DISSERTATIONES MEDICINAE UNIVERSITATIS TARTUENSIS
186

DISSERTATIONES MEDICINAE UNIVERSITATIS TARTUENSIS
186

TÕNU VOODER

Molecular differences and similarities between histological subtypes of non-small cell lung cancer

TARTU UNIVERSITY PRESS

Department of Cardiology, University of Tartu, Estonia

Dissertation is accepted for the commencement of the degree of Doctor of Philosophy (in medicine) on 15.06.2011 by the by the Council of the Faculty of Medicine, University of Tartu, Estonia

Supervisors: Andres Pulges MD, PhD
University of Tartu. Faculty of Medicine.
Clinic of Cardiology

Professor Andres Metspalu MD, PhD
University of Tartu. Institute of Molecular and Cell Biology.
Department of Biotechnology

Reviewers: Professor Sulev Kõks MD, PhD
University of Tartu. Faculty of Medicine.
Department of Physiology

Professor Andres Salumets PhD
University of Tartu. Faculty of Medicine.
Department of Obstetrics and Gynecology

Opponent: Professor Dan Grandér MD, PhD
Karolinska Institutet
Cancer Center Karolinska
Department of Oncology and Pathology

Commencement: A. Linkberg Auditorium, Puusepa Str. 8, Tartu 22.08.2011 at 10:00.

The publication of this dissertation is granted by the University of Tartu.

ISSN 1024–395X
ISBN 978–9949–19–824–5 (trükis)
ISBN 978–9949–19–825–2 (PDF)

Tartu Ülikooli Kirjastus
www.tyk.ee
Tellimus nr. 490

To my family

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

1. Välk K. *, **Vooder T. ***, Kolde R., Reintam MA., Petzold C., Vilo J., Metspalu A. Gene expression profiles of non-small cell lung cancer: survival prediction and new biomarkers. Oncology: 79(3–4), 283–292, 2010

2. **Vooder T***, Välk K*, Kolde R, Roosipuu R, Vilo J, Metspalu A. Gene Expression-Based Approaches in Differentiation of Metastases and Second Primary Tumour. Case Reports in Oncology: 21;3(2), 255–261, 2010

3. Urgard E*, **Vooder T***, Võsa U, Välk K, Mingming Liu, Cheng Luo, Hoti F, Roosipuu R, Annilo T, Laine J, Frenz CM, Liqing Zhang, Metspalu A. Metagenes associated with survival in NSCLC. Cancer Informatics: 10, 175–183, 2011

* These authors contributed equally to this work.

My contributions to the articles referred to in the current thesis are as follows:

Ref. I	Collection of samples and clinical data, participation in the experimental process and writing of the manuscript.
Ref. II	Study design, collection of samples and clinical data and manuscript writing.
Ref. III	Manuscript writing, participating in study design and collection of samples and clinical data.

LIST OF ABBREVIATIONS

AC	Adenocarcinoma
AAH	Atypical adenomatous hyperplasia
BAC	Bronchioloalveolar carcinoma
BRCA	Breast cancer susceptibility protein
CIS	*Carcinoma in situ*
CT	Computed tomography
CUP	Carcinoma with Unknown Primary
DRC	DNA Repair Capacity
EGFR	Epidermal Growth Factor Receptor
ERCC	Excision repair cross-complementation group
FNAB	Fine needle aspiration biopsy
GO	Gene Ontology
HE	Hematoxylin-eosin
HPV	Human Papilloma Virus
HRT	Hormone replacement therapy
IASLC	International Association for Study of Lung Cancer
ICL-R	Interstrand cross-link repair
KRAS	Kirsten rat sarcoma viral oncogene homolog
LC	Lung cancer
LCC	Large cell carcinoma
LCNEC	Large cell neuroendocrine carcinoma
NE	Neuroendocrine
NED	Neuroendocrine differentiation
NER	Nucleotide excision repair
NSCLC	Non-small Cell Lung Cancer
PCA	Principal component analysis
qRT-PCR	Quantitative real-time polymerase chain reaction
RIN	RNA integrity number
RFS	Recurrence-free survival
RRM	Ribonucleotide reductase messenger
SCLC	Small Cell Lung Cancer
SCC	Squamocellular cancer
SEER	Surveillance, Epidemiology and End Results Database
TRU	Terminal respiratory unit
WHO	World Health Organization

I. INTRODUCTION

Lung cancer (LC) is currently the leading cause of cancer death. This high mortality rate is partly due to delayed diagnoses and challenges in differentiating subgroups suitable for various treatment options. Moreover, it is widely recognized that the use of light microscopy, hematoxylin-eosin staining, immunohistochemistry and diagnoses according to the Tumor Node Metastasis (TNM) system are not sufficient for evaluating prognoses or for providing sufficient information regarding expected treatment response.

However, new methods are becoming available for the study and diagnosis of various cancers, including LC.

High throughput technologies for the study of single nucleotide polymorphisms, mRNA and microRNA expression, DNA methylation, alternative splicing and protein synthesis, represent useful tools for research studies and clinical evaluations of various types of cancer. Moreover, whole genome expression profiles can be obtained using gene microarrays. The application of these techniques have the potential to facilitate a differentiation of various cancer subtypes, the evaluation of prognosis, the selection of patient subgroups for different treatment modalities and the selection of novel markers or target molecules for further research.

Already, previous studies of gene expression have identified a large number of valuable molecular markers. Furthermore, it is hypothesized that different sets of markers would be useful for diagnostic, prediction and prognostic aims, since the use of a single molecular marker could be subject to variability and would not accurately represent prognoses or predictions. In addition, a single marker would be insufficient to describe the biology of an entire living organism, including cancer cells.

In the current study, gene expression profiles of non-small cell lung cancer (NSCLC) samples were obtained using genome-wide microarrays. Specifically, data associated with the differentiation of NSCLC subtypes and gene expression related to patient survival for each subtype, were analysed. These results demonstrate the possibility of staging and treating NSCLC, or other cancer types according to revised TNM staging system that incorporates gene expression details of tumour biology to improve patient diagnosis and prognosis.

II. LITERATURE REVIEW

Lung Cancer Epidemiology

LC is a malignant tumor of epithelial origin and exhibits classical cancer features such as uncontrolled growth, tissue invasion, and the ability to metastasize into lymph nodes and distant organs. The high incidence of this usually fatal disease is relatively new phenomenon. The number of deaths attributed to LC begun to rise in the late 1930s and then reached its highest levels by the early 1990s (Jemal et al., 2009).

In addition, the relationship between smoking as a main risk factor and LC as a disease was historically noted by the German Nazi doctors Eberhard Schairer and Erich Schöniger in 1943 and was subsequently published by Sir Richard Doll in non-nazi world in 1950 (Doll and Hill, 2004, Doll et al., 2004, Proctor, 1999, Doll, 1950, Doll and Hill, 1950, Schairer and Schoniger, 2001).

Typically LC disease affects patients usually in their 60s and 70s and has remained uncommon in younger patients (age ≤40 years). LC is characterized as a lifestyle-associated disease, which has a surprisingly poor prognosis even among patients whose tumor is discovered early and treated with radical surgery (Fry et al., 1999). Furthermore, LC currently affects males more often then females, yet this gap is narrowing. Moreover, this trend has been consistently observed in both United States and Europe. Main risk factors for LC include: cigarette, pipe, cigar, and cannabis smoking, occupational exposure to radon, asbestos, nickel, chromium, and arsenic as well as exposure to radiation, air pollution, and passive smoking (Alberg et al., 2007, Aldington et al., 2008). Moreover, there are approximately 60 carcinogens present in cigarette smoke, with at least 16 components in smokeless tobacco products having sufficient evidence of carcinogenity, and about 20 of them have strong association with LC according to studies in laboratory animal models (Hecht, 2003, Hecht, 1999).

Correspondingly, when smoking has been discontinued, the risk of LC diminishes significantly. However, carcinogenic damage of bronchial epithelium cells remains for years after the cessation of smoking (Spira et al., 2004, Sridhar et al., 2008).

LC has spread worldwide, and affects both genders depending on lifestyle and habits. Although the estimated number of cancer deaths and new cancer cases has slowly decreased, LC is main type of cancer diagnosed and the leading cause of cancer deaths for males in developed countries (Jemal et al., 2008, Jemal et al., 2009, Jemal et al., 2006, Jemal et al.).

Previously, a decrease in the incidence of LC was observed among men in the late 1980s, and among women in 1999 (Rivera and Stover, 2004, Jemal et al., 2005).

In Estonia, LC is second leading cause of cancer among men, and the seventh leading cause of cancer in women. Correspondingly, the number of new

LC cases diagnosed in men in 2006 was 509 and 154 in women. In general, the incidence of LC has decreased among men and stabilized among women compared to the previous decade. However, LC as a cause of death is not currently able to be evaluated for the Estonian people due to governmental regulations (Estonian Cancer Registry 2009, Eesti Arst 10; 2009).

Histological classification of lung cancer

According to currently valid World Health Organization (WHO) classification guidelines, LC can be divided into two main types, non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). NSCLC is the predominant histological cancer group accounting approximately 80–85% of all lung cancers while SCLC is the most typical neuroendocrine (NE) LC.

Moreover, there are four major most common histological types of LC, and these include adenocarcinoma (AC), squamocellular carcinoma (SCC), large cell carcinoma (LCC), and small cell carcinoma accounts for 99% of all LC cases.

For cases of SCLC these are distinguished by more aggressive features, they usually have metastases present at diagnosis and are associated with a substantially poorer prognosis than NSCLC (Brambilla et al., 2001).

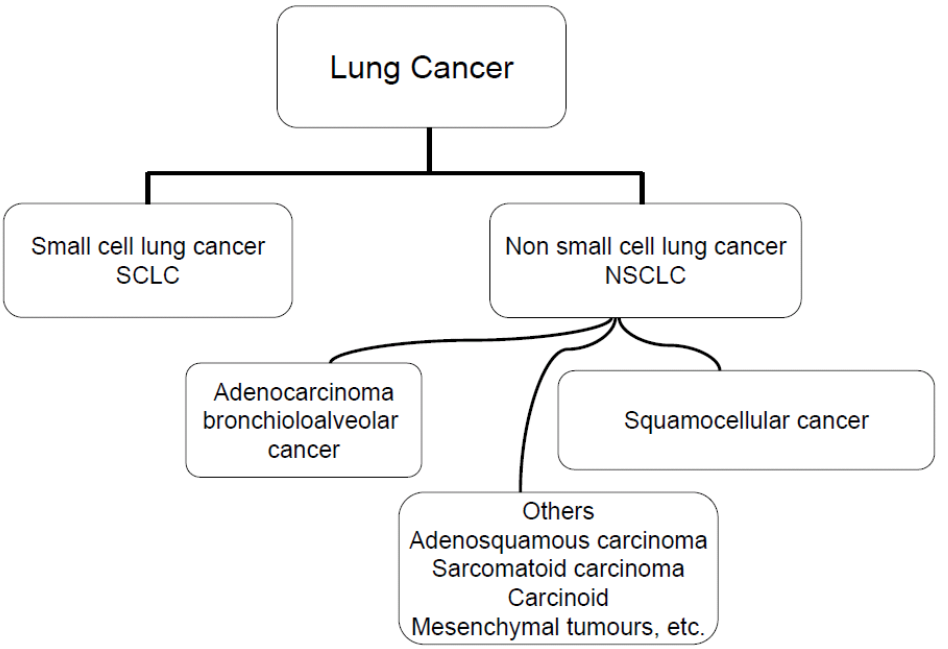

Figure 1. Classification of the histological forms of LC.

Histological forms of NSCLC

LC may arise from the epithelium of major bronchi, smaller bronchi, bronchioles, or certain alveolar cells. Depending on the epithelial source and histological structure, LC has been divided into various subclasses.

Squamocellular cancer s. epidermoid cancer (SCC) mainly exhibits central localization and originates in the bronchial metaplastic squamous epithelium. This tumor type is microscopically quite heterogenous, and is characterized depending on the differentiation grade determined according to keratinization present, intercellular bridges, and rarely, keratine pearl formation. Precursor lesions of this cancer type include bronchial squamous epithelium dysplasia and *carcinoma in situ* (CIS)(Kerr, 2001).

Adenocarcinoma (AC) tends to have more peripheral localisation, is slightly more common in females, and has a glandular epithelium origin. Exact histological structures and certain cell types of the respiratory tract that represent possible sources of AC remain unknown. However, a few relatively more evident, precursor lesions of AC include atypical adenomatous hyperplasia (AAH) of the peripheral airways, type II pneumocytes, and Clara cells, which express markers similar to AC of the lung (Yatabe et al., 2002, Chapman and Kerr, 2000).

AC is the most frequent histological subclass for both genders in many countries, and its frequency has grown in the last decade. Currently, it is the most prevalent form of LC in younger males and women of all ages, in non-smokers, and former smokers (Wahbah et al., 2007). Furthermore, most ACs are histologically heterogenous, with mixed type phenotypes (e.g. acinar, papillar, bronchioloalveolar, and solid AC with mucin formation) being more common than any single pattern type. In general, AC has been strongly associated with with a poorer prognosis than squamocellular cancer (SCC). One possible reason for this observation is the increased invasiveness and tendency for blood vessels invasion to be associated with AC, which can facilitate the occurrence of distant metastases (Bodendorf et al., 2009).

Bronchioloalveolar cancer (BAC) is a relatively rare subclass of AC which exhibits mostly non-invasive, or slightly invasive, features (Brambilla et al., 2001). Therefore, ACs with stromal, vascular, and pleural invasion are excluded from this subclass. Furthermore, ACs exhibiting these features are classified as AD mixed type with predominant bronchioloalveolar patterns, while BAC is more often spread from satellite nodules. A substantial computed tomography (CT) study of non-symptomatic patients showed that concomitant BAC nodules were detected in 47% of cases. (Vasquez et al., 2001) Moreover, pneumonitis-like tumors and metachronous-type tumors of BAC are also well-known forms of this cancer.

Patients with BAC tumours have also exhibited slightly better survival rates when diagnosed histologically, or when distinguished using gene expression profiles (Hayes et al., 2006).

Large cell carcinoma (LCC) represents a heterogenous group of undifferentiated cancers, including five histologically different variants. One of these variants, large neuroendocrine carcinoma (LNEC), is associated with neuroendocrine differentiation (NED). Moreover, LCNEC itself has four subclasses depending on the morphological and immunohistochemical parameters expressed (Brambilla et al., 2001, Travis, 2009). The main difference between LNEC and SCLC is the larger size of the anaplastic cells present and a lack of "salt and pepper" chromatin. Furthermore, a diagnosis of LCNEC is based on the presentation of at least one common NED marker during immunohistochemical evaluations.

A very rare precursor of NED tumours also includes diffuse idiopathic neuroendocrine cell hyperplasia.

NED is a common feature, and is present in 16–20% of other NSCLC-type cases (Sterlacci et al., 2009). While NED lung carcinomas do not exhibit the morphology associated with NE phenotype under light microscopy, NED can be detected in immunohistochemistry assays and/or electronmicroscopy studies.

Moreover, NED alone does not appear to be an independent prognostic marker for recurrence or patient survival in cases of NSCLC (Linnoila et al., 1994, Sterlacci et al., 2009, Howe et al., 2005), neither has it been found to be a predictive marker for any other cancer treatment modality known so far (Sterlacci et al., 2009).

Small cell lung cancer (SCLC) is the most common NE lung tumor diagnosed. Cases od SCLC present with the proliferation of small cells with morphological features that include minimal cytoplasm, ill-defined borders, finely granular „salt and pepper" chromatin, absent or inconspicuous nucleoli, frequent nuclear molding, and high mitotic activity. Until recently, no phenotypically identifiable lesion has been identified as a precursor of SCLC (Wistuba and Gazdar, 2006, Dacic, 2007). Taking into an account the similarity with SCLC, it has been proposed that it may originate from extensive and chronic damage present at the molecular level of otherwise microscopically normal bronchial epithelium (Wistuba and Gazdar, 2000, Wistuba et al., 2000).

Clinically, SCLC presents with a rapid and aggressive course, has a short-lived response to chemoradiotherapy, and is associated with 2-year and 5-year overall survival rates of 9% and 3% respectively (Paesmans et al., 2000). Although SCLC, in general, is not a disease suitable for conventional surgical treatment, in selected cases surgery has achieved a certain rate of success (Chandra et al., 2006). However, identifying predictive and prognostic parameters and markers of SCLC remain a challenge.

In the current study, due to differences in clinical behavior, therapeutic implications, and epidemiological features of LCNEC versus SCLC, these tumour types are referred to separately. Moreover, in the analysis of gene expression profiling performed, these two types of cancer are not considered.

Changes in histological distribution of lung cancer

Although LC is a relatively new disease, the distribution of different histological forms of it have changed over the past 25 years. For example, the incidence of AC has increased in both genders (Devesa et al., 2005), although previously, AC was the main histological form in women. Currently, AC is the most common form of LC in men, and accounts for approximately 36% and 46% of cases diagnosed in men and women respectively (Alberg et al., 2005, Rivera and Stover, 2004, Wahbah et al., 2007). Compared to corresponding studies published in the 1980s, this indicates a gain of approximately 10% in the number of cases for both genders. The same international study reported a decrease in the number of small cell and SCC cases for both genders, although SCLC remained the most common form of LC among women, with an incidence of 18% versus 13% in men. This observation may reflect the changing trend in smoking habits, since both, SCLC and SCC are forms of LC that are strongly related to smoking habits (Alberg et al., 2005). Furthermore, the incidence of LCC (undifferentiated) among men has increased slightly and currently accounts for about 18% of LC cases. The same cancer type in women has maintained a stable rate of incidence since 1980, and currently accounts for about 10% of LC cases.

Since 1979, the possible role of viral infections, primarily human papilloma virus (HPV), in the genesis of NSCLC has been investigated and mostly in relation of SCC (Roglic et al., 1975, Rubel and Reynolds, 1979, Inoue et al., 2001, Klein et al., 2009). For exemple, in a study of a population in Taiwan, non-smoking women were observed to develop AC of the lung in relation to the presence of oncogenic HPV (Cheng et al., 2001). However, a recent study reported contrasting results (Koshiol et al.). Therefore, the exact role of HPV in NSCLC remains unclear, and cannot be excluded as a potential co-carcinogen in the development of any NSCLC histological subtype (Syrjanen, 2002).

TNM classification of Lung Cancer

The currently valid 7th NSCLC staging system consideres the size and location of the primary tumor (T), the involvement of lymphnodes (N), and the presence of metastases (M) (e.g., TNM staging). Furthermore, it incorporates an analysis of data regarding clinical, surgical, pathological, and patient survival information from 100,869 primary NSCLC patients (Rami-Porta et al., 2009a, Rami-Porta et al., 2009b). However, despite the TNM staging criteria begin an essen-

tial aspect of LC management and currently the best method for the description of cancer spread, it requires further refinement. For example, it is also associated with a large number of unexpected relapses, and therefore, does not completely explain tumour behavior. The staging of the current study is based on Lymph Node classification for Lung Cancer staging (Mountain and Dresler, 1997) and TNM classification of malignant tumours (6^{th} edition).

Molecular biological staging

Currently, biological staging is not included in TNM staging system, although it has been proposed that the addition of molecular staging could further improve the TNM system (D'Amico, 2008).

Biological staging would identify target oncogenes, oncogenic protein products, growth factors, and their receptors, adhesion molecules, and other molecular markers which provide information regarding patient survival and response to different therapy modalities. Biological staging could also be applied to the characterization of primary tumours, lymph nodes, bone marrow, serum or metastases.

Diagnosis and prognosis of NSCLC

Sampling is essential for a histological diagnosis of cancer and most tumors are classified on the basis of light microscopy alone. Hematoxylin-eosin (HE) staining and immunohistochemical staining are then used to confirm histological findings. Currently, the most important expression markers used to characterize cancer tissue and to diagnose NSCLC include: AE1/AE3, CK5/CK6, CK7, CK20, TTF1 and some others (Chu et al., Camilo et al., 2006, Su et al., 2006). Despite these tools, however, the ability to diagnose LC in its early stages remain a challenge.

In general, the prognosis of NSCLC is poor, although survival differences are associated with different cancer stages and treatment modalities. For example, according to the Surveillance, Epidemiology, and End Results (SEER) database, there are survival differences between younger ($\leq$ 40 years) and older patients (> 40 years), except for stage IV cases where survival is poor independent of age (Subramanian et al.). Furthermore, for radically cured cases, the 5-year survival rate is 80% for stage Ia cases, 60% for stages Ib and IIa , 42% for stage IIb, 30% for stage IIIa, and ~ 20% for stages IIIb and IV (Goya et al., 2005) .

III. TREATMENT OF NSCLC

General considerations

Treatment of LC depends strongly on the extent of lymph node invasion and the presence of distant metastases. In the latter case, metastases occur more often in the later stages than in SCLC. Surgical treatment is the primary treatment option for NSCLC patients that are fit for surgery and have a resectable tumour. For example, radical treatment is possible for stage I cases with lobectomy or pneumonectomy with lymphadenectomy considered standard procedures and the best treatment option (Scott et al., 2007). For stage II – IIIa patients, a combination of surgery and chemo/chemoradiotherapy has been shown to be beneficial (Alberts, 2007).

Modalities of chemo- and radiotherapy

Conventional cisplatin-based chemotherapy is used for patients with stage IIa–IV LC. Chemotherapy is also considered for the treatment of stage Ib tumors that have a diameter T>4 cm. Adjuvant therapy is prescribed when all detectable disease appears to be removed, yet a risk for relapse due to occult disease remains.

In a meta-analyses of 4584 patients with NSCLC the benefit of adjuvant cisplatin-based chemotherapy was observed to largely depend on stage of disease being treated. For example patients with Ia experienced detrimental effects, while the effects on stage Ib cases are still being evaluated. For stage II and III cases, improved survival data has been obtained (Pignon et al., 2008).

For patients with stage IIIa N2 disease, cisplatin-based neoadjuvant chemotherapy is administred prior to surgery, with the goal of reducing the size of the tumour. To evaluate any restaging that may occurred as a result of chemotherapy, a CT scan and minimal invasive diagnostic interventions are used. In the case of persistent N2 disease, surgical treatment is not the best treatment, and additional treatment strategies have to be considered.

Radiotherapy can be a treatment option for patients that have not undergone a radical operation. Moreover, postoperative radiotherapy is not recommended for stage I–II and IIIa patients that have undergone radical resection, or patients with mediastinal lymphnodes involvment. In contrast, curative radiotherapy as a single treatment modality should be administered to patients with less-than-standard surgical resections, or for patients unfit to undergo standard surgery.

For patients with locally advanced, unresectable stage III NSCLC, concurrent chemotherapy and thoracic radiotherapy should be considered (D'Addario and Felip, 2008). In many cases, treatment results have also been observed to be superior in women versus men (Brahmer et al.).

Correspondingly, the Radiation therapy Oncology Group trials have observed an overall better treatment effect in women, even with radiotherapy alone (Werner-Wasik et al., 2000).

Treatment possibilities according to gene expression data

Currently, RNA expression microchips are used in clinical practice as part of treatment strategies for breast cancer, with three expression-based tests being commercially available. For example, Oncotype DX is the main predictive test that is used, and it is based on an expression analysis of 21 genes. This test is used to indicate which oestrogene-positive breast cancer patients can avoid adjuvant chemotherapy (Albain et al.). Mammaprint is another commercially available prognostic test which evaluates the expression of 70 genes in breast cancer tissues (Glas et al., 2006, van 't Veer et al., 2002). Lastly H/I is a two gene ratio assay that predicts patient an outcome in response to tamoxifen therapy, and is commercially produced by AvariaDx (Marchionni et al., 2008).

The application of chemotherapy to cases of NSCLC is mostly based on histological patterns of the tumor detected, while gene expression profiling is used to characterize histologically-specific features. For example, only non-SCC patients have recently been observed to be good candidates for pemetrexed and bevacizumab treatments. Moreover, gene expression assays detecting *TTF1, CK5, CK13* and *EGFR* have been associated with high specificity and sensitivity for effective discrimination of AC and SCC (Subramanian et al.).

IV. GENE EXPRESSION OF NSCLC

Gene expression is a process during which the heritable genetic information is transcribed and applied to the synthesis of a functional product of a gene. Correspondingly, there are two crucial steps during this process: transcription, whereby a copy of mRNA is produced using a DNA strand as a template and protein synthesis or translation, which results in the synthesis of a functional protein. However, there are some genes that encode a functional non-coding RNA. The flow of genetic information from DNA to RNA to protein is known as central dogma of molecular biology and gene expression is the most fundamental process for any type of cell. Following gene expression, the genotype of a cell becomes associated with corresponding phenotype, which allows the processes performed by a cell, and its status to be evaluated accordingly. Using whole genome microchips, gene expression profiles can be obtained, reflecting changes in cell function. The data obtained from microchip analyses can also be visualized graphically using a "heatmap", where up-regulated and down-regulated genes are coloured to distinguish these two gene groups. Furthermore, samples and genes are usually clustered hierarchically, thereby illustrating the similarity of the samples based on gene expression data (Figure 2).

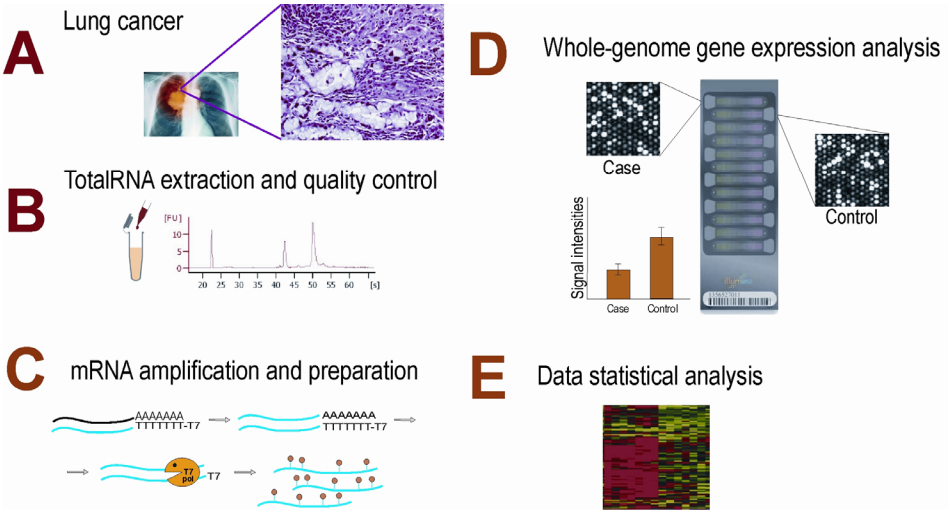

Figure 2. Principles of gene expression analysis.

