Faculty of Science and Technology

MASTER'S THESIS

Study program/ Specialization: Masters of Science in Biological Chemistry	**Spring semester, 2011** Open Access
Writer: Gurpartap Singh	………………………………………… (Writer's signature)
Faculty supervisor: Peter Ruoff **External supervisor(s):** Oddmund Nordgård	
Titel of thesis: Molecular Detection of Tumor Cells in Regional Lymph Nodes and Blood Samples from Patients Undergoing Surgery for Non-Small Cell Lung Cancer (NSCLC)	
Credits (ECTS): 60	
Key words: Non-Small Cell Lung Cancer, Lymph Node Metastases, Circulating Tumor Cells, RT-PCR,	**Pages: 63** + **Enclosure: 4** Stavanger, June 15/2011 Date/year

Molecular Detection of Tumor Cells in Regional Lymph Nodes and Blood Samples from Patients Undergoing Surgery for Non-Small Cell Lung Cancer (NSCLC)

Gurpartap Singh

Master's Thesis in Biological Chemistry

Faculty of Science and Technology

University of Stavanger

Autumn 2011

Acknowledgement

First of all I would like to convey my sincere gratitude to Almighty God for his grace in the completion of my project. I am indebted to the entire staff of Molecular Biology Lab at Stavanger University Hospital (SUS) for their kind help during this project. I owe my grateful acknowledgement to my teachers who made me confident enough to undertake and complete this project.

Thanks to University of Stavanger for accepting me as student for Masters of Science in biological Chemistry. Throughout my study period at UIS, I got familiar with the Norwegian Culture and Education System. I express my heartiest gratitude towards the entire faculty of Biological Chemistry at UIS for providing good education and knowledge. At UIS I got familiar with the research activities and research environment in Norway that is going to be a boon for my career.

I would like to convey my sincere gratitude to Oddmund Nordgård for accepting me as a project student in his research group. I will never forget his kindness, sincerity and punctuality that he showed towards me throughout this project. He had been a best supervisor to me and guided me wherever I needed help. He never ignored me or made delays for the availability of the things that I asked for. He always showed his cooperation to me while working in the lab. I will not forget to mention about Rune, Satu and Kjersti for helping me in the lab work.

This project would not have been a success without the collaboration of Brustgun and Helland. They were very much cooperative in providing the samples that we needed at various stages of the experiment phase and for providing us sufficient funds.

I wish to express special thanks to Prof. Peter Ruoff, Cathrine Lillo, Sigrun Reumann, and Lutz Eichacker for being good teachers to me.

Last but not the least, I express my deep sense of gratitude and indebtedness to my family for showing their moral and financial support throughout my tenure of study. Finally, I would like to give my heartiest apologies for all those helping hands names of whom I would have missed unintentionally.

GURPARTAP SINGH

Preface

This thesis is part of the course of Masters of Science in Biological Chemistry at University of Stavanger. This report is written in response to project done at Stavanger University Hospital under the supervision of Oddmund Nordgård. The project was a collaborated project between Oddmund Nordgård and Brustgun/Helland research group (Norwegian Radium Hospital, Oslo).

The project aims to detect molecular metastases in a group of patients undergoing surgery for Non-Small Cell Lung Cancer at Radium Hospital. All the patient samples tested in the project belongs to patients at Raium Hospital and were analyzed at Molecular Biology lab at Stavanger university Hospital (SUS). All the normal samples tested here were taken randomly at SUS.

A poster was presented about the findings of this study at AACR (American Association of Cancer Research) conference on April 2, 2011 and the work was highly admired by the visitors. This study will be published in few months as a research article.

Table of Contents

Contents **Page No.**

1. **INTRODUCTION** 8

1.1. Non-small cell Lung Cancer 8

1.1.1. Definition 8

1.1.2. Histopathology and Classification 10

1.1.3. Epidemiology 12

1.1.4. Detection and Treatment 12

1.2. Regional Lymph Node Metastases 13

1.2.1. Metastasis and Its impact 14

1.2.2. Detection Strategies 14

Routine analysis (H & E staining) 14

1.3. Circulating Tumor Cells in Blood 15

2. **Methods and Materials** 17

2.1. Methods 18

2.1.1. DNAse Treatment 18

2.1.2. Reverse Transcription 18

2.1.3. QPCR (Quantitative Polymerase Chain Reaction) 20

2.1.4. Relative Quantification of mRNA 23

2.1.5. Agarose gel electrophoresis of DNA 23

2.1.6. Purification of PCR Products 24

2.1.7. Sequencing of PCR Product 25

2.1.8. Cell Culturing 26

2.1.9. Resuscitation of Frozen Cells 27

2.1.10. Subculturing Cells 27
2.1.11. Cryopreservation of Cells 28
2.1.12. RNA isolation from cells 28
2.1.13. Measurement of nucleotide concentration using NanoDrop 29
2.2. Materials 30
3. **Results** 32
3.1. Selection of Candidate markers 33
3.1.1. Markers 34
3.1.2. Designing Primer Sets 35
3.1.3. Choosing the best primer set and Validation of Primers 35
3.1.4. Choosing a Calibrator 40
3.2. Marker validation in lung tumors, normal lymph nodes and normal blood 41
3.3. Marker validation in tumors, patient lymph nodes and patient blood 45
3.4. Comparison with Clinicopathological Data 49
4. **Discussion** 51
4.1. Relevance of RT-PCR and Molecular Markers 51
4.2. Lymph Node Metastases 52
4.3. Circulating Tumor Cells 55
4.4. Future Research 56
5. **Conclusion** 57
6. **Index of Figures and Tables** 58
7. **References** 60
Appendix 1 64
Appendix 2 66

Abstract

The spread of metastases to regional lymph nodes in patients undergoing surgery for non-small cell lung cancer (NSCLC) is routinely detected by histopathological examinations of tissue sections from retrieved regional lymph nodes. If the patient's lymph nodes do not show presence of metastases they are defined as LN- patients. In many cases the LN- patients get the disease back and die of it. This may be due to insensitivity of the routine investigation to detect micrometastases. The aim of the study was to find a group of patients where lymph node metastasis is not shown by routine investigation but they still have occult metastases in regional lymph nodes using more sensitive method that can be used on whole lymph nodes. Besides this the other aim was to detect circulating tumor cells (CTCs), the cells that are shed from primary tumor in to the circulation system, from blood.

The 5 out of 7 (CK19, SFTPA, SFTPB, SFTPC, EPCAM, CEACAM and PVA) mRNA molecules have been selected as potential markers for detection of micrometastases by RT-PCR. The markers were selected from scientific research articles and cancer databases followed by their optimization and validation to ensure pure PCR product. A calibrator cDNA was included in each plate to compare the results from different runs. Certain cell lines were cultured to select the calibrator expressing all marker mRNAs included in the study. The verification of the PCR product was done by agarose gel electrophoresis and DNA sequencing. In the initial phase 7 makers were quantified in 16 tumors, 16 normal lymph nodes and 12 blood samples to select best candidate markers. Some of these markers (CK19, CEACAM and PVA) have already been quantified in many previous studies but SFTPA and SFTPC are new markers and quantification method has been established for these markers in this study. The markers selected in initial phase were quantified in tumors, lymph nodes and blood samples for a cohort of 55 patients operated for non-small cell lung cancer (NSCLC) at Norwegian Radium Hospital. The relative quantification of each marker was determined using $2^{\Delta\Delta Ct}$ method.

1. INTRODUCTION

Lung cancer is one of the most common cancers we have and a large number of people die of this disease every year. The disease is often discovered in a late stage, but also in earlier stages lung cancer patients have worse outcome than patients with other cancers. Even without spreading to other organs, during stage I, the survival rate of lung cancer is under 70%. In comparison, for example, breast cancer there is 95% survival in stage I [1].

First location of spread in most of the lung cancer patients is lymph nodes. Pre-operative assessment of spread to lymph nodes is important for consideration of further treatment. Lymph nodes will be removed during surgery to be analyzed using light microscopy for review. If there are areas of tumor cells in lymph nodes, the patient is offered additional chemotherapy.

In most cases the spread of cancer to lymph nodes remain unnoticed due to insensitivity of the clinical methods and the patients have low survival rate even after the diagnosis. More advanced and sensitive methods are needed to help this out.

