showed intense bands were identified as ribosomal proteins. II–IV, Western blot analysis of HEK-293 cell lysate. Molecular weight control (M). Staining of endogenous survivin in HEK-293 cell lysate (15–16 kDa) with 1:2,500 diluted rabbit anti-survivin (S) and staining with blank control (BL). Antibody response to antigens in HEK-293 cell lysate in sera from seven nonsmoking control subjects (d, yellow box) and nine longitudinally collected sample sets (1–9), each set consisting of one case analyzed before diagnosis (b, dark blue box) and after diagnosis of NSCLC (a, red box) and one smoking-habit control matched to this case (c, green box). Immunodetected proteins at 16 kDa were observed for all samples, including smoking-habit control subjects and nonsmoking control subjects. No anti-survivin reactivity was detected in any of the patient sera.

consisting of one case (OD > median OD + 2 IQR of non-smoking controls) analyzed before and after diagnosis of NSCLC (89% adenocarcinoma; stage I) and one smoking-habit control matched to this case. Immunodetected proteins at 16 kDa were observed for all samples, including smoking-habit controls and non-smoking controls. Specific individual patterns of antibodies against HEK-293 proteins were identical for cases before and after diagnosis of lung cancer (Figure 3). Proteins that showed intense bands in the 16-kDa region after Coomassie Brilliant Blue staining, were identified by MS as ribosomal proteins (Supplementary Table S2 online); it is highly likely that some non-specific binding at these bands occurs. We conclude that no anti-survivin reactivity was detected in any of the patient sera.

DISCUSSION

No survivin-specific autoantibodies were detected by sandwich ELISA in sera from NSCLC cases, smoking-habit controls, or healthy non-smoking controls, even though background non-specific binding was 24% higher ($p < 0.001$) in NSCLC cases and smoking controls than in healthy non-smoking controls. Thus, if background is not subtracted, significantly higher OD values for cases and smokers can be misleading when compared only with non-smoking healthy controls. A study of autoantibodies to a panel of ten tumor-associated antigens reported that smokers—lung cancer patients and non-lung cancer patients alike—consis-

tently had significantly higher background binidng when assaying for autoantibodies than did healthy non-smoking controls.[18.]

Specific antibody reactivity for survivin was no higher in any of the 50 NELSON NSCLC patients than in smoking-habit controls. Positive rabbit control samples gave a strong response in the assay. Thus the assay was functioning correctly, and survivin *in vivo* had not induced significant antibody responses before or after diagnosis of NSCLC. This observation is supported by the absence of cytolytic responses against survivin peptides.[26] Western blot analysis revealed no antibody reactivity with the recombinant survivin, consistent with these ELISA results.

To our knowledge, the population we used is the best well-controlled population related to smoking habit in lung cancer case-control studies on survivin autoantibodies to date. Our inability to find survivin autoantibodies in NSCLC patients does not agree with the results reported by others,[11-18] who found survivin autoantibodies with an 8% to 52% prevalence. Although these inconsistent findings may have been due to differences in the type of tumor, the stage of lung cancer, or the source of antigen, the most likely explanation is the difference in assay methodology.

Forty-one (82%) of the 50 NSCLC patients in our study were diagnosed with adenocarcinoma and 45 (90%) were classified as early stage (I and II). This uneven distribution of pathology might explain the absence of survivin autoantibodies. Antibodies to survivin have nevertheless been found irrespective of tumor type and clinical stage of NSCLC.[15, 16] To exclude the possibility that our selection of patients had been unbalanced, we also investigated late-stage (IV) NSCLC samples. In these, too, sandwich-ELISA and western blot analysis revealed no positive samples.

To date, the method used most widely to detect survivin autoantibodies involved direct adsorption assays using immobilized recombinant survivin. If the antigen solution is not absolutely pure, contaminating antigens may be co-immobilized in much the same way as *Escherichia coli* proteins which are very often copurified with the recombinant protein.[27] Due to the high prevalence of *E. coli* infections in humans, background responses to the *E. coli* bacteria used in producing recombinant proteins have been found to be a major problem.[28, 29] We also detected these antibody responses to *E. coli* proteins in a western blot of *E. coli* extract with sera of lung cancer patients (data not shown). Although the specificity of the survivin recognition in the DAC-ELISAs was confirmed by decreased reactivity after pre-absorption of sera with soluble survivin, reactivity is also likely to be reduced by non-specific reactions with impurities in the survivin solution.

In solid-phase immunoassays, non-specific binding for autoantibodies has been found to correlate with increased concentrations of IgG and other inflammatory mediators.[30] Despite the fact that the DAC-ELISA showed higher apparent anti-survivin levels than the sandwich ELISA, in our hands this assay showed no significant difference between lung cancer patients previously reported[16] to be positive for anti-survivin and healthy non-smoking controls – in ad-

dition, any differences disappear when background is subtracted. In general, DAC-ELISA may incorrectly assess antibodies to survivin, and presumably to other tumor antigens, because it does not compensate for non-specific binding. Much more than DAC-ELISA, sandwich ELISA makes it possible to significantly reduce—and also monitor—non-specific binding. Due to better specificity, absorbance in the sandwich ELISA was an order of magnitude lower than in the DAC-ELISA.

To test if any differences in posttranslational modifications of endogenous survivin, compared to recombinant survivin, could be the cause of a lack of anti-survivin reactivity to recombinant survivin, we also used HEK-293 cell lysate as the source of endogenous survivin antigen. Although western blot analysis with HEK-293 cell lysate showed no autoantibody reactivity to endogenous survivin (16 kDa), it showed that the individual patterns of antibodies against HEK proteins before and after the diagnosis of lung cancer were almost identical in each NSCLC cancer patient.

In conclusion, we demonstrated that survivin autoantibody reactivity is not present in sera from NSCLC cases, smoking-habit matched controls, and healthy non-smoking controls. Higher apparent survivin autoantibody reactivity in serum of smokers than in non-smokers that has been previously reported is likely the result of non-specific binding in smokers. In general, autoantibodies to lung tumor antigens should be investigated using sandwich ELISA or another well-characterized technology in a well-controlled population, stratified for at least smoking habit.

ACKNOWLEDGMENTS

The authors thank Frank Santegoets and Roel Faber (Dept. of Public Health, Erasmus MC, Rotterdam, the Netherlands) for their contributions and maintenance of the database of the NELSON trial. They thank Pauline de Goeje (Department of Pulmonology, ErasmusMC) for providing the samples from the NVALT-12 study.

GRANT SUPPORT

This work was supported by a Zenith grant (93511034 to I. Broodman,M. M VanDuijn, and L.J.M. Dekker) from the Netherlands Organization for Scientific Research (NWO), which had no role in the study design, collection, analysis, or interpretation of the data, and no role in the preparation, review, or approval of the manuscript.

REFERENCES

1. International Early Lung Cancer Action Program I, Henschke CI, Yankelevitz DF, Libby DM, Pasmantier MW, Smith JP, et al. Survival of patients with stage I lung cancer detected on CT screening. *N Engl J Med*. 2006; 355: 1763-71.
2. Dunn GP, Old LJ and Schreiber RD. The immunobiology of cancer immunosurveillance and immunoediting. *Immunity*. 2004; 21: 137-48.
3. Andersen PS, Haahr-Hansen M, Coljee VW, Hinnerfeldt FR, Varming K, Bregenholt S, et al. Extensiverestrictions in the VH sequence usage of the human antibody response against the Rhesus D antigen. *Mol Immunol*. 2007; 44: 412-22.
4. Caron M, Choquet-Kastylevsky G and Joubert-Caron R. Cancer immunomics using autoantibody signatures for biomarker discovery. *Mol Cell Proteomics*. 2007; 6: 1115-22.
5. Reuschenbach M, von Knebel Doeberitz M and Wentzensen N. A systematic review of humoral immune responses against tumor antigens. *Cancer Immunol Immunother*. 2009; 58: 1535-44.
6. Farlow EC, Patel K, Basu S, Lee BS, Kim AW, Coon JS, et al. Development of a multiplexed tumor-associated autoantibody-based blood test for the detection of non-small cell lung cancer. *Clin Cancer Res*. 2010; 16: 3452-62.
7. Boyle P, Chapman CJ, Holdenrieder S, Murray A, Robertson C, Wood WC, et al. Clinical validation of an autoantibody test for lung cancer. *Ann Oncol*. 2011; 22: 383-9.
8. Guergova-Kuras M, Kurucz I, Hempel W, Tardieu N, Kadas J, Malderez-Bloes C, et al. Discovery of lung cancer biomarkers by profiling the plasma proteome with monoclonal antibody libraries. *Mol Cell Proteomics*. 2011; 10: M111 010298.
9. Chapman CJ, Healey GF, Murray A, Boyle P, Robertson C, Peek LJ, et al. EarlyCDT(R)-Lung test: improved clinical utility through additional autoantibody assays. *Tumour Biol*. 2012; 33: 1319-26.
10. Zhong L, Coe SP, Stromberg AJ, Khattar NH, Jett JR and Hirschowitz EA. Profiling tumor-associated antibodies for early detection of non-small cell lung cancer. *J Thorac Oncol*. 2006; 1: 513-9.
11. Rohayem J, Diestelkoetter P, Weigle B, Oehmichen A, Schmitz M, Mehlhorn J, et al. Antibody response to the tumor-associated inhibitor of apoptosis protein survivin in cancer patients. *Cancer Res*. 2000; 60: 1815-7.
12. Koziol JA, Zhang JY, Casiano CA, Peng XX, Shi FD, Feng AC, et al. Recursive partitioning as an approach to selection of immune markers for tumor diagnosis. *Clin Cancer Res*. 2003; 9: 5120-6.
13. Zhang JY, Casiano CA, Peng XX, Koziol JA, Chan EK and Tan EM. Enhancement of antibody detection in cancer using panel of recombinant tumor-associated antigens. *Cancer Epidemiol Biomarkers Prev*. 2003; 12: 136-43.
14. Megliorino R, Shi FD, Peng XX, Wang X, Chan EK, Tan EM, et al. Autoimmune response to anti-apoptotic protein survivin and its association with antibodies to p53 and c-myc in cancer detection. *Cancer Detect Prev*. 2005; 29: 241-8.
15. Yagihashi A, Asanuma K, Kobayashi D, Tsuji N, Shijubo Y, Abe S, et al. Detection of autoantibodies to livin and survivin in Sera from lung cancer patients. *Lung Cancer*. 2005; 48: 217-21.
16. Karanikas V, Khalil S, Kerenidi T, Gourgoulianis KI and Germenis AE. Anti-survivin antibody responses in lung cancer. *Cancer Lett*. 2009; 282: 159-66.

17. Ma L, Yue W, Zhang L, Wang Y, Zhang C and Yang X. [Clinical significance and diagnostic value of Survivin autoantibody in non-small cell lung cancer patients]. *Zhongguo Fei Ai Za Zhi*. 2010; 13: 706-12.

18. Rom WN, Goldberg JD, Addrizzo-Harris D, Watson HN, Khilkin M, Greenberg AK, et al. Identification of an autoantibody panel to separate lung cancer from smokers and nonsmokers. *BMC Cancer*. 2010; 10: 234.

19. Li F. Survivin study: what is the next wave? *J Cell Physiol*. 2003; 197: 8-29.

20. van Klaveren RJ, Oudkerk M, Prokop M, Scholten ET, Nackaerts K, Vernhout R, et al. Management of lung nodules detected by volume CT scanning. *N Engl J Med*. 2009; 361: 2221-9.

21. Horeweg N, van der Aalst CM, Vliegenthart R, Zhao Y, Xie X, Scholten ET, et al. Volumetric computed tomography screening for lung cancer: three rounds of the NELSON trial. *Eur Respir J*. 2013; 42: 1659-67.

22. ClinicalTrials.org [Internet]. Paclitaxel-Carboplatin-Bevacizumab +/- Nitroglycerin in Metastatic Non-Squamous-Non-Small Cell Lung Cancer (NVALT12). U.S. National Institutes of Health, 2015.

23. Maat P, VanDuijn M, Brouwer E, Dekker L, Zeneyedpour L, Luider T, et al. Mass spectrometric detection of antigen-specific immunoglobulin peptides in paraneoplastic patient sera. *J Autoimmun*. 2012; 38: 354-60.

24. ProteinAtlas.org [Internet]. Celline expression of BIRC5 - CAB004270 - The Human Protein Atlas. 2015.

25. Broodman I, de Costa D, Stingl C, Dekker LJ, VanDuijn MM, Lindemans J, et al. Mass spectrometry analyses of kappa and lambda fractions result in increased number of complementarity-determining region identifications. *Proteomics*. 2012; 12: 183-91.

26. Karanikas V, Soukou F, Kalala F, Kerenidi T, Grammoustianou ES, Gourgoulianis KI, et al. Baseline levels of CD8+ T cells against survivin and survivin-2B in the blood of lung cancer patients and cancer-free individuals. *Clin Immunol*. 2008; 129: 230-40.

27. Structural Genomics C, China Structural Genomics C, Northeast Structural Genomics C, Graslund S, Nordlund P, Weigelt J, et al. Protein production and purification. *Nat Methods*. 2008; 5: 135-46.

28. Chen JM, Yu M, Morrissy C, Zhao YG, Meehan G, Sun YX, et al. A comparative indirect ELISA for the detection of henipavirus antibodies based on a recombinant nucleocapsid protein expressed in Escherichia coli. *J Virol Methods*. 2006; 136: 273-6.

29. Warnes A, Fooks AR and Stephenson JR. Design and preparation of recombinant antigens as diagnostic reagents in solid-phase immunosorbent assays. *Methods Mol Med*. 2004; 94: 373-91.

30. Guven E, Duus K, Lydolph MC, Jorgensen CS, Laursen I and Houen G. Non-specific binding in solid phase immunoassays for autoantibodies correlates with inflammation markers. *J Immunol Methods*. 2014; 403: 26-36.

SUPPLEMENTARY DATA

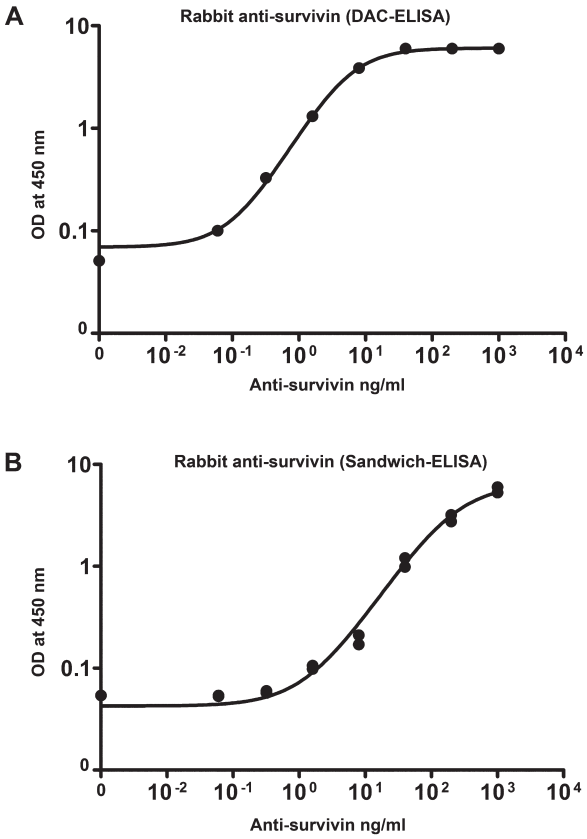


Figure S1. Standard curve of rabbit monoclonal antibody to recombinant survivin using DAC-ELISA (A) and sandwich ELISA (B).

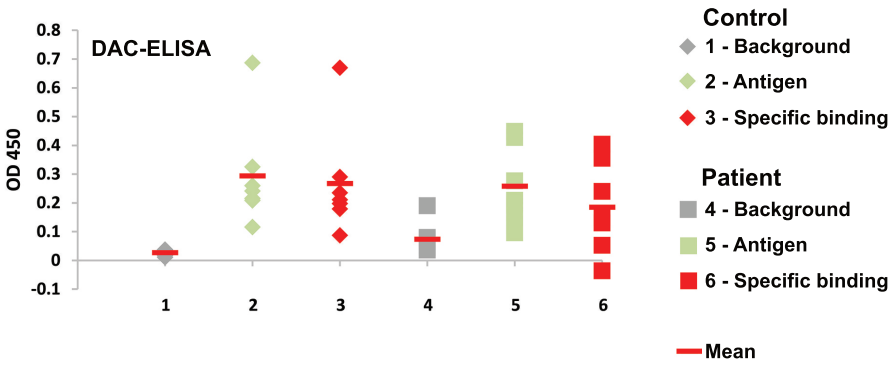


Figure S2. Antibody response to recombinant survivin was measured in seven NSCLC patient serum samples previously reported to be positive for survivin antibodies[16] and in serum samples from seven healthy non-smoking control subjects. Optical densities (ODs) were measured at a 1:100 dilution in DAC-ELISA that contained the antigen, survivin (green), and that without antigen (grey). Specific binding (red) was obtained by subtracting background ODs (grey) obtained in the absence of survivin. No significant difference was found between ODs with antigen from the NSCLC patients (square; median 0.209; IQR 0.154-0.427) and those of the healthy non-smoking controls (rhombus; median 0.241; IQR 0.209-0.325; Mann-Whitney U-test (MW), $p = 0.620$). In addition, no significant differences were found in specific binding (MW, $p = 0.535$).