Analysis of gene expression data is a method in molecular biology to evaluate thousands of known genes or transcripts simultaneously. For cancer, gene expression is suitable for distinguishing different cancers based on expression of a single gene (Llinares et al., 2004), or the profiling of a set of genes. Moreover, depending on the research question, gene expression data can be evaluated

in relation to a primary tumor, metastases, or a particular cell line (Toffalorio et al., 2009, Nevins and Potti, 2007, Hsu et al., 2009, Talbot et al., 2005). In some cases, gene expression profiles can be extremely similar in different parts of the same tumor, thereby suggesting that limited number of clinical samples that collected can represent tumor biology overall (Raponi et al., 2006). In addition, it is possible to perform RNA-based gene expression analyses using very small quantities of tissue (Lim et al., 2003).

The first report describing the possibility of using gene expression profiling for an evaluation of diffuse large B-cell lymphomas prognoses was published by Alizadeh and coauthors in 2000 (Alizadeh et al., 2000). This classical study was followed by similar studies of other malignant diseases including LC. For example, basic study of lung ACs was published by Bhattacharjee and coauthors, where different subtypes of AC and metastases associated with different primary tumours were detected using gene expression assays (Bhattacharjee et al., 2001). By using gene expression profiling to identify different subtypes of cancer, it is possible to find associations between various profiles and patient survival (Wigle et al., 2002, Garber et al., 2001).

Previously, the main subtypes of NSCLC were elegantly differentiated by Takeuchi and coauthors based on combination of expression data with *EGFR* and *K-ras* mutational status. As a result, two terminal respiratory unit related AC subtypes were revealed (Takeuchi et al., 2006, Dobbin et al., 2005). In addition, a significant correlation between *EGFR* mutations in terminal respiratory unit- type ACs and a poor prognosis was identified. In contrast, *K-ras* mutational status was not found to have an influence on AC patients outcame.

Reproducibility of microarray data for the purpose of differentiating LC subgroups is also challenge in NSCLC research. For example, in work by Hayes and coauthors, the applicability of expression based differentiation to three publicly available cohorts was demonstrated (Hayes et al., 2006).

Clinically, the value of NSCLC gene expression analyses is increasing. For example, it is now possible to present prognostic and predictive markers prior to treatment, to differentiate different types of cancers (Garber et al., 2001), distinguish metastases from primary tumors and to identify the source of metastases of unknown primary tumors.

A prognostic marker is a patient or tumor characteristic that identifies a better or worse outcome for a patient regardless of treatment modality. Moreover, a predictive marker is a patient or tumor characteristic that identifies the type of outcome expected, e.g., better or worse than if treatment was not performed (Coate et al., 2009).

There have been studies that not only provide evidence that gene expression profiling can be used to predict treatment outcome and establish molecular prognosticators, but can also be used to identify known and unknown genes differentially expressed in cancers and normal tissues, e.g., ASs and SCCs, as well as tumours associated with good or bad outcomes.

As mentioned previously, conventional NSCLC chemotherapy is performed using platinum based-medicines. Moreover, in the last few years a number of molecular biological studies have been performed to assist in the selection of chemoherapeutic agents based on gene expression profiles of the cancer present. Specifically, expression of *ERCC* and *RRM1* have been shown to predict cancer response to these types of drugs.

For gene expression research related to NSCLC, the aim has been to minimize the number of predictive and/or prognostic markers needed. Between 2002 and 2003, many of these studies were published, resulting in the identification of a 3 gene set (Figure 3) (Beer et al., 2002, Tomida et al., 2004, Lu et al., 2006, Bianchi et al., 2007, Chen et al., 2007, Lau et al., 2007, Zheng et al., 2007, Raponi et al., 2006). However, accumulating evidence has indicated that a single marker, or even a few markers, do not adequately describe the full extent of a cancer's biology. Furthermore, gene expression data is not sufficient to predict a prognosis due to the inconsistency of gene expression patterns in cancers from various patients (D'Amico, 2008).

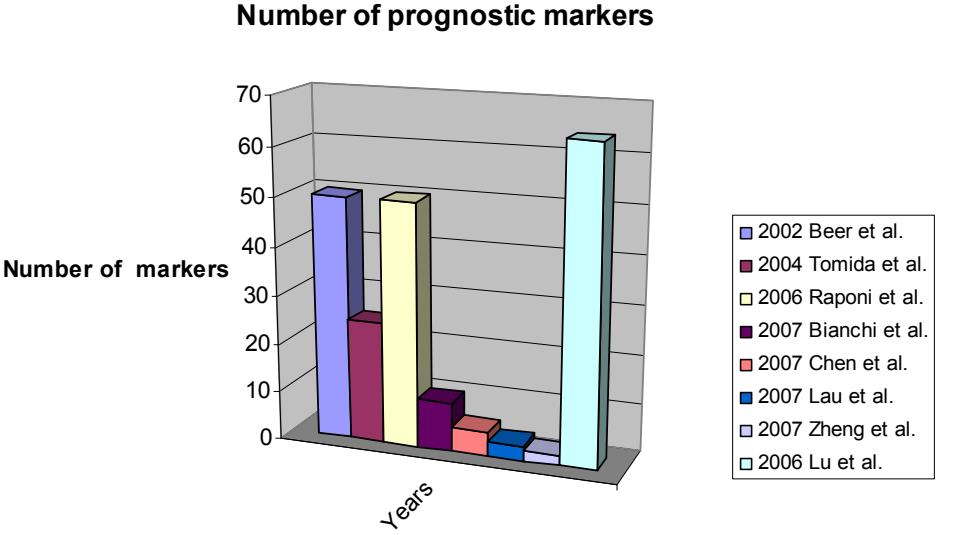

Figure 3. The number of NSCLC prognostic markers identified by studies performed between 2002 and 2007.

Molecular differences in gender, and in smokers versus non-smokers

The biology and clinical course of NSCLC has been shown to be gender-dependent (Mostertz et al., Cerfolio et al., 2006, Pauk et al., 2005). In addidtion, epidemiological evidence supports the observation that the progression of LC

patient survival and patient response to therapy differs between genders. For example, prognoses, response to all methods of conventional therapy for all stages, and histological forms, are more favorable in women (de Perrot et al., 2000, Cerfolio et al., 2006, Fu et al., 2005).

Women are also three times more likely to carry a mutated *KRAS* than men, and frequently, this mutation is associated with cases of lung AC in women (Thomas et al., 2005). However, although this mutation is relatively widespread, the mechanistic details that affect prognosis or prediction of treatment outcome remain unclear.

Gene expression analyses of NSCLC have only detected 7 genes, which exhibit significant differences in expression between men and women. All of these genes localized on the sex chromosomes. These include *JARID1D, RPS4Y1, DDX3Y, EIF1AY, USP9Y* and *UTY*, which are located in chromosome Y, and *XIST* in chromosome X. However, none of these genes have been associated with direct role in NSCLC(Planchard et al., 2009).

Women have been found to have a lower DNA repair capacity (DRC) than men which is molecularly detectable and contributes to tobacco carcinogen-induced carcinogenesis (Wei et al., 2000). In addition, mutated *KRAS* has been found in women with resected NSCLC more often than in men, if the data are normalized for exposure and histology (Nelson et al., 1999)

It is hypothesized that hormonal differences play a role as well. While these influences have been more extensively studied relation to exposure to smoking and hormone replacement therapy (HRT), the influence of HRT remains controversial (Ganti et al., 2006, Schabath et al., 2004). For example, early onset of menopause has been associated with a decreased risk of AC (Taioli and Wynder, 1994). In contrast, the administration of estrogen replacement therapy to patients who smoke, has been significantly associated with AC (Baik et al.).

Interestingly contributes the same feature to better common treatment outcomes in women in all clinical stages.

Usually, a smoking habit is the most common cause for the onset of LC in both women and men. However, the percentage of LC cases associated with non-smoking patients, patients that have never smoked, and passively smoking patients, is growing. In fact, LC that develops in non-smokers remains a unique and poorly understood condition. There are several studies that have shown that LC in smokers versus non-smokers is associated with significant molecular differences, or even represents different diseases (Miura et al., 2002, Powell et al., 2003, Sun et al., 2007). For cases of NSCLC in non-smokers versus smokers, these have primarily been investigated in relation to the mutational status of *KRAS, EGFR*, and *TP-53*, as a result of differences in clinical responses to tyrosine kinase inhibitors observed. Moreover, although significant expression patterns for LC associated with smokers versus non-smokers have not been identified in all studies (Powell et al., 2003) (Powell et al., 2003), higher levels of *MKI67* and EGFR are relatively common in patients with LC that are exposed

to smoke. In contrast, higher levels of *AKT1* and *CDKN1B* have been associated with LC cases involving non-smokers (Dutu et al., 2005).

Gene expression and cases of second primary NSCLC

Approximately 27% of radically treated stage I NSCLC patients experience recurrence. Similarly, approximately one-third of stage I cases involve a second primary cancer, ~30% of which are LCs (Martini et al., 1995). Currently, primary LC and metastasis are distinguished based on certain clinical and histological criteria, although the accuracy of these methods has been inconsistent in many cases.

In recent studies, the potential for gene expression data to classify tumour cell lines and tumours has been demonstrated. For example, Giordano and coauthors were able to distinguish primary ACs from different sources by profiling the expression of more than 6800 gene transcripts (Giordano et al., 2001).

In addition, they were able to identify two metastases that had previously been considered primary cancers. In the same study, principal component analysis (PCA) of the top genes was used as a means of visualizing the data obtained. Of particular interest has been the identification of the highly expressed, primary lung AC-specific marker, TTF1, which has been evaluated both individually and as part of a marker set (Giordano et al., 2001, Reis-Filho et al., 2000, Yatabe et al., 2002, Lazzaro et al., 1991).

Squamous cell cancers of the lung, or head and neck region, have been distinguished using the Affimetrix U133A chip containing 14,500 transcripts, in combination with hierarchical clustering. Based on these results, a ten-gene expression model was created (Vachani et al., 2007). Although there has been some success in identifying primary cancers using this gene set, there is no reliable or commercially available dataset for that. Furthermore, it has been particularly challenging to distinguish cancers of the same anatomical region and histological pattern, e.g., in the case of solitary pulmonary nodules.

The possibility of predicting survival using RNA microchips

Single oncogenic markers cannot be used to predict patient prognosis due to the frequency of aberrant expression, or inconsistent expression, of any one marker in most tumors (D'Amico, 2008).

Therefore, different sets of prognostic markers have been investigated. For example, Beer et al. (2002) and Bhattacharjee et al. (2001) identified correlations between molecular signatures of lung AC and patient prognosis. In other studies, patients have been stratified according to a good versus poor outcome, and the genes associated with each corresponding patient group were considered a training set of markers. As a result, microarray studies using training

sets of genes have revealed prognostic signatures for early stages of NSCLC (Tomida et al., 2004, Raponi et al., 2006, Lu et al., 2006, Guo et al., 2008, Larsen et al., 2007b, Larsen et al., 2007a, Shedden et al., 2008, Boutros et al., 2009). However, due to complications associated with the reproducibility of prognostic gene expression signatures, none of these training sets are still used. Differences between microchips (e.g., variances in the size of the transcripts on them), and differences in the mathematical models used for gene selection, are additional considerations for the lack of reproducibility associated with prognostic gene signatures.

In a study conducted to evaluate prognosis, or the predicted effectiveness of different drugs, gene expression profiling was used to create a metagene model (Potti et al., 2006).

A metagene model represents the dominant average pattern of expression of gene clusters obtained across tumour samples. The metagene model introduced by Potti and co-workers was designed to predict cancer recurrence and was applied to three additional sample sets. As a result, the model was associated with an accuracy rate of 72–80%, which is a higher accuracy rate than that of models which use clinical data alone (including patient age, patient gender, tumor diameter, stage of disease, histologic subtype, and smoking history) (Beane et al., 2009). Potti et al. are continuing to evaluate this metagene and its ability to predict treatment response with cases of stage Ib high-risk patients.

Unlike the analysis of cases of breast cancer, prognostic microchip or marker sets of NSCLC applicable for clinical work have not achieved widespread use. This is primarily due to the relatively high price associated with these analyses, the large number of markers required, and continuing challenges associated with data reproducibility.

Expression of well-known predictor markers of cancer

Well-known molecular markers that are used in the profiling of NSCLC cases include both *EGFR* and *KRAS*. These two genes are proto-oncogenes that are commonly mutated in lung ACs. While the role of *EGFR* expression remains unclear, the mutational status of *EGFR* is key. Mutations related to sensitivity to tyrosine kinase inhibitors have been shown to include deletion of exon 19 and L858R (Rosell et al., 2009), while Mitsudomi and coauthors also reported a role for L858R, which was confirmed with deletions near codons 746–750. (Mitsudomi et al., 2005). The mutation status of *EGFR* has also been associated with sensitivity to small molecule tyrosine kinase inhibitors such as gefitinib and erlotinib (Paez et al., 2004, Lynch et al., 2004, Mitsudomi et al., 2005). Moreover, *EGFR* mutations are significantly associated with the histology of AC, smoking status (particularly non-smokers and light smokers), female gender, and East Asian ethnic groups (Shigematsu et al., 2005).

KRAS is a Kirsten ras oncogene homolog from the mammalian ras gene family that encodes a small GTPase. Ras proteins are pivotal regulators of cellular proliferation, differentiation, motility, and apoptosis. A single amino acid substitution has been shown to be responsible for generating an activating mutation, and mutations in *KRAS* have been identified in 20–30% of NSCLC cases. As a result, mutations in *KRAS* have been proposed to represent possible prognostic and predictive markers. Correspondingly, protein products that result from *KRAS* mutations have been implicated in various cancers, including lung AC. Based on extensive investigations of *KRAS* mutations that have been performed, these mutations are considered to be a weak predictor of prognosis (Schiller et al., 2001). Currently, the mutational status of *KRAS* is also used to predict resistance to cetuximab in treatments of metastatic colon cancer. For NSCLC, it is being investigated as a potential marker for prognosis, and more importantly, as a predictor of tyrosine kinase inhibitor therapy effectiveness in combination with conventional chemotherapy. *KRAS* mutations are also more often associated with lung ACs, and are present in cases of SCC of smokers (Tam et al., 2006). Although this marker has been widely investigated, its role as a predictor or prognostic marker remains controversial (Mao et al., 2009). According to current knowledge and data from meta-analyses, mutated *KRAS* appears to be a marker of poor prognosis in cases of NSCLC, especially in cases of primary lung AC (Mascaux et al., 2005). Moreover, the combination of gene profiles obtained for *EGFR* and *KRAS* has revealed three reproducible subgroups to be associated with different prognoses (Takeuchi et al., 2006). Therefore, these markers can provide supporting information to gene expression data.

The nucleotide excision repair (NER) pathway is critical for the repair of DNA damage. Correspondingly, components of this pathway have been assessed in LC, as well as other cancers, thereby providing potential prognostic and predictive biomarkers.

ERCC1 is a rate-limiting protein in the NER pathway, and in interstrand cross-link repair (ICL-R) pathways. In the latter case, ERCC1 recognizes and removes platinium adducts by repairing interstrand cross-links (Vilmar and Sųrensen, 2009). Initially, investigations of ERCC1 were associated with controversial results. For example, in 2005, ERCC1 expression was identified as a potential prognostic marker (Simon et al., 2005), with overexpression of ERCC1 associated with a significantly longer survival for patients. These results were consistent with the results of the International Adjuvant Lung Cancer Biology Trial (IALT-Bio) that had been presented the year before (Arriagada et al., 2004). Moreover, overexpression of ERCC1 in combination with BRCA, MZF1, and RRM1 has been shown to confer a higher risk of relapse, and accordingly, a shorter survival time and poorer response to cisplatin therapy Although these findings were not significant, studies are ongoing with a larger cohort to confirm the findings. However, Lord et al. (2002) observed that the overall survival of patients with down-regulated ERCC1, but not overexpressed ERCC1, was significantly prolonged following the administration of

platinium-based chemotherapy. Moreover, this qualitative interaction was highly significant (Lord et al., 2002).

In combination, *RRM1* and *ERCC1* are also considered to be promising markers for the prediction of cisplatin-based treatments and prognoses (Gazdar, 2007, Toffalorio et al., 2009).

RRM1 is the regulatory component of ribonucleotide reductase, and catalyses the formation of desoxyribonucleotides from ribonucleotides participating in DNA repair. It also mediates the suppression of cell migration and tumor metastasis by inducing PTEN, a prominent tumor-suppressor gene responsible for the attenuation of growth factor pathway signalling. Overexpression of *RRM1* has been found to be a good prognostic factor in cases of stage I NSCLC (Zheng et al., 2007), while RRM1 is a target of the drug, gemcitabine (Gemzar) (Cerqueira et al., 2007). However, while expression of *RRM1* correlates with *ERCC*, it does not correlate with *PTEN*. Overexpression of *RRM1* was a good prognostic factor in case of stage I NSCLC.

BRCA1 is another factor involved in transcription-coupled nucleotide repair, and has a role in NSCLC. For example, increased levels of *BRCA1* mRNA have been shown to strongly correlate with poor patient survival (Rosell et al., 2007). Moreover, *BRCA1* has been found to be a marker of chemoresistance in various cell lines, as well as for locally advanced cases of NSCLC evaluated in randomized trials (Husain et al., 1998, Rosell et al., 2006).

V. AIMS OF THE PRESENT STUDY

The aims of the present study included: 1) to molecularly characterize NSCLC using gene expression profiling; and 2) to identify a set of gene markers that provide a differentiation of cancerous versus cancer-free tissue, and a discrimination of different NSCLC subtypes, with the goal of identifying a minimal gene set for that purpose.

Specifically, the objectives included:

- The creation of an Illumina-based NSCLC gene expression database for further studies.
- An evaluation of the gene expression profiles obtained from surgically resected samples of NSCLC patients treated at the Tartu University Hospital.
- The identification of differentially expressed, novel genes between NSCLC and controls, as well as within various cancer subtypes.
- The identification of a gene set that predicts patient survival.

VI. MATERIALS AND METHODS

Cohort description

A total of 146 patients with LC, or a peripheral malignant lesion, underwent surgery at the Centre of Thoracic Surgery of Tartu University Hospital between November 28, 2002 and April 1, 2006. Of these cases, 131 involved NSCLC. Histological diagnoses of all samples were confirmed by two pathologists. The cohort included 102 men and 29 women who ranged in age from 36 to 84 years (median, 66), with an average age of 64 and 72 years, respectively. Patients with non-malignant lesions, tumors with non-NSCLC histology, degraded RNA, and patients who had received preoperative chemoradiotherapy, were excluded.

Smokers were defined as: persons who had smoked more than 365 cigarettes in their lifetime, persons who currently smoke, or persons who had stopped smoking less than 12 months ago. Non-smokers were defined as persons who had smoked less than 365 cigarettes in their lifetime, while former smokers were persons who had smoked more than 365 cigarettes in their lifetime and had stopped more than 12 months ago.

The heatmap generated was based on gene expression data obtained from 81 patients (72 men, 9 women), who ranged in age from 36 to 84 years (median, 66), with an average age of 64 and 72 years, respectively.

The follow-up period for this cohort has been maintained since 28.11.2002.

Table 1. Patient characteristics associated with the heatmap generated.

		No. Patients (N = 81)	Percentage of samples
Histology			
	Adenocarcinoma	8	9.90%
	Bronchioloalveolar carcinoma	13	16.00%
	Squamous cell carcinoma	60	74.10%
	Lymph node		
	Positive	13	16.00%
	Negative	68	84.00%
Differentiation			
	Well/moderate	76	94.00%
	Poor/undifferentiated	5	6.00%
	Stage		
	Ia	13	16.10%
	Ib	46	56.80%
	IIa	1	1.20%
	IIb	3	3.70%
	IIIa	7	8.60%
	IIIb	6	7.40%
	IV	5	6.20%
	T1	15	18.50%
	T2	56	69.10%
	T3	5	6.20%
	T4	5	6.20%
Tumor size (mm)			
	< 30	36	44.40%
	> 30	45	55.60%
Surgical procedure			
	Wedge resection	6	7.40%
	Lobectomy	54	66.70%
	Bilobectomy	3	3.70%
	Pneumonectomy	18	22.20%
Gender			
	Female	9	11.10%
	Male	72	88.90%
Age, years			
	Range	38–81	
	Mean	65.8	
	Median	68	
	< 39	1	1.20%
	40–49	5	6.20%
	50–59	13	16.00%
	60–69	27	33.30%
	> 70	35	43.20%
Smoking status			
	Non-smoker	2	2.50%
	Smoker (PY)	79	97.50%
	Family history of cancer	9	11.00%
Occupational exposure			
	None	72	88.90%
	Possible	9	11.10%

Control cohort description

Twenty microscopically-determined, cancer-free lung specimens were used as control samples. They were obtained from patients suffering from either LC, tuberculoma, or a non-malignant pulmonary lesion. All samples were surgically removed and immediately snap-frozen. Characteristics of the control patients are listed in Table 2.

Table 2. Control cohort characteristics

Female/male	3/17
Age, years	
Range	53–84
Mean	68.5
Median	68
Smokers	18
Current	18
Ex	0
Pack years (mean)	51.5

Laboratory methods
RNA isolation and amplification

Tissue specimens were immediately cut into smaller pieces (e.g., 1 cm^3) before being submerged in liquid nitrogen and stored at –80°C. Typically, 50 mg tissue was used for the extraction and purification of each RNA sample using a Ribopure Kit (Ambion, Cat. #:AM1924) according to the manufacturer's protocol. If necessary, tissue samples were cut into smaller pieces in order to be treated with RNAlater®-ICE (Ambion, Cat#AM7030) prior to RNA extraction. For tissue disruption, an IKA Ultra-Turrax T8 homogenizer was used. RNA concentrations and quality were assessed using a NanoDrop-1000 spectrophotometer and Agilent Bioanalyzer Lab-on-a-Chip technology (Agilent RNA 6000 Nano Kit, cat# 5067–1511), respectively. An RNA Integrity Number (RIN) cut-off value of 7 was applied. An Illumina® TotalPrep RNA Amplification Kit (Ambion, Cat. #: AMIL1971) was used for RNA amplification and labeling. An Illumina Bead Array Scanner was also used.

Gene Expression Microarray

An Illumina (www.illumina.com) BeadChip platform and a HumanWG-6_V2 Expression Bead Chip containing 48,000 transcript probes were used for microarray gene expression experiments. According to the manufacturer's protocol, 1.5 μg of amplified cRNA was hybridized to each array. Slides were immediately scanned using a BeadArray reader (Illumina) following hybridi-

zation. Internal controls provided by Illumina and BeadStudio software were used to evaluate data consistency and quality control of the hybridization.

Microarray validation

To validate gene expression levels detected by microarray analysis, quantitative real-time PCR (qRT-PCR) was performed for the top four genes in pattern I.

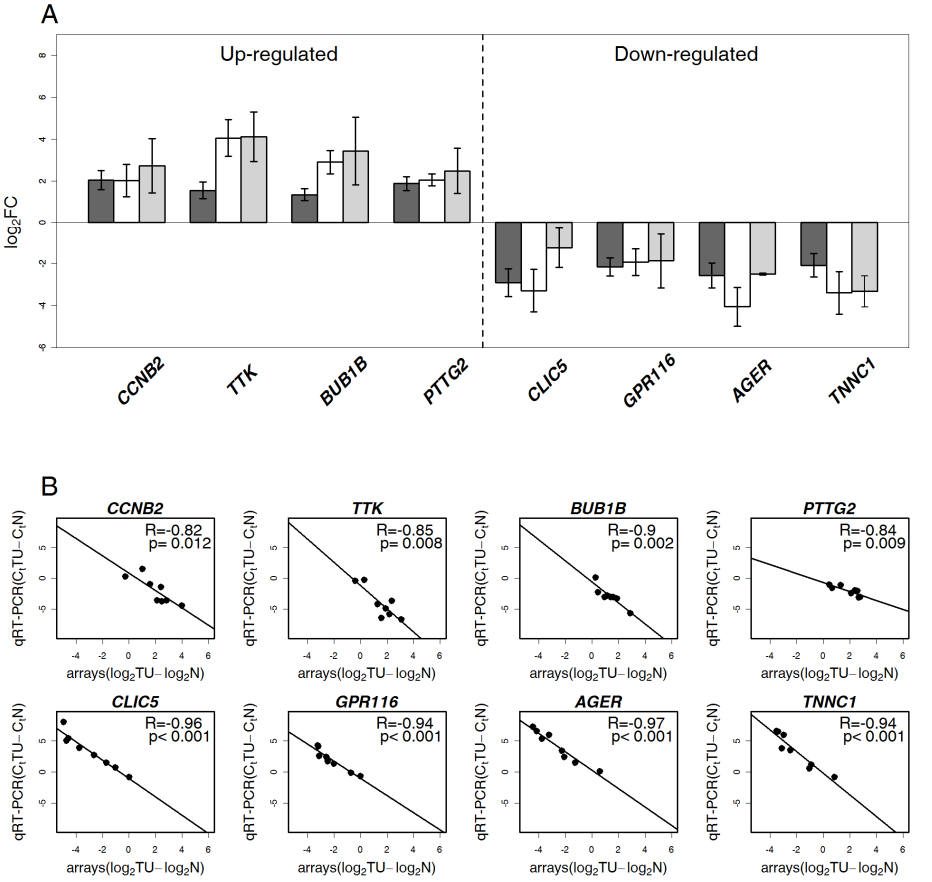

Figure 4. Validation of microarray data. A. Using qRT-PCR assays, the top four up-regulated and down-regulated genes identified from microarray data were analysed. Black bars represent the average log fold-change for paired LC samples (n = 21), white bars represent qRT-PCR averaged log fold-change values for the LC sample pairs (n = 8), and the grey bars represent qRT-PCR averaged log fold-change values for the LC sample pairs that were not presented on the microarray (n = 3). Error bars indicate the standard error of the mean (SEM). B. Correlation between array log2(signaltumor) - log2(signalnormal) and qRT-PCR ΔΔCt for validated genes using the same sample pairs as previous graph (n = 8). Pearson correlation coefficients (R), correlation test p-values, and best-fitting (least squares) lines are shown.