1.1. Non-Small Cell Lung Cancer

1.1.1. Definition

Lung Cancer, like all other cancers, results from an acquired abnormality in the body's basic unit of life, the cell. Normally, the body maintains a system of control mechanisms for cell growth, so that cells divide to produce new cells only when new cells are needed. Disruption of this control system results in an uncontrolled division and proliferation of cells that eventually forms a mass known as a tumor. Although it can arise in any part of the lung; 90%-95% of the lung cancers arise from epithelial cells, bronchi and bronchioles [2]. Sometimes it may also arise from the other supporting tissues within the lungs like blood vessels. Lung cancers arise through a multistep process involving many genetic and epigenetic changes that includes damage of many key cell-cycle genes [3]. The alterations may accumulate in bronchial epithelium leading to clonal cell expansion. In some cancer patients clonal cells does not exist only as malignant cells but also as histologically normal appearing areas adjacent to tumors [4].

Some cancer patients demonstrate chromosomal abnormalities that damage tumor suppressor genes or have mutations in oncogenes [5]. The mutations are common at chromosome regions 3p (that includes the FHIT, a tumor suppressor gene mutated in over 70% lung cancers), 9p (that includes p16INK4a, p19ARF genes, which are involved in the RB signaling pathway), 13q (RB) and 17p (TP53) [6].

Benign and Malignant Tumors: Tumors can be classified into Benign and Malignant.

Benign Tumors are those tumors that are not cancerous and are often localized without spreading to other parts of body or nearby tissues [7]. These tumors grow slowly and are less harmful. They can be easily removed and patients have less chances of getting it back.

Table 1: Classification of Benign Lung Tumors [9]

Type	**Example**
Laryngotracheobronchial	Adenoma Hamartoma
Parenchymal	Fibroma Hamartoma Leiomyoma

Malignant Tumors refer to cancer where cells grow aggressively and invade other tissues and organs of the body nearby and damage them. The cells of malignant tumors can pass into the bloodstream or lymphatic system causing the spread of a tumor. This process of spreading is termed metastasis.

Malignant tumors of the lung can be classified as follows:

Table 2: Classification of Primary Malignant Lung Tumors

Type	**Example**
Carcinoma	
Small Cell	Oat Cell Intermediate Cell Combined
Non Small Cell	Adenocarcinoma Acinar Bronchioloalveolar Papillary Adenosquamous

	Large Cell Squamous Cell Spindle Cell

1.1.2. Histopathology and Classification

Lung cancers are broadly classified into two types:

- Small cell lung cancers (SCLC)
- Non-small cell lung cancers (NSCLC)

Small cell lung cancer (SCLC): It is one of the most aggressive and rapidly growing lung cancers comprising 20% of all lung cancers [2]. This type of cancer is strongly related to cigarette smoking. SCLC often metastasizes rapidly to many sites and is discovered during late stages. These cancers have a specific cell appearance under the microscope, the cells being smaller than the cells of Non-Small Cell lung Cancer [10]. SCLC often remains central to the lung and grows along the wall of large bronchus [13]. The cells multiply quickly and form large tumors that spread throughout the body.

Non-small cell lung cancer (NSCLC): It is the most common type of lung cancers and accounts for about 80% of all lung cancers. NSCLC can be divided into three main types:

- **Adenocarcinomas:** This is found in the gland of the lung that produces mucous and is the most common type of NSCLC seen in Women and non smokers [10]. Adenocarcinomas comprise up to 50 % of Non Small Cell Lung cancers and it arises in the outer, or peripheral, areas of the lung. A subtype of it is Bronchioloalveolar Carcinoma that develops frequently at multiple sites in the lungs and spreads along the preexisting alveolar walls [2]. Sometimes adenocarcinomas arise around a scar tissue and are associated with asbestos exposure [11].

- **Squamous Cell Carcinomas:** These are also known as epidermoid carcinomas and accounts for about 30-40% of primary lung tumors. This type of cancer grows commonly in the central areas around major bronchi in a stratified or pseudoductal arrangement. The cells have an epithelial pearl formation with individual cell keratinization [11].

- **Large Cell Carcinomas:** The tumor cells are large and show no other specific morphological traits. Sometimes they are referred to as undifferentiated carcinomas, and they are the least common type of Non Small Cell Lung Cancer.

The prognosis and treatment options depend on how widespread the disease is when diagnosed. The TNM classification system is used to subgroup the patients according to the extent of the disease. The method classify patients based on the size of primary tumor (T), degree of spread to lymph nodes (N) or distant spread at the time of surgery (M). TNM classification is crucial for further treatment options and must be present before treatment is initiated.

T-stage

This stage considers mainly the size of the primary tumor. From TX (positive cytology, but unknown tumor) and T0 (not detected primary tumor) to T3 (tumor> 7 cm) and T4 (tumor invading surrounding organ areas).

N-stage

- NX - Regional lymph nodes cannot be assessed
- N0 - no lymph node metastases
- N1 – The cancer has spread to lymph nodes within the lung and/or around the area where the bronchus enters the lung (hilar lymph nodes) [12]. Metastasis to lymph nodes is on the same side as the primary tumor.
- N2 – The cancer has spread to lymph nodes around the carina (point where the windpipe splits into the left and right bronchi) or in the space behind the breast bone and in front of the heart (mediastinum) [12]. Metastasis to mediastinal nodes is on the same side as the primary tumor.
- N3 - Metastasis to nodes on the opposite side of the lungs.

M-stage

- MX - distant spread cannot be assessed
- M0 - no distant metastases
- M1 - distant metastases
- M1a - distant spread to the lung on the opposite side of the main tumor
- M1b - distant metastases.

1.1.3. Epidemiology

Lung Cancer is predominantly a disease of the elder persons. Nearly 70% of people diagnosed with lung cancer are above 65 years of age and less than 3% of lung cancers occur in the people below age of 45 years [2]. The incidence of lung cancer is strongly correlated with cigarette smoking, with 90 % of lung cancers arising because of tobacco use.

The five year survival rate is up to 65% among the patients of NSCLC where the disease is detected in early stages but the long term survival rate is 1% for those having metastases [13]. The prognosis for adenocarcinoma is poorer than for squamous cell carcinoma, whereas for large cell carcinoma it is poorest.

Statistics in Norway:

Lung cancer accounted for 10.2% of the annual number of new cancer cases in Norway for men and 9.1%for women in 2008 [14]. In 2008, 2,529 Norwegians got diagnosis of lung cancer, including 1422 men and 1107 women. In 2007, 2100 patients died of lung cancer. The incidence of lung cancer in men is declining, while for women, the number is doubled in the last 15 years. Today 22 percent of the Norwegian population smoke and the proportions are roughly equal for both sexes. Research shows that 80-90 percent of lung cancers are related to smoking.

1.1.4. Detection and Treatment

There are several symptoms connected with the presence of Lung Cancer that vary depending upon where and how widespread the tumor is:

Symptoms related to the primary tumor: The growth and invasion of cancer in the lung tissues and other surrounding areas may interfere with breathing that leads to some symptoms such as cough, shortness of breath, wheezing, chest pain and coughing up blood (hemoptysis). In case the cancer has invaded nerves it may cause shoulder pain that travels down the arm (Pan Coast's syndrome) o paralysis of the vocal cords leading to hoarseness. If it invades to esophagus it may cause difficulty in swallowing (dysphasia). Sometimes obstruction of large may occur leading to collapse of a portion of lung and it may cause pneumonia in the obstructed area [2].

Symptoms related to metastasis: if the lung cancer has spread to bones it may cause excruciating pain in the bones. In case of spreading to bones it causes number of neurologic symptoms that may include blurred vision, headaches, seizures, or, symptoms of stroke such as weakness or loss of sensation in parts of the body.

Paraneoplastic symptoms: Most frequently lung cancers are accompanied by symptoms that result from production of hormone-like substances by the tumor cells. The paraneoplastic syndromes occur most commonly with SCLC but in some cases it may also be seen with some other type of tumor. A release of parathyroid hormone like substance is the most frequent paraneoplastic syndrome seen with NSCLC. The release of this substance leads to elevated levels of calcium in the blood stream [2].

Nonspecific symptoms: some non specific symptoms may be seen in lung cancer such as weight loss, weakness, and fatigue. Other psychological symptoms like depression and mood changes are also common.