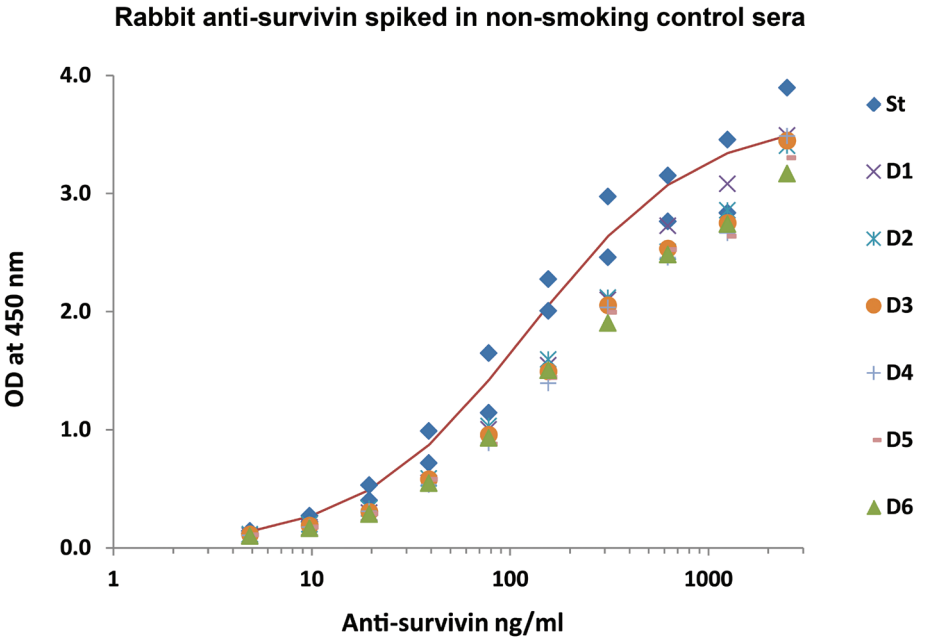


Figure S3. Standard curve of rabbit anti-survivin spiked in non-smoking control sera using sandwich ELISA. ODs of serial dilutions of rabbit anti-survivin in six healthy non-smoking control sera (D1-D6) were identical to those spiked in diluent (St). No sample matrix interference was found in the sandwich ELISA.

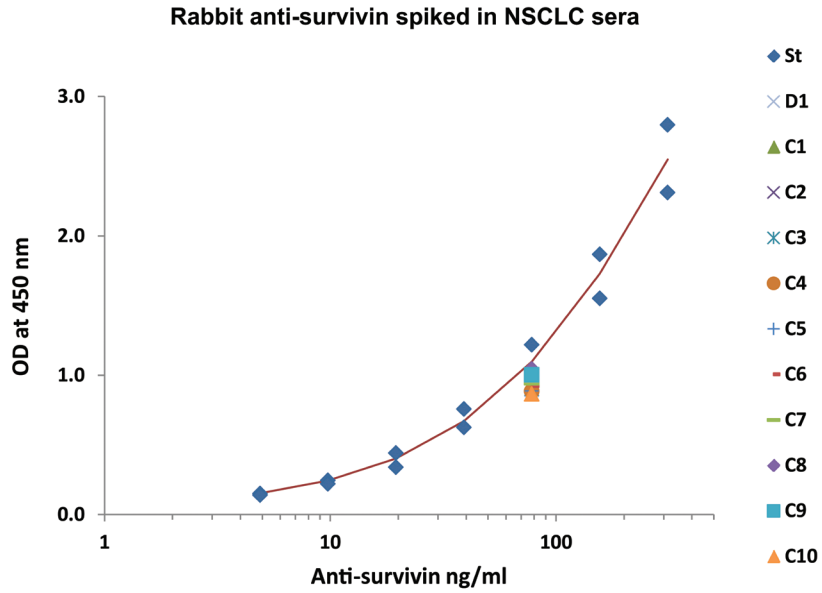


Figure S4. Standard curve of rabbit anti-survivin spiked in NSCLC patient sera using sandwich ELISA. ODs of serial dilutions of rabbit anti-survivin in ten NSCLC patient sera (C1-C10) were identical to those spiked in diluent (St). No sample matrix interference was found in the sandwich ELISA.

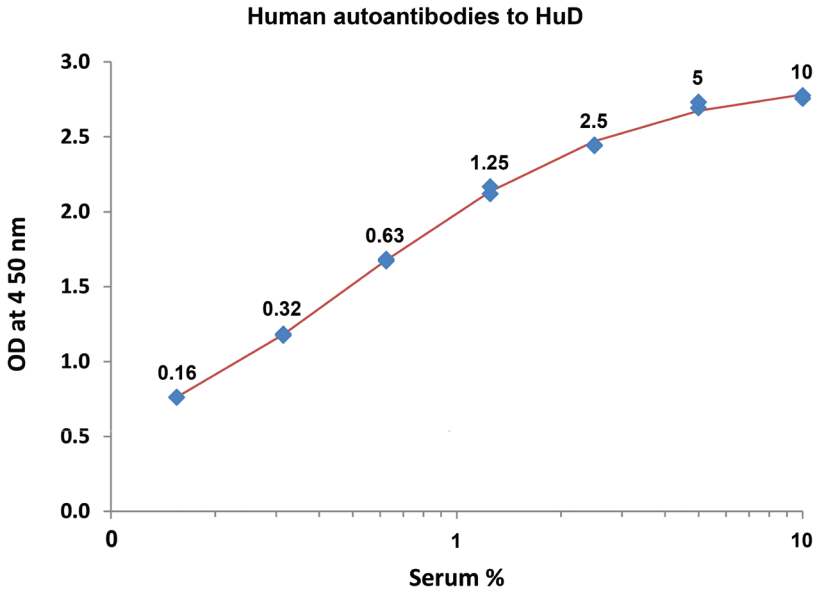


Figure S5. Standard curve of human autoantibodies to HuD using sandwich ELISA. Sandwich ELISA with HuD antigen and serial dilutions from 1:10 to 1:640 (0.16-10% serum) of an anti-HuD-positive serum from a patient with paraneoplastic neurological syndrome (PNS).

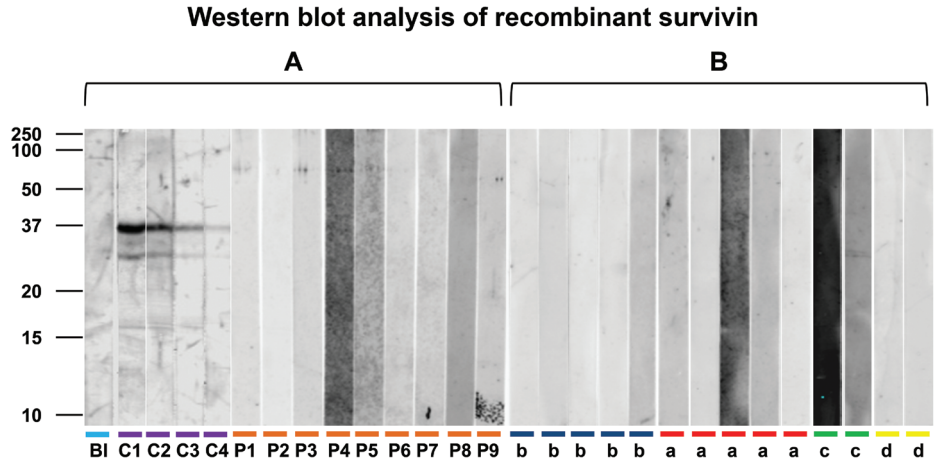


Figure S6. Results of sandwich ELISA were confirmed by western blot analysis.Examples of sample negatives measured by sandwich ELISA to recombinant survivin with CaM-tag (33 kDa; 50 ng/lane) analyzed by western blotting. Western blots (A and B) of blank control (5% BSA-TBST, Bl, light blue); positive control (Rabbit monoclonal to survivin, C1-C4; 1:5,000, 1:25,000, 1:125,000, 1:625,000, purple); sera from nine NSCLC patients previously found positive by DAC-ELISA (16) (P1-P9, orange); five cases analyzed before diagnosis of NSCLC (b, dark blue box) and after diagnosis (a, red box); two smoking-habit controls (c, green box); and two healthy non-smoking controls (d, yellow box). Western blot analysis on recombinant survivin was negative for survivin antibodies in patient sera.

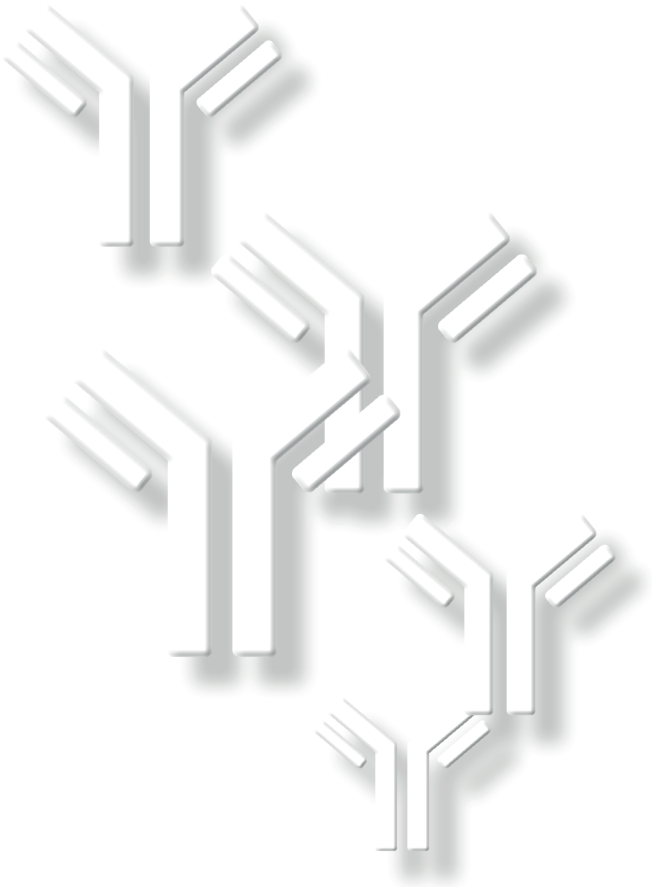
Table S1. Characteristics of NSCLC cases, smoking-habit controls and non-smoking controls

	NSCLC cases *n* = 50		Smoking-habit controls *n* = 50		*p*-value[a]	Non-smoking controls *n* = 50	
Gender					1.000		
Male	43	86.0%	43	86.0%		37	74.0%
Female	7	14.0%	7	14.0%		13	26.0%
Age (years)	62	(51-74)	63	(51-75)	0.646	59	(50-69)
Men	63	(51-74)	63	(51-75)	0.736	59	(52-69)
Women	61	(58-74)	61	(58-74)	0.903	60	(50-69)
Smoking status					0.548		
Current	28	56.0%	25	50.0%		0	0.0%
Former	22	44.0%	25	50.0%		0	0.0%
Smoking duration (years)					0.711		
26-40	22	44.0%	19	38.0%		0	0.0%
41-45	15	30.0%	18	36.0%		0	0.0%
> 45	13	26.0%	13	26.0%		0	0.0%
Cigarettes/day					0.265		
11-15	7	14.0%	14	28.0%		0	0.0%
16-20	16	32.0%	14	28.0%		0	0.0%
21-25	15	30.0%	10	20.0%		0	0.0%
> 25	12	24.0%	12	24.0%		0	0.0%

Table S1. Characteristics of NSCLC cases, smoking-habit controls and non-smoking controls (continued)

	NSCLC cases *n* = 50		Smoking-habit controls *n* = 50		*p*-value[a]	Non-smoking controls *n* = 50	
COPD					0.683		
Yes	22	44.0%	18	36.0%		0	0.0%
No	26	52.0%	29	58.0%		0	0.0%
Unknown	2	4.0%	3	6.0%		50	100.0%
Asbestos exposure					1.000		
Yes	9	18.0%	9	18.0%		0	0.0%
No	41	82.0%	41	82.0%		0	0.0%
Unknown	2	4.0%	3	6.0%		50	100.0%
Center					0.904		
Groningen	12	24.0%	13	26.0%		0	0.0%
Utrecht	16	32.0%	16	32.0%		0	0.0%
Haarlem	17	34.0%	18	36.0%		0	0.0%
Leuven	5	10.0%	3	6.0%		0	0.0%
Rotterdam	0	0.0%	0	0.0%		50	100.0%
Histology							
Adenocarcinoma	41	82.0%	-	-		-	-
Squamous cell carcinoma	7	14.0%	-	-		-	-
Other[b]	2	4.0%	-	-		-	-
Stage							
IA	35	70.0%	-	-		-	-
IB	7	14.0%	-	-		-	-
IIA	1	2.0%	-	-		-	-
IIB	2	4.0%	-	-		-	-
IIIA	4	8.0%	-	-		-	-
IV	1	2.0%	-	-		-	-
Time before diagnosis (years)[c]							
0-0.5	23	46.0%	-	-		-	-
0.5-1.5	13	26.0%	-	-		-	-
1.5-2.5	11	22.0%	-	-		-	-
2.5-3.5	3	6.0%	-	-		-	-
Time after diagnosis (years)[d]							
0-0.5	0	0.0%	-	-		-	-
0.5-1.5	14	28.0%	-	-		-	-
1.5-2.5	19	38.0%	-	-		-	-
2.5-3.5	12	24.0%					
3.5-4.5	5	10.0%	-	-		-	-

n, number of subjects; COPD, Chronic Obstructive Pulmonary Disease;[a] Significant differences in all characteristics between NSCLC cases and smoking-habit controls were analyzed by Pearson chi-square tests except for age (Mann-Whitney U-test). [b] NSCLC histology other than adenocarcinoma or squamous cell carcinoma. [c] Time of blood collection before diagnosis. [d] Time of blood collection after diagnosis.



Ingrid Broodman, Lennard J.M. Dekker, Christoph Stingl, Matthijs Oudkerk, Harry J. de Koning, Joachim G. Aerts, Dimitris Rizopoulos, Jan Lindemans, Theo M. Luider

Comment on "A blood-based proteomic classifier for the molecular characterization of pulmonary nodules"

CHAPTER 5

Submitted

INTRODUCTION

The observation of indeterminate pulmonary nodules on an imaging scan generally leads to subsequent diagnostic biopsy or surgery procedures which do have a serious risk profile.[1, 2] To avoid such invasive procedures for patients with benign nodules, predictive serum biomarkers will be of great value. Li and coworkers[3] developed and initially validated a 13-protein blood-based classifier using multiple-reaction-monitoring mass spectrometry (MRM)[4] that distinguishes benign from early-stage (IA) non-small cell lung cancer (NSCLC) nodules with a 90% negative predictive value (NPV) in a retrospective study. They suggested that this classifier might provide a diagnostic tool for physicians to rescue patients with benign lung nodules from unnecessary, invasive, and costly medical procedures.[5, 6] In a subsequent validation paper Vachani et al.,[7] including ten authors of the earlier paper,[3] state that a classifier composed of only five of the original 13 peptides could indeed be useful as a diagnostic tool to distinguish benign from malignant nodules in patients with indeterminate lung nodules.
As part of our study on prospective biomarkers in a longitudinal prospective cohort study on subjects with a smoking habit (NELSON trial),[8, 9] we performed a validation of both these classifiers[3, 7] and examined their utility as a diagnostic tool.

MATERIALS AND METHODS

The NELSON trial

In this NELSON trial (Dutch-Belgian Lung Cancer Screening trial), approved by the Dutch Ministry of Health and the Population Screening Act committee, participants were recruited between 2003 and 2005 by sending questionnaires to 548,489 individuals between 50–75 years of age. Current or former smokers with a smoking history of at least 15 cigarettes per day for at least 25 years or at least 10 cigarettes per day for at least 30 years were included in the trial. Individuals with malignancies other than primary lung cancer or diagnosed with lung cancer less than five years ago, were excluded. A total of 15,822 participants were randomized (1:1) to a screen or a control arm. The screen arm received computed tomography (CT) screening in years 1, 2 and 4, whereas the control arm received no CT screening (usual care).[10] Initial CT screening results were based on the lung nodule presence and volume. Participants with a positive screening result were referred to a pulmonologist for a diagnostic follow-up. If lung cancer was diagnosed, the participant was offered a treatment protocol and went off screening. Participants with an initial indeterminate screening result received a follow-up CT scan three months later to classify their final screening test result as negative or positive, based on nodule volume doubling time (VDT).[8, 11] Participants with a negative screening result underwent a second-round CT scan 12 months later.

Study population of the NELSON trial

For this validation study, we randomly selected 30 participants in the screen arm of the NELSON trial with nodules of which 16 had been proven benign and 14 malignant, stage IA-B NSCLC (64.3% adenocarcinoma, 21.4% squamous cell carcinoma, 14.3% other histology, 64.3% stage IA and 35.7% stage IB). Lung nodules were classified as benign or malignant based on radiological examination and histological confirmation. Because the original studies reported lung nodule size in diameter, we converted nodule volume at baseline screening into nodule diameter, assuming a spherical nodule. Indeterminate lung nodules (IPNs), nodules with a diameter size of 4 to 30 mm as described by Li et al., were correctly identified by CT in 81.2% of the benign and in 64.3% of the cancer subjects (Table 1). Benign subjects were matched to cancer subjects for age, gender, smoking status, smoking duration, number of cigarettes smoked per day, chronic obstructive pulmonary disease (COPD) status, asbestos exposure and blood collection center. No significant difference in these clinical characteristics were found between the benign and cancer subjects (Table 1). Written informed consent was obtained from all participants for the use of their samples. During the participants' visits to the center, one serum gel tube was collected per participant. The venous blood was allowed to clot, and was centrifuged for 10 min at 1400 x g and 4°C within 2 hours after collection. After centrifugation, the serum was stored immediately in aliquots at -80°C. In contrast to the above mentioned publications,[3, 7] we used available serum samples instead of plasma. Serum protein concentrations are very similar to plasma except for proteins related to coagulation such as fibrinogen, which is removed by conversion into a fibrin clot. Differences in the abundance of classifier peptides between serum and plasma from two non-smoking donors were determined by Mann-Whitney U-test. Except for fibrinogen, no differences (p>0.05) were observed for the peptides in the two classifier sets.