These included: *TTK* (Dual specificity protein kinase), *CCNB2* (Cyclin B2), *BUB1B* (budding uninhibited by benzimidazoles 1 homolog beta (yeast)), and *PTTG2* (pituitary tumor-transforming 2). Four genes from pattern II were also analysed: *CLIC5* (Chloride intracellular channel protein 5), *GPR116* (G-protein coupled receptor 116), *AGER* (advanced glycosylation end product-specific receptor), and *TNNC1* (troponin C type 1). *ESD* (esterase D/formylglutathione hydrolase) was selected as the endogenous reference for the qRT-PCR assays since it was previously identified as an invariant molecule in clinical LC specimens (Kuner et al., 2009, Saviozzi et al., 2006). Transcripts obtained were amplified using a Maxima SYBR Green /ROX qPCR Master mix (Fermentas) and sequenced (ABI Prism 7900HT, Applied Biosystems). Eight sample pairs (e.g., each tumor sample with corresponding normal lung sample) were present in each array, along with three sample pairs which were not, were used in the qRT-PCR experiment. Gene expression levels were calculated using the relative quantification method (Applied Biosystems).

Statistical methods

Differential gene expression analysis was performed using t-tests, with empirical Bayes correction applied, available in the Bioconductor Limma package (Smyth, 2004). Bonferroni correction for used for multiple testing, and a significance level of $\alpha = 0.05$ was used in all comparisons. Gene Ontology (GO) enrichments were calculated using the g:Profiler web toolkit (Reimand et al., 2007) and Genecodis 2.0 (Nogales-Cadenas et al., 2009, Carmona-Saez et al., 2007). Statistically significant, differentially expressed genes were clustered hierarchically using correlation distance, and were visualized using a heatmap. In addition to the statistical parameters described previously, a minimum two-fold change in expression was used to reduce the number of differentially expressed NSCLC-specific genes identified for further analysis.

To visualize the gene expression data of metastasis versus primary cancer and controls, additional correlation heatmaps and principal component analyses were performed.

The dimensionality of the data were reduced by clustering genes with similar profiles using complete linkage hierarchical clustering with correlation distance. Correspondingly, data were divided into 500 groups and the average profile of each cluster formed a metagene in the stage Ib patient group. Analyses performed used the sparse Bayesian probit model for binary response data (Hoti and Sillanpaa, 2006, Albert and Chib, 1993).

VII. RESULTS AND DISCUSSION

mRNA expression associated with NSCLC compared with cancer-free lung tissue

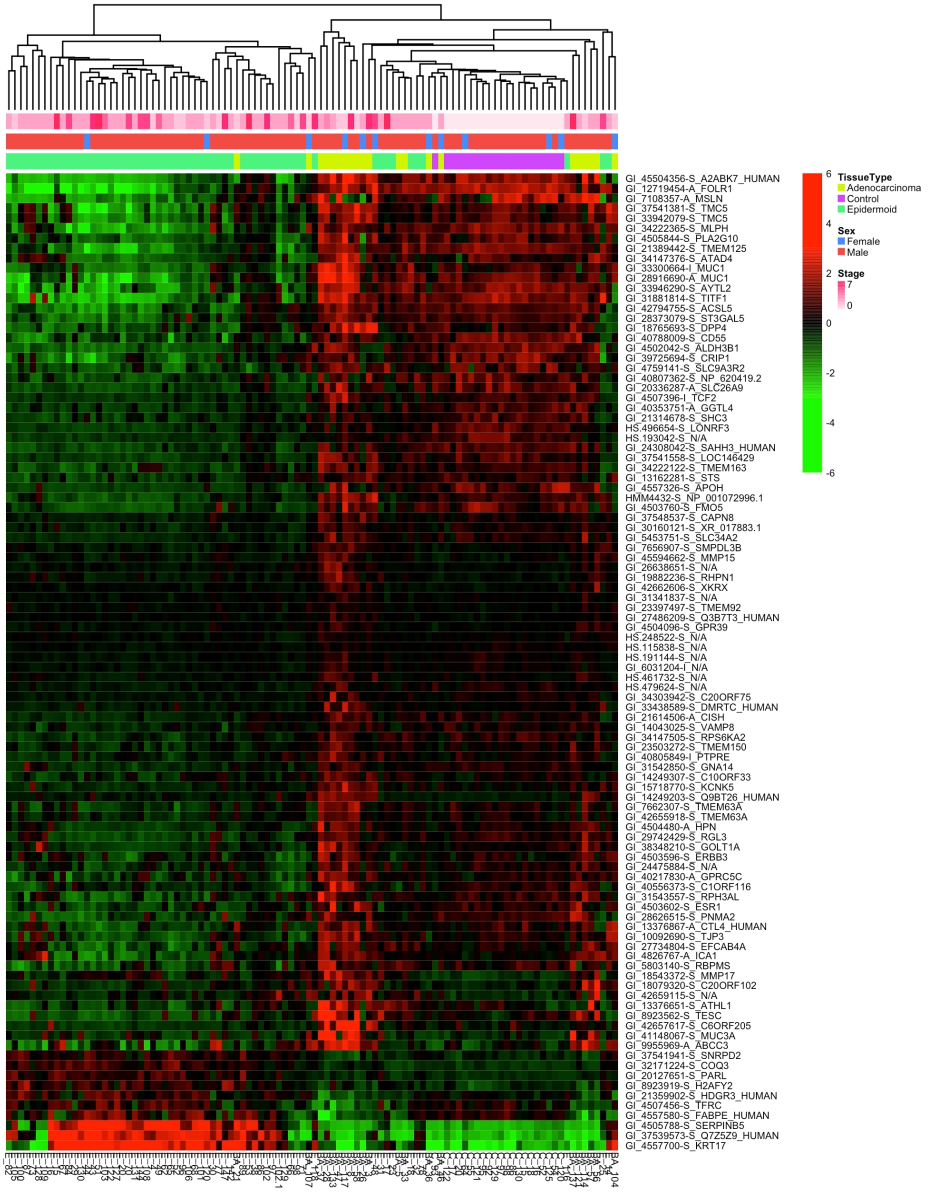

Figure 5. A heatmap of genes exhibiting a greater than 2-fold change and a p-value $< 10^{-6}$ for cancer-free versus tumor samples. Additional color coding is provided at the top of the heatmap which represents labeling of data according to tumor stage, patient gender, and tissue type.

Gene expression analysis of 48,000 transcripts using a fold-change of at least 2 and a p-value > 10^{-6} revealed 1103 genes that were up-regulated in NSCLC tissue versus cancer-free tissue, and 672 genes that were down-regulated. Moreover, unsupervised gene expression analysis clearly distinguished cancer tissue from cancer-free lung tissue as shown in the heatmap of Figure 5, as well as in the PCA shown in Figure 9.

GO analysis of the up-regulated genes identified processes related to cell cycle, cell division, mitosis, DNA replication, DNA repair, spindle organization, etc. (Supp. Table 1). In contrast, the variety and number of down-regulated processes identified in the GO analysis was much larger, and included the identification of various signal transduction pathways, the inflammatory response, cell-cell and cell-matrix adhesion, cell differentiation, immune response, etc. (Supp. Table 2).

An analysis of the differentially expressed genes revealed a large number of previously described NSCLC-associated genes, as well as several potentially novel biomarkers. The group of novel up-regulated genes associated with NSCLC included: *SPAG5, POLQ, KIF23, RAD54L, RAB26* and *ARHGEF19*, as well as four previously uncharacterized open reading frames. The group of novel down-regulated genes associated with NSCLC included: *SGCG, NLRC4, VAPA, SFTPA1B, MMRN1, SFTPD, SELPLG* and *PCDH17*.

Table 3. Gene ontology analysis of statistically significant up- and down-regulated genes between control and NSCLC samples.

Up-regulated genes in cancer			**Downregulated genes in cancer**		
P-value	**GO term**	**Biological process**	**P-value**	**GO term**	**Biological process**
1.05E-29	GO:0022403	cell cycle phase	1.79E-13	GO:0009605	response to external stimulus
1.81E-29	GO:0000279	M phase	3.37E-13	GO:0032502	developmental process
1.11E-26	GO:0000278	mitotic cell cycle	2.11E-12	GO:0009611	response to wounding
1.43E-25	GO:0051301	cell division	9.36E-12	GO:0048856	anatomical structure development
1.04E-15	GO:0007051	spindle organization and biogenesis	2.25E-10	GO:0022610	biological adhesion
2.61E-13	GO:0006996	organelle organization and biogenesis	2.25E-10	GO:0007155	cell adhesion
8.18E-12	GO:0010564	regulation of cell cycle process	2.84E-10	GO:0048731	system development
2.03E-10	GO:0051276	chromosome organization and biogenesis	3.22E-10	GO:0007275	multicellular organismal development
7.01E-10	GO:0006259	DNA metabolic process	4.07E-10	GO:0032501	multicellular organismal process
3.77E-09	GO:0000226	microtubule cytoskeleton organization and biogenesis	6.13E-10	GO:0048513	organ development
3.82E-09	GO:0007059	chromosome segregation	2.41E-08	GO:0065008	regulation of biological quality
4.09E-09	GO:0007017	Microtubule-based process	2.92E-08	GO:0035295	tube development
4.43E-09	GO:0006260	DNA replication	4.93E-08	GO:0009653	anatomical structure morphogenesis
1.76E-08	GO:0000070	mitotic sister chromatid segregation	5.99E-08	GO:0048869	cellular developmental process
2.50E-08	GO:0016043	cellular component organization and biogenesis	5.99E-08	GO:0030154	cell differentiation
3.83E-07	GO:0006323	DNA packaging	3.11E-07	GO:0006954	inflammatory response
5.29E-07	GO:0030261	chromosome condensation	7.47E-07	GO:0006952	defense response

Table 4. Potentially new biomarkers for NSCLC

Adjusted p-value	Average fold change	Gene symbol	Gene name and source
Up-regulated in cancer tissues			
5,41E-16	2,2	C6ORF129	chromosome 6 open reading frame 129, [source:uniprot/sptrembl;acc:q5t7f7]
9,50E-14	3,6	SPAG5	sperm-associated antigen 5, map126, deepest, [source:uniprot/swissprot;acc:q96r06]
1,11E-11	3,3	POLQ	dna polymerase theta, [source:uniprot/swissprot;acc:o75417]
5,42E-11	2,1	C6ORF125	uncharacterized protein c6orf125, [source:uniprot/swissprot;acc:q9brt2]
3,13E-10	2,9	KIF23	kinesin-like protein kif23, [source:uniprot/swissprot;acc:q02241]
1,01E-09	2,2	RAD54L	dna repair and recombination protein rad54-like, [source:uniprot/swissprot;acc:q92698]
1,24E-09	2,2	C12ORF48	upf0419 protein c12orf48. [source:uniprot/swissprot;acc:q9nws1]
2,04E-09	2,0	C16ORF33	u11/u12 snrnp 25 kda protein (minus-99 protein). [source:uniprot/swissprot;acc:q9bv90]
4,38E-09	2,0	RAB26	ras-related protein rab-26. [source:uniprot/swissprot;acc:q9ulw5]
4,94E-09	2,0	ARHGEF19	rho guanine nucleotide exchange factor 19. [source:uniprot/swissprot;acc:q8iw93]
Down-regulated in cancer tissues			
3,14E-20	2,4	SGCG	gamma-sarcoglycan, [source:uniprot/swissprot;acc:q13326]
2,22E-17	3,5	NLRC4	caspase recruitment domain-containing protein 12, [source:uniprot/swissprot;acc:q9npp4]
1,75E-16	2,1	VAPA	vesicle-associated membrane protein-associated protein, [source:uniprot/swissprot;acc:q9p0l0]
3,11E-15	9,5	SFTPA1B	pulmonary surfactant-associated protein a1 precursor, [source:uniprot/swissprot;acc:q8iwl2]
2,65E-12	2,0	MMRN1	multimerin-1 precursor (endothelial cell multimerin 1),[source:uniprot/swissprot;acc:q13201]
3,52E-08	10,9	SFTPD	pulmonary surfactant-associated protein d precursor, [source:uniprot/swissprot;acc:p35247]
6,44E-07	2,0	SELPLG	p-selectin glycoprotein ligand 1 precursor, [source:uniprot/swissprot;acc:q14242]
7,99E-07	2,1	PCDH17	protocadherin-17 precursor (protocadherin-68), [source:uniprot/swissprot;acc:o14917]

In one case, primary cancer, local recurrence, and a second primary tumor were able to be distinguished using PCA of gene expression data (Ref. II). Specifically, patient #17 was diagnosed with SCC LC with cavitation in the left lower lobe. During the follow-up period for this patient, a small tumour was also diagnosed and removed from the contralateral lung. Initially, this tumour was identified as a stage IV carcinoma. However, the clustering of gene expression data in the PCA that was performed indicated that this tumour was more likely to be a primary cancer. Subsequent survival data supported the stage I clustering obtained, and the presence of a second primary tumour in the contralateral lung, rather than a primary cancer with metastasis.

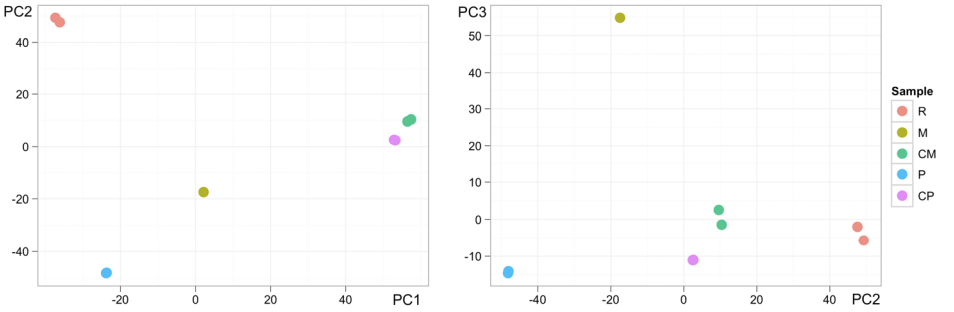

Figure 6. PCA of gene expression data from primary, recurrent, suggested metastasis, control samples of primary, and control sample of suggested metastasis samples. Replicate array data were available for all samples except the suggested metastasis sample.

Gene expression of different NSCLC subtypes

Analyses of gene expression data were able to effectively differentiate the main histological subgroups of NSCLC. For example, AC was able to be distinguished from SCC (Figure 7), although AC was not able to be distinguished from BAC. Furthermore, analysis of AC and BAC gene expression data did not reveal a single marker, or a combination of markers, that were significantly differentially expressed between the two cancer types, despite these cancer types being clustered relatively close to each other. Moreover, according to the current body of knowledge regarding BAC, this cancer type should exhibit a non-invasive phenotype and a possible multifocal appearance, while being associated with a slightly better prognosis and better response to chemotherapy.

Although AC and its subtypes clustered differently on a heatmap, distinguishing markers with acceptable p-values could not be identified. A possible reason for this is the small sample size of both groups that included 13 and 8 patients, respectively. In addition, a high degree of morphological similarity, as well as potential functional similarities, of these cancer types, and the role of host-related factors may also have contributed. However, future studies with a larger sample size would be expected to be able to distinguish BAC, or at least

AC, without invasive features. Similarly, Bhattacharjee and co-workers observed that AC and cancer-free tissue were more closely clustered than SCC (Bhattacharjee et al., 2001).

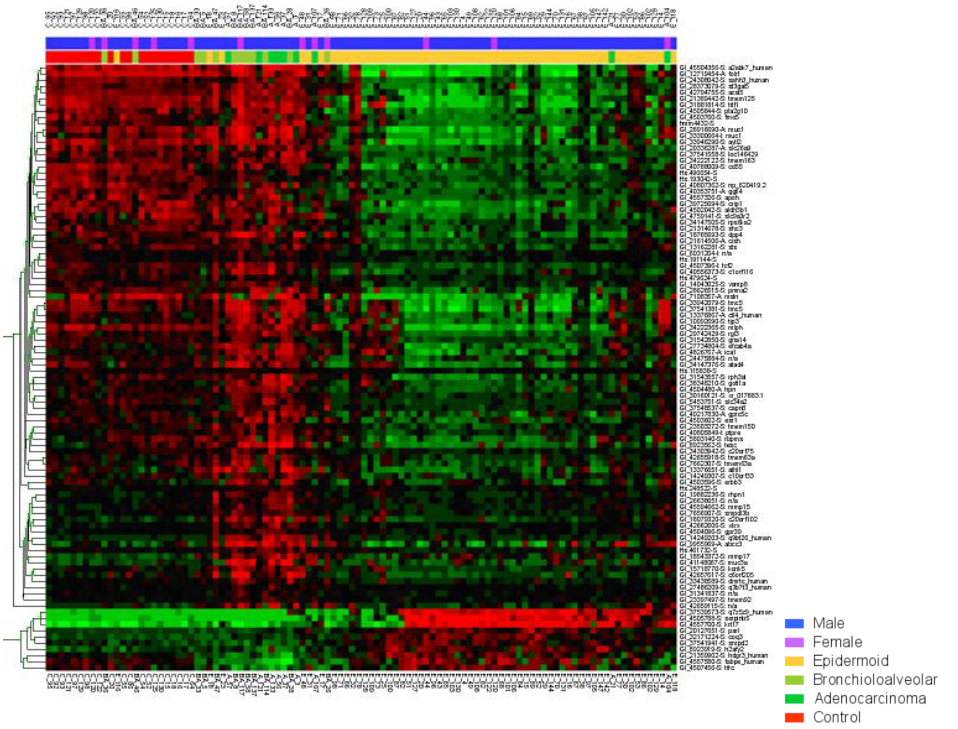

Figure 7. Heatmap of gene expression profiles obtained for different NSCLC subtypes using 1103 up-regulated genes and 672 down-regulated genes.

The up-regulated genes associated with SCC versus AC were primarily found to be related to keratinisation, keratinocytic development, epidermis development, and some metabolic processes. However, genes related to apoptosis, induction of apoptosis, and adhesion were also identified. For SCLC, well-known molecular markers were identified, particularly the cytokeratins, AE1/AE3, CK7, and CK 20. These were associated with large fold-changes in expression and high p-values (Table 5). Correspondingly, cytokeratin expression has previously found to be one of the best ways of differentiating SCLC. In the present research, the cytokeratins, KRT 6C and KRT 17, were found to be expressed in samples of AC (Figure 8). This expression peculiarity would be consistent with the presence of both AC and squamous cell patterns in the same sample, as previously observed with dimorphic and adenosquamous cancers (Bastide et al.).

Table 5. Up-regulated cytokeratins detected in SCC samples.

Gene ID	Fold change	Adj. p-value	Used in marker
KRT6A	15,00	4,14E-05	AE1/AE3
KRT17	11,02	4,62E-06	none
KRT5	10,33	1,47E-04	AE1/AE3
KRT6B	10,27	3,53E-04	AE1/AE3
KRT6C	7,46	1,33E-03	AE1/AE3
KRT13	6,07	2,04E-03	AE1/AE3
KRT16	5,54	1,49E-03	AE1/AE3
KRT14	4,64	6,78E-03	AE1/AE3
KRT15	4,50	5,41E-03	AE1/AE3

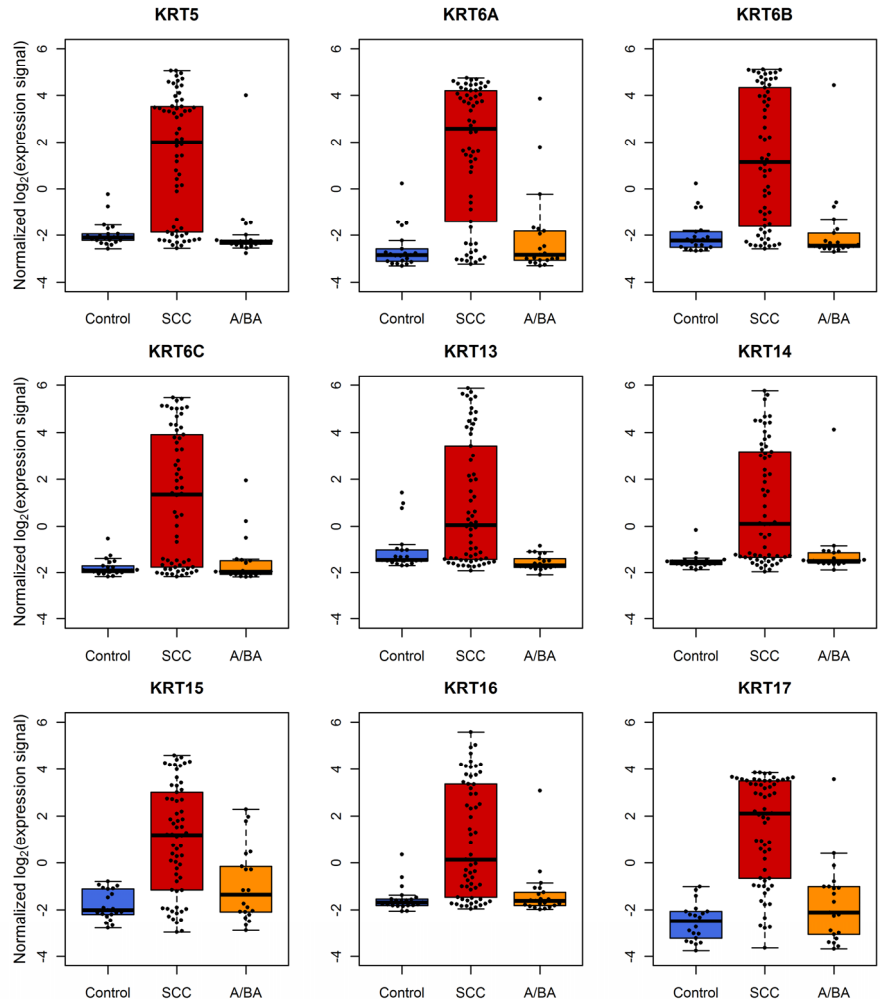

Figure 8. Boxplots show the expression of cytokeratins detected in three different tissue types: cancer-free tissue (blue), SCC (red), AC and BAC (yellow). Each point represents normalized and logarithmic expression signals.

Gene list of differentially expressed markers of SCC and AC.

Genes up-regulated in AC/BAC compared to SCC

Gene ID	FC	Adj.P-value
A2ABK7_HUMAN	6,88	2,17E-08
GAGD2_HUMAN	6,55	8,61E-04
FOLR1	5,61	4,14E-05
SFTA2_HUMAN	5,42	3,75E-03
LMO3	5,11	5,03E-05
PGC	5,09	1,11E-04
MUC1	4,69	2,35E-06
NKX2-1	4,60	4,62E-06
NP_689672,3	4,46	4,14E-05
SCGB3A1	4,38	6,59E-03
SFTPC	4,34	2,67E-02
AYTL2	4,19	1,06E-06
SFTPB	4,16	2,32E-02
CLDN3	3,97	8,77E-04
SPINK1	3,93	4,96E-04
C4BPA	3,93	9,78E-03
CEACAM6	3,90	1,45E-03
MLPH	3,81	1,04E-06
MSLN	3,80	1,36E-04
TESC	3,48	3,46E-06
SFTPD	3,33	3,50E-02
ABCC3	3,29	1,20E-04
ABCA3	3,28	3,35E-03
Q5SQA0_HUMAN	3,28	1,52E-05
TMEM125	3,25	1,02E-05
KRT7	3,24	1,55E-03
ACSL5	3,22	5,39E-07
DPP4	3,17	1,94E-06
TMC5	3,14	6,01E-06
TMPRSS2	3,07	1,08E-04
DMBT1	3,01	1,95E-02
SELENBP1	3,00	5,03E-05
GPRC5A	2,87	1,29E-03
SCGB3A2	2,85	4,03E-02
ATOH8	2,85	4,14E-05
AGR3	2,85	2,58E-03
SERPINA1	2,82	5,52E-03
PLA2G10	2,77	4,93E-06
FBP1	2,76	4,14E-04
ATAD4	2,74	1,41E-05
HOP_HUMAN	2,73	2,02E-03
SUSD2	2,70	2,95E-03
EEF1A2	2,70	6,39E-03
ST3GAL5	2,67	1,06E-06
ATHL1	2,62	1,73E-04
RNASE1	2,62	1,29E-03
TFPI2	2,62	1,02E-03
ALDH3B1	2,62	5,39E-07
CGN	2,60	2,24E-04

Genes down regulated in AC/BAC compared to SCC

Gene ID	FC	Adj. P-value
KRT6A	15,00	4,14E-05
KRT17	11,02	4,62E-06
KRT5	10,33	1,47E-04
KRT6B	10,27	3,53E-04
SERPINB5	9,33	3,77E-06
KRT6C	7,46	1,33E-03
CALML3	7,16	1,37E-03
S100A2	6,89	8,85E-04
KRT13	6,07	2,04E-03
KRT16	5,54	1,49E-03
AKR1B10	5,42	1,47E-02
TP63	5,22	3,54E-04
CLCA2	5,12	3,53E-04
LEG7_HUMAN	5,10	1,11E-03
DAPL1	5,07	9,22E-04
CSTA	4,97	1,08E-04
SERPINB13	4,78	3,67E-04
KRT14	4,64	6,78E-03
KRT15	4,50	5,41E-03
WDR72	4,38	1,08E-04
AKR1C2	4,19	1,90E-02
DSG3	4,16	1,34E-03
MMP10	3,96	3,24E-03
SERPINB3	3,90	9,22E-04
PNCK	3,88	5,60E-04
SPRR3	3,80	7,63E-03
CNTNAP2	3,67	1,93E-04
PI3	3,58	1,44E-02
SPRR1B	3,56	6,12E-03
COL7A1	3,47	6,01E-04
SLC2A1	3,44	1,83E-04
CTSL2	3,23	8,50E-05
NTS	3,19	2,90E-02
ANXA8L1	3,15	9,60E-03
S100A7	3,10	2,95E-02
MAGEA9B	3,08	3,13E-03
ADH7	2,95	2,51E-02
PERP	2,93	4,96E-04
GJB2	2,82	9,77E-03
FST	2,76	5,36E-04
MLLT11	2,68	1,63E-03
SPRR2D	2,68	1,74E-02
DSC3	2,66	3,48E-03
PTPRZ1	2,65	2,32E-03
SPRR2A	2,65	1,02E-02
IGFBP2	2,65	4,27E-03
TRIM29	2,61	1,04E-03
NDUFA4L2	2,57	2,58E-03
MMP1	2,55	3,81E-02

Molecular discrimination of clinical stages

The clinical stages analysed did not differ significantly from each other in the hierarchical clustering or PCA performed, as illustrated in the heatmap of Figure 5, and the PC analysis of stages in Figure 9. These results are consistent with the current strategy for clinical staging which primarily considers tumour spread, rather than tumour growth and the molecular features of cancerous tissue. However, molecular and cellular markers related to clinical stages have been investigated by Ju and coauthors, and limited success in using the expression of regulatory T cell markers as stage markers was reported (Ju et al., 2009). Unfortunately, this approach is currently not reliable, or sufficiently cost effective, to apply in the clinic. Therefore, a molecular description of either clinical or molecular stages of cancers remains a challenge.