No symptoms: in around 25 % of patients with lung cancer the disease is discovered on a routine chest X-ray or CT scan as a solitary mass (coin lesion). Some of these patients with small, single masses show up no symptoms at the time cancer is discovered.

There are several ways to treat lung cancer. The treatment depends on the type of lung cancer and how far it has spread. Treatments include surgery, chemotherapy, and radiation. People with lung cancer often get more than one kind of the following treatments.

- **Surgery:** Cancer tissues are removed by resection.
- **Chemotherapy:** This type of treatment involves the use of drugs to shrink or kill the cancer. The drugs could be tablets or medicines given through an IV (intravenous) tube.
- **Radiation:** Radiation uses high-energy rays (similar to X-rays) to try to kill the cancer cells. The rays are aimed at the part of the body where the cancer is.

Without treatment, lung cancer, almost without exception, eventually spread to other organs in the body, either via the lymphatic system (the system that produces, stores, and carries the cells that fight infections) or blood vessels. Spread to regional lymph nodes is the first and most common distribution system. Hematogen spread to other organs also occurs frequently, the most common localization is the other lung, bone, liver, adrenal glands and brain. Primary tumor can also spread by direct tumor growth into adjacent organs.

1.2. Regional Lymph Node Metastases

Metastasis means the spread of cancer. Cancer cells can break away from a primary tumor and enter the bloodstream or lymphatic system and spread to other parts of the body.

When cancer cells spread and form a new tumor in a different organ, the new tumor is called a metastatic tumor. The cells in the metastatic tumor come from the original tumor. This means, for example, that if breast cancer spreads to the lungs, the metastatic tumor in the lung is made up of cancerous breast cells (not lung cells). In this case, the disease in the lungs is metastatic breast cancer

(not lung cancer). Under a microscope, metastatic breast cancer cells generally look the same as the cancer cells in the breast.

1.2.1. Metastasis and Its impact

Regional lymph node involvement in lung cancer is heterogeneous. From micrometastases in intrapulmonary lymph nodes, coded as N1 (mi), to bulky contralateral nodal disease, coded as N3, the different situations in between vary in anatomic extent and prognosis. However, regardless of the amount of tumor burden in the regional lymph nodes, the present nodal staging of the tumor, node and metastases (TNM), described above, defines the extent of nodal involvement solely via anatomic location.

Several studies have found that within every N category, there are prognostic modifiers. Thus, for pathologically staged I tumors, the number of removed lymph nodes at thoracotomy seems to have prognostic impact [16]. The involvement of hilar (main bronchi) lymph nodes is closely associated with worse prognosis as compared to intrapulmonary lymph nodes in N1 patients. Other indicators of worse prognosis in these patients include macroscopic nodal involvement, involvement of multiple nodal and multiple nodal stations and metastatic involvement. Some other factors of adverse prognosis include multilevel N2 disease and bulky disease, involvement of highest mediastinal lymph node and an extra nodal extension.

1.2.2. Detection Strategies

Detection and Isolation of Lymph Nodes: Lymph nodes are collected by the pathologists in the resection specimen by palpation and visual inspection. Following their detection lymph node biopsy is used to remove lymph nodes. The specimens are then formalin fixed and paraffin embedded. In this process lymph nodes are cut during the surgery of patients having lung cancer. The isolated lymph nodes are further analyzed for the presence or absence of tumor cells.

Detection of metastases in Lymph Nodes: The sections of isolated lymph nodes are analyzed using various methods to follow the spread of the disease. Some of these methods are described below:

Routine analysis (H & E staining)

Hematoxylin-eosin staining (H & E staining) is based on staining of tissue sections by hematoxylin and eosin dyes, making it easier to distinguish different types of cells visually. The method is not sensitive enough to detect small metastases but it can distinguish cancer cells from healthy cells by looking at the cells' shape, size and growth manner. Very small occurrences of cancer cells or single cells may be difficult to detect.

Immunohistochemistry

Metastases in regional lymph nodes can also be detected using sensitive immunohistochemical (IHC) methods. The method is based on the binding of antibodies to epithelium-specific proteins on or inside the cancer cell. The cells are then stained and visually separated under microscope. Cancer cells are similar to the epithelial cells because they were originally epithelial cells, and these are not normally present in the lymph nodes. Previous experiments have shown that this method is more sensitive than routine analysis based on H & E staining, since the color gives a greater contrast that makes the cells be better separated from each other than with H & E staining. However, immunohistochemical methods are difficult; both labour and time requiring observation by a skilled worker for a reasonable objective evaluation. It is however a major advantage that IHC provides an opportunity to control tumor cell morphology visually.

Reverse-transcription PCR

Reverse transcription polymerase chain reaction (RT-PCR)-based detection of epithelium-specific mRNA is a fast and simple method for detecting metastases. The principle behind this method is that small amounts of cancer cells can be detected in clinical samples by amplifying specific mRNA that is expressed selectively in cancer cells, but not in the normal cells. The specific mRNA molecules that are amplified using RT-PCR are known as markers (see method section).

The use of molecular markers in combination with RT-PCR defines a sensitive technology to detect even a small amount of metastases in regional lymph nodes and blood.

1.3. Circulating Tumor Cells in Blood

Circulating tumor cells (CTCs) are the cancer cells that escape into circulator system from the primary tumor. While metastases directly lead to cancer recurrence, CTCs constitute a seed for metastases indicating the metastatic potential of the disease [17].

Analyses of CTCs allow earlier detection of metastasis at an early stage. The molecular characterization of CTCs may enable the treatment to be more effective and the removal of CTCs from circulation minimizes the potential of metastases after surgery or therapy. Some previous studies have shown presence of CTCs in peripheral blood and bone marrow has a prognostic significance from many tumors [17]. The measurement of burden of CTCs can be used to monitor treatment response or relapse.

Molecular characterization of CTCs is helpful for the development of new methodologies for CTC detection, enumeration, isolation and genetic analysis. The CTCs can be detected by the use of either indirect or direct methods [31].

In indirect Methods for detection of CTCs tumor specific genes circulating the blood were detected after amplification by polymerase chain reaction (PCR). These methods have an advantage of being highly sensitive but sometimes they may have less specificity and reproducibility.

The direct methods are more advantageous than the direct methods: they provide morphological confirmation of CTCs, allow quantitative analysis and isolation of CTCs for further analysis. Some of the direct methods include size-selection methods, flow-cytometric methods and positive or negative Immunomagnetic selection methods. In some cases these methods are also used in combination [31].

In both of the detection methods there is involvement of further two steps: an enrichment step and a detection step. The enrichment of CTCs can be performed by selecting the cells using tumor specific markers (immunoseparation) or on the basis of morphological traits such as cell size or density. As tumors display heterogeneity in cell size, density and marker expression, specificity and sensitivity is a big issue with both techniques of enrichment.

Enrichment Methods

Density Gradient Separation: in this method CTCs and mononuclear cells are separated from blood cells and granulocytes on the basis of their density. CTCs have higher density (density <1.077 g/ml) than blood cells [19].

Immunomagnetic Separation: the method relies on the positive selection of CTCs from blood samples through their binding of antibodies coupled to magnetic beads targeting epithelial-specific antigens or tumor specific cell surface antigens [19].