Reference sample

One reference donor sample (male; 59 years) was used as a quality control for each analysis step In accordance with the general guidelines of the Sanquin Blood Bank Rotterdam (the Netherlands), the healthy non-smoking donor of the reference sample gave written consent for the use of his serum for scientific research.

Proteomics analysis

Analysis was performed as previously described with minor modifications.[3, 7] All samples were blinded and randomized before analysis. Samples were prepared in two batches during depletion and in one batch during digestion and MRM-MS analysis. One reference donor sample was measured as quality control for each analysis step at the start, the middle and the end of the batch. To assess the technical variation of the MRM-MS analysis, one pool of digested samples was measured at the start, after every 10th sample and the end of the run. In brief, 60 µL of serum was diluted in 0.15 M $(NH_4)HCO_3$ to a final volume of 180 µL and

Table 1. Clinical characteristics of subjects and nodules

	Cancer *n* = 14		Benign *n* =16		*p-value**
Gender					0.63
Male	13	92.9%	14	87.5%	
Female	1	7.1%	2	12.5%	
Age (years)	63	(55-66)	64	(59-66)	0.55
Smoking status					0.23
Current	10	71.4%	8	50.0%	
Former	4	28.6%	8	50.0%	
Smoking duration (years)					0.65
26-40	7	50.0%	5	31.3%	
41-45	2	14.3%	6	37.5%	
> 45	5	35.7%	5	31.3%	
Cigarettes/day					0.33
0-15	3	21.4%	2	12.5%	
16-20	4	28.6%	5	31.2%	
21-25	5	35.7%	3	18.8%	
> 25	2	14.3%	6	37.5%	
COPD					0.94
Yes	7	50.0%	9	56.3%	
No	6	42.9%	6	37.5%	
Unknown	1	7.1%	1	6.2%	
Asbestos exposure					0.27
Yes	12	85.7%	11	68.8%	
No	2	14.3%	5	31.2%	
Center					0.51
Groningen	5	35.7%	4	25.0%	
Utrecht	2	14.3%	6	37.5%	
Haarlem	5	35.7%	5	31.3%	
Leuven	2	14.3%	1	6.2%	
Histology					
Adenocarcinoma	9	64.3%	-	-	
Squamous cell carcinoma	3	21.4%	-	-	
Other†	2	14.3%	-	-	
Stage					
IA	9	64.3%	-	-	
IB	5	35.7%	-	-	
Nodule size (mm)‡					0.01
< 4	0	0.0%	3	18.8%	
4-30	9	64.3%	13	81.2%	
> 30 and ≤ 50	5	35.7%	0	0.0%	

n, number of subjects; COPD, Chronic Obstructive Pulmonary Disease;* Significant differences in all characteristics between NSCLC cases and benign controls were analyzed by Pearson chi-square tests except for age (Mann-Whitney U-test). †NSCLC histology other than adenocarcinoma or squamous cell carcinoma. ‡ Nodule volume at baseline screening was converted into nodule diameter, assuming a spherical nodule.

filtered using a 0.2 µm AcroPrep 96-well filter plate (Pall Life Sciences). Immunodepletion of high and medium abundant proteins was performed on an IgY14-Supermix LC5 resin column (Sigma Aldrich) coupled to an Ultimate 3000 HPLC system (Thermo Fisher Scientific), comparable to the study of Li et al.[3] After immunodepletion, low-abundant protein fractions were pooled and concentrated in a vacuum concentrator. For tryptic digestion of the low-abundant proteins in peptides, we used our standard procedure because this procedure resulted in a more complete digestion compared to the procedure described by Li et al.[3] In brief, the pool of low-abundant proteins was divided into four equal aliquots. Aliquots were evaporated in a vacuum centrifuge (SPD 1010 Speedvac System; Thermo Fisher Scientific Inc., Waltham, MA) to complete dryness. Dried aliquots were immediately stored at -80°C until further analysis. One dried aliquot was reconstituted in 100 µL of 0.1% Rapigest™ SF (Waters, Milford, MA) in 50 mmol/L NH_4HCO_3. The proteins in the sample were reduced with 1 µL of 500 mmol/L dithiothreitol (DTT) in 50 mmol/L NH_4HCO_3 at 60°C for 30 min and alkylated with 5 µL of 300 mmol/L iodoacetamide in 50 mmol/L NH_4HCO_3 at room temperature in the dark for 30 min. Proteins were digested with 5 µL of 100 ng/µL trypsin in 50 mmol/L NH_4HCO_3 (Trypsin Gold, Promega, Leiden, the Netherlands) for 16 h at 37°C. Subsequently, digestion was terminated by addition of 20 µL of 50% (v/v) TFA to achieve a final pH <2 and incubation for 45 min at 37°C. Finally, the tryptic digested sample was evaporated in a vacuum centrifuge to complete dryness. The dried sample was immediately stored at -80°C until further analysis. The sample was dissolved in 25 µL with 0.1% (v/v) TFA in water by sonification for 5 min in an ultrasonic bath just prior to LC-MRM MS analysis. After centrifugation for 10 min at 10,000 g, the supernatant was spiked with stable isotope-labeled peptides and transferred into a labeled HPLC sample vial.

Six endogenous normalization peptides, as described by Li et al., were used to correct for the inter-sample variation in the depletion and digestion procedures.

MRM-MS analysis

The 13 classifier and the six normalization peptides were analyzed by MRM-MS. In contrast to the original study[3], we monitored by MRM-MS four transitions instead of one transition for each peptide to avoid false peptide identification from co-eluting transitions. We included stable isotope-labeled internal standard (SIS) peptides for each classifier and normalization peptide to control variation in the MRM-MS analysis as recommended by guidelines[12]. In the follow-up publication, Vachani and coworkers also used SIS-peptides.[7] In brief, SIS peptides were added to the tryptic digest prior to MRM analysis. Six microliters of this mixture were injected onto an Ultimate 3000 nano-RSLC system (Thermo Fisher Scientific). After pre-concentration and washing of the sample on a C18 trap column (5 mm × 300 µm ID), peptides were separated on a C18 PepMap column (250 mm × 75 µm ID) (Thermo Fisher Scientific) using a linear 90 min gradient (4–38% ACN/H_2O 0.1% formic acid) at a flow rate of 250 nL/min. The separation of the peptides was monitored by absorption at 214 nm. MRM-MS

analysis was performed on a Q-Exactive Plus (Thermo Fisher Scientific) mass spectrometer equipped with a nano-spray ion source. A targeted MS/MS method was developed for all endogenous and corresponding SIS peptides (Table S1). A quadrupole isolation window of 0.7 m/z units, an automatic gain control target of $1e^6$ ions, a maximum fill time of 250 ms and an Orbitrap resolving power of 17500 at 200 m/z were used. A normalized collision energy of 27 was used for each peptide precursor. Peak analysis of the MRM-MS data was performed with Skyline (MacCoss Lab, University of Washington). Abundances ratios of the endogenous peptides to the corresponding SIS peptides were normalized by the abundances ratios of the six normalization peptides according to the Inte-Quan method.[7, 13]

Statistical analysis

Statistical analyses were performed by Excel (Microsoft 2007) and SPSS (IBM SPSS Statistics 21). A *p*-value <0.05 was considered statistically significant.

RESULTS AND DISCUSSION

MRM-MS of classifier peptides

In our short imitation of the discovery study of Li et al.,[3] three (LRP1, FIBA and GSLG1) of the 13 peptides of the 13-protein classifier were not detectable or detected with a very low intensity by MRM-MS. These peptides could therefore not be properly assigned to the classifier. In agreement with this finding, Vachani and coworkers[7] in their later study were also not able to quantify these three peptides. More specifically, eight of the 13 classifier proteins gave unreliable results for the MRM-assay (BGH3, GRP78, GSLG1, ISLR, LRP1, PRDX1 and TETN) or had affinity to the depletion column (FIBA), and were eliminated from the classifier in their validation study.[7]

Statistical analysis of classifier peptides

Statistical analyses of normalized abundances of the classifier peptides were performed. The median technical coefficient of variation (CV) of four replicate MR-MS measurements of the classifier peptides in the sample pool was 3.5%. The median inter-sample CV of five sample preparations of the reference sample was 14.6%. For statistical analysis, we first used the Mann-Whitney U-test to determine the individual diagnostic performance of the remaining five classifier peptides. In accordance with the original study, no significant difference (p>0.05) between levels of the benign controls and NSCLC cases was found for the individual classifier peptides. Secondly, we applied logistic regression analysis to reproduce the 5-protein classifier of the transitions of the five remaining peptides using the logistic regression classification method as described previously.[3, 7] Logistic regression does not assume normally distributed independent variables, therefore transformation of the normalized abundances

of the classifier peptides was not needed. In addition, log transformation of the data as described by Li et al. did not significantly improve the performance of the logistic regression model. For better interpretation of the regression coefficients, we have calculated a 5-protein logistic regression classifier model based on standardized values (z-values). Classifier peptides and one interaction term composed of COIA1 and FRIL and their coefficients (β) in the logistic regression classifier model are shown in Table S2. Based on the logistic regression classifier model a classifier score between 0 and 1 was calculated for each sample. A cut-off value in this range, called "reference value", was selected to identify lung nodules with classifier scores at or below the reference value (a negative test result) as "benign", as described previously.[3, 7] In general, it is better to utilize the actual probabilities that come from the logistic regression model and not to dichotomize them in a classifier. However, our motivation in this commentary has been to replicate exactly the analysis of Li et al.

Classifier performance

Table 2 shows the performance of the classifiers in the discovery and validation studies for NSCLC prevalences with the selected reference values. To test the performance of the 5-protein classifier, we selected the reference value as described by Vachani et al.[7] Using their reference value of 0.36, the 5-protein classifier identified benign nodules with 88% NPV and 27% PPV at 31% specificity and 86% sensitivity in our validation set.

Li et al. and Vachani et al. found in their classifier validation areas under the receiver operating characteristics (ROC) curves (AUCs) of 0.60 and 0.62, which is too low to be clinically useful[14]. In addition, we found an AUC of 0.65 for our validation of the 5-protein classifier (Figure 1). As our cohort also comprised 36% stage IB nodules, whose larger nodule size represents a higher degree of malignancy,[15] we expected an even better association between prediction by imaging and true outcome.

Table 2. Performance of the classifier in discovery and validation studies.

Data set	AUC	Reference value	Sensitivity (%)	Specificity (%)	NPV (%)	PPV (%)
13-protein classifier at NSCLC prevalence of 15%						
***Li et al.*[3]**						
Discovery (*n* = 143)	0.82	0.60*	82	66	95	30
Validation (*n* = 104)	0.60	0.60*	71	44	90	18
5-protein classifier at NSCLC prevalence of 23%						
***Vachani et al.*[7]**						
Validation (*n* = 141)	0.62	0.36‡	90	17	85	25
Broodman et al.						
Discovery (*n* = 30)	0.65	0.36	86	31	88	27

*Reference value, cut-off value prioritizing a NPV of ≥95% for NSCLC prevalence of 15%.
‡Reference value, cut-off value prioritizing high sensitivity at a target NPV of 90% for NSCLC prevalence of 23%.

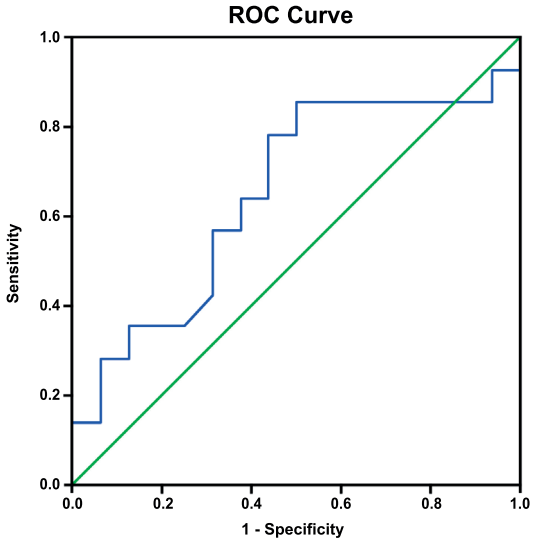


Figure 1. ROC curve of the 5-protein classifier. The area under the receiver operating characteristics (ROC) curve (AUC) of the 5-protein classifier was 0.65 (95% CI: 0.45-0.86).

Using the NSCLC prevalence of 23%, CT screening according to the NELSON volume-based protocol identified in our study benign nodules with 91% NPV and 31% PPV at 44% specificity and 86% sensitivity at baseline. According to this screening protocol, a repeat CT scan was performed for all indeterminate screening results (33%), three months later[8, 9]. After on average 95 days (range 91-112 days), we found 100% NPV and 54% PPV at 75% specificity and 100% sensitivity (Figure 2A). During a follow-up period of on average 4.3 years (range 3.0-5.7 years), none of the benign subjects were diagnosed with lung cancer. As such, NELSON CT screening at baseline identified benign nodules with a higher NPV and specificity than the 5-protein classifier in the validation study.

The performance of the 5-protein classifier in our prospective study shows that a NPV of 88% can be reached. However, we observed a sensitivity of 86% and a specificity of 31%. This means that the 5-protein classifier identifies accurately an optimistic one third (5/16) of the benign nodules in our cohort. However, at the cost of 69% (11/ 16) benign patients still receiving an unnecessary invasive follow-up procedure and, on the other hand 14% (2/14) cancer patients do not receive a necessary follow-up procedure (Figure 2B). From these results we conclude that the 5-protein classifier is not useful as a diagnostic tool to distinguish benign from malignant nodules in our NELSON cohort.

In a recent paper,[16] Vachani et al assessed the clinical utility of the 5-protein classifier in a prospective-retrospective study of 287 (81.3%) NSCLC and 66 (18.7%) benign patients with indeterminate nodules who had an invasive diagnostic procedure. All 287 cancer patients were correctly diagnosed after an invasive procedure and 66 benign patients had an unnecessary invasive procedure. The classifier identified benign nodules with NPV ≥ 84% at

32% specificity and 76% sensitivity. Using the classifier, 263 (74.5%) instead of 353 patients would receive an invasive procedure. At the cost of 69 (24.0%) cancer patients not receiving a necessary invasive procedure and 45 (68.2%) benign patients receiving an unnecessary invasive procedure. As such, the classifier does not reliably exclude malignancy which may cause harm to cancer patients by delay of treatment and decreases the number of invasive procedures only to a limited extent.

Based on our results and the results presented in the recent paper,[16] we conclude that the 5-protein classifier in its present form has no added value for rescuing patients with benign lung nodules from unnecessary invasive procedures in a clinical setting.

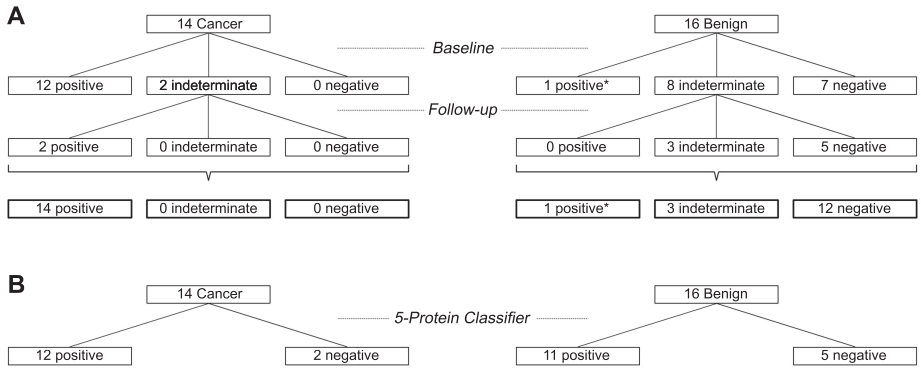


Figure 2. CT scan and 5-protein classifier results of the NELSON sample set.
A, Baseline CT scan results of the NELSON samples are shown at time of blood sampling. Follow-up CT scan results are shown for indeterminate screening results after approximately three months. The last row represents the numbers of positive, indeterminate and negative CT scan results, including baseline and follow-up scan results. B, 5-protein classifier results of the NELSON samples are shown at time of blood sampling. *Benign fibrotic nodule.