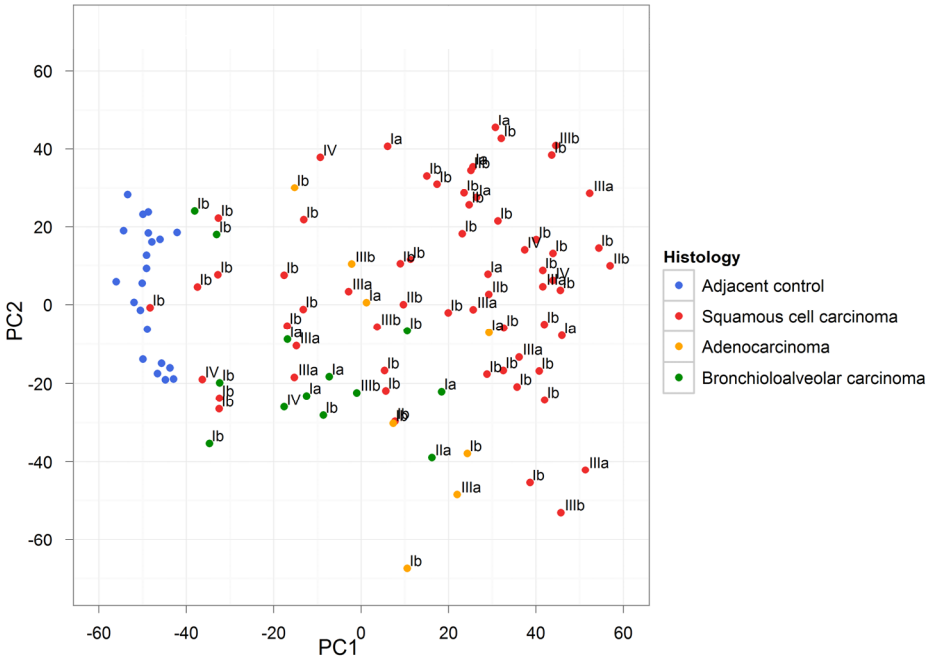

Figure 9. PCA of gene expression profiles obtained for all genes present on the array. The axes represent the first two principle components evaluated, while each point represents a single sample.

Survival and gene expression profiles

Two significantly different gene expression patterns were observed for – and – samples. These findings were associated with a significantly better prognostic value for AC or SCC histology, although they did not correlate with histological subtypes of NSCLC.

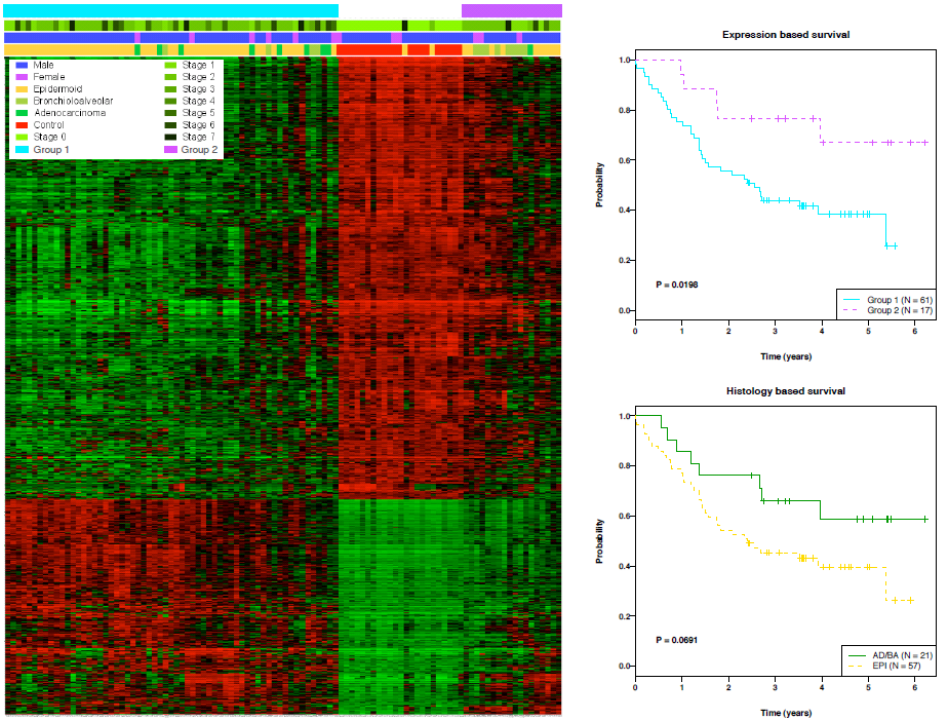

Figure 10. Gene expression profiling and Kaplan-Meier survival curves. The grouping of patients used for NSCLC patient survival curves was based on the gene expression profiles of significantly up- and down-regulated genes (n = 997). These survival curves provided better survival predictions than histological groupings.

Using a sparse Bayesian regression model for stage Ib outcome analysis, four out of 500 metagenes were found to be significantly related to survival (Ref. III), with two of the metagenes being clearly more prognostic. The first metagene contained MAGE A9B and MAGE A 10, which are located in Xq28 and belong to the melanoma antigene family. The third member of this metagene, NLRP2, belongs to the (NLR) family and localizes to 19q13.42. These results are consistent with previous characterizations of melanoma antigene family members in the context of LC. Moreover, these antigenes have been identified as potential vaccine targets (Mellstedt et al.).

The second metagene contained CYP3A5 (cytochrome 450 family 3 subfamily A polypeptide 5), the highly functional gene AKR 1B1 (aldoketo reductase family 1 member B1), and FAM 46B (family with sequence similarity 46 member B), which localize to 7q21.2, 7q35, and 1p36.11, respectively. In an analysis of survival patterns according to metagene expression, a time period of 1000 days was associated with the ability to distinguish a good versus bad prognosis (Figure 11A).

Expression of a single known, or previously unknown, marker was not found to predict survival in the current cohort (Ref. I and III). One possible reason for this is that a short follow-up period of less than 7 years was associated with many of the cases analysed. However, these results do not exclude the possibility that a survival marker will be discovered in future studies.

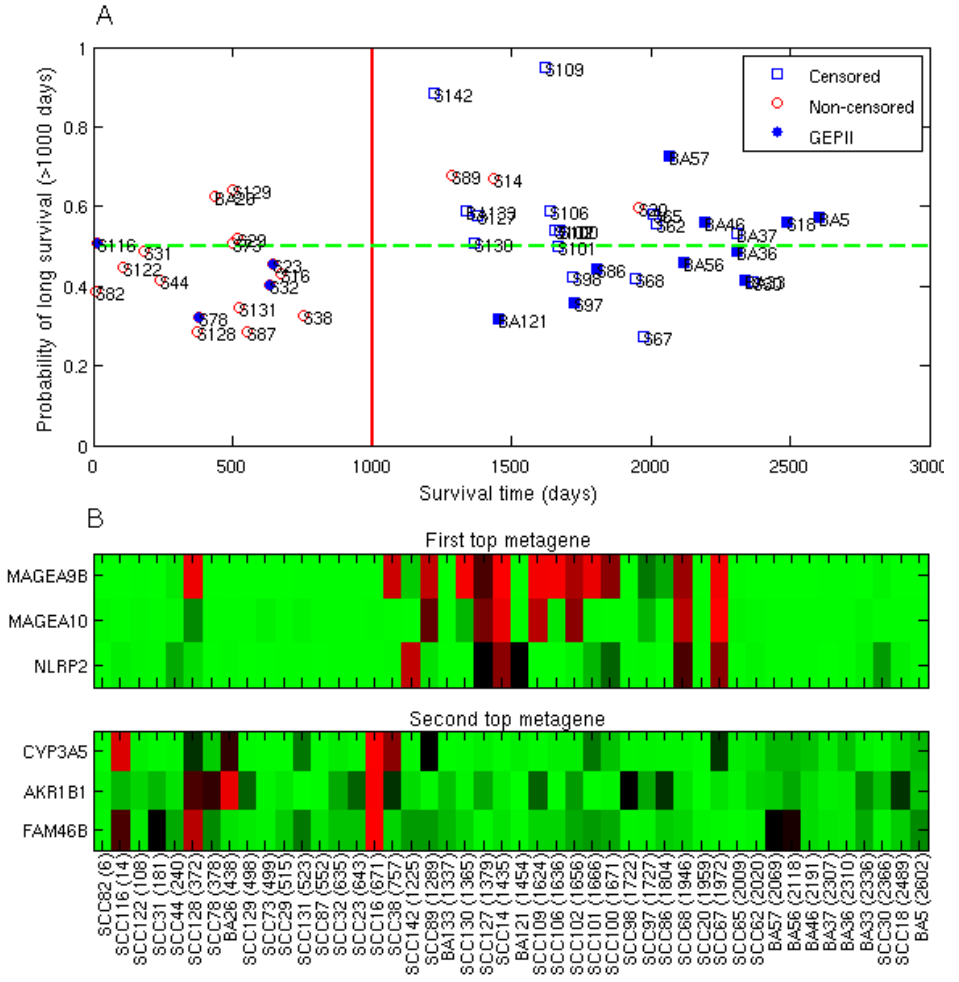

Figure 11. A , B The top two top metagenes related to patient survival.

Additional findings

Two tumors were identified with gene expression patterns that differed from the other tumor samples. One of them represented a metastasis of NSCLC, while the other tumour exhibited a different histological pattern. These results are consistent with those of Bhattacharjee and co-workers which demonstrated that gene expression profiles can serve as diagnostic tools to confirm and identify metastases to the lung, as well as the type of primary tumor present (Bhattacharjee et al., 2001).

In addition, sample #104 was diagnosed as AC histologically. Correspondingly, the gene methylation profile associated with this patient was found to differ from the other AC patient samples analysed. As a result, this case was identified as a potential epidermoid cancer. When this sample underwent additional histological evaluations, the sample was found to exhibit characteristics of both AC and SCLC.

Limitations of this study

1. Tumor tissues were frozen in liquid nitrogen at the time of surgical resection in order to extract the highest quality RNA. It is not known whether our results are generalizable to formalin-fixed tissues and cytology specimens, although similar investigations using an Ambion Recovery Kit and Affimetrix platform have recently been successfully performed (Jacobson et al.).
2. The amount of tissue used in the current investigation was relatively high. In order to introduce gene expression analyses into clinical practice, a protocol using clinically acceptable amounts of tissue would need to be developed.
3. Due to limited differences between AC and BAC, as well as a small sample size, BA and AC could not be effectively discriminated molecularly, which represents the larger challenge facing the field of NSCLC research.
4. This study was limited to patients with phenotypes more commonly associated with NSCLC. In future investigations, less common subtypes of NSCLC would be analyzed as well.
5. Possible reasons for de-regulation of gene expression were not investigated, including genomic rearrangements (deletions, duplications and translocations), changes in methylation profiles, and the influence of microRNA on gene expression, which represent considerations being addressed in ongoing research.

Further Research suggestions

Future studies will need to confirm the applicability of the different prognostic and predictive gene sets identified from the cohort evaluated. Comparison of the gene expression data obtained with methylation profiles, microRNA expression, and genome-wide association data also need to be performed. Continued collection of clinical data from patient follow-ups will also help further refine data regarding patient survival and patient response to different treatment modalities.

VII. SUMMARY AND CONCLUDING REMARKS

In this thesis, 48,000 known transcripts of the human genome were analyzed in NSCLC tissues in comparison with corresponding cancer-free lung tissue. For genes exhibiting at least a 2-fold change in expression, and a p-value of at least 10^{-6}, 1,775 markers were identified that distinguished NSCLC tissue from normal tissue. In addition, 18 potentially novel biomarkers were identified for NSCLC (Välk et al. 2010).

Two metagenes, associated with a prognostic value for stage Ib patients, were able to distinguish the entire cohort into two different survival groups; those surviving > 1000 days, and those surviving < 1000 days (Urgard et al. 2011). The contribution of this thesis to cancer biology includes the potential identification of novel biomarkers for NSCLC, and an improved understanding of the NSCLC microenvironment. The microarray data generated as part of this thesis is also a valuable contribution to the NSCLC field, and also demonstrates the complications associated with interpretation of this type of data.

The results of this thesis also have clinical value in that the design of array-based diagnostic, prognostic, and predictive tests were demonstrated for NSCLC. Integration of the gene expression data obtained with data regarding environmental risk factors, epidemiologic data, and clinical course (including prognostic and predictive data) will further help to elucidate mechanistic details of NSCLC, and these will need to be confirmed with additional studies.

Conclusions

1. Characterization of the NSCLC transcriptome has further defined the molecular processes underlying the tumorigenesis of this disease.
2. Gene expression is a powerful tool for the distinguishing of different cancers and cancer types which may be problematic in clinical practice.
3. Gene expression profiling using microarrays can provide a valuable and rapid screening method for the detection of novel, cancer-related genes, and their underlying pathways, for further research.
4. Analysis of gene expression profile may provide information about the source of a primary cancer.
5. Gene expression profiling may facilitate evaluations of patient prognosis in cases of cancer.
6. Gene expression data are not an endpoint for characterization of a cancer since it does adequately elucidate the details of the gene regulation that is detected, nor the qualitative and quantitative status of protein synthesis.

REFERENCES

Albain, K.S., W.E. Barlow, S. Shak, G.N. Hortobagyi, R.B. Livingston, I.T. Yeh, P. Ravdin, R. Bugarini, F.L. Baehner, N.E. Davidson et al. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. The Lancet Oncology 11: 55–65.

Alberg, A.J., M.V. Brock, and J.M. Samet. 2005. Epidemiology of lung cancer: looking to the future. J Clin Oncol 23: 3175–3185.

Alberg, A.J., J.G. Ford, and J.M. Samet. 2007. Epidemiology of lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition). Chest 132: 29S-55S.

Albert, J.H. and S. Chib. 1993. Bayesian Analysis of Binary and Polychotomous Response Data. Journal of the American Statistical Association 88: 669–679.

Alberts, W.M. 2007. Diagnosis and management of lung cancer executive summary: ACCP evidence-based clinical practice guidelines (2nd Edition). Chest 132: 1S-19S.

Aldington, S., M. Harwood, B. Cox, M. Weatherall, L. Beckert, A. Hansell, A. Pritchard, G. Robinson, R. Beasley, and G. on behalf of the Cannabis and Respiratory Disease Research. 2008. Cannabis use and risk of lung cancer: a caseã€“control study. European Respiratory Journal 31: 280–286.

Alizadeh, A.A., M.B. Eisen, R.E. Davis, C. Ma, I.S. Lossos, A. Rosenwald, J.C. Boldrick, H. Sabet, T. Tran, X. Yu et al. 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403: 503–511.

Arriagada, R., B. Bergman, A. Dunant, T. Le Chevalier, J.P. Pignon, and J. Vansteenkiste. 2004. Cisplatin-based adjuvant chemotherapy in patients with completely resected non-small-cell lung cancer. N Engl J Med 350: 351–360.

Baik, C.S., G.M. Strauss, F.E. Speizer, and D. Feskanich. Reproductive Factors, Hormone Use, and Risk for Lung Cancer in Postmenopausal Women, the Nurses' Health Study. Cancer Epidemiology Biomarkers & Prevention 19: 2525–2533.

Bastide, K., N. Ugolin, C. Levalois, J.F. Bernaudin, and S. Chevillard. Are adenosquamous lung carcinomas a simple mix of adenocarcinomas and squamous cell carcinomas, or more complex at the molecular level? Lung Cancer 68: 1–9.

Beane, J., A. Spira, and M.E. Lenburg. 2009. Clinical impact of high-throughput gene expression studies in lung cancer. J Thorac Oncol 4: 109–118.

Beer, D.G., S.L. Kardia, C.C. Huang, T.J. Giordano, A.M. Levin, D.E. Misek, L. Lin, G. Chen, T.G. Gharib, D.G. Thomas et al. 2002. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med 8: 816–824.

Bhattacharjee, A., W.G. Richards, J. Staunton, C. Li, S. Monti, P. Vasa, C. Ladd, J. Beheshti, R. Bueno, M. Gillette et al. 2001. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci U S A 98: 13790–13795.

Bianchi, F., P. Nuciforo, M. Vecchi, L. Bernard, L. Tizzoni, A. Marchetti, F. Buttitta, L. Felicioni, F. Nicassio, and P.P. Di Fiore. 2007. Survival prediction of stage I lung adenocarcinomas by expression of 10 genes. J Clin Invest.

Bodendorf, M.O., V. Haas, H.G. Laberke, G. Blumenstock, P. Wex, and T. Graeter. 2009. Prognostic value and therapeutic consequences of vascular invasion in non-small cell lung carcinoma. Lung Cancer 64: 71–78.

Boutros, P.C., S.K. Lau, M. Pintilie, N. Liu, F.A. Shepherd, S.D. Der, M.S. Tsao, L.Z. Penn, and I. Jurisica. 2009. Prognostic gene signatures for non-small-cell lung cancer. Proc Natl Acad Sci U S A 106: 2824–2828.

Brahmer, J.R., S.E. Dahlberg, R.J. Gray, J.H. Schiller, M.C. Perry, A. Sandler, and D.H. Johnson. Sex Differences in Outcome with Bevacizumab Therapy: Analysis of Patients with Advanced-Stage Non-small Cell Lung Cancer Treated with or without Bevacizumab in Combination with Paclitaxel and Carboplatin in the Eastern Cooperative Oncology Group Trial 4599. Journal of Thoracic Oncology 6: 103–108 110.1097/JTO.1090b1013e3181fa1098efd.

Brambilla, E., W.D. Travis, T.V. Colby, B. Corrin, and Y. Shimosato. 2001. The new World Health Organization classification of lung tumours. Eur Respir J 18: 1059–1068.

Camilo, R., V.L.z. Capelozzi, S.A. Coelho Siqueira, and F.o. Del Carlo Bernardi. 2006. Expression of p63, keratin 5/6, keratin 7, and surfactant-A in nonã€“small cell lung carcinomas. Human pathology 37: 542–546.

Carmona-Saez, P., M. Chagoyen, F. Tirado, J. Carazo, and A. Pascual-Montano. 2007. GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. Genome Biology 8: R3.

Cerfolio, R.J., A.S. Bryant, E. Scott, M. Sharma, F. Robert, S.A. Spencer, and R.I. Garver. 2006. Women with pathologic stage I, II, and III non-small cell lung cancer have better survival than men. Chest 130: 1796–1802.

Cerqueira, N.M., P.A. Fernandes, and M.J. Ramos. 2007. Understanding ribonucleotide reductase inactivation by gemcitabine. Chemistry 13: 8507–8515.

Chandra, V., M.S. Allen, F.C. Nichols, 3rd, C. Deschamps, S.D. Cassivi, and P.C. Pairolero. 2006. The role of pulmonary resection in small cell lung cancer. Mayo Clin Proc 81: 619–624.

Chapman, A.D. and K.M. Kerr. 2000. The association between atypical adenomatous hyperplasia and primary lung cancer. Br J Cancer 83: 632–636.

Chen, H.Y., S.L. Yu, C.H. Chen, G.C. Chang, C.Y. Chen, A. Yuan, C.L. Cheng, C.H. Wang, H.J. Terng, S.F. Kao et al. 2007. A five-gene signature and clinical outcome in non-small-cell lung cancer. N Engl J Med 356: 11–20.

Cheng, Y.W., H.L. Chiou, G.T. Sheu, L.L. Hsieh, J.T. Chen, C.Y. Chen, J.M. Su, and H. Lee. 2001. The association of human papillomavirus 16/18 infection with lung cancer among nonsmoking Taiwanese women. Cancer Res 61: 2799–2803.

Chu, P., E. Wu, and L.M. Weiss. Cytokeratin 7 and Cytokeratin 20 Expression in Epithelial Neoplasms: A Survey of 435 Cases. Mod Pathol 13: 962–972.

Coate, L.E., T. John, M.S. Tsao, and F.A. Shepherd. 2009. Molecular predictive and prognostic markers in non-small-cell lung cancer. Lancet Oncol 10: 1001–1010.

D'Addario, G. and E. Felip. 2008. Non-small-cell lung cancer: ESMO clinical recommendations for diagnosis, treatment and follow-up. Ann Oncol 19 Suppl 2: ii39–40.

D'Amico, T.A. 2008. Molecular biologic staging of lung cancer. Ann Thorac Surg 85: S737–742.

Dacic, S. 2007. Molecular profiling of lung carcinoma: identifying clinically useful tumor markers for diagnosis and prognosis. Expert Rev Mol Diagn 7: 77–86.

de Perrot, M., M. Licker, C. Bouchardy, M. Usel, J. Robert, and A. Spiliopoulos. 2000. Sex differences in presentation, management, and prognosis of patients with non-small cell lung carcinoma. J Thorac Cardiovasc Surg 119: 21–26.

Devesa, S.S., F. Bray, A.P. Vizcaino, and D.M. Parkin. 2005. International lung cancer trends by histologic type: male:female differences diminishing and adenocarcinoma rates rising. Int J Cancer 117: 294–299.

Dobbin, K.K., D.G. Beer, M. Meyerson, T.J. Yeatman, W.L. Gerald, J.W. Jacobson, B. Conley, K.H. Buetow, M. Heiskanen, R.M. Simon et al. 2005. Interlaboratory comparability study of cancer gene expression analysis using oligonucleotide microarrays. Clin Cancer Res 11: 565–572.

Doll, R. 1950. On the etiology of lung cancer. J Natl Cancer Inst 11: 638–640.

Doll, R. and A.B. Hill. 1950. Smoking and carcinoma of the lung; preliminary report. Br Med J 2: 739–748.

Doll, R. and A.B. Hill. 2004. The mortality of doctors in relation to their smoking habits: a preliminary report. 1954. Bmj 328: 1529–1533; discussion 1533.

Doll, R., R. Peto, J. Boreham, and I. Sutherland. 2004. Mortality in relation to smoking: 50 years' observations on male British doctors. Bmj 328: 1519.

Dutu, T., S. Michiels, P. Fouret, F. Penault-Llorca, P. Validire, S. Benhamou, E. Taranchon, L. Morat, D. Grunenwald, T. Le Chevalier et al. 2005. Differential expression of biomarkers in lung adenocarcinoma: a comparative study between smokers and never-smokers. Annals of Oncology 16: 1906–1914.

Fry, W.A., J.L. Phillips, and H.R. Menck. 1999. Ten-year survey of lung cancer treatment and survival in hospitals in the United States: a national cancer data base report. Cancer 86: 1867–1876.

Fu, J.B., T.Y. Kau, R.K. Severson, and G.P. Kalemkerian. 2005. Lung cancer in women: analysis of the national Surveillance, Epidemiology, and End Results database. Chest 127: 768–777.

Ganti, A.K., A.E. Sahmoun, A.W. Panwalkar, K.K. Tendulkar, and A. Potti. 2006. Hormone replacement therapy is associated with decreased survival in women with lung cancer. J Clin Oncol 24: 59–63.

Garber, M.E., O.G. Troyanskaya, K. Schluens, S. Petersen, Z. Thaesler, M. Pacyna-Gengelbach, M. van de Rijn, G.D. Rosen, C.M. Perou, R.I. Whyte et al. 2001. Diversity of gene expression in adenocarcinoma of the lung. Proc Natl Acad Sci U S A 98: 13784–13789.

Gazdar, A.F. 2007. DNA repair and survival in lung cancer – the two faces of Janus. N Engl J Med 356: 771–773.

Giordano, T.J., K.A. Shedden, D.R. Schwartz, R. Kuick, J.M. Taylor, N. Lee, D.E. Misek, J.K. Greenson, S.L. Kardia, D.G. Beer et al. 2001. Organ-specific molecular classification of primary lung, colon, and ovarian adenocarcinomas using gene expression profiles. Am J Pathol 159: 1231–1238.

Glas, A., A. Floore, L. Delahaye, A. Witteveen, R. Pover, N. Bakx, J. Lahti-Domenici, T. Bruinsma, M. Warmoes, R. Bernards et al. 2006. Converting a breast cancer microarray signature into a high-throughput diagnostic test. BMC Genomics 7: 278.

Goya, T., H. Asamura, H. Yoshimura, H. Kato, K. Shimokata, R. Tsuchiya, Y. Sohara, T. Miya, and E. Miyaoka. 2005. Prognosis of 6644 resected non-small cell lung cancers in Japan: a Japanese lung cancer registry study. Lung Cancer 50: 227–234.

Guo, N.L., Y.W. Wan, K. Tosun, H. Lin, Z. Msiska, D.C. Flynn, S.C. Remick, V. Vallyathan, A. Dowlati, X. Shi et al. 2008. Confirmation of gene expression-based prediction of survival in non-small cell lung cancer. Clin Cancer Res 14: 8213–8220.