2. Methods and Materials

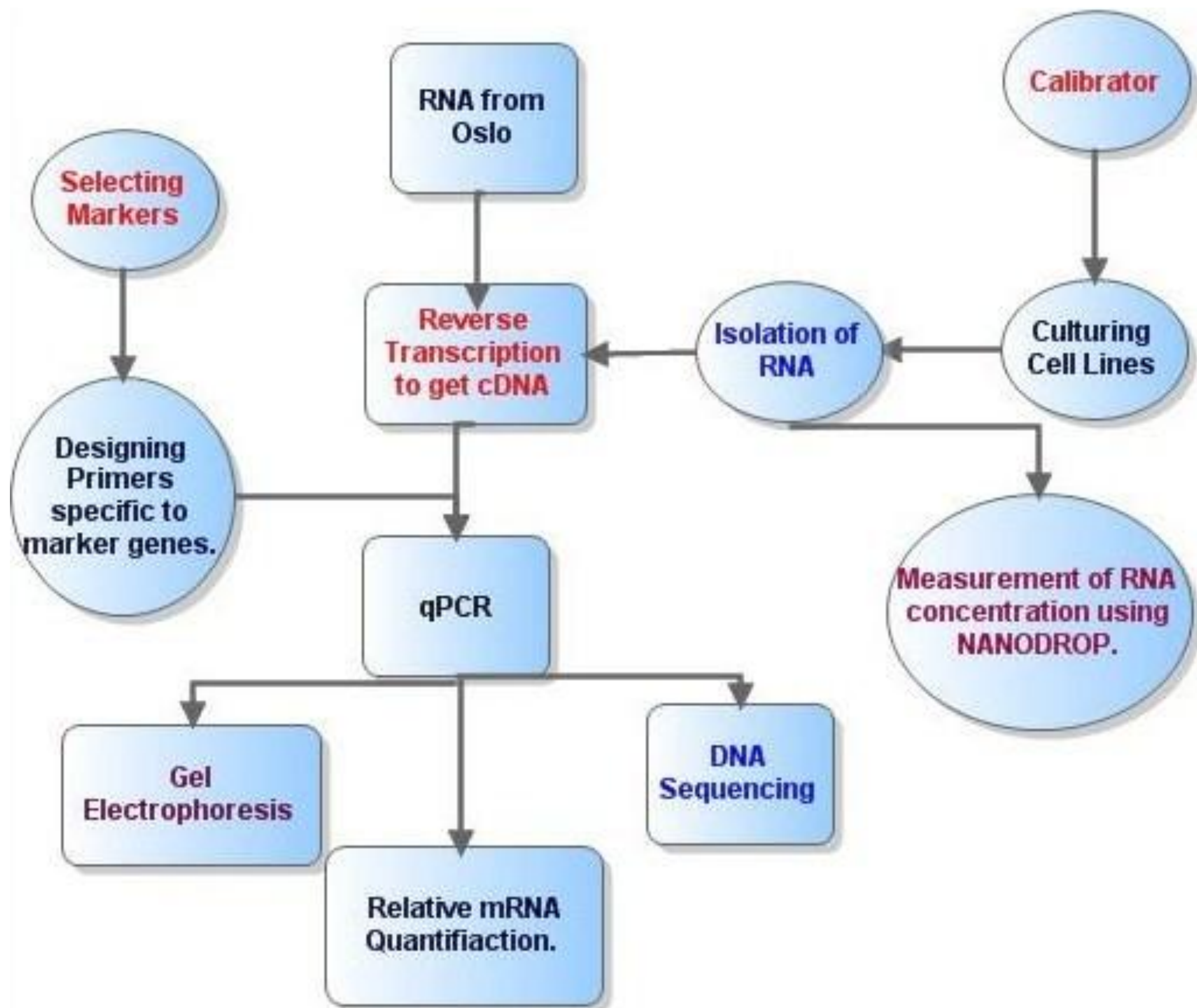

Figure 1 A Flow chart displaying the main steps of methodology. It describes the step wise illustration of all the methods performed at various stages during the project.

2.1. Methods

2.1.1. DNAse Treatment

In this project we have treated RNA with DNAse before reverse transcription; reason being to make RNA free of DNA that could be present. It is often that DNA can attach to RNA during isolation procedure from patients. To ensure that only cDNA is amplified during qPCR and not DNA, RQ1DNAse is used to degrade DNA.

RQ1DNAse is an endonuclease that degrades both the double stranded DNA and form 3'OH oligonucleotide. It is an RNAse free enzyme and is ideal to use when it is critical to preserve RNA in the sample. We also add RNAse inhibitor to avoid RNA degradation during the reaction.

Protocol:

1. Thaw RNA samples on crushed ice.
2. Make the master mix for all the proposed reactions[+1] by mixing the following per reaction:
   - 2µl 5x FSS (First Strand Synthesis) buffer
   - 1µl RQ1 DNAse
   - 0.25µl RNAse RNAseOUT inhibitor
3. Add 3µl of the master mix in 1.5ml RNAse free tubes. To this add 500ng RNA and up to7µl of RNAse free water (Total volume in the tube should be 10µl). Mix by pipetting up and down several times.
4. Incubate tubes at 37°C for 30 minutes using water bath.
5. Add 1.0µl RQ1 stop solution. Mix well and spin down.
6. Incubate at 65°C for 10 minutes. This is done to inactivate the enzyme. Finally spin down the tubes.

2.1.2. Reverse Transcription

After DNAse treatment mRNA samples are reverse transcribed. In this process mRNA is used as a template to get a single stranded complementary DNA (cDNA). This reaction is catalyzed by Reverse Transcriptase. The cDNA molecules generated in process can be directly used as a template for further amplification using PCR. Mostly the two processes are used in combination and the method is referred to as RT-PCR.

Reverse transcription followed by the polymerase chain reaction (PCR) is the technique of choice to analyze mRNA expression derived from various sources. In this project we have used RT-qPCR (Quantitative reverse transcription PCR). RT-qPCR is highly sensitive and allows quantification of rare transcripts and small changes in gene expression.

Protocol:

1. Make the master mix for all the proposed reactions [+1] by mixing the following per reaction:
    - 0.2 µl 1µg /µl random primer
    - 0.4 µl 25 mM dNTP
    - 0.4 µl DEPC H2O
2. Mix well and add 1.0 µl to each tube containing DNAse treated mRNA.
3. Incubate at 65°C for 5 minutes and immediately put on ice for minimum 2 minutes.
4. Make another master mix by mixing the following per reaction:
    - 2 µl 5X buffer FSS
    - 2 µl 0.1 M DTT
    - 1 µl RNAse OUT RNAse inhibitor
    - 2 µl DEPC H2O
5. Mix well and add 7 µl to each tube containing mRNA.
6. Incubate tubes at 37°C water bath for 2 minutes.
7. Add 1 µl of MMLV Reverse transcriptase (RT) to each tube except one that can be considered as –RT sample.
8. Keep the tubes at room temperature for 10 minutes.
9. Transfer the tubes to 37°C water bath and keep them for 1 hour.
10. Shift the tubes to 65°C and keep them for 15 minutes.
11. Add 30 µl of dH_2O to all tubes in order to give a final concentration of reverse transcribed RNA 10 ng/µl.
12. Freeze the tubes at -20°C.

2.1.3. QPCR (Quantitative Polymerase Chain Reaction)

PCR technology is widely used in quantifying DNA because the amplification of the target sequence allows for greater sensitivity of detection. In an optimized reaction, the target quantity will approximately double during each amplification cycle. In quantitative PCR (QPCR), the amount of amplified product is linked to fluorescence intensity using a fluorescent signal that is measured in order to calculate the initial template quantity at the end of the reaction (endpoint QPCR) or while the amplification is still progressing (real-time QPCR).

Real-time PCR is a quantitative method involving the use of fluorescent molecules that binds to double stranded DNA. SYBR Green I (used in this project) is a molecule that binds to all double stranded DNA. When lying freely in a solution the fluorescent molecule emits weak fluorescence, but when it binds to DNA fluorescence increases 1000 times. In each cycle of PCR reaction there occur increase in the amount of DNA and thus the intensity of fluorescence also, that is measured after each cycle. In the first cycle of real-time PCR the fluorescence generated from the amplification is higher than the background signals, called "Ct" or threshold cycle. This Ct value can be correlated directly to the initial concentration of DNA sample. The higher is the DNA concentration, the lower the Ct value.

PCR reaction performs by starting at a lower temperature followed by a gradual increase in the temperature. As the temperature increases, the double stranded DNA molecule denatures. Longer DNA molecules will denature at higher temperatures than shorter fragments. DNA fragments denatures by releasing SYBR Green 1 molecules, resulting in a sudden decrease in fluorescence, which is registered as a melting point. If you have more than one melting point of a PCR reaction, it has probably more than one type of PCR product.

Protocol:

1. PCR was performed using SYBR Green Core kit from Eurogentec.
2. Make a master mix for the desired number of reactions [+1], for each marker gene that will be amplified. Reagents are listed in Table 3, which shows the amount used per reaction. SYBR Green 1 was diluted 1:200 in DMSO solution provided with the kit.