REFERENCES

1. 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: 2418-29.
2. de Koning HJ, Meza R, Plevritis SK, ten Haaf K, Munshi VN, Jeon J, et al. Benefits and harms of computed tomography lung cancer screening strategies: a comparative modeling study for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2014; 160: 311-20.
3. Li XJ, Hayward C, Fong PY, Dominguez M, Hunsucker SW, Lee LW, et al. A blood-based proteomic classifier for the molecular characterization of pulmonary nodules. *Sci Transl Med*. 2013; 5: 207ra142.
4. Lange V, Picotti P, Domon B and Aebersold R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol*. 2008; 4: 222.
5. Midthun DE and Jett JR. Screening for lung cancer: the US studies. *J Surg Oncol*. 2013; 108: 275-9.
6. Wiener RS, Schwartz LM, Woloshin S and Welch HG. Population-based risk for complications after transthoracic needle lung biopsy of a pulmonary nodule: an analysis of discharge records. *Ann Intern Med*. 2011; 155: 137-44.
7. Vachani A, Pass HI, Rom WN, Midthun DE, Edell ES, Laviolette M, et al. Validation of a Multi-Protein Plasma Classifier to Identify Benign Lung Nodules. *J Thorac Oncol*. 2015.
8. van Klaveren RJ, Oudkerk M, Prokop M, Scholten ET, Nackaerts K, Vernhout R, et al. Management of lung nodules detected by volume CT scanning. *N Engl J Med*. 2009; 361: 2221-9.
9. Horeweg N, van der Aalst CM, Vliegenthart R, Zhao Y, Xie X, Scholten ET, et al. Volumetric computed tomography screening for lung cancer: three rounds of the NELSON trial. *Eur Respir J*. 2013; 42: 1659-67.
10. 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). *Int J Cancer*. 2007; 120: 868-74.
11. Xu DM, Gietema H, de Koning H, Vernhout R, Nackaerts K, Prokop M, et al. Nodule management protocol of the NELSON randomised lung cancer screening trial. *Lung Cancer*. 2006; 54: 177-84.
12. Han B and Higgs RE. Proteomics: from hypothesis to quantitative assay on a single platform. Guidelines for developing MRM assays using ion trap mass spectrometers. *Briefings in functional genomics & proteomics*. 2008; 7: 340-54.
13. Li X-j, Lee L, Hayward C, Brusniak M-Y, Fong P-Y, McLean M, et al. An integrated quantification method to increase the precision, robustness, and resolution of protein measurement in human plasma samples. *Clinical Proteomics*. 2015; 12: 3.
14. Taylor JM, Ankerst DP and Andridge RR. Validation of biomarker-based risk prediction models. *Clin Cancer Res*. 2008; 14: 5977-83.
15. Massion PP and Walker RC. Indeterminate pulmonary nodules: risk for having or for developing lung cancer? *Cancer Prev Res (Phila)*. 2014; 7: 1173-8.
16. Vachani A, Hammoud Z, Springmeyer S, Cohen N, Nguyen D, Williamson C, et al. Clinical Utility of a Plasma Protein Classifier for Indeterminate Lung Nodules. *Lung*. 2015.

Table S1. MRM assay of classifier and normalization proteins

Protein (human)	Peptide sequence	Precursor ion (m/z)	Charge	Retention time start (min)	Retention time end (min)	Fragment ions (m/z)
ALDOA	ALQASALK (light)	401.2451	2	15.55	23.55	617.3617+(y6), 489.3031+(y5), 331.2340+(y3), 260.1969+(y2)
ALDOA	ALQASALK (heavy)	405.2522	2	15.55	23.55	625.3759+(y6), 497.3173+(y5), 339.2482+(y3), 268.2111+(y2)
BGH3	LTLLAPLNSVFK (light)	658.4028	2	68.23	76.23	988.5826+(y9), 875.4985+(y8), 804.4614+(y7), 328.2231+(b3)
BGH3	LTLLAPLNSVFK (heavy)	662.4099	2	68.23	76.23	996.5968+(y9), 883.5127+(y8), 812.4756+(y7), 328.2231+(b3)
C163A	INPASLDK (light)	429.2400	2	18.77	26.77	744.3886+(y7), 630.3457+(y6), 533.2930+(y5), 462.2558+(y4)
C163A	INPASLDK (heavy)	433.2471	2	18.77	26.77	752.4028+(y7), 638.3599+(y6), 541.3072+(y5), 470.2700+(y4)
COIA1	AVGLAGTFR (light)	446.2560	2	33.19	41.19	721.3991+(y7), 664.3777+(y6), 551.2936+(y5), 480.2565+(y4)
COIA1	AVGLAGTFR (heavy)	451.2601	2	33.19	41.19	731.4074+(y7), 674.3860+(y6), 561.3019+(y5), 490.2648+(y4)
FIBA	NSLFEYQK (light)	514.7560	2	33.92	41.92	827.4298+(y6), 714.3457+(y5), 567.2773+(y4), 315.1663+(b3)
FIBA	NSLFEYQK (heavy)	518.7631	2	33.92	41.92	835.4440+(y6), 722.3599+(y5), 575.2915+(y4), 315.1663+(b3)
FRIL	LGGPEAGLGEYLFER (light)	804.4068	2	63.78	71.78	913.4414+(y7), 564.3140+(y4), 451.2300+(y3), 525.2667+(b6)
FRIL	LGGPEAGLGEYLFER (heavy)	809.4110	2	63.78	71.78	923.4497+(y7), 574.3223+(y4), 461.2382+(y3), 525.2667+(b6)
GELS	TASDFITK (light)	441.7320	2	21.43	29.43	781.4090+(y7), 710.3719+(y6), 623.3399+(y5), 508.3130+(y4)
GELS	TASDFITK (heavy)	445.7391	2	21.43	29.43	789.4232+(y7), 718.3861+(y6), 631.3541+(y5), 516.3272+(y4)
GRP78	TWNDPSVQQDIK (light)	715.8492	2	35.17	43.17	914.4942+(y8), 260.1969+(y2), 288.1343+(b2), 517.2041+(b4)
GRP78	TWNDPSVQQDIK (heavy)	719.8563	2	35.17	43.17	922.5084+(y8), 268.2111+(y2), 288.1343+(b2), 517.2041+(b4)
GSLG1	IIIQESALDYR (light)	660.8615	2	47.67	55.67	981.4636+(y8), 853.4050+(y7), 724.3624+(y6), 338.1823+(y2)
GSLG1	IIIQESALDYR (heavy)	665.8657	2	47.67	55.67	991.4719+(y8), 863.4133+(y7), 734.3707+(y6), 348.1905+(y2)
ISLR	ALPGTPVASSQPR (light)	640.8515	2	24.19	32.19	999.5218+(y10), 841.4526+(y8), 574.2944+(y5), 440.2504+(b5)
ISLR	ALPGTPVASSQPR (heavy)	645.8556	2	24.19	32.19	1009.5301(y10), 851.4609+(y8), 584.3026+(y5), 440.2504+(b5)

Table S1. MRM assay of classifier and normalization proteins (continued)

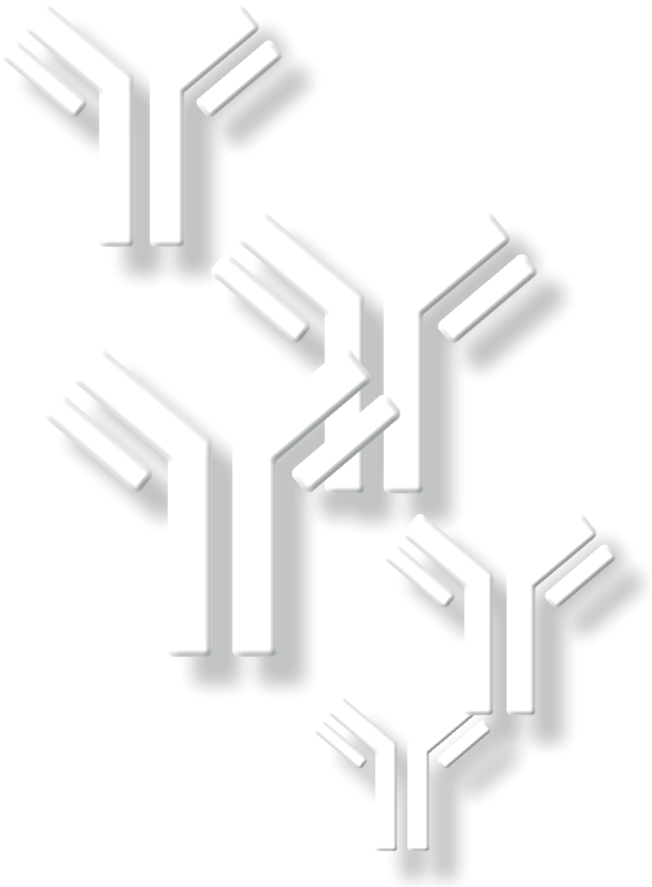
Protein (human)	Peptide sequence	Precursor ion (m/z)	Charge	Retention time start (min)	Retention time end (min)	Fragment ions (m/z)
LG3BP	VEIFYR (light)	413.7265	2	35.70	43.70	727.3774+(y5), 598.3348+(y4), 485.2507+(y3), 338.1823+(y2)
LG3BP	VEIFYR (heavy)	418.7307	2	35.70	43.70	737.3856+(y5), 608.3430+(y4), 495.2590+(y3), 348.1905+(y2)
LRP1	TVLWPNGLSLDIPAGR (light)	854.9727	2	68.04	76.04	1209.6586+(y12), 400.2303+(y4), 605.3329++(y12), 500.2867+(b4)
LRP1	TVLWPNGLSLDIPAGR (heavy)	859.9768	2	68.04	76.04	1219.6669+(y12), 410.2386+(y4), 610.3371++(y12), 500.2867+(b4)
LUM	SLEDLQLTHNK (light)	433.2297	3	28.74	36.74	740.4050+(y6), 612.3464+(y5), 499.2623+(y4), 398.2146+(y3)
LUM	SLEDLQLTHNK (heavy)	435.9011	3	28.74	36.74	740.4050+(y6), 612.3464+(y5), 499.2623+(y4), 398.2146+(y3)
MASP1	TGVITSPDFPNPYPK (light)	816.9170	2	52.17	60.17	862.4458+(y7), 715.3774+(y6), 504.2817+(y4), 258.1448+(b3)
MASP1	TGVITSPDFPNPYPK (heavy)	820.9241	2	52.17	60.17	870.4600+(y7), 723.3916+(y6), 512.2959+(y4), 258.1448+(b3)
PEDF	LQSLFDSPDFSK (light)	692.3432	2	57.14	65.14	942.4203+(y8), 795.3519+(y7), 593.2930+(y5), 329.1819+(b3)
PEDF	LQSLFDSPDFSK (heavy)	696.3503	2	57.14	65.14	950.4345+(y8), 803.3661+(y7), 601.3072+(y5), 329.1819+(b3)
PRDX1	QITVNDLPVGR (light)	606.3408	2	42.95	50.95	970.5316+(y9), 869.4839+(y8), 770.4155+(y7), 428.2616+(y4)
PRDX1	QITVNDLPVGR (heavy)	611.3449	2	42.95	50.95	980.5399+(y9), 879.4922+(y8), 780.4238+(y7), 438.2699+(y4)
PTPRJ	VITEPIPVSDLR (light)	669.8850	2	50.81	58.81	896.5200+(y8), 288.2030+(y2), 314.2074+(b3), 443.2500+(b4)
PTPRJ	VITEPIPVSDLR (heavy)	674.8891	2	50.81	58.81	906.5283+(y8), 298.2113+(y2), 314.2074+(b3), 443.2500+(b4)
TETN	LDTLAQEVALLK (light)	657.3874	2	62.88	70.88	871.5247+(y8), 800.4876+(y7), 543.3865+(y5), 330.1660+(b3)
TETN	LDTLAQEVALLK (heavy)	661.3945	2	62.88	70.88	879.5389+(y8), 808.5018+(y7), 551.4007+(y5), 330.1660+(b3)
TSP1	GFLLLASLR (light)	495.3108	2	65.55	73.55	559.3562+(y5), 446.2722+(y4), 375.2350+(y3), 318.1812+(b3)
TSP1	GFLLLASLR (heavy)	500.3149	2	65.55	73.55	569.3645+(y5), 456.2804+(y4), 385.2433+(y3), 318.1812+(b3)

Protein (human), UniProt (human) protein name; m/z, mass-to-charge ratio; light, endogenous peptide; heavy, stable isotope-labeled internal standard (SIS) peptide: Arginine (R), +10 Da, (13C6, 15N4); Lysine (K), +8 Da, (13C6, 15N2)

Table S2. The logistic regression classifiers

Protein (human)	Peptide	β
ALDOA	ALQASALK	-0.24
COIA1	AVGLAGTFR	0.32
FRIL	LGGPEAGLGEYLFER	0.08
LG3BP	VEIFYR	-0.45
TSP1	GFLLLASLR	0.07
COIA1 × FRIL	Interaction term	-0.40
Constant (α)		-0.16

Shown are the classifier peptides and one interaction term composed of COIA1 and FRIL and their corresponding coefficients (β) in the 5-protein logistic regression classifier model.



Summary and discussion

CHAPTER 6

SUMMARY AND DISCUSSION

Lung cancer is an aggressive disease that has the highest mortality rate among cancer patients in the world, in particular because most patients are only diagnosed at an advanced and non-curable stage. Survival rates of lung cancer patients may improve significantly with early detection and diagnosis. Early stage lung cancer patients have a better prognosis and are potentially curable.[1] In early-stage lung cancer patients are usually asymptomatic and at this stage the disease can only be detected by advanced imaging techniques. Improvement of lung cancer survival rates can therefore best be reached by screening of high risk individuals such as current and former smokers. Low-dose computed tomography (LDCT) is currently the usual detection method in clinical trials on high risk populations. However, the still high false-positive rate of LDCT may harm healthy individuals because of unnecessary follow-up scans and invasive follow-up procedures.

Non-invasive biomarkers, complementary to CT screening, could lower the false-positive rate of CT screening at baseline and thereby reduce the number of patients that need follow-up. On the other hand, additional biomarkers might also reduce potential false-negative and indeterminate CT results and thereby prevent delayed diagnosis and treatment of lung cancer patients.

Lung cancer tissue generates lung cancer-associated proteins to which the immune system might produce autoantibodies. These tumor-associated antigens or their autoantibodies can be measured in biofluid, and could serve as biomarkers for early detection of lung cancer. In this thesis we describe immunological and high-performance mass spectrometry techniques (MS) to identify lung cancer related proteins, especially sequences of autoantibodies. The main findings and conclusions of our studies in serum of high-risk individuals from a well-controlled multicenter population study (NELSON) are summarized and discussed in this chapter.

As described in **Chapter 1**, the immune system might respond to tumor-associated antigens by producing high-affinity autoantibodies. This autoantibody response starts in early stage lung cancer and may endure over years.[2-7] Antigen-specific sequences of antibodies may be shared among different lung cancer patients. Identification of these antigen-specific sequences as potential biomarkers could be an effective approach for early detection of lung cancer.

The estimated possible diversity in antibodies, also known as immunoglobulins, varies from 10^{13} to more than an unrealistic 10^{50}.[8, 9] Because of the vast diversity of immunoglobulins one could assume that antibody sequences are unique for each individual, and that these

sequences will not be shared among other individuals. However, different studies, including studies of our group, have shown that this assumption is not the whole story and that there is evidence for repertoire bias. This means that antibodies are subjected to selection pressures after rearrangement and affinity maturation, favoring certain genes from the germline repertoire that are used in response to a particular antigen.[10-15] For instance, similar CDR3 sequences were found in different humans after tetanus toxin immunization, and a sequencing project in zebra fish showed identical CDR3 sequences shared among several zebra fish.[13, 16] Patients with paraneoplastic neurological syndromes (PNS) are associated with aberrantly expression of autoantibodies to onconeural antigens such as HuD, Yo, Amp or CV2. In the study of Maat et al., primary amino acid structures of autoantibodies were uniquely identified in samples from one of the autoantibody characterized PNS groups.[17] These findings confirm that specific amino acid structures are shared among individuals with autoantibodies to the same onconeural antigens. Moreover, VanDuijn et al. demonstrated by proteomic analysis of affinity purified IgG that rats immunized with onconeural HuD antigen shared a selection of amino acid sequences.[18] This study showed that the development of antibodies during immune response is not a random process, but that selection pressures favor a selection of amino acid sequences in response to an antigen. We previously described a new approach for identification and quantification of antigen-specific sequences of the purified IgG antigen-binding fragment (Fab) by mass spectrometry.[19] An overlap of 83% in MS signals of IgG Fab derived peptides showed that the immunoglobulin repertoire of seven healthy donors is very similar. In this study, we demonstrated that our approach is able to detect, with high reproducibility and recovery, qualitative (17%) and quantitative (4%) differences in Fab sequences between healthy donor serum samples by mass spectrometry. We were also able to isolate and identify several complementary determining regions (CDRs) by mass spectrometry. From these results we conclude that with this approach it is technically feasible to identify CDRs of lung cancer related antibody fragments without the requirement of prior knowledge of the antigen involved.

The possibility to find differences between lung cancer cases and controls depends on the number of CDRs identified. Identification of CDRs by mass spectrometry may be significantly improved by molecular dissection. Molecular dissection reduces the complexity of the immunoglobulin molecule, which in turn reduces ion suppression and leads to a significantly higher sensitivity to detect these very low-abundant CDR peptides by mass spectrometry.