Hayes, D.N., S. Monti, G. Parmigiani, C.B. Gilks, K. Naoki, A. Bhattacharjee, M.A. Socinski, C. Perou, and M. Meyerson. 2006. Gene expression profiling reveals

reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. J Clin Oncol 24: 5079–5090.
Hecht, S.S. 1999. Tobacco smoke carcinogens and lung cancer. J Natl Cancer Inst 91: 1194–1210.
Hecht, S.S. 2003. Tobacco carcinogens, their biomarkers and tobacco-induced cancer. Nat Rev Cancer 3: 733–744.
Hoti, F. and M.J. Sillanpaa. 2006. Bayesian mapping of genotype [times] expression interactions in quantitative and qualitative traits. Heredity 97: 4–18.
Howe, M.C., A. Chapman, K. Kerr, M. Dougal, H. Anderson, and P.S. Hasleton. 2005. Neuroendocrine differentiation in non-small cell lung cancer and its relation to prognosis and therapy. Histopathology 46: 195–201.
Hsu, Y.C., S. Yuan, H.Y. Chen, S.L. Yu, C.H. Liu, P.Y. Hsu, G. Wu, C.H. Lin, G.C. Chang, K.C. Li et al. 2009. A four-gene signature from NCI-60 cell line for survival prediction in non-small cell lung cancer. Clin Cancer Res 15: 7309–7315.
Husain, A., G. He, E.S. Venkatraman, and D.R. Spriggs. 1998. BRCA1 up-regulation is associated with repair-mediated resistance to cis-diamminedichloroplatinum(II). Cancer Res 58: 1120–1123.
Inoue, Y., M. Oka, H. Ishii, K. Kimino, M. Kishikawa, M. Ito, T. Ito, M. Nakashima, and S. Kohno. 2001. A solitary bronchial papilloma with malignant changes. Intern Med 40: 56–60.
Jacobson, T.A., J. Lundahl, H. Mellstedt, and A. Moshfegh. Gene expression analysis using long-term preserved formalin-fixed and paraffin-embedded tissue of non-small cell lung cancer. Int J Oncol 38: 1075–1081.
Jemal, A., F. Bray, M.M. Center, J. Ferlay, E. Ward, and D. Forman. Global cancer statistics. CA Cancer J Clin.
Jemal, A., R. Siegel, E. Ward, Y. Hao, J. Xu, T. Murray, and M.J. Thun. 2008. Cancer statistics, 2008. CA Cancer J Clin 58: 71–96.
Jemal, A., R. Siegel, E. Ward, Y. Hao, J. Xu, and M.J. Thun. 2009. Cancer statistics, 2009. CA Cancer J Clin 59: 225–249.
Jemal, A., R. Siegel, E. Ward, T. Murray, J. Xu, C. Smigal, and M.J. Thun. 2006. Cancer statistics, 2006. CA Cancer J Clin 56: 106–130.
Jemal, A., E. Ward, and M.J. Thun. 2005. Contemporary lung cancer trends among U.S. women. Cancer Epidemiol Biomarkers Prev 14: 582–585.
Ju, S., H. Qiu, X. Zhou, B. Zhu, X. Lv, X. Huang, J. Li, Y. Zhang, L. Liu, Y. Ge et al. 2009. CD13+CD4+CD25hi regulatory T cells exhibit higher suppressive function and increase with tumor stage in non-small cell lung cancer patients. Cell Cycle 8: 2578–2585.
Kerr, K.M. 2001. Pulmonary preinvasive neoplasia. J Clin Pathol 54: 257–271.
Klein, F., W.F. Amin Kotb, and I. Petersen. 2009. Incidence of human papilloma virus in lung cancer. Lung Cancer 65: 13–18.
Koshiol, J., M. Rotunno, M.L. Gillison, L.-J. Van Doorn, A.K. Chaturvedi, L. Tarantini, H. Song, W.G.V. Quint, L. Struijk, A.M. Goldstein et al. Assessment of Human Papillomavirus in Lung Tumor Tissue. Journal of the National Cancer Institute.
Kuner, R., T. Muley, M. Meister, M. Ruschhaupt, A. Buness, E.C. Xu, P. Schnabel, A. Warth, A. Poustka, H. Sultmann et al. 2009. Global gene expression analysis reveals specific patterns of cell junctions in non-small cell lung cancer subtypes. Lung Cancer 63: 32–38.

Larsen, J.E., S.J. Pavey, L.H. Passmore, R. Bowman, B.E. Clarke, N.K. Hayward, and K.M. Fong. 2007. Expression profiling defines a recurrence signature in lung squamous cell carcinoma. Carcinogenesis 28: 760–766.

Larsen, J.E., S.J. Pavey, L.H. Passmore, R.V. Bowman, N.K. Hayward, and K.M. Fong. 2007. Gene expression signature predicts recurrence in lung adenocarcinoma. Clin Cancer Res 13: 2946–2954.

Lau, S.K., P.C. Boutros, M. Pintilie, F.H. Blackhall, C.Q. Zhu, D. Strumpf, M.R. Johnston, G. Darling, S. Keshavjee, T.K. Waddell et al. 2007. Three-gene prognostic classifier for early-stage non small-cell lung cancer. J Clin Oncol 25: 5562–5569.

Lazzaro, D., M. Price, M. de Felice, and R. Di Lauro. 1991. The transcription factor TTF-1 is expressed at the onset of thyroid and lung morphogenesis and in restricted regions of the foetal brain. Development 113: 1093–1104.

Lim, E.H., A. Aggarwal, T. Agasthian, P.S. Wong, C. Tan, E. Sim, L. Tan, P.S. Goh, S.C. Wang, K.L. Khoo et al. 2003. Feasibility of using low-volume tissue samples for gene expression profiling of advanced non-small cell lung cancers. Clin Cancer Res 9: 5980–5987.

Linnoila, R.I., S. Piantadosi, and J.C. Ruckdeschel. 1994. Impact of neuroendocrine differentiation in non-small cell lung cancer. The LCSG experience. Chest 106: 367S–371S.

Llinares, K., F. Escande, S. Aubert, M.P. Buisine, C. de Bolos, S.K. Batra, B. Gosselin, J.P. Aubert, N. Porchet, and M.C. Copin. 2004. Diagnostic value of MUC4 immunostaining in distinguishing epithelial mesothelioma and lung adenocarcinoma. Mod Pathol 17: 150–157.

Lord, R.V., J. Brabender, D. Gandara, V. Alberola, C. Camps, M. Domine, F. Cardenal, J.M. Sanchez, P.H. Gumerlock, M. Taron et al. 2002. Low ERCC1 expression correlates with prolonged survival after cisplatin plus gemcitabine chemotherapy in non-small cell lung cancer. Clin Cancer Res 8: 2286–2291.

Lu, Y., W. Lemon, P.Y. Liu, Y. Yi, C. Morrison, P. Yang, Z. Sun, J. Szoke, W.L. Gerald, M. Watson et al. 2006. A gene expression signature predicts survival of patients with stage I non-small cell lung cancer. PLoS Med 3: e467.

Lynch, T.J., D.W. Bell, R. Sordella, S. Gurubhagavatula, R.A. Okimoto, B.W. Brannigan, P.L. Harris, S.M. Haserlat, J.G. Supko, F.G. Haluska et al. 2004. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 350: 2129–2139.

Mao, C., L.X. Qiu, R.Y. Liao, F.B. Du, H. Ding, W.C. Yang, J. Li, and Q. Chen. 2009. KRAS mutations and resistance to EGFR-TKIs treatment in patients with non-small cell lung cancer: A meta-analysis of 22 studies. Lung Cancer.

Marchionni, L., R.F. Wilson, A.C. Wolff, S. Marinopoulos, G. Parmigiani, E.B. Bass, and S.N. Goodman. 2008. Systematic Review: Gene Expression Profiling Assays in Early-Stage Breast Cancer. Annals of Internal Medicine 148: 358–369.

Martini, N., M.S. Bains, M.E. Burt, M.F. Zakowski, P. McCormack, V.W. Rusch, and R.J. Ginsberg. 1995. Incidence of local recurrence and second primary tumors in resected stage I lung cancer. J Thorac Cardiovasc Surg 109: 120–129.

Mascaux, C., N. Iannino, B. Martin, M. Paesmans, T. Berghmans, M. Dusart, A. Haller, P. Lothaire, A.P. Meert, S. Noel et al. 2005. The role of RAS oncogene in survival of patients with lung cancer: a systematic review of the literature with meta-analysis. Br J Cancer 92: 131–139.

Mellstedt, H., J. Vansteenkiste, and N. Thatcher. Vaccines for the treatment of non-small cell lung cancer: Investigational approaches and clinical experience. Lung Cancer 73: 11–17.

Mitsudomi, T., T. Kosaka, H. Endoh, Y. Horio, T. Hida, S. Mori, S. Hatooka, M. Shinoda, T. Takahashi, and Y. Yatabe. 2005. Mutations of the epidermal growth factor receptor gene predict prolonged survival after gefitinib treatment in patients with non-small-cell lung cancer with postoperative recurrence. J Clin Oncol 23: 2513–2520.

Miura, K., E.D. Bowman, R. Simon, A.C. Peng, A.I. Robles, R.T. Jones, T. Katagiri, P. He, H. Mizukami, L. Charboneau et al. 2002. Laser capture microdissection and microarray expression analysis of lung adenocarcinoma reveals tobacco smoking- and prognosis-related molecular profiles. Cancer Res 62: 3244–3250.

Mostertz, W., M. Stevenson, C. Acharya, I. Chan, K. Walters, W. Lamlertthon, W. Barry, J. Crawford, J. Nevins, and A. Potti. Age- and sex-specific genomic profiles in non-small cell lung cancer. Jama 303: 535–543.

Mountain, C.F. and C.M. Dresler. 1997. Regional lymph node classification for lung cancer staging. Chest 111: 1718–1723.

Nelson, H.H., D.C. Christiani, E.J. Mark, J.K. Wiencke, J.C. Wain, and K.T. Kelsey. 1999. Implications and prognostic value of K-ras mutation for early-stage lung cancer in women. J Natl Cancer Inst 91: 2032–2038.

Nevins, J.R. and A. Potti. 2007. Mining gene expression profiles: expression signatures as cancer phenotypes. Nat Rev Genet 8: 601–609.

Nogales-Cadenas, R., P. Carmona-Saez, M. Vazquez, C. Vicente, X. Yang, F. Tirado, J.M.a. Carazo, and A. Pascual-Montano. 2009. GeneCodis: interpreting gene lists through enrichment analysis and integration of diverse biological information. Nucleic Acids Research 37: W317–W322.

Paesmans, M., J.P. Sculier, J. Lecomte, J. Thiriaux, P. Libert, R. Sergysels, G. Bureau, G. Dabouis, O. Van Cutsem, P. Mommen et al. 2000. Prognostic factors for patients with small cell lung carcinoma: analysis of a series of 763 patients included in 4 consecutive prospective trials with a minimum follow-up of 5 years. Cancer 89: 523–533.

Paez, J.G., P.A. Janne, J.C. Lee, S. Tracy, H. Greulich, S. Gabriel, P. Herman, F.J. Kaye, N. Lindeman, T.J. Boggon et al. 2004. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 304: 1497–1500.

Pauk, N., A. Kubik, P. Zatloukal, and E. Krepela. 2005. Lung cancer in women. Lung Cancer 48: 1–9.

Pignon, J.P., H. Tribodet, G.V. Scagliotti, J.Y. Douillard, F.A. Shepherd, R.J. Stephens, A. Dunant, V. Torri, R. Rosell, L. Seymour et al. 2008. Lung adjuvant cisplatin evaluation: a pooled analysis by the LACE Collaborative Group. J Clin Oncol 26: 3552–3559.

Planchard, D., Y. Loriot, A. Goubar, F. Commo, and J.C. Soria. 2009. Differential expression of biomarkers in men and women. Semin Oncol 36: 553–565.

Potti, A., S. Mukherjee, R. Petersen, H.K. Dressman, A. Bild, J. Koontz, R. Kratzke, M.A. Watson, M. Kelley, G.S. Ginsburg et al. 2006. A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. N Engl J Med 355: 570–580.

Powell, C.A., A. Spira, A. Derti, C. DeLisi, G. Liu, A. Borczuk, S. Busch, S. Sahasrabudhe, Y. Chen, D. Sugarbaker et al. 2003. Gene expression in lung adenocarcinomas of smokers and nonsmokers. Am J Respir Cell Mol Biol 29: 157–162.

Proctor, R.N. 1999. Why did the Nazis have the world's most aggressive anti-cancer campaign? Endeavour 23: 76–79.
Rami-Porta, R., K. Chansky, and P. Goldstraw. 2009. Updated lung cancer staging system. Future Oncol 5: 1545–1553.
Rami-Porta, R., J.J. Crowley, and P. Goldstraw. 2009. The revised TNM staging system for lung cancer. Ann Thorac Cardiovasc Surg 15: 4–9.
Raponi, M., Y. Zhang, J. Yu, G. Chen, G. Lee, J.M. Taylor, J. Macdonald, D. Thomas, C. Moskaluk, Y. Wang et al. 2006. Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. Cancer Res 66: 7466–7472.
Reimand, J., M. Kull, H. Peterson, J. Hansen, and J. Vilo. 2007. g:Profiler – a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res 35: W193–200.
Reis-Filho, J.S., C. Carrilho, C. Valenti, D. Leitao, C.A. Ribeiro, S.G. Ribeiro, and F.C. Schmitt. 2000. Is TTF1 a good immunohistochemical marker to distinguish primary from metastatic lung adenocarcinomas? Pathol Res Pract 196: 835–840.
Rivera, M.P. and D.E. Stover. 2004. Gender and lung cancer. Clin Chest Med 25: 391–400.
Roglic, M., S. Jukic, and I. Damjanov. 1975. Cytology of the solitary papilloma of the bronchus. Acta Cytol 19: 11–13.
Rosell, R., M. Cobo, D. Isla, C. Camps, and B. Massuti. 2006. Pharmacogenomics and gemcitabine. Ann Oncol 17 Suppl 5: v13–16.
Rosell, R., T. Moran, C. Queralt, R. Porta, F. Cardenal, C. Camps, M. Majem, G. Lopez-Vivanco, D. Isla, M. Provencio et al. 2009. Screening for epidermal growth factor receptor mutations in lung cancer. N Engl J Med 361: 958–967.
Rosell, R., M. Skrzypski, E. Jassem, M. Taron, R. Bartolucci, J.J. Sanchez, P. Mendez, I. Chaib, L. Perez-Roca, A. Szymanowska et al. 2007. BRCA1: A Novel Prognostic Factor in Resected Non-Small-Cell Lung Cancer. PLoS ONE 2: e1129.
Rubel, L. and R.E. Reynolds. 1979. Cytologic description of squamous cell papilloma of the respiratory tract. Acta Cytol 23: 227–231.
Saviozzi, S., F. Cordero, M. Lo Iacono, S. Novello, G.V. Scagliotti, and R.A. Calogero. 2006. Selection of suitable reference genes for accurate normalization of gene expression profile studies in non-small cell lung cancer. BMC Cancer 6: 200.
Schabath, M.B., X. Wu, R. Vassilopoulou-Sellin, A.A. Vaporciyan, and M.R. Spitz. 2004. Hormone replacement therapy and lung cancer risk: a case-control analysis. Clin Cancer Res 10: 113–123.
Schairer, E. and E. Schoniger. 2001. Lung cancer and tobacco consumption. Int J Epidemiol 30: 24-27; discussion 30–21.
Schiller, J.H., S. Adak, R.H. Feins, S.M. Keller, W.A. Fry, R.B. Livingston, M.E. Hammond, B. Wolf, L. Sabatini, J. Jett et al. 2001. Lack of prognostic significance of p53 and K-ras mutations in primary resected non-small-cell lung cancer on E4592: a Laboratory Ancillary Study on an Eastern Cooperative Oncology Group Prospective Randomized Trial of Postoperative Adjuvant Therapy. J Clin Oncol 19: 448–457.
Scott, W.J., J. Howington, S. Feigenberg, B. Movsas, and K. Pisters. 2007. Treatment of non-small cell lung cancer stage I and stage II: ACCP evidence-based clinical practice guidelines (2nd edition). Chest 132: 234S–242S.
Shedden, K., J.M.G. Taylor, S.A. Enkemann, M.-S. Tsao, T.J. Yeatman, W.L. Gerald, S. Eschrich, I. Jurisica, T.J. Giordano, D.E. Misek et al. 2008. Gene expression-

based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. Nat Med 14: 822–827.
Shigematsu, H., L. Lin, T. Takahashi, M. Nomura, M. Suzuki, Wistuba, II, K.M. Fong, H. Lee, S. Toyooka, N. Shimizu et al. 2005. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. J Natl Cancer Inst 97: 339–346.
Simon, G.R., S. Sharma, A. Cantor, P. Smith, and G. Bepler. 2005. ERCC1 expression is a predictor of survival in resected patients with non-small cell lung cancer. Chest 127: 978–983.
Smyth, G.K. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 3: Article3.
Spira, A., J. Beane, V. Shah, G. Liu, F. Schembri, X. Yang, J. Palma, and J.S. Brody. 2004. Effects of cigarette smoke on the human airway epithelial cell transcriptome. Proc Natl Acad Sci U S A 101: 10143–10148.
Sridhar, S., F. Schembri, J. Zeskind, V. Shah, A.M. Gustafson, K. Steiling, G. Liu, Y.M. Dumas, X. Zhang, J.S. Brody et al. 2008. Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium. BMC Genomics 9: 259.
Sterlacci, W., M. Fiegl, W. Hilbe, J. Auberger, G. Mikuz, and A. Tzankov. 2009. Clinical relevance of neuroendocrine differentiation in non-small cell lung cancer assessed by immunohistochemistry: a retrospective study on 405 surgically resected cases. Virchows Arch 455: 125–132.
Su, Y.-C., Y.-C. Hsu, and C.-Y. Chai. 2006. Role Of TTF-1, CK20, And CK7 Immunohistochemistry for Diagnosis of Primary and Secondary Lung Adenocarcinoma. The Kaohsiung Journal of Medical Sciences 22: 14–19.
Subramanian, J., L. Corrales, D. Soulieres, D. Morgensztern, and R. Govindan. Summary of Presentations from the 46th Annual Meeting of the American Society of Clinical Oncology (2010) Focus on Tumor Biology and Biomarkers Related to Lung Cancer. Journal of Thoracic Oncology 6: 399–403 310.1097/JTO.1090b1013e318200f318972.
Subramanian, J., D. Morgensztern, B. Goodgame, M.Q. Baggstrom, F. Gao, J. Piccirillo, and R. Govindan. Distinctive characteristics of non-small cell lung cancer (NSCLC) in the young: a surveillance, epidemiology, and end results (SEER) analysis. J Thorac Oncol 5: 23–28.
Sun, S., J.H. Schiller, and A.F. Gazdar. 2007. Lung cancer in never smokers [mdash] a different disease. Nat Rev Cancer 7: 778–790.
Syrjanen, K.J. 2002. HPV infections and lung cancer. J Clin Pathol 55: 885–891.
Taioli, E. and E.L. Wynder. 1994. Re: Endocrine Factors and Adenocarcinoma of the Lung in Women. Journal of the National Cancer Institute 86: 869–870.
Takeuchi, T., S. Tomida, Y. Yatabe, T. Kosaka, H. Osada, K. Yanagisawa, T. Mitsudomi, and T. Takahashi. 2006. Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors. J Clin Oncol 24: 1679–1688.
Talbot, S.G., C. Estilo, E. Maghami, I.S. Sarkaria, D.K. Pham, O.c. P, N.D. Socci, I. Ngai, D. Carlson, R. Ghossein et al. 2005. Gene expression profiling allows distinction between primary and metastatic squamous cell carcinomas in the lung. Cancer Res 65: 3063–3071.
Tam, I.Y., L.P. Chung, W.S. Suen, E. Wang, M.C. Wong, K.K. Ho, W.K. Lam, S.W. Chiu, L. Girard, J.D. Minna et al. 2006. Distinct epidermal growth factor receptor

and KRAS mutation patterns in non-small cell lung cancer patients with different tobacco exposure and clinicopathologic features. Clin Cancer Res 12: 1647–1653.
Thomas, L., L.A. Doyle, and M.J. Edelman. 2005. Lung cancer in women: emerging differences in epidemiology, biology, and therapy. Chest 128: 370–381.
Toffalorio, F., E. Giovannetti, T. De Pas, D. Radice, G. Pelosi, M. Manzotti, D. Minocci, L. Spaggiari, G. Spitaleri, C. Noberasco et al. 2009. Expression of gemcitabine- and cisplatin-related genes in non-small-cell lung cancer. Pharmacogenomics J.
Tomida, S., K. Koshikawa, Y. Yatabe, T. Harano, N. Ogura, T. Mitsudomi, M. Some, K. Yanagisawa, T. Takahashi, H. Osada et al. 2004. Gene expression-based, individualized outcome prediction for surgically treated lung cancer patients. Oncogene 23: 5360–5370.
Travis, W.D. 2009. Lung tumours with neuroendocrine differentiation. Eur J Cancer 45 Suppl 1: 251–266.
Vachani, A., M. Nebozhyn, S. Singhal, L. Alila, E. Wakeam, R. Muschel, C.A. Powell, P. Gaffney, B. Singh, M.S. Brose et al. 2007. A 10-gene classifier for distinguishing head and neck squamous cell carcinoma and lung squamous cell carcinoma. Clin Cancer Res 13: 2905–2915.
van ’t Veer, L.J., H. Dai, M.J. van de Vijver, Y.D. He, A.A. Hart, M. Mao, H.L. Peterse, K. van der Kooy, M.J. Marton, A.T. Witteveen et al. 2002. Gene expression profiling predicts clinical outcome of breast cancer. Nature 415: 530–536.
Vasquez, M., D. Flieder, D. Yankelevitz, C. Henschke, and 2001. Undetected pathologic lesions in lobectomy specimens of CT detected bronchioloalveolar carcinoma (BAC). Modern Pathol 2001: 227A.
Vilmar, A. and J.B. Sųrensen. 2009. Excision repair cross-complementation group 1 (ERCC1) in platinum-based treatment of non-small cell lung cancer with special emphasis on carboplatin: A review of current literature. Lung Cancer 64: 131–139.
Wahbah, M., N. Boroumand, C. Castro, F. El-Zeky, and M. Eltorky. 2007. Changing trends in the distribution of the histologic types of lung cancer: a review of 4,439 cases. Ann Diagn Pathol 11: 89–96.
Wei, Q., L. Cheng, C.I. Amos, L.E. Wang, Z. Guo, W.K. Hong, and M.R. Spitz. 2000. Repair of tobacco carcinogen-induced DNA adducts and lung cancer risk: a molecular epidemiologic study. J Natl Cancer Inst 92: 1764–1772.
Werner-Wasik, M., C. Scott, J.D. Cox, W.T. Sause, R.W. Byhardt, S. Asbell, A. Russell, R. Komaki, and J.S. Lee. 2000. Recursive partitioning analysis of 1999 Radiation Therapy Oncology Group (RTOG) patients with locally-advanced non-small-cell lung cancer (LA-NSCLC): identification of five groups with different survival. Int J Radiat Oncol Biol Phys 48: 1475–1482.
Wigle, D.A., I. Jurisica, N. Radulovich, M. Pintilie, J. Rossant, N. Liu, C. Lu, J. Woodgett, I. Seiden, M. Johnston et al. 2002. Molecular profiling of non-small cell lung cancer and correlation with disease-free survival. Cancer Res 62: 3005–3008.
Wistuba, II, J. Berry, C. Behrens, A. Maitra, N. Shivapurkar, S. Milchgrub, B. Mackay, J.D. Minna, and A.F. Gazdar. 2000. Molecular changes in the bronchial epithelium of patients with small cell lung cancer. Clin Cancer Res 6: 2604–2610.
Wistuba, II and A.F. Gazdar. 2000. Molecular pathology of lung cancer. Verh Dtsch Ges Pathol 84: 96–105.
Wistuba, II and A.F. Gazdar. 2006. Lung cancer preneoplasia. Annu Rev Pathol 1: 331–348.

Yatabe, Y., T. Mitsudomi, and T. Takahashi. 2002. TTF-1 expression in pulmonary adenocarcinomas. Am J Surg Pathol 26: 767–773.
Zheng, Z., T. Chen, X. Li, E. Haura, A. Sharma, and G. Bepler. 2007. DNA synthesis and repair genes RRM1 and ERCC1 in lung cancer. N Engl J Med 356: 800–808.

SUMMARY IN ESTONIAN

Mitteväikerakulise kopsuvähi histoloogiliste alatüüpide molekulaarsed erinevused ja sarnasused

Käesoleval ajal kasutuses olevad vähi diagnostika meetodid nagu valgusmikroskoopia, preparaadi hematoksüliin-eosiin värvimine, immuunohistokeemia ja TNM süsteemil põhinev toopiline diagnoosimine, on paljudel juhtudel haiguse prognoosi ja eeldatava raviefekti hindamisel ebapiisavad.

Kopsuvähi geenieksрessiooni analüüsimine kõiki teadaolevaid kodeerivaid geene hõlmavate kiipidega teadusuuringutes ja kindlaid geenimarkereid kasutades kliinilistes uuringutes on järgmiseks loogiliseks sammuks. Geenieksрessiooni analüüsi tuntuimad rakendused oleksid vähi erinevate alatüüpide eristamine, prognoosi hindamine, ravimeetodite sobivuse hindamine ja teadus- ning kliinilisteks uuringuteks sobivate uute markerite leidmine.

Igapäevaselt on diagnostilistel, prediktiivsetel ja prognostilistel eesmärkidel ratsionaalne kasutada markerite komplekte.

Käesolevas uuringus analüüsiti mitteväikerakulise kopsuvähi geenieksрessiooni, kasutades selleks ülegenoomi mikrokiipe. Saadud andmetega teostati mitteväikerakulise kopsuvähiga haigete elulemuse analüüs ning eristati peamisi vähi alatüüpe.