Table 3: Concentration of reaction mixtures for the markers

REAGENTS	BCR	CK19	SFTPA	SFTPB	SFTPC	EPCAM	CEA	PVA
10 x PCR buffer	2.5 µl	2.5 µl	2.5 µl	2.5 µl	2.5 µl	2.5 µl	2.5 µl	2.5 µl
50 mM MgCl2	1.0 µl	1.0 µl	0.625 µl	1.0 µl	0.625 µl	0.625 µl	0.625 µl	0.875 µl
5 mM dNTP-U mix	1.0 µl	1.0 µl	1.0 µl	1.0 µl	1.0 µl	1.0 µl	1.0 µl	1.0 µl
10µM F- primer	0.75 µl	0.75 µl	0.75 µl	0.75 µl	0.75 µl	0.75 µl	0.5 µl	0.75 µl
10µM R- primer	0.75 µl	0.75 µl	0.75 µl	0.75 µl	0.75 µl	0.75 µl	0.5 µl	0.75 µl
1:200 SYBR Green1	0.75 µl	0.75 µl	0.75 µl	0.75 µl	0.75 µl	0.75 µl	0.75 µl	0.75 µl
Hot Gold Star PCR enzyme	0.125 µl	0.125 µl	0.125 µl	0.125 µl	0.125 µl	0.125 µl	0.125 µl	0.125 µl
dH2O	16.125 µl	16.125 µl	16.500 µl	16.125 µl	16.500 µl	16.500 µl	17.000 µl	16.250 µl
Total volume	23 µl	23 µl	23 µl	23 µl	23 µl	23 µl	23 µl	23 µl

3. Master Mix is mixed well and spun down shortly. Distribute master mix into desired number of wells in a 96 well PCR plate, pipetting 23 µl in each.

Table 4: Validated Primer Sets For each Marker

Primer Set	Forward Primer	Reverse Primer
SFTPA-F	5'-ttggaggcagagacccaagcag-3'	5’-ggctccaagaaatcagcgaccc-3’
SFTPB-E	5'-gtccagccctctccagtgtatc-3'	5'-gcccgtctcacttggcttttc-3'
EPCAM-B	5'-cgcagctcaggaagaatgtg-3'	5'-tgaagtacactggcattgacg-3'
CEACAM5-C	5’-gggacctatgcctgttttgtctc-3’	5’-gagcaaccccaaccagcac-3’
PVA-B	5'-ggcaaaaacgtgaatgggtga-3'	5'-gggttgcttggtaatctgaagta-3'

4. Thaw cDNA templates on ice in a separate lab. Move the PCR plate to template lab. Vortex the tubes containing cDNA shortly and spin them down.

5. Add 2 µl of 10 ng/µl reverse transcribed RNA template into two wells for each marker gene. Also add 2 µl of –RT sample into the wells described for NO RT sample. Add 2 µl of dH2O in the wells labeled for NTC. An example plate set up is shown in figure 2.

All	1	2	3	4	5	6	7	8	9	10	11	12
	MARKER 1		MARKER 2		MARKER 3		MARKER 4		MARKER 5		MARKER 6	
A	SAMPLE 1		FAM	FAM	FAM	FAM	FAM	FAM	FAM	FAM	FAM	FAM
B	SAMPLE 2		Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM
C	SAMPLE 3		Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM
D	SAMPLE 4		Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM
E	SAMPLE 5		Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM
F	SAMPLE 6		Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM	Unknown ROX FAM
G	NO RT		No RT ROX FAM	No RT ROX FAM	No RT ROX FAM	No RT ROX FAM	No RT ROX FAM	No RT ROX FAM	No RT ROX FAM	No RT ROX FAM	No RT ROX FAM	No RT ROX FAM
H	NTC		NTC ROX FAM	NTC ROX FAM	NTC ROX FAM	NTC ROX FAM	NTC ROX FAM	NTC ROX FAM	NTC ROX FAM	NTC ROX FAM	NTC ROX FAM	NTC ROX FAM

Figure 2: A picture showing view of Plate Set up used for qPCR run

6. Cover the PCR plate with foil lid and move it to the room where qPCR instrument is placed.

7. Adjust the qPCR instrument according the program mentioned in the table 5.

Table 5: PCR Program

Activation of Polymerase	95ºC	10 min
40 Cycles	95ºC	30 s
	60ºC (colored FAM fluorescence with stop)	60 s
Melting Curve	95ºC	60 s
	55ºC	30 s
	-> 95ºC continuous colored fluorescence	

8. Run the temperature program.

9. Inspect the melting curve to determine the melting points and purity of the amplified PCR-products.

2.1.4. Relative Quantification of mRNA

Relative concentrations of DNA can be calculated in different ways, but we used $2^{\Delta\Delta Ct}$ method [15]. This method is useful for analysis of relative changes in gene expression. One or more "Housekeeping" gene transcripts (genes that are believed to have the same expression in all cells at any time) must be included in the analysis, in this case BCR. Expression of the gene under study will be measured as relative to the housekeeping gene. In addition, a reference sample (calibrator) is included in the analysis. The method can be explained by the formula as under:

$$\textbf{R (Relative Concentration)} = 2^{\Delta\Delta Ct}$$

$$\mathbf{\Delta\Delta Ct} = (\text{Ctsample - Ctcal})_{BCR} - (\text{Ctsample - Ctcal})_{Ck19}$$

Where, $(\text{Ctsample - Ctcal})_{BCR}$ means Ct value for BCR in the sample minus Ct Value for BCR in the Calibrator and

$(\text{Ctsample - Ctcal})_{Ck19}$ means Ct value for gene under study (for e.g. CK19 in this case) minus Ct value for Calibrator for the gene of interest (CK19 in this case).

The $2^{\Delta\Delta Ct}$ is applied to all the markers in order to determine the relative concentration of mRNA in the sample.

2.1.5. Agarose Gel Electrophoresis of DNA

Agarose gel electrophoresis is the easiest and commonest way of separating and analyzing DNA fragments. The purpose of the gel electrophoresis might be to determine the fragment sizes, to estimate relative amounts, or to isolate a particular fragment. The DNA is visualized in the gel by addition of ethidium bromide. This reagent is fluorescent; it absorbs invisible UV light and transmits the energy as visible orange light. Ethidium bromide binds strongly to DNA by intercalating between the bases and produces increased fluorescence upon binding.

The method is based on the movement of electrically charged DNA molecules in an electric field. In this method 2% agarose is used to make a porous gel with a number of wells at one end. The gel is placed in a buffer solution in an electrophoresis apparatus which sets up an electric field with a negative pole at

one end where the wells are located, and a positive pole at the other end. Because DNA molecules are negatively charged, these will start to move toward the positively charged electrode. Larger DNA molecules will move more slowly through pores in the gel than smaller molecules and after a given time the molecules separate by size. Separated DNA fragments can be visualized by adding ethidium bromide to the gel during casting and exposing the gel to UV light after electrophoresis. DNA fragments that are analyzed are compared with fragments with a size standard.

Protocol:

Casting of 2% Agarose Gel

1. 1.0g of agarose is mixed with 50 ml 0.5x TBE buffer in an Erlenmeyer flask.
2. Heat it for 2 minutes using microwave oven.
3. Let it cool to 50°C and add 2.5 µl of 10 µg/ µl ethidium bromide solution.
4. Mix it well and pour it in gel pouring chamber.
5. Place one or two combs in the chamber and let it cool for about 30 minutes.

Electrophoresis

1. Fill the electrophoresis chamber with 0.5x TBE buffer.
2. Mix the DNA sample and loading buffer in the appropriate dilution (1:3). Do the same for the ladder.
3. Load the samples and ladder to the wells of the gel. Do not overfill the wells.
4. Run electrophoreses at 60V for (approx.) 90 minutes (depending upon the size of the product. Take a picture of the gel illuminated by the UV light.

2.1.6. Purification of PCR Products

Principle

Prior to sequencing, the PCR product needs to be cleaned of buffers and other components from the PCR reaction solution. This purification is done to get only those products that are amplified during the PCR cycle and to filter the unwanted fragments. For cleaning and purifying the PCR product QIA quick PCR purification kit is used. The system combines the "spin-column" technology with the selective binding properties of a uniquely designed silica membrane. DNA adsorbs to silica membrane in the presence of high salt concentration, while other components run through the column. We did this

purification as we want to perform DNA sequencing for the PCR products of all markers. In order to ensure the correct results from the Sequencing Data, purification of PCR products is a necessary.