In **Chapter 2** we present a new method to further reduce the complexity of the immunoglobulin by molecular dissection of IgG into kappa (κ) and lambda (λ) fragments and thereby identify significantly more CDRs than by just Fab purification. We performed proteomic analysis on purified Fab, Fab-κ, Fab-λ, IgG-k and IgG-λ fragments from 10 stage I non-small cell lung cancer (NSCLC) patients and 10 matched controls from current and former smokers

of the NELSON trial. The molecular dissection method is reproducible (CV <5%) with a high total recovery (>90%) of IgG-κ and IgG-λ for IgG and Fab-κ and Fab-λ for Fab. More precisely, high-resolution mass spectrometry analysis of the fractions demonstrated a sufficient purification with preserved normal κ to λ ratio in the Fab and light chain of the healthy donor and the Fab of the lung cancer cases and controls. Since the whole Fab fraction comprises heavy and both kappa and lambda light chains, we identified more CDRs in this fraction than in the other individual fractions. Moreover, the lower number of CDRs identified in IgG-κ and IgG-λ can be explained by the fact that these fractions are missing the CDRs, including the highly diverse CDR3, of the heavy chain.[20] This was also demonstrated by the approximately four times higher ratio of CDR3 (CDR1: CDR2:CDR3) found in Fab, Fab-κ and Fab-λ compared to IgG-κ and IgG-λ. We identified twice as many CDRs when Fab-κ, Fab-λ, IgG-κ and IgG-λ fractions were combined than in the Fab fraction alone. No significant difference between lung cancer cases and matched controls for these numbers were found. We identified more CDRs and more significantly different CDRs in the Fab-κ fraction than in the Fab-λ, IgG-κ and IgG-λ fractions. This can be explained by the fact that during B-cell differentiation the heavy chain genes are rearranged first, followed by the κ chain genes. Only if κ chain rearrangements do not result in a functional κ chain, the λ chain genes do start to rearrange.[9] Immunoglobulins that are expressed during tumor development have been described to consist mainly of heavy and κ chains.[21] Because of these findings, we recommend to use both Fab-κ and Fab purifications as the best combination. Mass spectrometry measurements of both Fab and Fab-κ purifications requires twice as much measurement time, but makes the effort worthwhile because of the additional 50% CDRs that can be identified. Furthermore, identification of more CDRs may increase the likelihood of finding lung cancer-related CDR sequences.

In **Chapter 3** we applied our previously described IgG Fab purification method on a case-control study.[19] The aim of our study was to find specific peptide sequences in the antigen-binding fragments (Fab) of antibodies that distinguish lung cancer patients from controls without the need of previously known antigens. We applied proteomic analysis on purified IgG Fab fragments from 44 NSCLC cases and 49 matched controls of the NELSON trial. A model of 12 antibody peptide sequences was able to distinguish lung cancer patients from controls in this high-risk population. This antibody peptide model consists not only of peptide sequences at the CDR regions of the immunoglobulin, but also of peptide sequences at the framework regions of the immunoglobulin. The fact that the antibody peptide model not exclusively consists of antibody peptides of the hypervariable CDR regions but also of antibody peptides of framework regions can be explained by their abundance in the immunoglobulin pool. Peptides with only few mutations compared to the germline, such as framework peptides, are more likely to exist in various clones and therefore have a higher abundance.[22] Because of their higher abundance these peptides have more chance of being detected by mass spectrometry in a high complex immunoglobulin sample. Reduction of the

complexity of the immunoglobulin by molecular dissection into smaller protein fragments such as Fab-κ or Fab-λ, or by purifying only fragments containing the variable regions of the immunoglobulin, could improve the mass spectrometry measurement of the low-abundant CDR peptides. Besides this complexity problem, it is possible that hypermutated CDRs are not commonly shared among patients.[22] Instead of hypermutated CDRs, moderately mutated peptides may have the best overall properties in respect to abundance, specificity and sharing as a potential biomarker for lung cancer. Lung cancer is a heterogeneous disease which causes high variability between patients and may elicit several immune responses to several tumor-associated antigens.[4-6, 23-25] Therefore, it is not unexpected that we were not able to find one single antibody peptide in our data set that could distinguish lung cancer patients from controls. This finding is supported by other studies as well. The studies of Brichory et al. for instance showed low sensitivities ranging from 14%-33% for antibodies to single TAAs.[26, 27] In contrast, studies of Zhong et al. and Lowe et al. showed validated autoantibody panels to TAAs for lung cancer screening with sensitivity ranging from 83%-86% and specificities ranging from 78%-88%.[4, 7] Using a multivariate model we were able to distinguish lung cancer patients from controls with a sensitivity of 96% and a specificity of 100% in the discovery set (NELSON 1). In the independent validation set (NELSON 2) we observed 84% sensitivity and 90% specificity. Because of experimental and biological variation we needed to recalibrate our antibody peptide model for each data set. We evaluated the statistical background of these results for an effect of random selection of the data and thereby the likelihood of discovering a similar model by chance. This background evaluation showed that our 12 antibody peptide model performed significantly better than a model based on permutated data. Up to now, only age and smoking history have been used as inclusion criteria for enrolment of high-risk individuals in screening trials. Additional biomarkers might stratify high-risk individuals more accurately when combined with the inclusion criteria age and smoking history in screening trials. CT screening is able to detect lung cancer with high sensitivity and specificity at different screening rounds after approximately one year of baseline screening.[28] Nevertheless, about 27% of the high-risk individuals in the NELSON trial received unnecessary invasive and expensive follow-up procedures that revealed benign disease at baseline CT screening.[28] CT screening test performance improves after follow-up scans, but only after a relatively long period, on average one year after baseline screening. Therefore, additional diagnostic tests are needed to improve the diagnostic value of CT screening at baseline. For instance, the group of Massion reported the added diagnostic value of a serum proteomic signature in the evaluation of CT screening results of indeterminate lung nodules.[29] In our study, we were not only able to detect lung cancer with an antibody peptide panel at an early stage, but also at an earlier stage than CT screening, on average one year before the final diagnosis was determined. Our results show that autoantibody profiling could be a valuable additional test for early detection of lung cancer in

CT screening. However, still technical challenges have to be overcome before this method is applicable in the clinical setting.

Several case-control studies have suggested that autoantibodies to the apoptosis inhibitor protein 'survivin' are potential biomarkers for early detection of lung cancer. These were detected using *direct antigen-coating* ELISA (DAC-ELISA) with recombinant survivin as capture protein. Survivin autoantibody levels in lung cancer cases were compared with healthy blood-donor controls. In **Chapter 4** we wanted to test the hypothesis that survivin autoantibodies can be detected equally well by our usual *sandwich* ELISA, in a well-controlled population stratified for smoking habit (NELSON), before radiological diagnosis of NSCLC. Our preference for sandwich ELISA is based on its generally higher antigen-specificity than the DAC-ELISA. We therefore first compared the DAC-ELISA with our sandwich ELISA. Because of the more robust results of the sandwich ELISA, we used this assay to measure survivin autoantibodies in serum from 50 stage I-II NSCLC cases, both before and after diagnosis, and 50 smoking-habit matched controls from the NELSON trial. In addition, we measured also 20 late stage NSCLC cases from the Dutch Association of Pulmonologists for Lung Diseases and Tuberculosis (NVALT)-12 study and 50 healthy nonsmoking controls. No specific autoantibodies to survivin were detected in sera from any of the early NSCLC cases, but a remarkably higher (+24%, p<0,001) nonspecific binding was found in both NSCLC cases and smoking controls, than in healthy nonsmoking controls. Hence, without subtraction of background binding, significantly higher binding for NSCLC patients and smokers will be found when compared only with healthy nonsmokers. This misleading finding is supported by a study of autoantibodies to 10 tumor-associated antigens.[30] This study demonstrated that smokers, including lung cancer patients and non-lung cancer patients, had significantly higher background binding in assays for autoantibodies than healthy non-smokers.[30] We found no specific antibody reactivity to survivin in any lung cancer patient, either before or after diagnosis, taking apparently healthy smokers as control. Thus, survivin *in vivo* had not elicited an antibody responses before or after diagnosis of lung cancer. This observation is also supported by the absence of cytolytic T-cell immune responses against survivin peptides in lung cancer.[31] We confirmed our ELISA results with Western blot analysis of recombinant survivin, which also revealed no anti-survivin antibody reactivity. To our knowledge, we used the best well-controlled population related to smoking habit in lung cancer case-control studies on anti-survivin autoantibodies to date. However, the absence of anti-survivin antibodies in NSCLC patient sera reported by our study is in disagreement with the results reported by other studies, which reported the presence of anti-survivvin antibodies in NSCLC patient sera to range from 8% to 52%.[30, 32-38] These inconsistent results may have been caused by differences in the type of tumor, the stage of NSCLC, or the source of antigen, but the most likely reason is the difference in the method of the assay. Survivin autoantibodies have been reported in spite of tumor type or stage of NSCLC.[36, 37] However, the uneven distribution

of pathology of the 50 NSCLC patients of the NELSON study (82% adenocarcinoma, 90% stage I/II) might explain the absence of survivin autoantibodies. To exclude this possibility, we also investigated late stage (IV) NSCLC patients of the NVALT study. In these patients too, sandwich ELISA and Western blot analysis revealed no antibody reactivity to survivin. Up to now, the most common assays for detection of survivin antibodies are direct antigen-coating assays using immobilized recombinant survivin solution. The disadvantage of these assays is that if the antigen solution is not absolute pure, contaminating proteins may be immobilized as well. For instance, E. coli proteins are frequently copurified with the recombinant protein, and because of the high prevalence of E. coli infections of humans, background antibody response to E.coli proteins in assays is a major problem.[39-41] We also detected these antibody responses to E. coli proteins in Western blot analysis of lung cancer sera. Another disadvantage is the non-specific binding in direct antigen-coating (DAC) assays for autoantibodies due to increased concentrations of IgG and other inflammatory mediators.[42] Generally, DAC-ELISA may incorrectly assess antibodies to survivin and to other tumor antigens because of non-specific binding. Non-specific binding can be monitored and significantly reduced in a sandwich ELISA. Differences in posttranslational modifications of survivin, compared with recombinant survivin, could be the reason of the lack of anti-survivin reactivity to recombinant survivin. However, sandwich ELISA results and Western blot analysis of endogenous survivin (HEK-293) also confirmed the lack of survivin antibody reactivity in patient sera. In conclusion, we demonstrated that specific survivin autoantibody reactivity is not present in sera before or after diagnosis of lung cancer. Higher apparent survivin antibody reactivity in smokers than in nonsmokers reported by other studies is likely caused by nonspecific binding in smokers. In conclusion, autoantibodies to lung tumor antigens should be examined using a sandwich ELISA or another well-defined assay in a population that is stratified for smoking.

The discovery of indeterminate pulmonary nodules on a CT scan may lead to invasive diagnostic follow-up procedures such as biopsy or surgery, which may imply a considerable burden and medical risk for the patient. To avoid unnecessary invasive procedures for patients with benign nodules, there is need for a non-invasive predictive tool to distinguish patients with benign nodules from patients with malignant nodules. **Chapter 5** describes our validation study of a blood-based 13-protein and a 5-protein classifier, which have been described by Li et al. and Vachani et al. respectively as a diagnostic tool to distinguish benign from early-stage malignant nodules in patients with indeterminate lung nodules. For our validation study, we used serum samples from 14 stage IA-B NSCLC cases and 16 well-matched benign controls from the NELSON trial. Indeterminate lung nodules (IPNs) with a diameter size of 4 to 30 mm were correctly identified by CT in 81.2% of the benign and in 64.3% of the cancer subjects. In analogy, we used immunodepletion on IgY14-Supermix resin columns and MRM-MS analysis to analyze the classifier peptides and six endogenous normalization peptides.[43] In contrast, to the original study of Li et al., we monitored four transitions instead

of one transition for each peptide, and included stable isotope-labeled internal standard (SIS) peptides for each classifier and normalization peptide. In our study, three of the 13 peptides of the 13-protein classifier could not be quantified. In the later study of Vachani et al., these three peptides and five other peptides could either not be quantified and were therefore eliminated from the classifier in their validation study.[44] In line with the original study, no significant difference ($p>0.05$) was found between normalized abundances of the benign controls and NSCLC cases for single classifier peptides. We applied logistic regression analysis to reproduce the 5-protein classifier of the transitions of the five remaining peptides using the logistic regression classification method as described by Vachani et al.[44] Using their reference value of 0.36 for a NSCLC prevalence of 23%, the 5-protein classifier identified benign nodules with 88% NPV and 27% PPV at 31% specificity and 86% sensitivity in our validation set. Li et al. and Vachani et al. found in their validation study ROC curves with AUCs of 0.60 and 0.62, which are too low to be medically useful.[45] Furthermore, we found an AUC of 0.65 for our validation of the 5-protein classifier. Since our cohort also comprised 36% stage IB nodules, whose larger nodule size represents a higher degree of malignancy, we expected an even better relation between prediction and true outcome. Using the NSCLC prevalence of 23%, CT screening according to the NELSON protocol identified in our cohort benign nodules with 91% NPV and 31% PPV at 44% specificity and 86% sensitivity at baseline. As such, the NELSON CT screening at baseline identified benign nodules with a higher specificity and NPV than the 5-protein classifier in our validation study. The 5-protein classifier identified accurately an optimistic one third (5/16) of the benign nodules in our cohort. However, at the cost of 69% (11/ 16) benign patients still receiving an unnecessary invasive follow-up procedure and, on the other hand 14% (2/14) cancer patients do not receive a necessary follow-up procedure. Recently, Vachani et al. tested the clinical utility of the 5-protein classifier in a prospective-retrospective study of patients with IPNs who had an invasive diagnostic procedure.[46] Using the 5-protein classifier, 263 (74.5%) instead of 353 patients would receive an invasive procedure. However, 69 (24.0%) malignant patients would not receive a necessary invasive procedure and 45 (68.2%) benign patients would receive an unnecessary invasive procedure. Based on these results and our results, we conclude that this 5-protein classifier has no added value for rescuing patients with benign lung nodules from unnecessary invasive procedures in the clinic.

Future perspectives

Serum-biomarkers for lung cancer might be useful for early detection and monitoring of the disease in individuals at high risk. In a complex and heterogeneous disease like lung cancer a variety of proteins is involved. Therefore, it is more logical to search for a panel of biomarkers to achieve high sensitivity and specificity than for a single biomarker. A panel of specific peptide sequences in the antigen-binding fragments could be a model for sensitive and specific biomarkers in lung cancer as we described in this thesis.

We used a bottom-up approach to identify the 12 putative sequences of this antibody peptide model by label-free high-resolution mass spectrometry. Despite the highly sensitive mass spectrometry this peptide-based approach has some limitations. Digestion of proteins into peptides increases the complexity of the sample for mass spectrometry measurement. This peptide complexity can be reduced before protein digestion by molecular dissection of the immunoglobulin protein into smaller protein fragments or by depletion of the constant regions, or by enrichment of the variable regions of the immunoglobulin protein. This will reduce the complexity of the sample and thereby significantly improve the detection of the low-abundant CDRs peptides by mass spectrometry. In addition, because of the tryptic digestion it is difficult to identify the highly diverse CDR3 peptides. The CDR3 peptide sequence has often at the beginning a lysine (K) or arginine (R) amino acid, which are the cleavage sites of trypsin. As a result, their tryptic-digested peptides often contain either a mostly conserved V-region fragment or a highly diverse fragment, which is difficult to align to the germline sequence. The use of alternative enzymes could help to improve this CDR3 identification. An even better option for CDR identification would be the measurement of larger fragments by using middle-down or top-down proteomics. Top-down (protein-based) proteomics using high-resolution mass spectrometry allows the measurement of larger fragments than the antigen-binding site of antibodies and thereby may facilitate the identification of the longer segments of CDR sequences.[47, 48]

For clinical use as biomarker panel we need to know the full sequence of the 12 antibody peptides. However, the full sequences of these antibodies are not present in databases. Proteogenomics is an emerging field that provides an opportunity to identify peptides by searching mass spectrometry data of peptides against genomic databases.[49, 50] Proteogenomics in combination with the emerging development of targeted next generation sequencing (NGS) allow us to develop databases derived from RNA-sequencing data containing full sequences of rearranged antibody peptides.[51, 52] For this purpose, antibodies produced by tumor-infiltrating cells will be characterized by mass spectrometry and NGS in order to identify common autoantibodies specific for lung cancer. While the NGS data will reveal antibodies that are locally produced by tumor-infiltrating cells in lung tissue, this tissue may also contain antibodies that are produced elsewhere, in lymphoid tissue, in peripheral circulation or in bone marrow. Therefore, the NGS data will be combined with proteomics data in order to reveal a complete inventory of antibodies produced locally in lung tissue and elsewhere in the body. In brief, lung tissue and sera will be collected from lung cancer patients and from COPD patients as controls, since these patients have the same history of smoking and inflammation, which are known to provoke also substantial auto-immune response. Immune cells such as T-cells, B-cells, dendritic cells and macrophages often infiltrate the lung tumor. These tumor-infiltrating lymphocytes may be organized into tertiary lymphoid structures (TLS), also called tumor-induced bronchus-associated lymphoid tissues (T)i-BALTs, which are highly

organized lymphoid aggregates that also can be found in lymph nodes.[53-57] In these TLS, T-cells and B-cells closely interact with tumor-associated antigens presented by dendritic cells, resulting in further development of the variable regions in their antigen-specific sites.[58] After staining for infiltrating B-cells, the TLS sections of the lung tissues will be obtained by laser capture microdissection. Total RNA from these sections will be extracted using RNA isolation kits and subsequently be transcribed to cDNA using primers specific for the constant regions of immunoglobulins. The cDNA will be amplified by PCR with a universal primer set specific for the set of human immunoglobulin variable regions. After this PCR step the PCR products will be ligated to Illumina adapters for subsequently sequencing of the variable regions of immunoglobulins. After sequencing, this NGS data will be filtered for unique sequences. Proteomic analysis by high-resolution mass spectrometry will be performed of purified IgG in serum and lung tissue samples from the same cohort. Mass spectra data as fingerprints for immunoglobulin peptides in serum and lung tissue samples will be searched against a database based on the NGS data. After data analysis and integration of mass spectrometry data and NGS data of these tissue and serum samples the most promising autoantibody sequences will be selected. Further identification of the complete variable region sequence by using targeted NGS data of lung carcinoma tissue enables us to synthesize exact and reliable sequences of our 12 autoantibody panel or a newly discovered autoantibody panel. Subsequently, reliable quantification of these antibody sequences by parallel reaction monitoring (PRM) on high resolution mass spectrometers using stable isotope-labeled standard peptides (SIS) as references may improve this autoantibody panel. This autoantibody panel will be validated in sera of early stage lung cancer patients and matched controls of an independent cohort from the well-controlled prospective NELSON study. After validation, we may be able to develop a serum test of an antibody panel that combined with CT screening at baseline, reaches the desired sensitivity and specificity of at least 95% for clinical use. For clinical use of this serum test high-throughput multiplex assays can be developed such as bead-based immunoassays (e.g. Luminex®), high-affinity assays using commercially available aptamers (e.g. Avacta Life Science®, Olink Bioscience® and SomaLogic®) or even fully automated immunoassays (e.g. by Roche Diagnostics®).