Uurimistöö eesmärgid

Käesoleva uuringu põhiliseks eesmärgiks oli kasutada geenieksрessiooni profiile Tartu Ülikooli Kliinikumis kirurgiliselt ravitud kopsuvähi patsientide vähiproovide ja samade haigete vähivaba kontrollkoe hindamiseks.
Sellele lisaks

- luua edasisteks uuringuteks Illumina platvormi põhine mitteväikerakulise kopsuvähi geenieksрessiooni andmebaas.
- tuua välja normaalset kopsukude, kopsu adenokartsinoomi, bronhioloalveolaarset vähki ja kopsu lamerakulist vähki eristavad geenid.
- leida elulemust määrav geenikomplekt või üksikmarker ning uusi kopsuvähiga seoses olevaid geene.
- leida TNM klassifikatsiooni staadiumeid eristavaid markereid või markerite komplekt.

Materjalid ja meetodid

Uuringuks kasutati Tartu Ülikooli Kliinikumi Kardiovaskulaar ja torakaalkirurgia Kliinikus ning Torakaalkirurgia Keskuses ajavahemikul 2002–2007 mitteväikerakulise kopsuvähi tõttu opereeritud 147 patsiendi koematerjali ning fenotüübi ja elulemuse andmeid. Kohordi koeproovidest eraldati Ambioni

RiboPure Kit'i kasutades kvaliteetne RNA 81 juhust. Sellega viidi läbi ekspressiooni kiibikatse Illumina Human-6 ülegenoomi kiibiga mis sisaldab rohkem kui 48000 transkripti. Geenieksprессiooni katsest saadud andmete analüüsiks kasutati t-testi koos empiirilise Bayesi korrektsiooniga Bioconductori Limma teegist. Olulisustõenäosuste korrektsioon mitmese testimise vastu viidi läbi kasutades Bonferroni meetodit ja kõik testid viidi läbi usaldusnivool 0,05. Leitud markerite arvu vähendamiseks säilitati geenid mis näitasid vähemalt kahekordset ekspressiooni erinevust vähivabast normkoest. Geeniontoloogia analüüsiks kasutati võrgu tarkvara g:Profiler. Bioloogiliste protsesside analüüsiks kasutati võrgus vabavarana kättesaadavat programmi Genecodis 2.0. Statistiliselt oluliselt ekspresseerunud geenid on hierarhiliselt klasterdatud ja visualiseeritud *heatmap'* ile kasutades R-paketti. Erinevate vähitüüpide ja staadiumite molekulaarsete profiilide erinevuste visualiseerimiseks kasutati peakomponentanalüüsi. Metageeni leidmiseks piirati andmestiku suurust, klasterdades sarnase ekspressiooniga geene hierarhiliselt täieliku aheldatuse meetodil, kasutades sarnasusmõõduna Pearsoni korrelatsiooni. I b staadiumi patsientide geenieksprессiooni andmed jagati 500 gruppi ning iga klastri profiili keskmine moodustas metageeni. Analüüsideks kasutati Bayesi binomiaalset mudelit (sparse Bayesian probit model for binary response data) .

Uurimistöö peamised tulemused

1. Uuringu tulemusena leiti mitteväikerakulises kopsuvähis normaalsest kopsukoest vähemalt kahekordse ekspressiooni erinevusega 672 alla- ja 1103 ülesekspresseerunud geeni. Kõikide nende geenide juhusliku esinemise tõenäosus on alla 5%.
2. Geenieksprессiooni profiili peakomponentanalüüs aitas eristada erinevaid vähitüüpe teineteisest ja mõnel juhul ka algkollet metastaasist. Samas ei eristu mitteväikerakulise kopsuvähi geenieksprессiooni profiilil ega ka selle peakomponentanalüüsil TNM il põhinevad vähkide kliinilised staadiumid.
3. Kasutades p väärtust 10^{-6} ei õnnestunud veenvalt leida adenokartsinoomi ja bronhioloalveolaarset vähki eristavaid markereid.
4. Ib staadiumi patsientide prognoosi hindamiseks leiti kaks statistilise olulisuse ja aktsepteeritava veapiiriga metageeni mis jaotas haigusjuhud ekspressiooni profiilide järgi kahte prognostilisse gruppi elulemusega vastavalt <757 ja >1225 päeva. Kahe grupi eristamiseks valisime kokkuleppeliselt 1000 päeva piiri.

Järeldused

1. Geeniekspressiooni hindamine on sobilik vähikoe ja vähivaba koe eristamiseks aga ka vähi tähtsamate subgruppide eristamiseks.
2. Vähi geeniekspressiooni andmete põhjal saab piisava suurusega kohordi korral leida prognostilisi metageene.
3. Tulenevalt kohordi iseärasustest ei saanud antud töös teha üldistusi mitteväikerakulise kopsuvähi geeniekspressiooni profiilide sooliste ega ka diferentseerumisastmetel põhinevate erinevuse kohta.

Kokkuvõte

Käesolev uurimistöö lisab suures mahus uut informatsiooni mitteväikerakulise kopsuvähi kahe peamise histoloogilise vormi, lamerakulise vähi ning adenokartsinoomi ja bronhioloalveolaarse vähi geeniekspressiooni kohta kogu transkriptoomi ulatuses. Uuringu tulemusena toodi välja hulgaliselt vähivaba kude ja vähki, ning adenokartsinoomi ja lamerakulist vähki eristavaid markereid, kuid vähitüüpide sarnasuse tõttu ei suudetud molekulaarselt eristada adenokartsinoomi ja selle alaliiki bronhioloalveolaarset vähki. Samuti ei leitud TNM klassifikatsiooni erinevatele staadiumitele iseloomulikke molekulaarseid markereid.

Kasutades geeniekspressiooni andmeid toodi välja kaks statistiliselt olulist metageeni mille ekspressiooni järgi saab jaotada Ib staadiumi kopsuvähki kahte prognostilisse gruppi.

Uuringu osana loodud mitteväikerakulise kopsuvähi geeniekspressiooni andmebaas on vajalik järgnevate vähiuuringute läbiviimiseks ning võimalike diagnostiliste ja prognostiliste kiipide koostamiseks.

ACKNOWLEDGMENTS

This work was carried out at the Department of Vascular and Thoracic Surgery in The Clinic Cardiovascular and Thoracic Surgery of Tartu University Hospital, at the Institute of Molecular Biology of Tartu University, and at the Institute of Computer Science of Tartu University in Estonia between 2002 and 2011. Our research group expresses their appreciation to all of the patients that participated in this study.

I also want to express my sincere gratitude to all who made this work possible, especially:

Professor Andres Metspalu, my supervisor, for introducing me to the field of cancer genetics, and for providing necessary discussions, mental support, encouragement, and an enthusiastic attitude during my studies.

Docent Andres Pulges, my supervisor who directed me into scientific research and provided valuable ideas for publishing and completing this study.

Docent Aavo Lang for continous invisible support, friendly interest, and guidance long before the beginning of my official PhD studies, as well as during these studies, and at the end of this process.

Professor Sulev Kõks and professor Andres Salumets, the official reviewers, for invaluable and constructive comments and suggestions for improving this thesis.

Professor Alan Altraja for providing unofficial critical scientific comments, psychological support, translation of outside unscientific comments, and for necessary explanations in difficult days.

Mariann Reintam for providing basic knowledge of lung cancer pathology, for being a fundamental coauthor of the paper, and for surprising psychological support and help in understanding the endstage problems associated with thesis writing.

Retlav Roosipuu for his critical review and remarks of the histological aspect of this study, and his fundamental coauthorship in research papers of the current thesis.

Professor Ants Kurg for providing the first real introductive colloquium in the field of biotechnology.

Kristjan Välk and Raivo Kolde as partners in this research and main coauthors of all research papers.

All members of the Lung Cancer research group: Triin “Trixu” Umbleja,

Priit Palta, Tanel Kaart, Märt Möls, Egon Urgard, Kaie Kirotar, Neeme Tõnisson, Marina Solovjova, Tarmo Annilo, Urmo Võsa, Tõnu Esko, and Tiit Nikopensius.

All coauthors of papers belonging to the current thesis.

Viljo Soo and Agnes Osa for excellent assistance in all scientific and less scientific lab activities.

Heljo Luts, Margit Paas, Sille Tamm, and especially Krista Liiv and Merike Leego, for outstanding technical support and assistance.

Merje Tikk, Margit Mägi, Piret Truu, and Mare Puusepp for excellent statistical support in the collection and processing of clinical data.

Vahur Makke, Marianna Frik-Amelin, and Vladimir Järv for improving my radiological knowledge and for their kind help in completing radiological illustrations of papers.

Boris Malikov for valuable humorous discussions and for sharing his original approach to science, scientists, and scientific research.

I express my best thanks to colleagues from the Department of Pulmonology and Thoracic Surgery, the Operating Theatre, and the Clinical and Biochemical Laboratory of Tartu University Hospital contributed to this research.

Joel Starkopf, Peeter Tähepõld, and Karl Kull for continous favourable interest in this research process and mental support in difficult days.

Tarmo Annilo and Tarmo Puurand for valuable ideas and consultations during literary work.

Friends, coursemates, and people who shed light on life and science: Andres Sild, Innar Tõru, Arno Ruusalepp, Leili Kütt, Kaido Beljaev, Argo Lätt, Kaarel Krjutškov, Jaak Lehtsaar, Kuldar Kaljurand, Kaiu Prikk, Jana Jaal, Andrus Arak, Ragnar Heido, Aalo and Maret Eller, Aare and Tiiu Märtson.

Dear friends Viljar and Imbi Jaks and Katre Maasalu for scientific discussions and memorable indoor and outdoor activities.

Mart Kalder, Aavo Metsmägi, Eevi Lellep, Ruth Valdre, Tähti Saar, and Maarika Ojala for friendly interest. Toomas Ellervee for interest and support.

Riin Tamm, Mari Nelis, and Eva Reinmaa for friendly interest, support, necessary remarks in scientific writing, and pleasant company in the lab.

Liis Leitsalu for linguistic help during writing of this manuscript

I express my best thanks also to Captain Arvo Saare from Kihnu Island whose hospitable house was a great place for writing papers and this thesis.

Special thanks to my family: parents, Kaarli and Mall, and my brother, Margus.

This work has been supported by ETF Grant nr: 6465. Targeted Financing from the Estonian Ministry of Education and Research (SF0180142s08) and by the European Union via the European Regional Development Fund grant to the Centre of Excellence in Genomics, Estonian Biocentre and University of Tartu, FP7 and grants OPENGENE (#245536), ECOGENE (#205419, EBC).

I also received the Harry and Hilda Marie Mägi scholarship from the Tartu University Foundation in 2010.

Tartu
2011

PUBLICATIONS

Supplemental Table 1. List of the top processes affected by genes down-regulated in NSCLC.

No. Genes	NGR	NG	Hyp	Hyp*	Annotations
79	1871(29095)	79(557)	3.10407e-11	1.03986e-08	GO:0007165 :signal transduction (BP)
23	227(29095)	23(557)	9.83307e-11	1.64704e-08	GO:0006954 :inflammatory response (BP)
34	512(29095)	34(557)	4.30599e-10	4.80836e-08	GO:0007155 :cell adhesion (BP)
16	164(29095)	16(557)	1.19151e-07	9.97887e-06	GO:0007166 :cell surface receptor linked signaling pathway (BP)
30	531(29095)	30(557)	1.68703e-07	1.13031e-05	GO:0055114 :oxidation reduction (BP)
5	12(29095)	5(557)	1.78998e-06	9.99403e-05	GO:0045730 :respiratory burst (BP)
4	6(29095)	4(557)	1.93345e-06	9.25296e-05	GO:0007186 :G-protein coupled receptor protein signaling pathway (BP) GO:0006898 :receptor-mediated endocytosis (BP)
39	898(29095)	39(557)	2.13334e-06	8.93335e-05	GO:0007275 :multicellular organismal development (BP)
26	483(29095)	26(557)	2.72696e-06	0.000101504	GO:0030154 :cell differentiation (BP)
9	67(29095)	9(557)	5.16144e-06	0.000172908	GO:0016337 :cell-cell adhesion (BP)
3	3(29095)	3(557)	6.97931e-06	0.000212552	GO:0045944 :positive regulation of transcription from RNA polymerase II promoter (BP) GO:0009409 :response to cold (BP)
5	16(29095)	5(557)	9.26443e-06	0.000258632	GO:0001937 :negative regulation of endothelial cell proliferation (BP)
24	459(29095)	24(557)	1.06457e-05	0.000274332	GO:0006955 :immune response (BP)
7	41(29095)	7(557)	1.16046e-05	0.000277682	GO:0006898 :receptor-mediated endocytosis (BP)
5	17(29095)	5(557)	1.29181e-05	0.000288503	GO:0032870 :cellular response to hormone stimulus (BP)
8	58(29095)	8(557)	1.41809e-05	0.000296912	GO:0007165 :signal transduction (BP) GO:0007166 :cell surface receptor linked signaling pathway (BP)
6	29(29095)	6(557)	1.56494e-05	0.000308386	GO:0007585 :respiratory gaseous exchange (BP)
6	30(29095)	6(557)	1.92466e-05	0.0003582	GO:0009612 :response to mechanical stimulus (BP)
10	99(29095)	10(557)	2.06746e-05	0.000364526	GO:0045087 :innate immune response (BP)
5	19(29095)	5(557)	2.35171e-05	0.000393912	GO:0019370 :leukotriene biosynthetic process (BP)
7	46(29095)	7(557)	2.54364e-05	0.000405771	GO:0043627 :response to estrogen stimulus (BP)
3	4(29095)	3(557)	2.75185e-05	0.000341434	GO:0007204 :elevation of cytosolic calcium ion concentration (BP)

NRG-number of annotated genes in the reference list (Total number of genes in the reference list).**NG** Number of annotated genes in the input list (Total number of genes in the input list).**Hyp** Hypergeometric pValue. **Hyp*** Corrected hypergeometric pValue

Supplemental Table 2. List of the top processes affected by genes up-regulated in NSCLC

No. Genes	NGR	NG	Hyp	Hyp*	Annotations
95	459(29095)	95(945)	1.49942e-48	6.92734e-46	GO:0007049 :cell cycle (BP)
66	240(29095)	66(945)	3.63945e-42	8.40714e-40	GO:0051301 :cell division (BP)
61	223(29095)	61(945)	6.82931e-39	1.05171e-36	GO:0007049 :cell cycle (BP) GO:0051301 :cell division (BP)
55	191(29095)	55(945)	1.80626e-36	2.08623e-34	GO:0007067 :mitosis (BP)
49	162(29095)	49(945)	9.67375e-34	8.93854e-32	GO:0007067 :mitosis (BP) GO:0007049 :cell cycle (BP)
48	155(29095)	48(945)	1.30519e-33	1.005e-31	GO:0007067 :mitosis (BP) GO:0051301 :cell division (BP)
46	151(29095)	46(945)	6.66111e-32	4.39633e-30	GO:0007067 :mitosis (BP) GO:0007049 :cell cycle (BP) GO:0051301 :cell division (BP)
37	146(29095)	37(945)	8.66637e-23	5.00483e-21	GO:0006260 :DNA replication (BP)
22	63(29095)	22(945)	2.11458e-17	1.08549e-15	GO:0031145 :anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process (BP)
34	188(29095)	34(945)	3.49717e-16	1.61569e-14	GO:0006281 :DNA repair (BP)
21	65(29095)	21(945)	6.99777e-16	2.93907e-14	GO:0051437 :positive regulation of ubiquitin-protein ligase activity during mitotic cell cycle (BP)
20	62(29095)	20(945)	3.55251e-15	1.36772e-13	GO:0051436 :negative regulation of ubiquitin-protein ligase activity during mitotic cell cycle (BP)
19	59(29095)	19(945)	1.80384e-14	6.41057e-13	GO:0051437 :positive regulation of ubiquitin-protein ligase activity during mitotic cell cycle (BP) GO:0031145 :anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process (BP)
19	60(29095)	19(945)	2.56006e-14	8.4482e-13	GO:0051436 :negative regulation of ubiquitin-protein ligase activity during mitotic cell cycle (BP) GO:0031145 :anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process (BP)
9	9(29095)	9(945)	3.87672e-14	1.19403e-12	GO:0007049 :cell cycle (BP) GO:0007051 :spindle organization (BP)

					GO:0051436 :negative regulation of ubiquitin-protein ligase activity during mitotic cell cycle (BP)
18	59(29095)	18(945)	2.54905e-13	7.3604e-12	GO:0051437 :positive regulation of ubiquitin-protein ligase activity during mitotic cell cycle (BP)
10	13(29095)	10(945)	3.26427e-13	8.87115e-12	GO:0007051 :spindle organization (BP)
					GO:0007049 :cell cycle (BP)
					GO:0051301 :cell division (BP)
8	8(29095)	8(945)	1.20344e-12	3.08882e-11	GO:0007051 :spindle organization (BP)
					GO:0051436 :negative regulation of ubiquitin-protein ligase activity during mitotic cell cycle (BP)
					GO:0051437 :positive regulation of ubiquitin-protein ligase activity during mitotic cell cycle (BP)
17	57(29095)	17(945)	1.78104e-12	4.33074e-11	GO:0031145 :anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process (BP)
31	211(29095)	31(945)	2.30946e-12	5.33486e-11	GO:0006412 :translation (BP)
					GO:0007067 :mitosis (BP)
					GO:0007049 :cell cycle (BP)
7	7(29095)	7(945)	3.73194e-11	7.83707e-10	GO:0007051 :spindle organization (BP)
					GO:0007049 :cell cycle (BP)
					GO:0048015 :phosphoinositide-mediated signaling (BP)
7	7(29095)	7(945)	3.73194e-11	7.83707e-10	GO:0007051 :spindle organization (BP)
					GO:0051301 :cell division (BP)
10	18(29095)	10(945)	4.3067e-11	8.65084e-10	GO:0031145 :anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process (BP)
					GO:0007049 :cell cycle (BP)
13	37(29095)	13(945)	7.13642e-11	1.37376e-09	GO:0006260 :DNA replication (BP)
19	90(29095)	19(945)	7.17261e-11	1.3255e-09	GO:0006364 :rRNA processing (BP)
9	15(29095)	9(945)	1.62681e-10	2.89072e-09	GO:0006541 :glutamine metabolic process (BP)
9	16(29095)	9(945)	3.61091e-10	6.17867e-09	GO:0006164 :purine nucleotide biosynthetic process (BP)
					GO:0051301 :cell division (BP)
9	17(29095)	9(945)	7.45137e-10	1.18708e-08	GO:0051437 :positive regulation of ubiquitin-protein ligase activity during mitotic cell cycle (BP)

					GO:0007049 :cell cycle (BP)
					GO:0051301 :cell division (BP)
9	17(29095)	9(945)	7.45137e-10	1.18708e-08	GO:0031145 :anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process (BP)
					GO:0006281 :DNA repair (BP)
13	44(29095)	13(945)	8.42407e-10	1.29731e-08	GO:0006260 :DNA replication (BP)
					GO:0007049 :cell cycle (BP)
					GO:0051301 :cell division (BP)
					GO:0048015 :phosphoinositide-mediated signaling (BP)
6	6(29095)	6(945)	1.15611e-09	1.61855e-08	GO:0007051 :spindle organization (BP)
					GO:0007067 :mitosis (BP)
6	6(29095)	6(945)	1.15611e-09	1.61855e-08	GO:0007049 :cell cycle (BP)

NRG-number of annotated genes in the reference list (Total number of genes in the reference list).**NG** Number of annotated genes in the input list (Total number of genes in the input list).**Hyp** Hypergeometric pValue. **Hyp*** Corrected hypergeometric pValue

Supplemental table nr. 3 List of Up-regulated genes in NSCLC according to fold-change.

ID	logFC	P.Value	Adj.P.Val	Fold change	Gene name	Description
GI_38146097-S	3,868875736	9,81E-21	4,99E-18	14,60991355	SPP1	osteopontin precursor (bone sialoprotein-1) (secreted phosphoprotein 1) (spp-1) (urinary stone protein) (nephropontin) (uropontin). [source:uniprot/swissprot;acc:p10451]
GI_13027798-S	3,474904397	1,87E-13	1,54E-11	11,11860894	MMP1	interstitial collagenase precursor (ec 3.4.24.7) (matrix metalloproteinase-1) (mmp-1) (fibroblast collagenase) [contains: 22 kda interstitial collagenase; 27 kda interstitial collagenase]. [source:uniprot/swissprot;acc:p03956]
GI_4557700-S	3,264423691	5,45E-08	1,11E-06	9,609249052	KRT17	keratin, type i cytoskeletal 17 (cytokeratin-17) (ck-17) (keratin-17) (k17) (39.1). [source:uniprot/swissprot;acc:q04695]
GI_18375521-A	3,047479032	1,91E-12	1,23E-10	8,267659838	COL11A1	collagen alpha-1(xi) chain precursor. [source:uniprot/swissprot;acc:p12107]
GI_34147546-S	2,944625229	1,98E-20	9,48E-18	7,698755368	CTHRC1	collagen triple helix repeat-containing protein 1 precursor (nmtc1 protein). [source:uniprot/swissprot;acc:q96cg8]
GI_45269153-S	2,940428026	8,36E-07	1,25E-05	7,676390087	S100A2	protein s100-a2 (s100 calcium-binding protein a2) (protein s-100l) (can19). [source:uniprot/swissprot;acc:p29034]
GI_4557436-S	2,806949962	4,61E-16	7,08E-14	6,998035396	CDC20	cell division cycle protein 20 homolog (p55cdc). [source:uniprot/swissprot;acc:q12834]
GI_13027795-S	2,806878838	1,24E-13	1,07E-11	6,997690405	MMP11	stromelysin-3 precursor (ec 3.4.24.-) (st3) (sl-3) (matrix metalloproteinase-11) (mmp-11). [source:uniprot/swissprot;acc:p24347]
GI_34304346-A	2,783185017	1,85E-10	7,27E-09	6,883703781	TMPRSS4	transmembrane protease, serine 4 (ec 3.4.21.-) (membrane-type serine protease 2) (mt-sp2). [source:uniprot/swissprot;acc:q9nrs4]
GI_19913405-S	2,636941047	1,13E-16	1,98E-14	6,220114109	P11388-2	top2a_human isoform 2 of p11388 - homo sapiens (human) [source:uniprot/varsplic;acc:p11388-2]
GI_17738300-S	2,551031594	1,11E-10	4,56E-09	5,860531838	COL7A1	collagen alpha-1(vii) chain precursor (long-chain collagen) (lc collagen). [source:uniprot/swissprot;acc:q02388]
GI_31341566-S	2,54527524	9,11E-10	2,98E-08	5,837194902	GJB2	gap junction beta-2 protein (connexin-26) (cx26). [source:uniprot/swissprot;acc:p29033]
GI_4505206-S	2,468197963	3,46E-10	1,26E-08	5,533521759	MMP12	macrophage metalloelastase precursor (ec 3.4.24.65) (hme) (matrix metalloproteinase-12) (mmp-12) (macrophage elastase) (me). [source:uniprot/swissprot;acc:p39900]

GI_24797094-A	2,424334274	2,14E-20	1,00E-17	5,367812484	PYCR1	pyrroline-5-carboxylate reductase 1 (ec 1.5.1.2) (p5cr 1) (p5c reductase 1). [source:uniprot/swissprot;acc:p32322]
GI_10938017-S	2,357405903	2,86E-17	5,82E-15	5,124481019	CCNB2	g2/mitotic-specific cyclin-b2. [source:uniprot/swissprot;acc:o95067]
GI_5730050-S	2,286332455	2,90E-10	1,08E-08	4,878144367	SLC2A1	solute carrier family 2, facilitated glucose transporter member 1 (glucose transporter type 1, erythrocyte/brain) (glut-1) (hepg2 glucose transporter). [source:uniprot/swissprot;acc:p11166]
GI_29789063-S	2,252198917	1,80E-15	2,30E-13	4,76408422	SULF1	extracellular sulfatase sulf-1 precursor (ec 3.1.6.-) (hsulf-1). [source:uniprot/swissprot;acc:q8iwu6]
GI_45598382-S	2,221059641	9,79E-11	4,08E-09	4,66235753	C20ORF42	unc-112-related protein 1 (kindlin-1) (kindlerin) (kindlin syndrome protein). [source:uniprot/swissprot;acc:q9bql6]
GI_33356546-S	2,179829529	7,35E-14	6,61E-12	4,53100012	MCM2	dna replication licensing factor mcm2 (minichromosome maintenance protein 2 homolog) (nuclear protein bm28). [source:uniprot/swissprot;acc:p49736]
GI_19923110-S	2,176999457	1,01E-12	6,98E-11	4,522120567	IGFBP3	insulin-like growth factor-binding protein 3 precursor (igfbp-3) (ibp- 3) (igf-binding protein 3). [source:uniprot/swissprot;acc:p17936]
GI_4826835-S	2,160150846	2,05E-11	1,01E-09	4,469615864	MMP9	matrix metalloproteinase-9 precursor (ec 3.4.24.35) (mmp-9) (92 kda type iv collagenase) (92 kda gelatinase) (gelatinase b) (gelb) [contains: 67 kda matrix metalloproteinase-9; 82 kda matrix metalloproteinase-9]. [source:uniprot/swissprot;acc:p14780]
GI_4507518-S	2,157429839	3,59E-13	2,78E-11	4,461193854	TK1	thymidine kinase, cytosolic (ec 2.7.1.21). [source:uniprot/swissprot;acc:p04183]
GI_4502144-S	2,118250224	6,95E-15	7,88E-13	4,341670452	BIRC5	baculoviral iap repeat-containing protein 5 (apoptosis inhibitor survivin) (apoptosis inhibitor 4). [source:uniprot/swissprot;acc:o15392]
GI_4759177-S	2,10530225	1,27E-14	1,36E-12	4,302878937	AURKB	serine/threonine-protein kinase 12 (ec 2.7.11.1) (aurora-b) (aurora- and ipl1-like midbody-associated protein 1) (aim-1) (aurora/ipl1- related kinase 2) (aurora-related kinase 2) (stk-1). [source:uniprot/swissprot;acc:q96gd4]
GI_13699832-S	2,103098415	8,59E-16	1,22E-13	4,296310954	KIF2C	kinesin-like protein kif2c (mitotic centromere-associated kinesin) (mcak) (kinesin-like protein 6). [source:uniprot/swissprot;acc:q99661]
GI_14719826-S	2,094811916	8,92E-09	2,23E-07	4,271704693	COL1A1	collagen alpha-1(i) chain precursor (alpha-1 type i collagen). [source:uniprot/swissprot;acc:p02452]

GI_21536289-S	2,068560103	3,26E-15	3,93E-13	4,194678101	COL1A2	collagen alpha-2(i) chain precursor (alpha-2 type i collagen). [source:uniprot/swissprot;acc:p08123]
GI_17981703-S	2,057382533	4,71E-15	5,50E-13	4,162304562	CDKN3	cyclin-dependent kinase inhibitor 3 (ec 3.1.3.48) (ec 3.1.3.16) (cdk2- associated dual-specificity phosphatase) (kinase-associated phosphatase) (cyclin-dependent kinase-interacting protein 2) (cyclin- dependent kinase interactor 1). [source:uniprot/swissprot;acc:q16667]
GI_34147658-S	2,042024045	1,98E-07	3,46E-06	4,118228971	UCHL1	ubiquitin carboxyl-terminal hydrolase isozyme l1 (ec 3.4.19.12) (ec 6.-.-.-) (uch-l1) (ubiquitin thioesterase l1) (neuron cytoplasmic protein 9.5) (pgp 9.5) (pgp9.5). [source:uniprot/swissprot;acc:p09936]
GI_39725649-S	2,036900839	2,20E-13	1,79E-11	4,103630517	ECE2	endothelin-converting enzyme 2 (ec 3.4.24.71) (ece-2). [source:uniprot/swissprot;acc:o60344]
GI_24475622-S	2,03599116	1,71E-11	8,62E-10	4,101043823	MDK	midkine precursor (mk) (neurite outgrowth-promoting protein) (midgestation and kidney protein) (amphiregulin-associated protein) (arap) (neurite outgrowth-promoting factor 2). [source:uniprot/swissprot;acc:p21741]
GI_34304372-S	2,028440179	1,26E-15	1,69E-13	4,079635275	CCNB1	g2/mitotic-specific cyclin-b1. [source:uniprot/swissprot;acc:p14635]
GI_42544164-A	2,024418644	4,62E-13	3,48E-11	4,068279064	FOXM1	forkhead box protein m1 (forkhead-related protein fkhl16) (hepatocyte nuclear factor 3 forkhead homolog 11) (hnf-3/fork-head homolog 11) (hfh-11) (winged-helix factor from ins-1 cells) (m-phase phosphoprotein 2) (mpm-2 reactive phosphoprotein 2) (transcri [source:uniprot/swissprot;acc:q08050]

Supplemental Table 4. List of genes down-regulated in NSCLC according to fold-change.