Protocol:

1. Add 5 parts PBI buffer to 1 part of PCR product solution.
2. Check the color of the solution, if yellow it means pH < 7.5 if orange or purple then add 10 µl 3M sodium acetate with pH 5.0 to make the solution yellow.
3. Transfer the contents into QIA quick column placed in a 2ml collecting tube.
4. Centrifuge the column for 1 minute at 1300 rpm.
5. Discard the contents in the collecting tube. And place the column back into the tube.
6. Centrifuge again for 2 minutes.
7. Transfer the column on to a new 1.5 ml tube and add 40 µl of dH2O.
8. Wait for 1 minute and again centrifuge.

2.1.7. Sequencing of PCR Product

Principle: Sanger Method (dideoxy method)

DNA sequencing enables us to perform a thorough analysis of DNA by providing the most basic information of all the sequence of nucleotides. It can be used to locate regulatory and gene sequences, to compare between homologous genes and to identify mutations.

Sanger's method is also referred to as dideoxy sequencing or chain termination method. In this method dideoxynucleotides (ddNTP's) are also used in addition to normal nucleotides (NTP's). As the DNA synthesizes, nucleotides are added on to the growing chain by the DNA polymerase. At the moment a dideoxynucleotide is incorporated into the chain in place of normal nucleotide, the chain is terminated.

The key to this method is that all the reactions start from the same nucleotide and end with a specific base. Thus in a solution where the same chain of DNA is being synthesized over and over again, the new chain will terminate at all positions where the nucleotide do not has the potential to be added because of the integration of the dideoxynucleotides. In this way, bands of different lengths are produced.

We performed DNA sequencing in order to verify the identity of PCR products. The verification is done to ensure that correct PCR product is amplified during the PCR cycle for all the markers.

Protocol:

1. According to the Big-Dye protocol we should use 1-3 ng PCR products if it has a size of 100-200 bp. The reaction solution was prepared according to Big Dye protocol shown in the table 6.

Table 6: Reaction mixture for Sequencing

Reagents	**Quantity**
Big-Dye Version 3.1	1 µl
Sequencing Buffer (store at 4oC)	1 µl
Template (see table above)	200 ng
Primer	3.2 pmol
Deionised water	q.s.
Total	10 µl

2. Perform cycle sequencing on thermal cycler as per the program described in table 7.

Table 7: Thermal Cycler Program

1.		96 °C, 5 min
2.	25 Cycles`	96 °C, 10 sec
		50 °C, 5 sec
		60 oC, 4 min
3.		4 oC,

3. After this step we sent the reaction mixtures to the Sequencing Center in Bergen for purification and capillary electrophoresis. The major and final steps of DNA sequencing are performed there.

2.1.8. Cell Culturing

Cell culture is the maintenance and growth of the cells of multicellular organisms outside the body in specially designed containers and under precise conditions of temperature, humidity, nutrition, and freedom from contamination. Cells are grown and maintained at an appropriate temperature and gas mixture (typically, 37°C, 5% CO_2 for mammalian cells) in a cell incubator. Culture conditions vary widely

for each cell type, and variation of conditions for a particular cell type can result in different phenotypes being expressed. Aside from temperature and gas mixture, the most commonly varied factor in culture systems is the growth medium. Recipes for growth media can vary in pH, glucose concentration, growth factors, and the presence of other nutrients. The growth factors used to supplement media are often derived from animal blood, such as calf serum.

In this project we performed cell culturing while growing certain cell lines in search for the cDNA that can act as calibrator cDNA.

2.1.9. Resuscitation of Frozen Cells

Protocol:

1. Prepare a T25 flask appropriately labelled with Cell line name, passage number and date.
2. Add 12 ml of media in each flask.
3. Take the cryoculture tube out of liquid nitrogen and thaw them quickly using water bath at 37 ^{o}C.
4. Wipe of the tube with a 70% ethanol cloth. Add the cells into the flasks containing media.
5. Mix the contents of flask by moving a little.
6. Incubate the flasks at 37°C until sufficient confluency for subculturing.

2.1.10. Subculturing Cells

Protocol:

1. Heat PBS and trypsin at 37°C for 30 minutes.
2. Take cell culture flasks out of incubator.
3. Remove the media using pipette.
4. Wash the cells using PBS.
5. Add trypsin to the flask. Keep it at 37°C incubator for 3-4 minutes (until all the cells detach from the bottom). This is called trypsinization and it removes adherence of the cells from the surface of the flask.
6. Now add 25 ml media to a new50 ml flask.
7. Add 5ml media to the flask containing trypsinized cells. This will stop the action of trypsin and thereby protecting the cells.

8. Mix the contents by pipetting up and down several times.
9. Put the necessary volume of the trypsinized cells into the 50 ml flask containing media, typically 1/5 of the cells. Incubate at 37°C for 48 hours.

2.1.11. Cryopreservation of Cells

Protocol:

1. Prepare freezing media using EMEM, 20%FBS, 10%DMSO.
2. Harvest the cells using PBS and trypsin.
3. Count the cells.
4. Add the freezing media and store them in small aliquots using slowly cooling box at -80oC.
5. After 24 hours transfer the cell vials into liquid nitrogen cylinder.

2.1.12. RNA isolation from cells

The RNeasy procedure represents a well-established technology for RNA purification. This technology combines the selective binding properties of a silica-based membrane with the speed of microspin technology. A specialized high-salt buffer system allows up to 100 µg of RNA longer than 200 bases to bind to the RNeasy silica membrane. Biological samples are first lysed and homogenized in the presence of a highly denaturing guanidine-thiocyanate–containing buffer, which immediately inactivates RNases to ensure purification of intact RNA. Ethanol is added to provide appropriate binding conditions, and the sample is then applied to an RNeasy Mini spin column, where the total RNA binds to the membrane and contaminants are efficiently washed away. High-quality RNA is then eluted in 30–100 µl water.

It is essential to use the correct amount of starting material in order to obtain optimal RNA yield and purity. The minimal amount is generally 100 cells, while the maximum depends upon the RNA content of the cell type, the binding capacity of the RNA easy spin column and the volume of buffer RLT used for efficient lysis.

Protocol:

1. Trypsinize cells as described in the previous section and spin down using centrifuge.
2. Disrupt the cells by adding buffer RLT.
3. Pipette the lysate into QIAshredder spin column placed in a 2ml collection tube and centrifuge for 2 minutes at full speed.

4. Add 1 volume of 70% ethanol to the homogenized lysate and mix well by pipetting.

5. Transfer up to 700 µl of the sample to the RNA easy spin column placed in 2ml collection tube. Close the lid and centrifuge at >8000 x g for 15 seconds. Discard the flow through.

6. Add 700 µl buffer RW1 to the RNA easy spin column and centrifuge it. Discard the flow thorough.

7. Add 500 µl buffer RPE to the RNA easy spin column and centrifuge it. Repeat this step twice. Discard the flow thorough.

8. Place the RNA easy spin column in new collection tube and add 30-50 µl of RNase free water. Centrifuge it shortly. In this step we will get the RNA in the collection tube.

2.1.13. Measurement of nucleotide concentration using NanoDrop

There are several methods to establish the concentration of a solution of nucleic acids, including spectrophotometric quantification and UV fluorescence in presence of a DNA dye. Nucleic acids absorb ultraviolet light in a specific pattern. In a spectrophotometer, a sample is exposed to ultraviolet light at 260 nm, and a photo-detector measures the light that passes through the sample. The more light absorbed by the sample, the higher the nucleic acid concentration in the sample.

The Thermo Scientific NanoDrop 2000 is the only micro-volume spectrophotometer with a patented sample retention technology that allows for sample volumes as small as 0.5 µL.

Protocol:

1. Clean the tip of NanoDrop where the sample is placed using a tissue paper and a drop of distilled water.

2. Set up blank using 2µl of distilled water.

3. Take the first reading using 2µl of distilled water same as in previous step. The A260 value should be around Zero. If it is more, then consider at as error correction value for further readings.

4. Take more readings using 2 µl of nucleotide (DNA or RNA) sample.

5. Note down A260 and A260/A280 values for the successive readings.

6. Do not forget to clean the NanoDrop tip after taking each reading.

2.2. Materials

Like every other experiments, materials remained an important part in this project as well. The materials used here are listed in table 8.