REFERENCES

1. Society AC. Lung Cancer | American Cancer Society. 2015.
2. Anderson KS and LaBaer J. The sentinel within: exploiting the immune system for cancer biomarkers. *J Proteome Res*. 2005; 4: 1123-33.
3. Finn OJ. Immune response as a biomarker for cancer detection and a lot more. *N Engl J Med*. 2005; 353: 1288-90.
4. Zhong L, Coe SP, Stromberg AJ, Khattar NH, Jett JR and Hirschowitz EA. Profiling tumor-associated antibodies for early detection of non-small cell lung cancer. *J Thorac Oncol*. 2006; 1: 513-9.
5. Chapman CJ, Murray A, McElveen JE, Sahin U, Luxemburger U, Tureci O, et al. Autoantibodies in lung cancer: possibilities for early detection and subsequent cure. *Thorax*. 2008; 63: 228-33.
6. Tan HT, Low J, Lim SG and Chung MC. Serum autoantibodies as biomarkers for early cancer detection. *The FEBS journal*. 2009; 276: 6880-904.
7. Lowe FJ, Shen W, Zu J, Li J, Wang H, Zhang X, et al. A novel autoantibody test for the detection of pre-neoplastic lung lesions. *Mol Cancer*. 2014; 13: 78.
8. Saada R, Weinberger M, Shahaf G and Mehr R. Models for antigen receptor gene rearrangement: CDR3 length. *Immunol Cell Biol*. 2007; 85: 323-32.
9. Murphy K. TP, Walport M. *Janeway's immunobiology*. 7th ed.: Garland Science, 2008.
10. Baranzini SE, Jeong MC, Butunoi C, Murray RS, Bernard CC and Oksenberg JR. B cell repertoire diversity and clonal expansion in multiple sclerosis brain lesions. *J Immunol*. 1999; 163: 5133-44.
11. Andersen PS, Haahr-Hansen M, Coljee VW, Hinnerfeldt FR, Varming K, Bregenholt S, et al. Extensive restrictions in the VH sequence usage of the human antibody response against the Rhesus D antigen. *Mol Immunol*. 2007; 44: 412-22.
12. Foreman AL, Van de Water J, Gougeon ML and Gershwin ME. B cells in autoimmune diseases: insights from analyses of immunoglobulin variable (Ig V) gene usage. *Autoimmun Rev*. 2007; 6: 387-401.
13. Poulsen TR, Meijer PJ, Jensen A, Nielsen LS and Andersen PS. Kinetic, affinity, and diversity limits of human polyclonal antibody responses against tetanus toxoid. *J Immunol*. 2007; 179: 3841-50.
14. Scheid JF, Mouquet H, Ueberheide B, Diskin R, Klein F, Oliveira TY, et al. Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science*. 2011; 333: 1633-7.
15. Henry Dunand CJ and Wilson PC. Restricted, canonical, stereotyped and convergent immunoglobulin responses. *Philosophical transactions of the Royal Society of London Series B, Biological sciences*. 2015; 370.
16. Weinstein JA, Jiang N, White RA, 3rd, Fisher DS and Quake SR. High-throughput sequencing of the zebrafish antibody repertoire. *Science*. 2009; 324: 807-10.
17. Maat P, Vanduijn M, Brouwer E, Dekker L, Zeneyedpour L, Luider T, et al. Mass spectrometric detection of antigen-specific immunoglobulin peptides in paraneoplastic patient sera. *J Autoimmun*. 2012.
18. VanDuijn MM, Dekker LJ, Zeneyedpour L, Smitt PA and Luider TM. Immune responses are characterized by specific shared immunoglobulin peptides that can be detected by proteomic techniques. *J Biol Chem*. 2010; 285: 29247-53.

19. de Costa D, Broodman I, Vanduijn MM, Stingl C, Dekker LJ, Burgers PC, et al. Sequencing and quantifying IgG fragments and antigen-binding regions by mass spectrometry. *J Proteome Res*. 2010; 9: 2937-45.

20. Xu JL and Davis MM. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity*. 2000; 13: 37-45.

21. Chen Z, Qiu X and Gu J. Immunoglobulin expression in non-lymphoid lineage and neoplastic cells. *Am J Pathol*. 2009; 174: 1139-48.

22. Robins HS, Campregher PV, Srivastava SK, Wacher A, Turtle CJ, Kahsai O, et al. Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood*. 2009; 114: 4099-107.

23. Hanash S. Harnessing immunity for cancer marker discovery. *Nat Biotechnol*. 2003; 21: 37-8.

24. Caron M, Choquet-Kastylevsky G and Joubert-Caron R. Cancer immunomics using autoantibody signatures for biomarker discovery. *Mol Cell Proteomics*. 2007; 6: 1115-22.

25. Qiu J and Hanash S. Autoantibody profiling for cancer detection. *Clin Lab Med*. 2009; 29: 31-46.

26. Brichory F, Beer D, Le Naour F, Giordano T and Hanash S. Proteomics-based identification of protein gene product 9.5 as a tumor antigen that induces a humoral immune response in lung cancer. *Cancer Res*. 2001; 61: 7908-12.

27. Brichory FM, Misek DE, Yim AM, Krause MC, Giordano TJ, Beer DG, et al. An immune response manifested by the common occurrence of annexins I and II autoantibodies and high circulating levels of IL-6 in lung cancer. *Proc Natl Acad Sci U S A*. 2001; 98: 9824-9.

28. van Klaveren RJ, Oudkerk M, Prokop M, Scholten ET, Nackaerts K, Vernhout R, et al. Management of lung nodules detected by volume CT scanning. *N Engl J Med*. 2009; 361: 2221-9.

29. Pecot CV, Li M, Zhang XJ, Rajanbabu R, Calitri C, Bungum A, et al. Added value of a serum proteomic signature in the diagnostic evaluation of lung nodules. *Cancer Epidemiol Biomarkers Prev*. 2012; 21: 786-92.

30. Rom WN, Goldberg JD, Addrizzo-Harris D, Watson HN, Khilkin M, Greenberg AK, et al. Identification of an autoantibody panel to separate lung cancer from smokers and nonsmokers. *BMC Cancer*. 2010; 10: 234.

31. Karanikas V, Soukou F, Kalala F, Kerenidi T, Grammoustianou ES, Gourgoulianis KI, et al. Baseline levels of CD8+ T cells against survivin and survivin-2B in the blood of lung cancer patients and cancer-free individuals. *Clin Immunol*. 2008; 129: 230-40.

32. Rohayem J, Diestelkoetter P, Weigle B, Oehmichen A, Schmitz M, Mehlhorn J, et al. Antibody response to the tumor-associated inhibitor of apoptosis protein survivin in cancer patients. *Cancer Res*. 2000; 60: 1815-7.

33. Koziol JA, Zhang JY, Casiano CA, Peng XX, Shi FD, Feng AC, et al. Recursive partitioning as an approach to selection of immune markers for tumor diagnosis. *Clin Cancer Res*. 2003; 9: 5120-6.

34. Zhang JY, Casiano CA, Peng XX, Koziol JA, Chan EK and Tan EM. Enhancement of antibody detection in cancer using panel of recombinant tumor-associated antigens. *Cancer Epidemiol Biomarkers Prev*. 2003; 12: 136-43.

35. Megliorino R, Shi FD, Peng XX, Wang X, Chan EK, Tan EM, et al. Autoimmune response to anti-apoptotic protein survivin and its association with antibodies to p53 and c-myc in cancer detection. *Cancer Detect Prev*. 2005; 29: 241-8.

36. Yagihashi A, Ohmura T, Asanuma K, Kobayashi D, Tsuji N, Torigoe T, et al. Detection of autoantibodies to survivin and livin in sera from patients with breast cancer. *Clin Chim Acta*. 2005; 362: 125-30.

37. Karanikas V, Khalil S, Kerenidi T, Gourgoulianis KI and Germenis AE. Anti-survivin antibody responses in lung cancer. *Cancer Lett*. 2009; 282: 159-66.

38. Ma L, Yue W, Zhang L, Wang Y, Zhang C and Yang X. [Clinical significance and diagnostic value of Survivin autoantibody in non-small cell lung cancer patients]. *Zhongguo Fei Ai Za Zhi*. 2010; 13: 706-12.

39. Warnes A, Fooks AR and Stephenson JR. Design and preparation of recombinant antigens as diagnostic reagents in solid-phase immunosorbent assays. *Methods Mol Med*. 2004; 94: 373-91.

40. Chen JM, Yu M, Morrissy C, Zhao YG, Meehan G, Sun YX, et al. A comparative indirect ELISA for the detection of henipavirus antibodies based on a recombinant nucleocapsid protein expressed in Escherichia coli. *J Virol Methods*. 2006; 136: 273-6.

41. Structural Genomics C, China Structural Genomics C, Northeast Structural Genomics C, Graslund S, Nordlund P, Weigelt J, et al. Protein production and purification. *Nat Methods*. 2008; 5: 135-46.

42. Guven E, Duus K, Lydolph MC, Jorgensen CS, Laursen I and Houen G. Non-specific binding in solid phase immunoassays for 2 autoantibodies correlates with inflammation markers. *J Immunol Methods*. 2013.

43. Li XJ, Hayward C, Fong PY, Dominguez M, Hunsucker SW, Lee LW, et al. A blood-based proteomic classifier for the molecular characterization of pulmonary nodules. *Sci Transl Med*. 2013; 5: 207ra142.

44. Vachani A, Pass HI, Rom WN, Midthun DE, Edell ES, Laviolette M, et al. Validation of a Multi-Protein Plasma Classifier to Identify Benign Lung Nodules. *J Thorac Oncol*. 2015.

45. Taylor JM, Ankerst DP and Andridge RR. Validation of biomarker-based risk prediction models. *Clin Cancer Res*. 2008; 14: 5977-83.

46. Vachani A, Hammoud Z, Springmeyer S, Cohen N, Nguyen D, Williamson C, et al. Clinical Utility of a Plasma Protein Classifier for Indeterminate Lung Nodules. *Lung*. 2015.

47. Tsybin YO, Fornelli L, Stoermer C, Luebeck M, Parra J, Nallet S, et al. Structural analysis of intact monoclonal antibodies by electron transfer dissociation mass spectrometry. *Anal Chem*. 2011; 83: 8919-27.

48. Dekker L, Wu S, Vanduijn M, Tolic N, Stingl C, Zhao R, et al. An integrated top-down and bottom-up proteomic approach to characterize the antigen-binding fragment of antibodies. *Proteomics*. 2014; 14: 1239-48.

49. Alfaro JA, Sinha A, Kislinger T and Boutros PC. Onco-proteogenomics: cancer proteomics joins forces with genomics. *Nat Methods*. 2014; 11: 1107-13.

50. Nagaraj SH, Waddell N, Madugundu AK, Wood S, Jones A, Mandyam RA, et al. PGTools: A Software Suite for Proteogenomic Data Analysis and Visualization. *J Proteome Res*. 2015; 14: 2255-66.

51. Benichou J, Ben-Hamo R, Louzoun Y and Efroni S. Rep-Seq: uncovering the immunological repertoire through next-generation sequencing. *Immunology*. 2012; 135: 183-91.

52. Greiff V, Menzel U, Haessler U, Cook SC, Friedensohn S, Khan TA, et al. Quantitative assessment of the robustness of next-generation sequencing of antibody variable gene repertoires from immunized mice. *BMC Immunol*. 2014; 15: 40.

53. Dieu-Nosjean MC, Antoine M, Danel C, Heudes D, Wislez M, Poulot V, et al. Long-term survival for patients with non-small-cell lung cancer with intratumoral lymphoid structures. *J Clin Oncol*. 2008; 26: 4410-7.

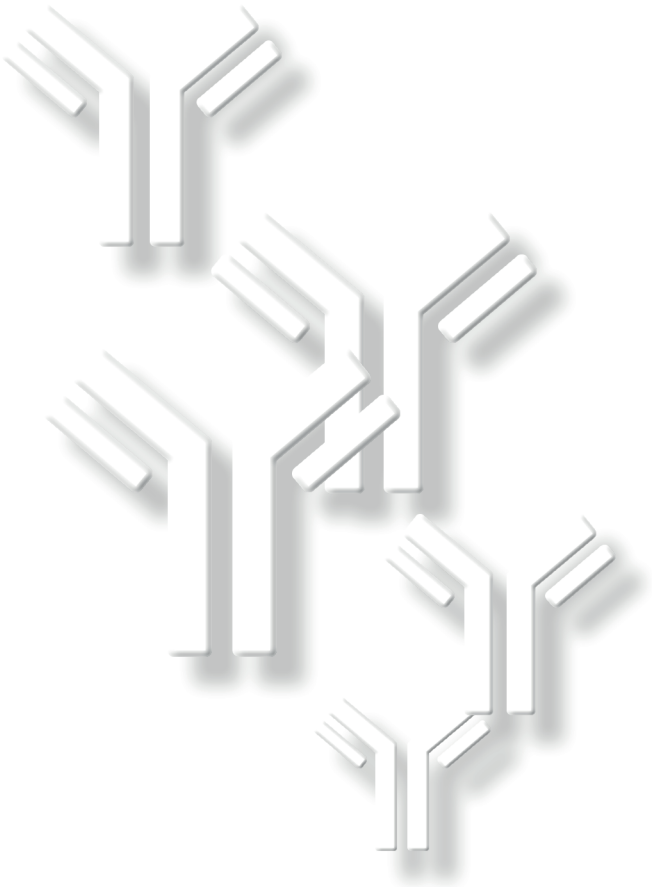
54. Dieu-Nosjean MC, Goc J, Giraldo NA, Sautes-Fridman C and Fridman WH. Tertiary lymphoid structures in cancer and beyond. *Trends Immunol*. 2014; 35: 571-80.

55. Germain C, Gnjatic S, Tamzalit F, Knockaert S, Remark R, Goc J, et al. Presence of B cells in tertiary lymphoid structures is associated with a protective immunity in patients with lung cancer. *Am J Respir Crit Care Med*. 2014; 189: 832-44.

56. Goc J, Fridman WH, Hammond SA, Sautes-Fridman C and Dieu-Nosjean MC. Tertiary lymphoid structures in human lung cancers, a new driver of antitumor immune responses. *Oncoimmunology*. 2014; 3: e28976.

57. Goc J, Germain C, Vo-Bourgais TK, Lupo A, Klein C, Knockaert S, et al. Dendritic cells in tumor-associated tertiary lymphoid structures signal a Th1 cytotoxic immune contexture and license the positive prognostic value of infiltrating CD8+ T cells. *Cancer Res*. 2014; 74: 705-15.

58. Germain C, Gnjatic S and Dieu-Nosjean MC. Tertiary Lymphoid Structure-Associated B Cells are Key Players in Anti-Tumor Immunity. *Front Immunol*. 2015; 6: 67.



Samenvatting

CHAPTER 7

SAMENVATTING

Longkanker is een agressieve ziekte met het hoogste sterftecijfer onder kankerpatiënten in de wereld, hoofdzakelijk doordat bij de meeste patiënten de ziekte pas in een vergevorderd en ongeneeslijk stadium wordt gediagnosticeerd. De overlevingskansen van longkankerpatienten kunnen aanzienlijk verbeteren door vroegtijdige opsporing en diagnose. Patiënten in een vroeg stadium van longkanker hebben een betere prognose en een aannemelijke kans op genezing. In een vroeg stadium zijn longkankerpatiënten echter vaak asymptomatisch en de ziekte kan in dit stadium enkel worden opgespoord door geavanceerde beeldtechnieken. Het verbeteren van de overlevingskans bij longkanker kan het best worden bereikt door screening van individuen met een verhoogd risico zoals huidige en voormalige rokers. Lage dosis CT (computed tomography) is momenteel de gebruikelijke detectiemethode bij klinische studies met hoog-risico groepen. Echter kan het nog steeds hoge percentage vals-positieve uitslagen bij LDCT schade toebrengen aan gezonde individuen als gevolg van de onnodige vervolgscans en invasieve vervolgprocedures (follow-up).

Aanvullend op CT screening kunnen niet-invasieve biomarkers mogelijk het aantal vals-positieven bij aanvang van CT screening (baseline screening) verlagen en daardoor het aantal patiënten dat follow-up procedures nodig heeft verminderen. Daarnaast kunnen aanvullende biomarkers waarschijnlijk ook het aantal vals-negatieve en ondefinieerbare CT resultaten verminderen en daarmee de vertraagde diagnose en behandeling van longkankerpatiënten voorkomen.