ID	logFC	P.Value	Adj.P.Val	Fold change	Gene name	Description
GI_42476334-S	-4,622269	9,66E-12	5,23E-10	24,6287	SFTPC	pulmonary surfactant-associated protein c precursor (sp-c) (sp5) (pulmonary surfactant-associated proteolipid spl(val)). [source:uniprot/swissprot;acc:p11686]
GI_27754777-A	-4,117448	1,26E-20	6,32E-18	17,35702	FCN3	ficolin-3 precursor (collagen/fibrinogen domain-containing protein 3) (collagen/fibrinogen domain-containing lectin 3 p35) (hakata antigen). [source:uniprot/swissprot;acc:o75636]
GI_34577060-S	-4,096174	3,56E-16	5,61E-14	17,10296	ADH1C	alcohol dehydrogenase 1b (ec 1.1.1.1) (alcohol dehydrogenase beta subunit). [source:uniprot/swissprot;acc:p00325]
GI_26787961-A	-4,039442	6,10E-23	4,81E-20	16,44346	RAGE_HUMAN	advanced glycosylation end product-specific receptor precursor (receptor for advanced glycosylation end products). [source:uniprot/swissprot;acc:q15109]
GI_4557578-S	-3,784403	7,63E-17	1,39E-14	13,77903	FABP4	fatty acid-binding protein, adipocyte (afabp) (adipocyte lipid-binding protein) (albp) (a-fabp). [source:uniprot/swissprot;acc:p15090]
GI_8922795-S	-3,65272	3,17E-26	6,80E-23	12,57703	TMEM100	transmembrane protein 100. [source:uniprot/swissprot;acc:q9nv29]
GI_27885012-S	-3,633958	2,65E-23	2,24E-20	12,41453	C19ORF59	mast cell-expressed membrane protein 1. [source:uniprot/swissprot;acc:q8ix19]
GI_13435388-S	-3,49263	9,60E-15	1,06E-12	11,25606	CYP4B1	cytochrome p450 4b1 (ec 1.14.14.1) (cypivb1) (p450-hp). [source:uniprot/swissprot;acc:p13584]
GI_8393146-S	-3,460629	1,34E-24	1,63E-21	11,00913	CLIC5	chloride intracellular channel protein 5. [source:uniprot/swissprot;acc:q9nza1]
GI_42476329-S	-3,442	1,09E-09	3,52E-08	10,86789	SFTPD	pulmonary surfactant-associated protein d precursor (sp-d) (psp-d) (lung surfactant protein d). [source:uniprot/swissprot;acc:p35247]
GI_13346505-S	-3,423087	1,45E-08	3,43E-07	10,72634	SFTA2_HUMAN	pulmonary surfactant-associated protein a2 precursor (sp-a2) (sp-a) (psp-a) (pspa) (alveolar proteinosis protein) (35 kda pulmonary surfactant-associated protein). [source:uniprot/swissprot;acc:q8iwl1]
GI_4507556-S	-3,373061	3,56E-26	7,32E-23	10,36078	CLEC3B	tetranectin precursor (tn) (c-type lectin domain family 3 member b) (plasminogen kringle 4-binding protein). [source:uniprot/swissprot;acc:p05452]

GI_39725696-S	-3,346781	1,13E-09	3,63E-08	10,17376	SCGB1A1	uteroglobin precursor (secretoglobin family 1a member 1) (clara cell phospholipid-binding protein) (ccpbp) (clara cells 10 kda secretory protein) (cc10) (urinary protein 1) (urine protein 1) (up1). [source:uniprot/swissprot;acc:p11684]
GI_32698947-S	-3,317739	1,62E-23	1,47E-20	9,971004	GKN2	gastrokine-2 precursor (blottin) (trefoil factor interactions(z) 1) (down-regulated in gastric cancer). [source:uniprot/swissprot;acc:q86xp6]
GI_38888174-S	-3,243645	1,36E-17	3,11E-15	9,471844	SFTPA1B	pulmonary surfactant-associated protein a1 precursor (sp-a1) (sp-a) (psp-a) (pspa) (alveolar proteinosis protein) (35 kda pulmonary surfactant-associated protein). [source:uniprot/swissprot;acc:q8iwl2]
GI_16915927-A	-3,222937	5,54E-24	5,83E-21	9,336857	CPB2	carboxypeptidase b2 precursor (ec 3.4.17.20) (carboxypeptidase u) (cpu) (thrombin-activable fibrinolysis inhibitor) (tafi) (plasma carboxypeptidase b) (pcpb). [source:uniprot/swissprot;acc:q96iy4]
GI_40538801-S	-3,173949	3,62E-19	1,24E-16	9,025136	MAMDC2	mam domain-containing protein 2 precursor. [source:uniprot/swissprot;acc:q7z304]
GI_26787960-I	-3,156301	4,46E-29	2,48E-25	8,915408	AGER	advanced glycosylation end product-specific receptor precursor (receptor for advanced glycosylation end products). [source:uniprot/swissprot;acc:q15109]
GI_19923305-S	-3,154618	2,01E-18	5,83E-16	8,905016	INMT	indolethylamine n-methyltransferase (ec 2.1.1.49) (aromatic alkylamine n-methyltransferase) (indolamine n-methyltransferase) (arylamine n- methyltransferase) (amine n-methyltransferase). [source:uniprot/swissprot;acc:o95050]
GI_6005923-S	-3,129144	1,56E-24	1,84E-21	8,749157	FAM107A	protein fam107a (down-regulated in renal cell carcinoma 1) (protein tu3a). [source:uniprot/swissprot;acc:o95990]
GI_7705960-S	-3,024849	1,39E-26	3,29E-23	8,138985	CLDN18	claudin-18. [source:uniprot/swissprot;acc:p56856]
GI_34147646-S	-3,007476	2,16E-21	1,28E-18	8,041564	FHL1	four and a half lim domains protein 1 (fhl-1) (skeletal muscle lim- protein 1) (slim 1) (slim). [source:uniprot/swissprot;acc:q13642]
Hs.187628-S	-2,929668	2,09E-27	6,18E-24	7,61935	BTNL9	butyrophilin-like protein 9 precursor. [source:uniprot/swissprot;acc:q6uxg8]
GI_11496886-S	-2,923917	1,80E-10	7,06E-09	7,58904	ADH1A	alcohol dehydrogenase 1a (ec 1.1.1.1) (alcohol dehydrogenase alpha subunit). [source:uniprot/swissprot;acc:p07327]

GI_21359975-S	-2,919592	5,78E-09	1,52E-07	7,566321	SCGB3A2	secretoglobin family 3a member 2 precursor (uteroglobin-related protein 1) (pneumo secretory protein 1) (pnsp-1). [source:uniprot/swissprot;acc:q96pl1]
GI_20149581-S	-2,871656	5,60E-14	5,19E-12	7,319048	MARCO	macrophage receptor marco (macrophage receptor with collagenous structure) (scavenger receptor class a member 2). [source:uniprot/swissprot;acc:q9uew3]
GI_4557726-S	-2,869183	8,65E-19	2,74E-16	7,306513	LPL	lipoprotein lipase precursor (ec 3.1.1.34) (lpl). [source:uniprot/swissprot;acc:p06858]
GI_31542938-S	-2,836943	4,21E-16	6,55E-14	7,145046	HPGD	15-hydroxyprostaglandin dehydrogenase [nad+] (ec 1.1.1.141) (pgdh) (prostaglandin dehydrogenase 1). [source:uniprot/swissprot;acc:p15428]
GI_14043068-S	-2,770403	5,91E-13	4,35E-11	6,822987	HBA2	hemoglobin subunit alpha (hemoglobin alpha chain) (alpha-globin). [source:uniprot/swissprot;acc:p69905]
GI_4507614-S	-2,751585	8,41E-24	8,40E-21	6,734564	TNNC1	troponin c, slow skeletal and cardiac muscles (tn-c). [source:uniprot/swissprot;acc:p63316]
GI_5803016-S	-2,732752	9,63E-11	4,02E-09	6,647224	N/A	n/a
hmm10338-S	-2,730711	4,31E-13	3,26E-11	6,637828	SUSD2	sushi domain-containing protein 2 precursor. [source:uniprot/swissprot;acc:q9ugt4]
GI_14456711-S	-2,727248	7,81E-13	5,53E-11	6,621914	HBA2	hemoglobin subunit alpha (hemoglobin alpha chain) (alpha-globin). [source:uniprot/swissprot;acc:p69905]
GI_4501848-S	-2,720477	2,62E-10	9,90E-09	6,590906	ABCA3	atp-binding cassette sub-family a member 3 (atp-binding cassette transporter 3) (atp-binding cassette 3) (abc-c transporter). [source:uniprot/swissprot;acc:q99758]
GI_4505756-S	-2,700621	2,07E-08	4,73E-07	6,500816	PGC	gastricsin precursor (ec 3.4.23.3) (pepsinogen c). [source:uniprot/swissprot;acc:p20142]
GI_33589822-S	-2,700408	6,10E-17	1,15E-14	6,499857	PDK4	[pyruvate dehydrogenase [lipoamide]] kinase isozyme 4, mitochondrial precursor (ec 2.7.11.2) (pyruvate dehydrogenase kinase isoform 4). [source:uniprot/swissprot;acc:q16654]
GI_21359850-S	-2,673893	5,53E-07	8,66E-06	6,381488	C4BPA	c4b-binding protein alpha chain precursor (c4bp) (proline-rich protein) (prp). [source:uniprot/swissprot;acc:p04003]
GI_4503756-S	-2,620741	1,17E-18	3,58E-16	6,15066	FMO2	dimethylaniline monooxygenase [n-oxide-forming] 2 (ec 1.14.13.8) (pulmonary flavin-containing monooxygenase 2) (fmo 2) (dimethylaniline oxidase 2) (fmo 1b1). [source:uniprot/swissprot;acc:q99518]

GI_38373687-A	-2,582989	5,18E-17	9,81E-15	5,991798	AQP4	aquaporin-4 (aqp-4) (wch4) (mercurial-insensitive water channel) (miwc). [source:uniprot/swissprot;acc:p55087]
GI_20665033-A	-2,563877	4,80E-17	9,19E-15	5,912947	CP21A_HUMAN	cytochrome p450 21 (ec 1.14.99.10) (cytochrome p450 xxi) (steroid 21- hydroxylase) (21-ohase) (p450-c21) (p-450c21) (p450-c21b). [source:uniprot/swissprot;acc:p08686]
GI_38455384-S	-2,563045	9,56E-14	8,42E-12	5,909536	LAMP3	lysosome-associated membrane glycoprotein 3 precursor (lamp-3) (lysosomal-associated membrane protein 3) (dc-lysosome-associated membrane glycoprotein) (dc lamp) (protein tsc403) (cd208 antigen). [source:uniprot/swissprot;acc:q9uqv4]
GI_12719454-A	-2,553403	4,14E-08	8,75E-07	5,870175	FOLR1	folate receptor alpha precursor (fr-alpha) (folate receptor 1) (folate receptor, adult) (adult folate-binding protein) (fbp) (ovarian tumor- associated antigen mov18) (kb cells fbp). [source:uniprot/swissprot;acc:p15328]
GI_45580687-S	-2,525787	1,32E-09	4,13E-08	5,758877	C7	complement component c7 precursor. [source:uniprot/swissprot;acc:p10643]
GI_4505244-S	-2,497818	4,12E-15	4,88E-13	5,648305	MRC1L1	mannose receptor, c type 1-like 1 [source:refseq_peptide;acc:np_001009567]

CURRICULUM VITAE

Tõnu Vooder

Citizenship: Estonian
Born: 04.10.1969
Address: Tamme puiestee 69a, Tartu, Estonia
Phone: +372 53226232
E-mail: tonu.vooder@ut.ee

Education

1977–1988 Tartu Secondary School No 5
1988–1994 Faculty of Medicine, University of Tartu

Special courses

04.10–06.10.2010: Battlefield Advanced Trauma Life Support (BATLS) Course
01.11–30.11.2006: Prince of Wales Hospital in Hong Kong: General thoracic surgery and VATS (Prof. A.P Yim)
08.10–16.10.2005: European School of Cardiothoracic Surgery in Bergamo, Italy C level
10.11–12.11.2004: European Surgical Institute in Hamburg: VATS practical training on animals (Dr.J.Brenner)
18.10–23.10.2004: European School of Cardiothoracic Surgery in Bergamo, Italy B level (EACTS scholarship).
02.09–02.10.2004: Prince of Wales Hospital in Hong Kong: General thoracic surgery and VATS (Prof. A.P Yim)
07.06–25.06.2004: Further training in general thoracic surgery at Düsseldorf University, Center of General Thoracic Surgery in Duisburg (Priv.dots. A.Granetzny).
31.03–05.04.2003: European School of Cardiothoracic Surgery in Bergamo, Italy A level
03.10–06.10.2002: Review course of cardiothoracic surgery Birmingham, UK
06.11–27.11.2000: Further training in general thoracic surgery in Clinic Heckeshorn, Berlin (Prof. Kaiser)
07.04–24.05.2000: Practical training of vascular surgery at Semmelweiss University of Budapest, Institute of Cardiovascular Surgery (Prof. Csaba Dzsinich).
15.10–20.12.1994: Laboratory Medicine in Leipzig (Dr.D.Lehmann, r.rer.nat. W.Ackermann)
21.02–31.03.1994: Practical training in traumatology at Göttingen University Hospital (Prof. Stankovitš)
01.07–30.07.1991: Practice of General surgery at University Hospital in Turkku, Finland (Prof. T. Havia)

Professional Employment

1994–1996	Tartu Maarjamõisa Hospital internship
1996–1997	Assistant doctor at Tartu Lung Clinic
1997–2002	Resident of surgery at surgical departments of Tartu University Hospital
2002–2010	General thoracic surgeon at Tartu University Hospital
2002–2011	PhD student at the Clinic of Cardiovascular and Thoracic Surgery, Lung Clinic and Institute of Molecular and Cell biology of Tartu University.

Scientific Work

Main scientific work focuses on molecular biology and treatment of Non-Small Cell Lung Cancer.

Eleven scientific publications.

I am a member of the Estonian Medical Association, the Estonian Society of Thoracic Surgery, the Estonian Society of Oncologists, the Society of German-Baltic Doctors, and the International Association for the Study of Lung Cancer.

ELULOOKIRJELDUS

Tõnu Vooder

Kodakondsus: Eesti
Sündinud: 04.10.1969
Aadress: Tamme puiestee 69a, Tartu
E-mail: tonu.vooder@ut.ee

Haridus

1977–1988 Tartu 5. Keskkool
1988–1994 Tartu Ülikooli arstiteaduskonna ravi osakond

Erialane täiendus

04.10–06.10.2010: Battlefield Advanced Trauma Life Support (BATLS) kursus
01.11–30.11.2006: Prince of Wales Hospital Hong Kongis: Üldtorakaalkirurgia ja VATS (Prof. A.P Yim)
08.10–16.10.2005: Euroopa Kardiotorakaalkirurgia kool Bergamos, Itaalias C tase
10.11–12.11.2004: European Surgical Institute Hamburgis: VATS praktiline treening loommudelitel (Dr.J.Brenner)
18.10–23.10.2004: Euroopa Kardiotorakaalkirurgia kool Bergamos, Itaalias B tase (EACTS-I stipendium)
02.09–02.10.2004: Prince of Wales Hospital Hong Kongis: Üldtorakaalkirurgia ja VATS-I täiendus (Prof. A.P Yim)
07.06–25.06.2004: Üldtorakaalkirurgia täiendus Düsseldorfi Ülikooli Duisburgi Torakaalkirurgia Keskuses (Priv.dots. A. Granetzny).
31.03–05.04.2003: Euroopa Kardiotorakaalkirurgia kool Bergamos, Itaalias A tase
03.10–06.10.2002: Kardiotorakaalkirurgia ülevaatekursus Birminghamis, UK
06.11–27.11.2000: Üldtorakaalkirurgia täiendus Heckeshorni haiglas, Berliinis (Prof. Kaiser)
07.04–24.05.2000: Veresoontekirurgia praktiline koolitus Budapestis Semmelweissi Ülikooli Kardiovaskulaarkirurgia instituudis (Prof. Csaba Dzsinich).
15.10–20.12.1994: Laborimeditsiini täiendus Leipzigis (Dr.D.Lehmann, Dr.rer.nat.Ackermann)
21.02–31.03.1994: Praktiline koolitus traumatoloogias Göttingeni Ülikooli traumatoloogia kliinikus (Prof. Stankovitš)
01.07–30.07.1991: Üldkirurgia praktika Turu Ülikooli Kliinikus Soomes (T. Havia)

Erialane teenistuskäik

1994–1996	Tartu Maarjamõisa Haigla intern
1996–1997	Nooremarst Tartu Kopsukliinikus
1997–2002	Kirurgia resident Tartu Ülikooli Kliinikumi kirurgia osakondades
2002–2010	Üldtorakaalkirurgia arst-õppejõud Tartu Ülikooli Kliinikumis
2002–2011	Doktorant TÜ Arstiteaduskonna Kardiovaskulaar ja torakaalkirurgia Kliinikus, Kopsukliinikus ja TÜ Molekulaar- ja Rakubioloogia Instituudis.

Teadustegevus

Peamiseks uurimisvaldkonnaks on mitteväikerakulise kopsuvähi molekulaarbioloogia ja kaasaegne ravi.

Kuulun järgmistesse erialastesse ja teadusseltsidesse: Eesti Arstide Liit, Eesti Torakaalkirurgia Selts, Eesti Onkoloogide Selts, Balti-Saksa Arstide Selts, Rahvusvaheline Kopsuvähiuurimise Assotsiatsioon.

Publikatsioonid

1. Võsa U, **Vooder T**, Kolde R, Välk K, Tõnisson N, Roosipuu R, Vilo J, Metspalu A, Annilo T. MicroRNA expression profiling in patients with early-stage non-small cell lung cancer identifies miR-374a as a prognostic marker for survival. Genes, Chromosomes and Cancer. In Press
2. **Vooder T**, Metspalu A. Investigating gene expression profile of non-small cell lung cancer. Central European Journal of Medicine. In Press
3. Luo C, Urgard E, **Vooder T**, Metspalu A. The Role of COX-2 and Nrf2/ARE in anti-inflammation and antioxidative stres: Aging and antiaging. Medical Hypothesis. In press
4. Urgard E, **Vooder T**, Võsa U, Välk K, Mingming Liu, Cheng Luo, Hoti F, Roosipuu R, Annilo T, Laine J, Frenz CM, Liqing Zhang, Metspalu A Metagenes associated with survival in NSCLC. Cancer Informatics, 2011. 10, 175–183.
5. McKay JD, Truong T, Gaborieau V, Chabrier A, Chuang SC, Byrnes G, Zaridze D, Shangina O, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, Fabianova E, Bucur A, Bencko V, Holcatova I, Janout V, Foretova L, Lagiou P, Trichopoulos D, Benhamou S, Bouchardy C, Ahrens W, Merletti F, Richiardi L, Talamini R, Barzan L, Kjaerheim K, Macfarlane GJ, Macfarlane TV, Simonato L, Canova C, Agudo A, Castellsagué X, Lowry R, Conway DI, McKinney PA, Healy CM, Toner ME, Znaor A, Curado MP, Koifman S, Menezes A, Wünsch-Filho V, Neto JE, Garrote LF, Boccia S, Cadoni G, Arzani D, Olshan AF, Weissler MC, Funkhouser WK, Luo J, Lubiński J, Trubicka J, Lener M, Oszutowska D, Schwartz SM, Chen C, Fish S, Doody DR, Muscat JE, Lazarus P, Gallagher CJ, Chang SC, Zhang ZF, Wei Q, Sturgis EM, Wang LE, Franceschi S, Herrero R, Kelsey KT, McClean MD, Marsit CJ, Nelson HH, Romkes M, Buch S, Nukui T, Zhong S, Lacko M, Manni JJ, Peters WH, Hung RJ, McLaughlin J, Vatten L, Njølstad I, Goodman GE, Field JK, Liloglou T, Vineis P, Clavel-Chapelon F, Palli D, Tumino R, Krogh V, Panico S, González CA, Quirós JR, Martínez C, Navarro C, Ardanaz E, Larrañaga N, Khaw KT, Key T, Bueno-de-Mesquita HB, Peeters PH, Trichopoulou A, Linseisen J, Boeing H, Hallmans G, Overvad K, Tjønneland A, Kumle M, Riboli E, Välk K, **Vooder T**, Metspalu A, Zelenika D, Boland A, Delepine M, Foglio M, Lechner D, Blanché H, Gut IG, Galan P, Heath S, Hashibe M, Hayes RB, Boffetta P, Lathrop M, Brennan P. A Genome-Wide Association Study of Upper Aerodigestive Tract Cancers Conducted within the INHANCE Consortium. PLoS Genet. 2011 Mar;7(3):e1001333.
6. Välk K, **Vooder T**, Kolde R, Reintam MA, Petzold C, Vilo J, et al. Gene Expression Profiles of Non-Small Cell Lung Cancer: Survival Prediction and New Biomarkers. Oncology.2010.79(3–4):283–92.

7. **Vooder T**, Valk K, Kolde R, Roosipuu R, Vilo J, Metspalu A. Gene Expression-Based Approaches in Differentiation of Metastases and Second Primary Tumour. Case Rep Oncol.2010. 3(2):255–61.
8. Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, Mirabello L, Jacobs K, Wheeler W, Yeager M, Bergen AW, Li Q, Consonni D, Pesatori AC, Wacholder S, Thun M, Diver R, Oken M, Virtamo J, Albanes D, Wang Z, Burdette L, Doheny KF, Pugh EW, Laurie C, Brennan P, Hung R, Gaborieau V, McKay JD, Lathrop M, McLaughlin J, Wang Y, Tsao MS, Spitz MR, Wang Y, Krokan H, Vatten L, Skorpen F, Arnesen E, Benhamou S, Bouchard C, Metsapalu A, **Vooder T**, Nelis M, Välk K, Field JK, Chen C, Goodman G, Sulem P, Thorleifsson G, Rafnar T, Eisen T, Sauter W, Rosenberger A, Bickeböller H, Risch A, Chang-Claude J, Wichmann HE, Stefansson K, Houlston R, Amos CI, Fraumeni JF Jr, Savage SA, Bertazzi PA, Tucker MA, Chanock S, Caporaso NE. A Genome-wide Association Study of Lung Cancer Identifies a Region of Chromosome 5p15 Associated with Risk for AdenocarcinomaAm J Hum Genet. 2009 Nov;85(5):679–91
9. Soh J, Okumura N, Lockwood WW, Yamamoto H, Shigematsu H, Zhang W, Chari R, Shames DS, Tang X, MacAulay C, Varella-Garcia M, **Vooder T**, Wistuba II, Lam S, Brekken R, Toyooka S, Minna JD, Lam WL, Gazdar AF. Oncogene mutations, copy number gains and mutant allele specific imbalance (MASI) frequently occur together in tumor cells.PLoS One. 2009 Oct 14;4(10):e7464
10. Laisaar T, **Vooder T**, Umbleja T.Thoracoscopic resection of a solitary pulmonary nodule in patients with a known history of malignancy. Thorac Cardiovasc Surg. 2008 Oct;56(7):418–21
11. Laisaar T, Palmiste V, **Vooder T**, Umbleja T. Life expectancy of patients with malignant pleural effusion treated with video-assisted thoracoscopic talc pleurodesis. Interact Cardiovasc Thorac Surg. 2006 Jun;5(3):307–10.