Table 8: Materials and their Suppliers

Materials	Supplier
qPCR(Quantitative PCR machine)	Stratagene
Nano Drop Spectrophotometer 2000	Thermo Scientific
96 well PCR plates	Sarstedt
Biosphere RNAse- and DNAse-free tubes	Sarstedt
SYBR GREEN core PCR kit	Eurogenetec
RNA easy Mini Kit for RNA isolation	QIAGEN
BIG-DYE Version 3.1 Sequencing kit	Applied Biosystems
RNA free tubes	BIOSPHERE
MicroAmp Reaction Tube with Cap	Applied Biosystems
QIA PCR purification kit	QIAGEN
Centrifuge	Thermo Scientific
Solutions and Chemicals:	
5X FSS buffer (First strand Synthesis)	Invitrogen
RQ1 RNAse free DNAse	Promega
RNase OUT RNase inhibitor	Invitrogen
RQ1 DNAse Stop Solution	Promega
Random Primers	Promega
0.1 M DTT	Invitrogen
MMLV Reverse Transcriptase	Invitrogen
Agarose NA	GE Health Care Biosciences

Ethidium Bromide	VWR International
1X TBE Buffer 54 g Trizma Base (Sigma Aldrich) 27.5 g Boric Acid (Calbiochem) 20 ml 0.5 M EDTA dH_2O upto 1000 ml	
Cell Culturing Medium 50 ml FBS (Fetal Bovine serum) 5ml Non Essential Amino acids 5ml Sodium Pyruvate Sodium Bicarbonate 5ml Glutamine 2ml Pencillin	
Freezing Medium EMEM (Eagle's Minimal Essential Medium) 20% FBS 10% DMSO	

3. Results

The main objective of the study was to find a set of gene transcripts (mRNA) that can be used to detect tumor cells in regional lymph nodes and peripheral blood samples from patients operated for non-small cell lung cancer (NSCLC).

Secondary objectives for the implementation of main goals:

- Establish qPCR reaction setup for quantification of selected tumor cell marker mRNA
- Quantitate markers in tumor samples, normal lymph nodes and normal blood to evaluate the potential markers.
- Quantitate markers in lymph nodes and blood from patients who were operated for lung cancer at the Radium Hospital.

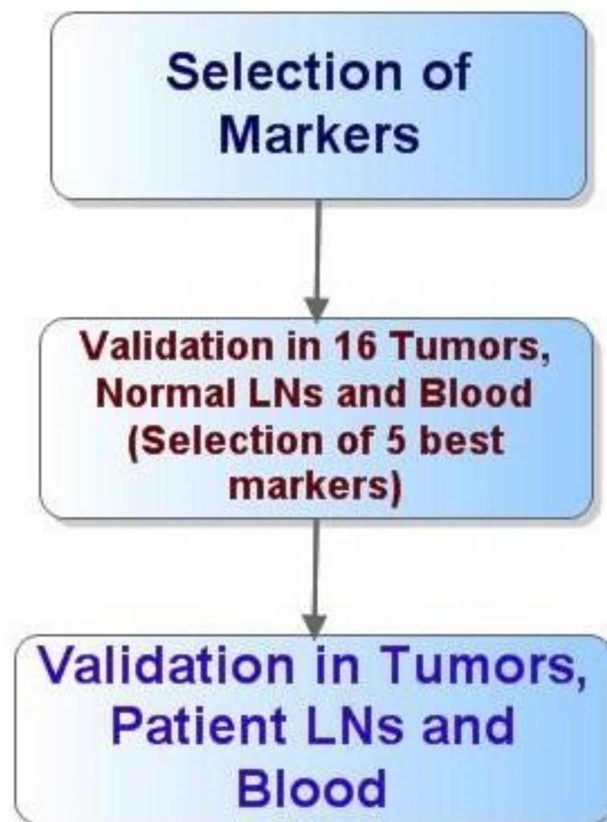

Figure 3: A flow chart showing the different phases of Result Section

The figure describes the main phases of the project. The primary phase was to select molecular markers that can be quantified in the patient and normal samples to reveal the presence of metastases. The

secondary was phase consists of two sections. In the first section the selected tumors are quantified in 16 samples of tumors and normal lymph nodes and 12 samples of blood. On the basis of the results from this section a combination of best markers was selected to determine the relative quantification of the markers in 55 samples of tumors, patient lymph nodes and patient blood.

3.1. Selection of Candidate markers

The makers are used to detect the presence of specific gene or fragment in the DNA molecule. The primers are designed specific to the marker mRNA and added to the PCR reaction mixture in order to detect the expression of corresponding gene. The markers are highly specific in detection of expression of a particular gene. The criteria adopted in the selection of markers in this project as follows:

- Markers should have high expression in the lung tumors.
- Markers should be absent or have little expression in normal lymph nodes and blood.
- Markers should be highly stable in terms of reproducibility.

Previous scientific research articles and DGED (Digital Gene Expression Displayer) database searches are certain sources that helped us in marker selection. CGAP (Cancer Genome Anatomy Project) helped us a lot in this. The NCI's Cancer Genome Anatomy Project provides the information regarding the gene expression profiles of normal, precancer and cancer cells. We selected 7 different markers (Table 9) for our study.

Table 9: Marker genes and their ID

Gene name	NCBI gene ID
CK19	3880
SFTPA	171327
SFTPB	6439
SFTPC	6440
EPCAM/TACSTD1	4072
CEA/CEACAM5	1048
PVA/DSG3	1830

3.1.1. Markers

CK19 (Cytokeratin 19)

Cytokeratins (CKs) constitute the largest intermediate filament protein subgroup and represent a multigene family with more than 20 different types of polypeptides that are divided into acidic type I (CK9-CK20) and basic type II (CK1-CK8) keratins.

SFTPA

Surfactant protein A (SFTPA) is an abundant, multifunctional protein that is secreted by airway epithelial cells and functions as part of the innate immune response. SFTPA may be critical in protecting the lungs from infectious agents and environmental exposures early in life before the acquisition of specific immunity [20].

SFTPB AND SFTPC

Surfactant proteins B and C are the genes critical for the function of pulmonary surfactant, a surface-active material that lines the lung alveoli. These genes are expressed in the developing lung epithelium and in alveolar type II and bronchiolar Clara cells in the adult lung [21].

EPCAM (Epithelial Cell Adhesion Molecule)

This antigen is expressed on most normal epithelial cells and gastrointestinal carcinomas and functions as a homotypic calcium-independent cell adhesion molecule [22]. This gene encodes a carcinoma-associated antigen and is a member of a family that includes at least two typeI membrane proteins.

CEACAM5 (Carcinoembryonic Antigen-related Cell Adhesion Molecule 5)

CEACAM5, also known as CEA or D66e, belongs to the large CEACAM subfamily of immunoglobulin superfamily. CEACAM5 is expressed primarily by epithelial cells, and is synthesized as a glycoprotein with a MW of 180 kDa comprising 60% carbohydrate.

PVA

Pemphigus vulgaris antigen (PVA or Dsg3) belongs to the desmoglein subfamily of the cadherin gene superfamily. However, in contrast to most cadherins it does not bind alpha- or beta-catenins. PVA is a calcium-binding transmembrane glycoprotein component of desmosomes in vertebrate epithelial cells Desmosome is structure formed at the site of adhesion between two cells, consisting of a dense plate in each adjacent cell separated by a thin layer of extracellular material [23].

3.1.2. Designing Primer Sets

The designing of primer sets was an important step in the experimental phase because we want to quantify the marker mRNAs. We took help of in-silico tools to design the primer sets. Primer3 and Primer Bisearch were mainly used for this. The primer sets were also verified manually.

To start with we designed two sets of primers for each marker and tested them. For some of the markers (like SFTPA, SFTPB and CEACAM) we had to design more primer sets because the first primer sets did not performed well in terms of a pure and efficient PCR reaction.

Finally we have 6 sets for SFTPA, 6 sets for SFTPB, 2 sets for EPCAM, 4 sets for CEACAM and 2 sets for PVA. SFTPC and CK19 primers were already validated in the previous study in the lab.