Longkankerweefsel genereert aan longkanker gerelateerde eiwitten waartegen het immuunsysteem autoantistoffen kan produceren. Deze tumor gerelateerde antigenen of hun autoantistoffen kunnen in biovloeistoffen worden gemeten en zouden kunnen dienen als biomarkers voor vroegtijdige opsporing van longkanker. In dit proefschrift beschrijven we immunologische en hoogwaardige massaspectrometrie technieken (MS) om longkanker gerelateerde eiwitten te identificeren en in het bijzonder de sequenties van autoantistoffen. De belangrijkste bevindingen en conclusies uit onze studies omtrent serum van hoog-risico individuen uit een goed gecontroleerd multicenter bevolkingsonderzoek, de NELSON trial, zijn in dit hoofdstuk samengevat.

Zoals beschreven in **hoofdstuk 1**, kan het immuunsysteem op aan tumor gerelateerde antigenen reageren door het produceren van autoantistoffen met hoge binding voor deze antigenen. Deze autoimmuunrespons begint in een vroeg stadium van longkanker en kan meerdere jaren voortduren. Antigeen-specifieke sequenties van antistoffen kunnen gemeenschappelijk voorkomen in verschillende longkankerpatiënten. Identificatie van deze

antigeen-specifieke sequenties als mogelijke biomarkers zou een effectieve aanpak kunnen zijn voor de vroegtijdige opsporing van longkanker.

Antistoffen, ook wel bekend als immuunglobulinen, bestaan uit constante en variabele delen. Het variabele deel in het antigeen-bindend fragment (Fab) van een antistof bevat drie "complementary determining regions" (CDRs). Deze CDRs zijn hypervariabel en zorgen voor de specifieke herkenning en binding van het antigen. De geschatte mogelijke variatie in antistoffen varieert van 10^{13} tot meer dan een onrealistische 10^{50}. Vanwege de enorme verscheidenheid in immunoglobulinen zou men kunnen aannemen dat aminozuursequenties van antistoffen uniek zijn voor elk individu en dat deze sequenties niet gemeenschappelijk voorkomen in andere individuen. Verschillende studies, met inbegrip van studies van onze groep, hebben echter aangetoond dat deze veronderstelling niet het hele verhaal is en dat er bewijs is voor een "repertoire bias". Dit betekent dat antistoffen worden onderworpen aan een bepaalde selectiedruk na herschikking van de immunoglobuline-genen (V, D en J-genen) en bindingsrijping, ten gunste van bepaalde genen uit het repertoire van de immunoglobuline-genen die worden gebruikt in respons op een bepaald antigeen. Soortgelijke CDR3 sequenties werden bijvoorbeeld bij verschillende mensen gevonden na immunisatie met tetanus toxine en een sequencing project met zebravissen toonde identieke CDR3 sequenties in verschillende vissen. Patiënten met paraneoplastische neurologische syndromen (PNS) worden geassocieerd met abnormale expressie van autoantistoffen tegen onconeurale antigenen zoals HuD, Yo, Amp of CV2. In de studie van Maat et al. werden primaire aminozuurstructuren van autoantistoffen exclusief geïdentificeerd in monsters behorende tot één specifieke autoantistof getypeerde PNS groep. Deze bevindingen bevestigen dat gemeenschappelijke aminozuurstructuren voorkomen bij personen met autoantistoffen tegen dezelfde onconeurale antigenen. Bovendien kon VanDuijn et al. door proteomics analyse van affiniteitsgezuiverde IgG van ratten na immunisatie met het onconeurale HuD antigeen, aantonen dat deze groep ratten gemeenschappelijke aminozuursequenties hadden ontwikkeld. Deze studie laat ook zien dat de ontwikkeling van antistoffen tijdens de immuunrespons niet een willekeurig proces is, maar dat de selectiedruk een bepaalde selectie van aminozuursequenties begunstigt in respons op een antigeen. Eerder beschreven we een nieuwe aanpak voor de identificatie en kwantificering van antigeen-specifieke sequenties van het gezuiverde IgG antigeen-bindend fragment (Fab) door massaspectrometrie. Een overlapping van 83% in MS signalen van IgG Fab gezuiverde peptiden liet zien dat het immunoglobuline repertoire van zeven gezonde donoren zeer vergelijkbaar is. In deze studie toonden we aan dat onze aanpak in staat is om met hoge reproduceerbaarheid en opbrengst, kwalitatieve (17%) en kwantitatieve (4%) verschillen in Fab sequenties tussen gezonde donor serummonsters door massaspectrometrie te detecteren. Eveneens waren wij in staat om de verschillende CDRs (complementarity determining regions) met massaspectrometrie te isoleren en te identificeren. Uit deze resultaten concluderen wij dat het met deze aanpak technisch haalbaar is om

CDRs van longkanker gerelateerde antistoffragmenten te identificeren zonder vereiste en voorafgaande kennis van het betrokken antigeen.

De mogelijkheid om verschillen tussen longkankerpatiënten en controles te vinden is afhankelijk van het aantal geïdentificeerde CDRs. Identificatie van CDRs door massaspectrometrie kan aanzienlijk worden verbeterd door moleculaire dissectie. Moleculaire dissectie verlaagt de complexiteit van de immunoglobulinemolecule, waardoor de ionsuppressie in de massaspectrometrie vermindert wat uiteindelijk leidt tot een aanzienlijk hogere gevoeligheid om deze zeer laag abundante CDR-peptiden te detecteren.

In **hoofdstuk 2** presenteren we een nieuwe methode voor het verder verlagen van de complexiteit van de immunoglobuline door moleculaire dissectie van IgG in kappa (κ) en lambda (λ) fragmenten. We voerden een proteomics analyse uit op de gezuiverde Fab, Fab-κ, Fab-λ, IgG-k en IgG-λ fragmenten van 10 NSCLC (non-small cell lung cancer) patiënten met stadium I en 10 gematchte controles van huidige en voormalige rokers van de NELSON studie. De moleculaire dissectie methode is goed reproduceerbaar (CV <5%) en geeft hoge totale zuiveringsopbrengsten (>90%) voor IgG-κ en IgG-λ gezuiverd uit IgG en voor Fab-κ en Fab-λ gezuiverd uit Fab. Deze methode liet ook een behouden normale κ:λ ratio in de Fab van zowel longkankerpatiënten als controles zien. Aangezien de gehele Fab fractie bestaat uit zowel de zware ketens als de beide lichte ketens, kappa (κ) en lambda (λ), identificeerden we meer CDRs in deze fractie dan in de andere afzonderlijke fracties. Daarnaast kan het lager aantal geïdentificeerde CDRs in IgG-κ en IgG-λ worden verklaard door het feit dat in deze fracties de CDRs van de zware keten ontbreken, inclusief de zeer variabele CDR3. Dit werd ook aangetoond door de ongeveer vier maal hogere verhouding van CDR3 (CDR1:CDR2:CDR3) gevonden in Fab, Fab-κ en Fab-λ ten opzichte van IgG-κ en IgG-λ. We identificeerden tweemaal zoveel CDRs in de gecombineerde Fab-κ, Fab-λ, IgG-κ en IgG-λ fracties dan in de afzonderlijke Fab fractie. Er werd geen significant verschil voor deze aantallen gevonden tussen de longkankerpatiënten en de gematchte controles. Fab-k leverde naast meer CDRs ook meer significant verschillende CDRs op dan de Fab-λ, IgG-κ en IgG-λ fracties. Immunoglobulinen die tijdens de ontwikkeling van de tumor tot expressie komen blijken voornamelijk uit zware en κ-ketens te bestaan. Vanwege deze bevindingen raden we massa spectrometrie van zowel de Fab en de Fab-κ fracties aan. Massaspectrometrie metingen van zowel de totale Fab als de afzonderlijke Fab-κ fractie vereist weliswaar twee keer zoveel tijd, maar levert wel 50% meer geïdentificeerde CDRs op en vergroot daarmee de kans op het vinden van longkanker gerelateerde CDR sequenties.

In **hoofdstuk 3** pasten we onze eerder beschreven IgG Fab zuivering methode toe op een patiënt-controlestudie (case-controle studie). Het doel van onze studie was het vinden van specifieke peptidesequenties in de antigeen-bindende fragmenten (Fab) van antistoffen die

longkankerpatiënten en controles van elkaar kunnen onderscheiden zonder de noodzaak om op voorhand de antigenen al te kennen. We voerden een proteomics analyse uit op gezuiverde IgG Fab-fragmenten van 44 NSCLC patiënten en 49 gematchte controles van de NELSON studie. Een model van 12 peptidesequenties van antistoffen was in staat om in deze hoog-risico groep de longkankerpatiënten van de controles onderscheiden. Dit antistof-peptide model bestaat niet enkel uit peptidesequenties van de hypervariabele CDR-regio's van de immunoglobuline, maar ook uit peptidesequenties van de framework- regio's van de immunoglobuline. Deze framework peptiden met slechts een paar mutaties hebben meer kans om in verschillende klonen voor te komen en zijn daardoor in hogere mate aanwezig (abundantie). Door hun abundantie is de kans groter dat deze peptiden in een zeer complex immunoglobuline monster door massaspectrometrie worden gedetecteerd. Naast dit complexiteitsprobleem is het ook mogelijk dat hypergemuteerde CDRs niet gemeenschappelijk voorkomen in patiënten. In plaats van de hypergemuteerde CDRs hebben de matig gemuteerde peptiden wellicht de beste algemene eigenschappen met betrekking tot abundantie, specificiteit en het gemeenschappelijk voorkomen onder patiënten als een potentiële biomarker voor longkanker. Longkanker is een heterogene ziekte die grote verschillen tussen patiënten veroorzaakt en verschillende immuunresponsen tegen verschillende tumor-geassocieerde antigenen kan ontlokken. Daarom is het niet onverwacht dat er niet één enkele antistof-peptide in onze dataset is gevonden die longkankerpatiënten van controles kan onderscheiden. Deze bevinding wordt ook door meerdere studies ondersteund. Daarom zijn we op zoek gegaan naar een model dat uit verschillende antistof-peptiden bestaat. Met behulp van een multivariaat antistof-peptide model waren we in staat om longkankerpatiënten en controles in de eerste set (NELSON 1) met een sensitiviteit van 96% en een specificiteit van 100% van elkaar te onderscheiden. In de tweede onafhankelijke validatie set (NELSON 2) vonden we een sensitiviteit van 84% en een specificiteit van 90%. Voor het neutraliseren van de experimentele en biologische verschillen hebben we het model voor elke dataset opnieuw gekalibreerd. Om de kans op het vinden van een vergelijkbaar model gebaseerd op willekeurige selectie van data uit te sluiten, hebben we een achtergrond evaluatie uitgevoerd. Deze achtergrond evaluatie liet zien dat het door ons gevonden antistof-peptide model significant beter is dan een model gebaseerd op permutatie van de data. Tot nu toe zijn alleen leeftijd en rookhistorie als selectiecriteria gebruikt voor het includeren van hoog-risico individuen in een screeningtrial. Het nauwkeuriger selecteren van hoog-risico individuen voor een screening trial kan waarschijnlijk worden bereikt door leeftijd en rookhistorie met aanvullende biomarkers te combineren. CT screening kan longkanker met hoge sensitiviteit en specificiteit na meerdere screening rondes en ongeveer één jaar na baseline screening (beginpunt screening) detecteren. Niettemin ondergaat ongeveer 27% van de deelnemers aan de NELSON trial onnodige invasieve en kostbare follow-up procedures die resulteren in goedaardige longziekten. CT screening resultaten verbeteren na een aantal follow-up scans, echter alleen na een betrekkelijk lange periode van gemiddeld één jaar na baseline screening.

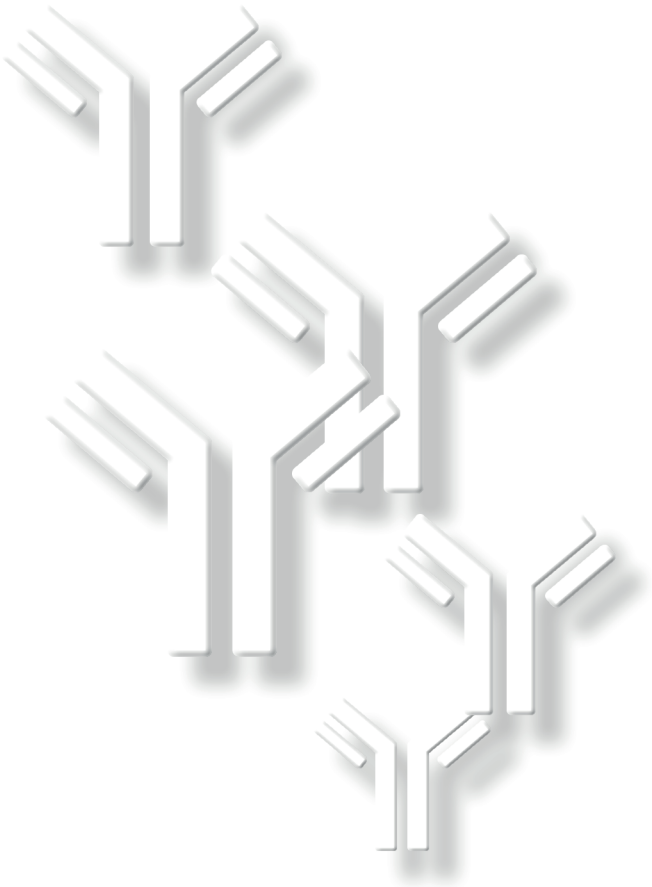
Daarom zijn er aanvullende diagnostische tests nodig om de diagnostische waarde van CT screening op baseline te verbeteren. Onze studie laat zien dat we niet alleen in staat zijn om longkanker met een antistof-peptide panel in een vroeg stadium te detecteren, maar ook in een eerder stadium dan CT screening, gemiddeld een jaar voordat de uiteindelijke diagnose is vastgesteld. Onze resultaten tonen aan dat autoantistof profilering een waardevolle aanvullende test zou kunnen zijn voor de vroegtijdige detectie van longkanker bij CT screening. Echter moeten er nog een aantal technische uitdagingen worden overwonnen voordat deze methode toegepast kan worden in de kliniek.

Diverse patiënt-controle studies hebben gesuggereerd dat autoantistoffen tegen het apoptose-remmer eiwit 'survivin', potentiele biomarkers kunnen zijn voor de vroegtijdige detectie van longkanker. Survivin autoantistof concentraties werden in deze studies gedetecteerd met behulp van een directe antigeen-coating ELISA (DAC-ELISA) met recombinant survivin als "capture" eiwit en vergeleken tussen longkankerpatiënten en gezonde bloeddonor controles. In 8% tot 52% van de longkankerpatiënten werden deze autoantistoffen gedetecteerd. In **hoofdstuk 4**, testten we de hypothese dat deze survivin autoantistoffen even goed te detecteren zijn met onze gebruikelijke sandwich ELISA vóór de radiologische diagnose in patiënten in een vroeg stadium van NSCLC. We vergeleken eerst de DAC-ELISA met onze sandwich-ELISA. Vanwege de meer robuuste resultaten van de sandwich ELISA, gebruikten we deze test voor het meten van survivin autoantistoffen in het serum van 50 NSCLC patiënten, zowel vóór als na de diagnose, en 50 gematchte controles voor rookgedrag van huidige en voormalige rokers uit de NELSON trial. Daarnaast hebben we deze antistoffen ook in serum van 20 laat stadium (IV) NSCLC patiënten van de NVALT-12 studie en 50 gezonde niet-rokers controles gemeten. In geen enkel serum van de NSCLC patiënten werden specifieke autoantistoffen tegen survivin gedetecteerd. Echter werd er zowel in de NSCLC patiënten als in de gematchte controles voor rookgedrag een opmerkelijke hogere (+24%, p <0,001) niet-specifieke binding ten opzichte van de gezonde niet-rokers controles gevonden. Dus zonder correctie voor niet-specifieke achtergrondbinding kan er een significant hogere binding voor NSCLC patiënten en rokers ten opzichte van niet-rokers als controles worden gevonden. Deze niet-specifieke binding kan in een sandwich ELISA beter gecontroleerd en aanzienlijk gereduceerd worden dan in een DAC-ELISA. Ook in western blotting werden geen autoantistoffen tegen recombinant survivin in de patiënten sera gevonden. Verschillen in de posttranslationale modificaties in survivin ten opzichte van recombinant survivin zou het ontbreken van anti-survivin reactiviteit tegen recombinant survivin kunnen verklaren. Echter ook western blotting met endogene survivin (HEK-293) toonde het ontbreken van anti-survivin reactiviteit in patiënten sera aan. We concluderen dan ook dat specifieke anti-survivin autoantistof reactiviteit hoogstwaarschijnlijk niet aanwezig is in serum van longkankerpatiënten zowel voor of na diagnose van longkanker. Hogere schijnbare anti-survivin autoantistof reactiviteit in serum van rokers ten opzichte van niet-rokers die door andere

studies worden gerapporteerd, wordt zeer waarschijnlijk veroorzaakt door niet-specifieke binding in rokers. Autoantistoffen tegen longkanker antigenen dienen met behulp van een sandwich ELISA of een andere goed gedefinieerde assay in een goed gecontroleerde en op rookgedrag geselecteerde patiëntengroep te worden onderzocht.