DISSERTATIONES MEDICINAE UNIVERSITATIS TARTUENSIS

1. **Heidi-Ingrid Maaroos.** The natural course of gastric ulcer in connection with chronic gastritis and *Helicobacter pylori*. Tartu, 1991.
2. **Mihkel Zilmer.** Na-pump in normal and tumorous brain tissues: Structural, functional and tumorigenesis aspects. Tartu, 1991.
3. **Eero Vasar.** Role of cholecystokinin receptors in the regulation of behaviour and in the action of haloperidol and diazepam. Tartu, 1992.
4. **Tiina Talvik.** Hypoxic-ischaemic brain damage in neonates (clinical, biochemical and brain computed tomographical investigation). Tartu, 1992.
5. **Ants Peetsalu.** Vagotomy in duodenal ulcer disease: A study of gastric acidity, serum pepsinogen I, gastric mucosal histology and *Helicobacter pylori*. Tartu, 1992.
6. **Marika Mikelsaar.** Evaluation of the gastrointestinal microbial ecosystem in health and disease. Tartu, 1992.
7. **Hele Everaus.** Immuno-hormonal interactions in chronic lymphocytic leukaemia and multiple myeloma. Tartu, 1993.
8. **Ruth Mikelsaar.** Etiological factors of diseases in genetically consulted children and newborn screening: dissertation for the commencement of the degree of doctor of medical sciences. Tartu, 1993.
9. **Agu Tamm.** On metabolic action of intestinal microflora: clinical aspects. Tartu, 1993.
10. **Katrin Gross.** Multiple sclerosis in South-Estonia (epidemiological and computed tomographical investigations). Tartu, 1993.
11. **Oivi Uibo.** Childhood coeliac disease in Estonia: occurrence, screening, diagnosis and clinical characterization. Tartu, 1994.
12. **Viiu Tuulik.** The functional disorders of central nervous system of chemistry workers. Tartu, 1994.
13. **Margus Viigimaa.** Primary haemostasis, antiaggregative and anticoagulant treatment of acute myocardial infarction. Tartu, 1994.
14. **Rein Kolk.** Atrial versus ventricular pacing in patients with sick sinus syndrome. Tartu, 1994.
15. **Toomas Podar.** Incidence of childhood onset type 1 diabetes mellitus in Estonia. Tartu, 1994.
16. **Kiira Subi.** The laboratory surveillance of the acute respiratory viral infections in Estonia. Tartu, 1995.
17. **Irja Lutsar.** Infections of the central nervous system in children (epidemiologic, diagnostic and therapeutic aspects, long term outcome). Tartu, 1995.
18. **Aavo Lang.** The role of dopamine, 5-hydroxytryptamine, sigma and NMDA receptors in the action of antipsychotic drugs. Tartu, 1995.
19. **Andrus Arak.** Factors influencing the survival of patients after radical surgery for gastric cancer. Tartu, 1996.
20. **Tõnis Karki.** Quantitative composition of the human lactoflora and method for its examination. Tartu, 1996.

21. **Reet Mändar.** Vaginal microflora during pregnancy and its transmission to newborn. Tartu, 1996.
22. **Triin Remmel.** Primary biliary cirrhosis in Estonia: epidemiology, clinical characterization and prognostication of the course of the disease. Tartu, 1996.
23. **Toomas Kivastik.** Mechanisms of drug addiction: focus on positive reinforcing properties of morphine. Tartu, 1996.
24. **Paavo Pokk.** Stress due to sleep deprivation: focus on $GABA_A$ receptor-chloride ionophore complex. Tartu, 1996.
25. **Kristina Allikmets.** Renin system activity in essential hypertension. Associations with atherothrombogenic cardiovascular risk factors and with the efficacy of calcium antagonist treatment. Tartu, 1996.
26. **Triin Parik.** Oxidative stress in essential hypertension: Associations with metabolic disturbances and the effects of calcium antagonist treatment. Tartu, 1996.
27. **Svetlana Päi.** Factors promoting heterogeneity of the course of rheumatoid arthritis. Tartu, 1997.
28. **Maarike Sallo.** Studies on habitual physical activity and aerobic fitness in 4 to 10 years old children. Tartu, 1997.
29. **Paul Naaber.** *Clostridium difficile* infection and intestinal microbial ecology. Tartu, 1997.
30. **Rein Pähkla.** Studies in pinoline pharmacology. Tartu, 1997.
31. **Andrus Juhan Voitk.** Outpatient laparoscopic cholecystectomy. Tartu, 1997.
32. **Joel Starkopf.** Oxidative stress and ischaemia-reperfusion of the heart. Tartu, 1997.
33. **Janika Kõrv.** Incidence, case-fatality and outcome of stroke. Tartu, 1998.
34. **Ülla Linnamägi.** Changes in local cerebral blood flow and lipid peroxidation following lead exposure in experiment. Tartu, 1998.
35. **Ave Minajeva.** Sarcoplasmic reticulum function: comparison of atrial and ventricular myocardium. Tartu, 1998.
36. **Oleg Milenin.** Reconstruction of cervical part of esophagus by revascularised ileal autografts in dogs. A new complex multistage method. Tartu, 1998.
37. **Sergei Pakriev.** Prevalence of depression, harmful use of alcohol and alcohol dependence among rural population in Udmurtia. Tartu, 1998.
38. **Allen Kaasik.** Thyroid hormone control over β-adrenergic signalling system in rat atria. Tartu, 1998.
39. **Vallo Matto.** Pharmacological studies on anxiogenic and antiaggressive properties of antidepressants. Tartu, 1998.
40. **Maire Vasar.** Allergic diseases and bronchial hyperreactivity in Estonian children in relation to environmental influences. Tartu, 1998.
41. **Kaja Julge.** Humoral immune responses to allergens in early childhood. Tartu, 1998.
42. **Heli Grünberg.** The cardiovascular risk of Estonian schoolchildren. A cross-sectional study of 9-, 12- and 15-year-old children. Tartu, 1998.

43. **Epp Sepp.** Formation of intestinal microbial ecosystem in children. Tartu, 1998.
44. **Mai Ots.** Characteristics of the progression of human and experimental glomerulopathies. Tartu, 1998.
45. **Tiina Ristimäe.** Heart rate variability in patients with coronary artery disease. Tartu, 1998.
46. **Leho Kõiv.** Reaction of the sympatho-adrenal and hypothalamo-pituitary-adrenocortical system in the acute stage of head injury. Tartu, 1998.
47. **Bela Adojaan.** Immune and genetic factors of childhood onset IDDM in Estonia. An epidemiological study. Tartu, 1999.
48. **Jakov Shlik.** Psychophysiological effects of cholecystokinin in humans. Tartu, 1999.
49. **Kai Kisand.** Autoantibodies against dehydrogenases of α-ketoacids. Tartu, 1999.
50. **Toomas Marandi.** Drug treatment of depression in Estonia. Tartu, 1999.
51. **Ants Kask.** Behavioural studies on neuropeptide Y. Tartu, 1999.
52. **Ello-Rahel Karelson.** Modulation of adenylate cyclase activity in the rat hippocampus by neuropeptide galanin and its chimeric analogs. Tartu, 1999.
53. **Tanel Laisaar.** Treatment of pleural empyema — special reference to intrapleural therapy with streptokinase and surgical treatment modalities. Tartu, 1999.
54. **Eve Pihl.** Cardiovascular risk factors in middle-aged former athletes. Tartu, 1999.
55. **Katrin Õunap.** Phenylketonuria in Estonia: incidence, newborn screening, diagnosis, clinical characterization and genotype/phenotype correlation. Tartu, 1999.
56. **Siiri Kõljalg.** *Acinetobacter* – an important nosocomial pathogen. Tartu, 1999.
57. **Helle Karro.** Reproductive health and pregnancy outcome in Estonia: association with different factors. Tartu, 1999.
58. **Heili Varendi.** Behavioral effects observed in human newborns during exposure to naturally occurring odors. Tartu, 1999.
59. **Anneli Beilmann.** Epidemiology of epilepsy in children and adolescents in Estonia. Prevalence, incidence, and clinical characteristics. Tartu, 1999.
60. **Vallo Volke.** Pharmacological and biochemical studies on nitric oxide in the regulation of behaviour. Tartu, 1999.
61. **Pilvi Ilves.** Hypoxic-ischaemic encephalopathy in asphyxiated term infants. A prospective clinical, biochemical, ultrasonographical study. Tartu, 1999.
62. **Anti Kalda.** Oxygen-glucose deprivation-induced neuronal death and its pharmacological prevention in cerebellar granule cells. Tartu, 1999.
63. **Eve-Irene Lepist.** Oral peptide prodrugs – studies on stability and absorption. Tartu, 2000.
64. **Jana Kivastik.** Lung function in Estonian schoolchildren: relationship with anthropometric indices and respiratory symptomas, reference values for dynamic spirometry. Tartu, 2000.

65. **Karin Kull.** Inflammatory bowel disease: an immunogenetic study. Tartu, 2000.
66. **Kaire Innos.** Epidemiological resources in Estonia: data sources, their quality and feasibility of cohort studies. Tartu, 2000.
67. **Tamara Vorobjova.** Immune response to *Helicobacter pylori* and its association with dynamics of chronic gastritis and epithelial cell turnover in antrum and corpus. Tartu, 2001.
68. **Ruth Kalda.** Structure and outcome of family practice quality in the changing health care system of Estonia. Tartu, 2001.
69. **Annika Krüüner.** *Mycobacterium tuberculosis* – spread and drug resistance in Estonia. Tartu, 2001.
70. **Marlit Veldi.** Obstructive Sleep Apnoea: Computerized Endopharyngeal Myotonometry of the Soft Palate and Lingual Musculature. Tartu, 2001.
71. **Anneli Uusküla.** Epidemiology of sexually transmitted diseases in Estonia in 1990–2000. Tartu, 2001.
72. **Ade Kallas.** Characterization of antibodies to coagulation factor VIII. Tartu, 2002.
73. **Heidi Annuk.** Selection of medicinal plants and intestinal lactobacilli as antimicrobil components for functional foods. Tartu, 2002.
74. **Aet Lukmann**. Early rehabilitation of patients with ischaemic heart disease after surgical revascularization of the myocardium: assessment of health-related quality of life, cardiopulmonary reserve and oxidative stress. A clinical study. Tartu, 2002.
75. **Maigi Eisen.** Pathogenesis of Contact Dermatitis: participation of Oxidative Stress. A clinical – biochemical study. Tartu, 2002.
76. **Piret Hussar.** Histology of the post-traumatic bone repair in rats. Elaboration and use of a new standardized experimental model – bicortical perforation of tibia compared to internal fracture and resection osteotomy. Tartu, 2002.
77. **Tõnu Rätsep.** Aneurysmal subarachnoid haemorrhage: Noninvasive monitoring of cerebral haemodynamics. Tartu, 2002.
78. **Marju Herodes.** Quality of life of people with epilepsy in Estonia. Tartu, 2003.
79. **Katre Maasalu.** Changes in bone quality due to age and genetic disorders and their clinical expressions in Estonia. Tartu, 2003.
80. **Toomas Sillakivi.** Perforated peptic ulcer in Estonia: epidemiology, risk factors and relations with *Helicobacter pylori*. Tartu, 2003.
81. **Leena Puksa.** Late responses in motor nerve conduction studies. F and A waves in normal subjects and patients with neuropathies. Tartu, 2003.
82. **Krista Lõivukene**. *Helicobacter pylori* in gastric microbial ecology and its antimicrobial susceptibility pattern. Tartu, 2003.
83. **Helgi Kolk.** Dyspepsia and *Helicobacter pylori* infection: the diagnostic value of symptoms, treatment and follow-up of patients referred for upper gastrointestinal endoscopy by family physicians. Tartu, 2003.

84. **Helena Soomer.** Validation of identification and age estimation methods in forensic odontology. Tartu, 2003.
85. **Kersti Oselin.** Studies on the human MDR1, MRP1, and MRP2 ABC transporters: functional relevance of the genetic polymorphisms in the *MDR1* and *MRP1* gene. Tartu, 2003.
86. **Jaan Soplepmann.** Peptic ulcer haemorrhage in Estonia: epidemiology, prognostic factors, treatment and outcome. Tartu, 2003.
87. **Margot Peetsalu.** Long-term follow-up after vagotomy in duodenal ulcer disease: recurrent ulcer, changes in the function, morphology and *Helicobacter pylori* colonisation of the gastric mucosa. Tartu, 2003.
88. **Kersti Klaamas.** Humoral immune response to *Helicobacter pylori* a study of host-dependent and microbial factors. Tartu, 2003.
89. **Pille Taba.** Epidemiology of Parkinson's disease in Tartu, Estonia. Prevalence, incidence, clinical characteristics, and pharmacoepidemiology. Tartu, 2003.
90. **Alar Veraksitš**. Characterization of behavioural and biochemical phenotype of cholecystokinin-2 receptor deficient mice: changes in the function of the dopamine and endopioidergic system. Tartu, 2003.
91. **Ingrid Kalev.** CC-chemokine receptor 5 (CCR5) gene polymorphism in Estonians and in patients with Type I and Type II diabetes mellitus. Tartu, 2003.
92. **Lumme Kadaja.** Molecular approach to the regulation of mitochondrial function in oxidative muscle cells. Tartu, 2003.
93. **Aive Liigant**. Epidemiology of primary central nervous system tumours in Estonia from 1986 to 1996. Clinical characteristics, incidence, survival and prognostic factors. Tartu, 2004.
94. **Andres, Kulla.** Molecular characteristics of mesenchymal stroma in human astrocytic gliomas. Tartu, 2004.
95. **Mari Järvelaid.** Health damaging risk behaviours in adolescence. Tartu, 2004.
96. **Ülle Pechter.** Progression prevention strategies in chronic renal failure and hypertension. An experimental and clinical study. Tartu, 2004.
97. **Gunnar Tasa.** Polymorphic glutathione S-transferases – biology and role in modifying genetic susceptibility to senile cataract and primary open angle glaucoma. Tartu, 2004.
98. **Tuuli Käämbre.** Intracellular energetic unit: structural and functional aspects. Tartu, 2004.
99. **Vitali Vassiljev.** Influence of nitric oxide syntase inhibitors on the effects of ethanol after acute and chronic ethanol administration and withdrawal. Tartu, 2004.
100. **Aune Rehema.** Assessment of nonhaem ferrous iron and glutathione redox ratio as markers of pathogeneticity of oxidative stress in different clinical groups. Tartu, 2004.
101. **Evelin Seppet.** Interaction of mitochondria and ATPases in oxidative muscle cells in normal and pathological conditions. Tartu, 2004.

102. **Eduard Maron.** Serotonin function in panic disorder: from clinical experiments to brain imaging and genetics. Tartu, 2004.
103. **Marje Oona.** *Helicobacter pylori* infection in children: epidemiological and therapeutic aspects. Tartu, 2004.
104. **Kersti Kokk.** Regulation of active and passive molecular transport in the testis. Tartu, 2005.
105. **Vladimir Järv.** Cross-sectional imaging for pretreatment evaluation and follow-up of pelvic malignant tumours. Tartu, 2005.
106. **Andre Õun.** Epidemiology of adult epilepsy in Tartu, Estonia. Incidence, prevalence and medical treatment. Tartu, 2005.
107. **Piibe Muda.** Homocysteine and hypertension: associations between homocysteine and essential hypertension in treated and untreated hypertensive patients with and without coronary artery disease. Tartu, 2005.
108. **Külli Kingo.** The interleukin-10 family cytokines gene polymorphisms in plaque psoriasis. Tartu, 2005.
109. **Mati Merila.** Anatomy and clinical relevance of the glenohumeral joint capsule and ligaments. Tartu, 2005.
110. **Epp Songisepp**. Evaluation of technological and functional properties of the new probiotic *Lactobacillus fermentum* ME-3. Tartu, 2005.
111. **Tiia Ainla.** Acute myocardial infarction in Estonia: clinical characteristics, management and outcome. Tartu, 2005.
112. **Andres Sell.** Determining the minimum local anaesthetic requirements for hip replacement surgery under spinal anaesthesia – a study employing a spinal catheter. Tartu, 2005.
113. **Tiia Tamme.** Epidemiology of odontogenic tumours in Estonia. Pathogenesis and clinical behaviour of ameloblastoma. Tartu, 2005.
114. **Triine Annus**. Allergy in Estonian schoolchildren: time trends and characteristics. Tartu, 2005.
115. **Tiia Voor.** Microorganisms in infancy and development of allergy: comparison of Estonian and Swedish children. Tartu, 2005.
116. **Priit Kasenõmm.** Indicators for tonsillectomy in adults with recurrent tonsillitis – clinical, microbiological and pathomorphological investigations. Tartu, 2005.
117. **Eva Zusinaite.** Hepatitis C virus: genotype identification and interactions between viral proteases. Tartu, 2005.
118. **Piret Kõll.** Oral lactoflora in chronic periodontitis and periodontal health. Tartu, 2006.
119. **Tiina Stelmach.** Epidemiology of cerebral palsy and unfavourable neurodevelopmental outcome in child population of Tartu city and county, Estonia Prevalence, clinical features and risk factors. Tartu, 2006.
120. **Katrin Pudersell.** Tropane alkaloid production and riboflavine excretion in the field and tissue cultures of henbane (*Hyoscyamus niger* L.). Tartu, 2006.

121. **Külli Jaako.** Studies on the role of neurogenesis in brain plasticity. Tartu, 2006.
122. **Aare Märtson.** Lower limb lengthening: experimental studies of bone regeneration and long-term clinical results. Tartu, 2006.
123. **Heli Tähepõld.** Patient consultation in family medicine. Tartu, 2006.
124. **Stanislav Liskmann.** Peri-implant disease: pathogenesis, diagnosis and treatment in view of both inflammation and oxidative stress profiling. Tartu, 2006.
125. **Ruth Rudissaar.** Neuropharmacology of atypical antipsychotics and an animal model of psychosis. Tartu, 2006.
126. **Helena Andreson.** Diversity of *Helicobacter pylori* genotypes in Estonian patients with chronic inflammatory gastric diseases. Tartu, 2006.
127. **Katrin Pruus.** Mechanism of action of antidepressants: aspects of serotoninergic system and its interaction with glutamate. Tartu, 2006.
128. **Priit Põder.** Clinical and experimental investigation: relationship of ischaemia/reperfusion injury with oxidative stress in abdominal aortic aneurysm repair and in extracranial brain artery endarterectomy and possibilities of protection against ischaemia using a glutathione analogue in a rat model of global brain ischaemia. Tartu, 2006.
129. **Marika Tammaru.** Patient-reported outcome measurement in rheumatoid arthritis. Tartu, 2006.
130. **Tiia Reimand.** Down syndrome in Estonia. Tartu, 2006.
131. **Diva Eensoo.** Risk-taking in traffic and Markers of Risk-Taking Behaviour in Schoolchildren and Car Drivers. Tartu, 2007.
132. **Riina Vibo.** The third stroke registry in Tartu, Estonia from 2001 to 2003: incidence, case-fatality, risk factors and long-term outcome. Tartu, 2007.
133. **Chris Pruunsild.** Juvenile idiopathic arthritis in children in Estonia. Tartu, 2007.
134. **Eve Õiglane-Šlik.** Angelman and Prader-Willi syndromes in Estonia. Tartu, 2007.
135. **Kadri Haller.** Antibodies to follicle stimulating hormone. Significance in female infertility. Tartu, 2007.
136. **Pille Ööpik.** Management of depression in family medicine. Tartu, 2007.
137. **Jaak Kals.** Endothelial function and arterial stiffness in patients with atherosclerosis and in healthy subjects. Tartu, 2007.
138. **Priit Kampus.** Impact of inflammation, oxidative stress and age on arterial stiffness and carotid artery intima-media thickness. Tartu, 2007.
139. **Margus Punab.** Male fertility and its risk factors in Estonia. Tartu, 2007.
140. **Alar Toom**. Heterotopic ossification after total hip arthroplasty: clinical and pathogenetic investigation. Tartu, 2007.
141. **Lea Pehme.** Epidemiology of tuberculosis in Estonia 1991–2003 with special regard to extrapulmonary tuberculosis and delay in diagnosis of pulmonary tuberculosis. Tartu, 2007.
142. **Juri Karjagin.** The pharmacokinetics of metronidazole and meropenem in septic shock. Tartu, 2007.

143. **Inga Talvik.** Inflicted traumatic brain injury shaken baby syndrome in Estonia – epidemiology and outcome. Tartu, 2007.
144. **Tarvo Rajasalu.** Autoimmune diabetes: an immunological study of type 1 diabetes in humans and in a model of experimental diabetes (in RIP-B7.1 mice). Tartu, 2007.
145. **Inga Karu.** Ischaemia-reperfusion injury of the heart during coronary surgery: a clinical study investigating the effect of hyperoxia. Tartu, 2007.
146. **Peeter Padrik.** Renal cell carcinoma: Changes in natural history and treatment of metastatic disease. Tartu, 2007.
147. **Neve Vendt**. Iron deficiency and iron deficiency anaemia in infants aged 9 to 12 months in Estonia. Tartu, 2008.
148. **Lenne-Triin Heidmets.** The effects of neurotoxins on brain plasticity: focus on neural Cell Adhesion Molecule. Tartu, 2008.
149. **Paul Korrovits.** Asymptomatic inflammatory prostatitis: prevalence, etiological factors, diagnostic tools. Tartu, 2008.
150. **Annika Reintam.** Gastrointestinal failure in intensive care patients. Tartu, 2008.
151. **Kristiina Roots.** Cationic regulation of Na-pump in the normal, Alzheimer's and CCK_2 receptor-deficient brain. Tartu, 2008.
152. **Helen Puusepp.** The genetic causes of mental retardation in Estonia: fragile X syndrome and creatine transporter defect. Tartu, 2009.
153. **Kristiina Rull.** Human chorionic gonadotropin beta genes and recurrent miscarriage: expression and variation study. Tartu, 2009.
154. **Margus Eimre.** Organization of energy transfer and feedback regulation in oxidative muscle cells. Tartu, 2009.
155. **Maire Link.** Transcription factors FoxP3 and AIRE: autoantibody associations. Tartu, 2009.
156. **Kai Haldre.** Sexual health and behaviour of young women in Estonia. Tartu, 2009.
157. **Kaur Liivak.** Classical form of congenital adrenal hyperplasia due to 21-hydroxylase deficiency in Estonia: incidence, genotype and phenotype with special attention to short-term growth and 24-hour blood pressure. Tartu, 2009.
158. **Kersti Ehrlich.** Antioxidative glutathione analogues (UPF peptides) – molecular design, structure-activity relationships and testing the protective properties. Tartu, 2009.
159. **Anneli Rätsep.** Type 2 diabetes care in family medicine. Tartu, 2009.
160. **Silver Türk.** Etiopathogenetic aspects of chronic prostatitis: role of mycoplasmas, coryneform bacteria and oxidative stress. Tartu, 2009.
161. **Kaire Heilman.** Risk markers for cardiovascular disease and low bone mineral density in children with type 1 diabetes. Tartu, 2009.
162. **Kristi Rüütel.** HIV-epidemic in Estonia: injecting drug use and quality of life of people living with HIV. Tartu, 2009.
163. **Triin Eller.** Immune markers in major depression and in antidepressive treatment. Tartu, 2009.

164. **Siim Suutre.** The role of TGF-β isoforms and osteoprogenitor cells in the pathogenesis of heterotopic ossification. An experimental and clinical study of hip arthroplasty. Tartu, 2010.
165. **Kai Kliiman.** Highly drug-resistant tuberculosis in Estonia: Risk factors and predictors of poor treatment outcome. Tartu, 2010.
166. **Inga Villa.** Cardiovascular health-related nutrition, physical activity and fitness in Estonia. Tartu, 2010.
167. **Tõnis Org.** Molecular function of the first PHD finger domain of Autoimmune Regulator protein. Tartu, 2010.
168. **Tuuli Metsvaht.** Optimal antibacterial therapy of neonates at risk of early onset sepsis. Tartu, 2010.
169. **Jaanus Kahu.** Kidney transplantation: Studies on donor risk factors and mycophenolate mofetil. Tartu, 2010.
170. **Koit Reimand.** Autoimmunity in reproductive failure: A study on associated autoantibodies and autoantigens. Tartu, 2010.
171. **Mart Kull.** Impact of vitamin D and hypolactasia on bone mineral density: a population based study in Estonia. Tartu, 2010.
172. **Rael Laugesaar.** Stroke in children – epidemiology and risk factors. Tartu, 2010.
173. **Mark Braschinsky.** Epidemiology and quality of life issues of hereditary spastic paraplegia in Estonia and implemention of genetic analysis in everyday neurologic practice. Tartu, 2010.
174. **Kadri Suija.** Major depression in family medicine: associated factors, recurrence and possible intervention. Tartu, 2010.
175. **Jarno Habicht.** Health care utilisation in Estonia: socioeconomic determinants and financial burden of out-of-pocket payments. Tartu, 2010.
176. **Kristi Abram.** The prevalence and risk factors of rosacea. Subjective disease perception of rosacea patients. Tartu, 2010.
177. **Malle Kuum.** Mitochondrial and endoplasmic reticulum cation fluxes: Novel roles in cellular physiology. Tartu, 2010.
178. **Rita Teek.** The genetic causes of early onset hearing loss in Estonian children. Tartu, 2010.
179. **Daisy Volmer.** The development of community pharmacy services in Estonia – public and professional perceptions 1993–2006. Tartu, 2010.
180. **Jelena Lissitsina.** Cytogenetic causes in male infertility. Tartu, 2011.
181. **Delia Lepik.** Comparison of gunshot injuries caused from Tokarev, Makarov and Glock 19 pistols at different firing distances. Tartu, 2011.
182. **Ene-Renate Pähkla**. Factors related to the efficiency of treatment of advanced periodontitis. Tartu, 2011.
183. **Maarja Krass.** L-Arginine pathways and antidepressant action. Tartu, 2011.
184. **Taavi Lai.** Population health measures to support evidence-based health policy in Estonia. Tartu, 2011.

185. **Tiit Salum.** Similarity and difference of temperature-dependence of the brain sodium pump in normal, different neuropathological, and aberrant conditions and its possible reasons. Tartu, 2011.