De ontdekking van ondefinieerbare nodules, "vlekjes", op een CT-scan van de longen kan leiden tot invasieve diagnose follow-up procedures, zoals een biopsie of een operatie die een aanzienlijke belasting en een medisch risico voor de patiënt kan betekenen. Om deze onnodige invasieve procedures voor patiënten met benigne (goedaardige) nodules te voorkomen is er behoefte aan een niet-invasief diagnostisch instrument om de benigne nodules van de maligne nodules (kwaadaardige) te kunnen onderscheiden. **Hoofdstuk 5** beschrijft onze validatiestudie van een 13-eiwit classificator en een 5-eiwit classificator in bloed, die zijn beschreven door Li et al. en Vachani et al. respectievelijk als een diagnostisch instrument met een hoge negatief voorspellende waarde (NPV) om de benigne van de vroeg stadium (IA) maligne nodules in patiënten met ondefinieerbare long nodules (IPNs) te kunnen onderscheiden. Voor onze validatiestudie gebruikten we serummonsters van 14 vroeg stadium IA-B NSCLC patiënten (64,3 % IPNs) en 16 goed gematchte benigne controles (81,2% IPNs) van een prospectieve cohortstudie (NELSON). In overeenstemming met de beschreven studies gebruikten we immunodepletie en MRM-massaspectrometrie om deze classificatie peptiden te analyseren. In tegenstelling tot de originele studie van Li et al. monitorden we vier transities in plaats van één transitie en gebruikten we stabiele isotoop gelabelde interne standaard (SIS) peptiden voor de meting van elke peptide. In onze studie konden drie van de dertien peptiden van de 13-eiwit classificator niet worden gekwantificeerd. In de vervolgstudie van Vachani et al. konden deze drie peptiden en ook vijf andere peptiden niet worden gekwantificeerd en werden daarna in hun validatiestudie verwijderd uit de classificator. Na validatie identificeerde onze 5-eiwit classificator benigne nodules met een NPV van 88% en een positief voorspellende waarde (PPV) van 27% bij een specificiteit van 31% en een sensitiviteit van 86%. Wij vonden in onze validatiestudie een ROC curve met een AUC van 0.65. Aangezien onze cohort ook bestond uit 36% stadium IB NSCLC nodules, waarvan de nodule grootte een hogere graad van maligniteit vertegenwoordigt, hadden we een betere relatie tussen voorspelling en ware uitkomst verwacht. Recent hebben Vachani et al. de klinische toepasbaarheid van hun 5-eiwit classificator getest in een prospectieve-retrospectieve studie van patiënten met IPNs die een invasieve diagnostische procedure hadden ondergaan. Hieruit bleek dat 69 (24.0%) van de maligne patiënten niet een noodzakelijke vervolgprocedure zouden ontvangen, terwijl 45 (68.2%) van de benigne patiënten een onnodige invasieve procedure zouden ondergaan. Op basis van deze resultaten en onze resultaten concluderen we dat deze 5-eiwit classificator klinisch niet bruikbaar is.

In een complexe en heterogene ziekte zoals longkanker is een scala aan eiwitten betrokken. Daarom heeft een panel van eiwitten meer kans dan een enkel eiwit om een hoge gevoeligheid en specificiteit als biomarker voor longkanker te bereiken. De conclusie van dit proefschrift is dat een panel van specifieke peptidesequenties in de antigeen-bindende fragmenten van antistoffen een sensitieve en specifieke niet-invasieve biomarker voor de vroegtijdige detectie van longkanker kan zijn. Verder onderzoek is nodig om de klinische toepassing als biomarker te realiseren. In **hoofdstuk 6** beschrijf ik hoe dit onderzoek tot stand kan komen.



List of abbreviations
Dankwoord
List of publications
PhD portfolio
About the author

APPENDICES

List of abbreviations

AAB	tumor-associated autoantibody
AAH	atypical adenomatous hyperplasia
ADC	adenocarcinoma
ALC%	average local confidence score
AUC	area under the curve
BSA	bovine serum albumin
CAD	collisional-activated dissociation
CaM	calmodulin
CDR	complementarity-determining regions
CID	collision-induced dissociation
COPD	chronic obstructive pulmonary disease
CT	computed tomography
CV	coefficient of variation
CXR	chest X-ray
D	discovery set
DAC	direct antigen-coating
DANTE	Detection and Screening of Early Lung cancer by Novel Imaging Technology and Molecular Essays
DLCST	Danish Lung Cancer Screening Trial
Dx	diagnosis
ELISA	enzyme-linked immuno sorbent assay
ES	early stage
ETD	electron-transfer dissociation
Fab	fragment antigen binding
Fc	fragment crystallizable
FMAT	fluorometric microvolume assay technology
H	heavy chain
HCD	higher energy collisional dissociation
HEK	human embryonic kidney
HRP	horseradish peroxidase
IAP	inhibitor apoptosis protein
IASLC	International Associations for the Study of Lung
IgG	immunoglobulin G
IPN	indeterminate pulmonary nodule
IQR	interquartile range
ITALUNG	Italian lung study
κ	kappa

λ	lambda
LC	lung cancer
LCC	large cell carcinoma
LCMS	liquid chromatography mass spectrometry
LDCT	low-dose spiral computed tomography
LS	late stage
LUSI	German Lung Screening and Intervention trial
mAb	monoclonal antibody
MILD	Multicentric Italian Lung Detection trial
MRI	magnetic resonance imaging
MRM	multiple reaction monitoring
MS	mass spectrometry
N	patient number
NA	not applicable
NELSON	Dutch-Belgian Lung Cancer Screening Trial (Dutch acronym)
NLST	National Lung Screening Trial
NSCLC	non-small cell lung cancer
NVALT	Dutch Association of Pulmonologists for Lung Diseases and Tuberculosis
OD	optical density
PBS	phosphate buffered saline
PBST	phosphate buffered saline containing 0.05% Tween-20
PET	positron emission tomography
PNS	paraneoplastic neurological syndromes
PTM	posttranslational modifications
PVDF	polyvinylidene difluoride
SC	smoking controls
SCD	squamous cell carcinoma dysplasia
SCLC	small cell lung cancer
scFv	single chain fragment variable antibodies
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SRM	selective reaction monitoring
SCC	squamous cell carcinoma
TAA	tumor-associated antigen
TBS	tris buffered saline
TBST	tris buffered saline containing 0.05% Tween-20
TSA	tumor-specific antigen
TMB	3,3',5,5'-tetramethylbenzidine
TNM	tumor, node and metastases
Tx	tumor stage unknown

UKLS	UK Lung Screening trial
V	validation set
VDT	volume doubling time
yrs	years

Dankwoord

Eindelijk is mijn proefschrift klaar!
Met het schrijven van dit dankwoord sluit ik een belangrijke fase in mijn leven af. Graag wil ik een ieder bedanken die heeft bijgedragen aan deze mooie, enerverende en soms ook pittige periode. Daarbij wil ik een aantal personen in het bijzonder bedanken.

Mijn copromotor, dr. Theo Luider. Beste Theo, graag wil ik je bedanken voor je vertrouwen en de kansen die je mij tijdens mijn promotieonderzoek hebt geboden. Ik herinner mij nog goed dat je tijdens een bespreking het vertrouwen in mij uitsprak en mij vroeg of ik wilde promoveren. Je altijd positieve houding en begeleiding tijdens moeizame periodes heeft mij zeer geholpen om mijn promotie te volbrengen.

Mijn promotor, prof. dr. Jan Lindemans. Beste Jan, ik ben je zeer dankbaar voor het vertrouwen in mij en de kans die je mij hebt geboden om te kunnen promoveren. Mede door jouw kritische en objectieve input zijn de hoofdstukken van dit proefschrift tot stand gekomen. Hartelijk dank voor je enorme steun en prettige samenwerking.

Dr. Rob van Klaveren. Beste Rob, in de korte tijd dat je mij kon begeleiden dank ik je hartelijk voor je betrokkenheid, ondersteuning en je feedback op mijn manuscripten.

Prof. dr. Peter Sillevis Smitt. Beste Peter, dank voor je constructieve en kritische opmerkingen tijdens mijn presentaties. Ik heb dit zeer gewaardeerd.

De leden van de kleine commissie. Joachim Aerts, Rainer Bischoff en Dimitris Rizopoulos, wil ik vriendelijk bedanken voor de bereidheid om plaats te nemen in de kleine commissie en de tijd en moeite die jullie hebben besteed aan het beoordelen van mijn proefschrift.

Natuurlijk gaat mijn dank ook uit naar alle co-auteurs. Mede door jullie bijdrage en prettige samenwerking zijn er mooie publicaties tot stand gekomen.

Graag wil ik al mijn collega's en oud-collega's van de Proteomics groep bedanken voor de inspirerende omgeving en de fijne sfeer waarin ik mijn promotieonderzoek kon uitvoeren. Beste collega's, Diana, Lennard, Christoph, Coskun, Lona, Marcel, Martijn, Peter Burgers, Nick, en oud-collega's Arzu, Azadeh, Dominique, Dana, Evert-Jan, Gero, Giovanny, Halima, Henk, Jeroen, Linda, Peter Maat, Robert-Jan, Roland en Vaibav bedankt voor de uiterst aangename samenwerking, de gezellige congressen, borrels en praatjes. In het bijzonder wil ik Christoph en Martijn bedanken.

Beste Christoph, ik heb je leren kennen als een bijzonder vriendelijke en hulpvaardige collega. Ik wil je oprecht bedanken voor het delen van je diepgaande kennis van massaspectrometrie en de behulpzame discussies over de proteomic data analyse.

Beste Martijn, ik heb door jou veel extra kennis opgedaan over antibodies en de bijbehorende experimenten. Thanks!

Collega's van de Neuro-oncologie bedankt voor de prettige sfeer op het lab en tijdens werkbesprekingen. Beste Esther, hartelijk bedankt voor je hulp op het lab en de gezellige lunches. Beste Eric, dank voor het redden van mijn computer data.

Mijn collega's van de Klinische chemie. Graag wil ik mijn PhD collega's bedanken. Beste Chérina, Maja en Pooja bedankt voor de fijne gesprekken en het delen van al het promotie lief en leed. Beste collega's Anne Marie, Johanna, Madeleine, Miranda, Monique, Paul en Richard bedankt voor jullie interesse voor mijn promotieonderzoek.

Beste Roel en Frank, allebei bedankt voor het aanleveren van de klinische data van de NELSON studie en de fijne medewerking.

Lieve paranimfen, Diana en Lennard, ik voel mij gesterkt dat jullie tijdens mijn verdediging naast mij staan. Dank voor al jullie steun ter voorbereiding van deze bijzondere dag. Lennard, dank je wel voor alle hulp tijdens mijn promotieonderzoek.

Lieve familie en vrienden. Alhoewel het voor jullie niet altijd duidelijk was wat ik allemaal deed tijdens mijn onderzoek hadden jullie altijd een luisterend oor voor mijn verhalen en toonden interesse voor mijn promotieonderzoek. Lieve ouders en schoonouders, ik weet dat jullie trots op mij zijn. Jolanda en René, bedankt voor alle saunabezoekjes en het gastronomisch entertainment bij jullie thuis. Rosaline, dank je wel voor je begripvolle en gezellige gesprekken. Dankzij jullie vond ik de nodige afleiding en ontspanning.

Tot slot, de meest belangrijke en dierbare persoon in mijn leven. Lieve Harro, door jouw creativiteit heeft mijn boekje een mooie cover gekregen. Dank je voor je liefde, geduld, enorme steun en aanmoediging tijdens mijn studie. Ook als het soms wel erg laat in de avond werd. Je hielp mij tijdens twijfels en moeizame momenten verder. Ik houd van je uit het diepst van mijn hart.

List of publications

Broodman I, Lindemans J, Sten HJM, Bischoff R, Luider TM. Serum protein markers for the early detection of lung cancer. *Provisionally accepted by J Proteome Res.*

Broodman I, Dekker LJM, Stingl C, Oudkerk M, de Koning HJ, Aerts JG, Lindemans J, Luider TM. Comment on "A blood-based proteomic classifier for the molecular characterization of pulmonary nodules." *Submitted*

Broodman I, VanDuijn MM, Stingl C, Dekker LJM, Germenis AE, de Koning HJ, van Klaveren RJ, Aerts JG, Lindemans J, Luider TM. Survivin autoantibodies are not elevated in lung cancer when assayed controlling for specificity and smoking status. *Cancer Immunology Research*. 2015; 4(2): 1–8.

de Costa D, **Broodman I**, Calame W, Stingl C, Dekker LJM, Vernhout RM, de Koning HJ, Hoogsteden HC, Sillevis Smitt PAE, van Klaveren RJ, Luider TM, VanDuijn MM. Peptides from the variable region of specific antibodies are shared among lung cancer patients. *PLoS One*. 2014; 9: e96029.

Broodman I, de Costa D, Stingl C, Dekker LJM, Vanduijn MM, Lindemans J, van Klaveren RJ, Luider TM. Mass spectrometry analyses of kappa and lambda fractions result in increased number of complementarity determining regions identifications. *Proteomics*. 2012; 12: 183–191.

de Costa D, **Broodman I**, Vanduijn MM, Stingl C, Dekker LJM, Burgers PC, Hoogsteden HC, Sillevis Smitt PAE, van Klaveren RJ, Luider TM. Sequencing and quantifying IgG fragments and antigen-binding regions by mass spectrometry. *J Proteome Res*. 2010; 9: 2937-45.

PhD portfolio

Name	Ingrid Broodman
	Erasmus MC, department Clinical Chemistry
Research school	Erasmus Postgraduate school Molecular Medicine
PhD period	2009-2014
Promotor	Prof. dr. J. Lindemans
Copromotor	Dr. T.M. Luider

PhD training	Year	Workload (ECTs)
General academic skills		
English Biomedical Writing and Communication	2010	4.0
Presenting Skills	2012	1.0
Photoshop and Illustrator CS5	2013	0.3
Research skills		
Principles of Research in Medicine and Epidemiology	2009	0.7
Introduction to Data Analysis	2009	1.0
Regression Analysis for Clinicians	2012	1.9
Basic Introduction Course on SPSS	2013	0.8
In-depth courses		
Advanced course on Molecular Immunology	2009	3.0
Mass Spectrometry of Peptides and Proteins, Conference of the American Society for Mass Spectrometry, Philadelphia, USA	2009	0.6
Case Studies in Quantitative Proteomics, Conference of the American Society for Mass Spectrometry, Denver, USA	2011	0.6
Presentations		
Mass Spectrometry Analyses of Kappa and Lambda Fractions of IgG Result in Increased Number of Complementary Determining Regions Identifications, ASMS, Denver, USA	2011	1.0
Mass Spectrometry Analyses of Kappa and Lambda Fractions of IgG Result in Increased Number of Complementary Determining Regions Identifications, IASLC, Amsterdam, the Netherlands	2011	1.0
Scientific meetings of JNI, Neuro-Oncology and Clinical Chemistry departments, Erasmus MC, the Netherlands	2009-2014	5.5
(Inter)national conferences		
The American Society For Mass Spectrometry, Philadelphia, USA	2009	1.5
Proteomics and Metabolics Biomarker Discovery in CSF, Rotterdam, the Netherlands	2009	0.5
The American Society For Mass Spectrometry, Denver, USA	2011	1.5

14th World Conference on Lung Cancer, IASLC, Amsterdam, the Netherlands	2011	1.0
Attended meetings, seminars and workshops		
Seminars, workshops and PhD-days, Erasmus MC	2009-2012	1.0
Pubmed workshop, Erasmus MC	2009	0.2
Methodologie van Patiëntgebonden Onderzoek en Voorbereiding van Subsidieaanvragen, Erasmus MC	2013	0.2
Weekly Josephine Nefkens Institute scientific meetings	2009-2014	2.0
Weekly Neuro-oncology/Proteomics meeting	2009-2014	2.0
Symposia		
NELSON Symposium, Utrecht, the Netherlands	2009	0.4
NELSON Symposium, Rotterdam, the Netherlands	2011	0.4
The Increasing Role of Mass Spectrometry in Laboratory Medicine, Symposium LUMC, Leiden, the Netherlands	2012	0.4
Interview		
Interview Oncologie TV,14th World Conference on Lung Cancer, IASLC, Amsterdam, the Netherlands	2011	0.3

About the Author

Ingrid Broodman was born on January 18th, 1965 in Rotterdam, the Netherlands. In 1985 she graduated in Clinical Chemistry at the "Van 't Hoff Institute" in Rotterdam. From June till October 1985 she worked as a technician at the Sanquin Blood Supply Rotterdam. In October 1985 she started as a technician at the department of Clinical Chemistry at the Erasmus University Medical Center Rotterdam. In 1987 she obtained the Post-graduate Immunology degree at the "Van 't Hoff Institute" Rotterdam. In 1998 she was promoted to senior technician. In that role she was regularly deployed for special research projects in the field of Lipids, Cancer, Allergy and Immunology, which thanks to her contribution resulted in different publications. In that position she completed different in-depth courses and trainings. In 2007, she seconded as research technician to the Laboratory of Neuro-Oncology, Clinical & Cancer Proteomics, of the department of Neurology under the leadership of Dr. Theo Luider, to provide support at a PhD research project of the NELSON trial. In the course of this research project she was offered a PhD promotion study. In 2009 she was appointed as PhD student to the Department of Clinical Chemistry in a collaboration study between the departments of Clinical Chemistry, Neurology and Pulmonology at the Erasmus University Medical Center Rotterdam under supervision of Prof. Dr. J. Lindemans, Dr. T.M. Luider and Dr. R.J. van Klaveren. The results of the studies performed during this period are presented in this thesis.