3.1.3. Choosing the Best Primer Set and Validation of Primers

To choose the best primer set for each marker we adopted specific criteria that involve:

- High amplification efficiency (Low Ct Value for a certain cDNA template)
- Amplification of a pure PCR product of correct size
- No unwanted PCR products in the NTC (No Template Control)
- Higher reproducibility

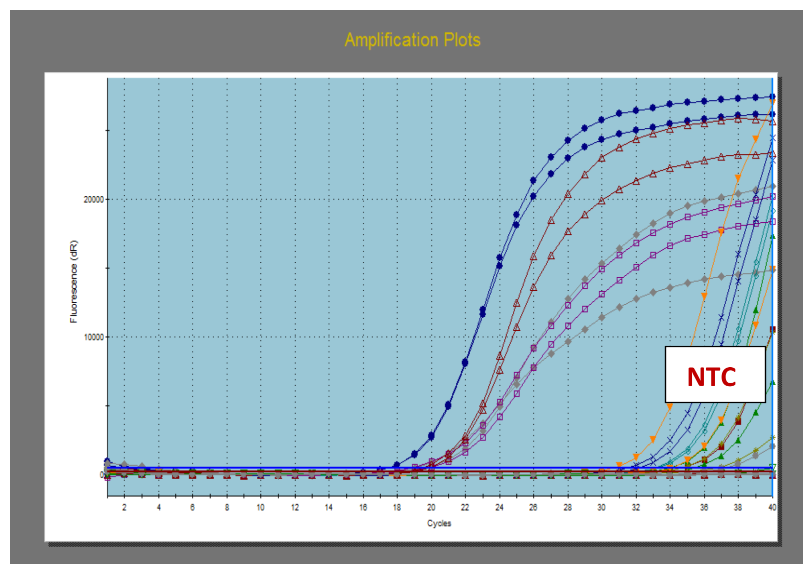

Figure 4: Amplification Curves for different primer sets of SFTPA. It may be seen that there are many unwanted products present in NTC.

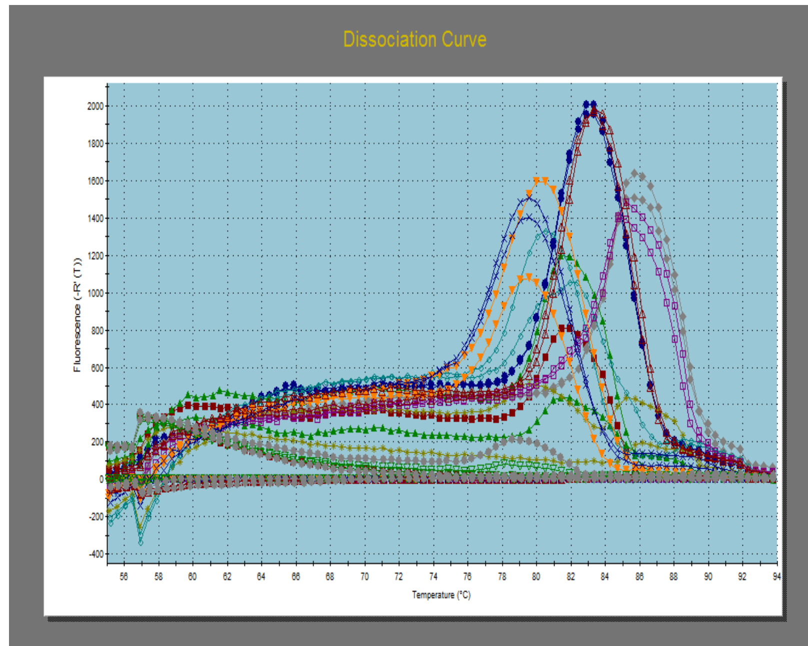

Figure 5: Dissociation curves for different primer sets of SFTPA. It may be seen that the unwanted products have different melting points than the correct SFTPA.

During the selection of best primer set for each marker we used identical reaction set up for all the primer sets of each marker. The only difference in the reaction mixture was the different primer set (table 10).

Table 10: *Primer sequences of various primer sets for all markers*

MARKERS		FORWARD PRIMER	REVERSE PRIMER
SFTPA	A	5'-ccacactccacgacttyagacatc-3'	5'-gactgcccattgctggagaagac-3'
	B	5'-ccacactccacgacttyagacatc-3'	5'-ggcatcaaaagtgatggactgccc-3'
	C	5'-gagcctgaaaagaaggagcagcgac-3'	5'-ttcacttcgcacgcagcaccagag-3'
	D	5'-tggagagtgtggagagaagg-3'	5'-agtcgtggagtgtggcttg-3'
	E	5'-ttggaggcagagacccaagcag-3'	5'-atcagcgacccacacacagag-3'
	F	5'-ttggaggcagagacccaagcag-3'	5’-ggctccaagaaatcagcgaccc-3’
	A	5'-tgaggacatcgtccacatcc-3'	5'-ccaggaacttcctcatcgtgt-3'
	B	5'-acatgtgggagccgatgac-3'	5'-cctccttggccatcttgttaag-3'

SFTPB	C	5'-gccaaggaggccattttccagg-3'	5'-tgagcagcttcaaggggaggac-3'
	D	5'-acagccccgacctttgatgagaac-3'	5'-cccgtctcacttggcttttcctttg-3'
	E	5'-gtccagccctctccagtgtatc-3'	5'-gcccgtctcacttggcttttc-3'
EPCAM	A	5'-ataacctgctctgagcgagtg-3'	5'-tgaagtgcagtccgcaaact-3'
	B	5'-cgcagctcaggaagaatgtg-3'	5'-tgaagtacactggcattgacg-3'
CEACAM5	A	5'-agacaatcacagtctctgcgga-3'	5'-atccttgtcctccacgggtt-3'
	B	5'-tttctccctatgtggtcgctccag-3'	5'-agcagatttttattgaacttgtgc-3'
	C	5'-gggacctatgcctgttttgtctc-3'	5'-gagcaaccccaaccagcac-3'
	D	5'-gaggctcctgctcacagcc-3'	5'-tcaatagtgagcttggcagtgg-3'
PVA	A	5'-ggcagtctggaaccatgagaa-3'	5'-tcctggccatcgtcttcct-3'
	B	5'-ggcaaaaacgtgaatgggtga-3'	5'-gggttgcttggtaatctgaagta-3'

On the basis of the criteria adopted we selected one primer set for each marker.

- SFTPA-F
- SFTPB-E
- EPCAM-B
- CEACAM5-C
- PVA-B

Later on the reaction set up for the selected primer sets were optimized using MgCl2 and different primer concentrations. For example: in case of SFTPA we made reaction 4 reaction mixtures having different concentration of MgCl2 (i.e. 0.25 µM, 0.5 µM, 0.75 µM and 1.0 µM). Similarly the primer set for CEACAM was optimized using different concentrations of Primer in three reaction mixtures (0.1 µM, 0.2 µM and 0.3 µM)

The motive behind the optimization was to get clear NTC and high reproducibility. The final concentrations used are described in method section (Table 3).

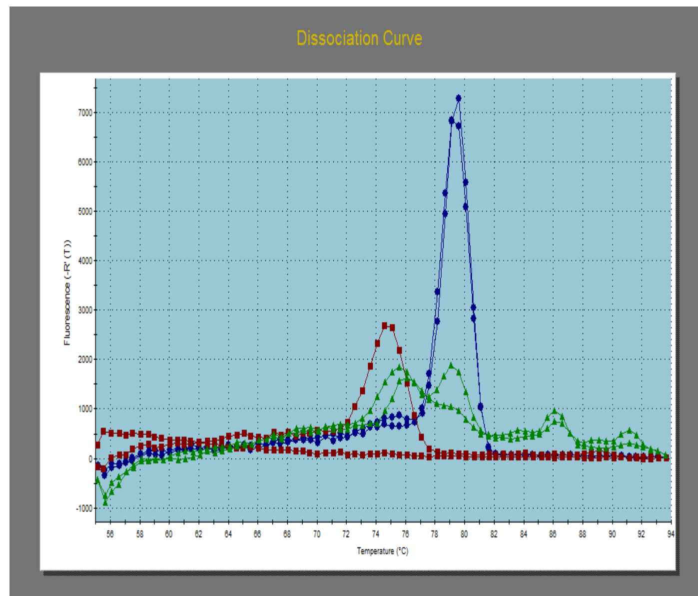

Figure 6: Dissociation curves before $MgCl_2$ titration. There are some unwanted products in the NTC.

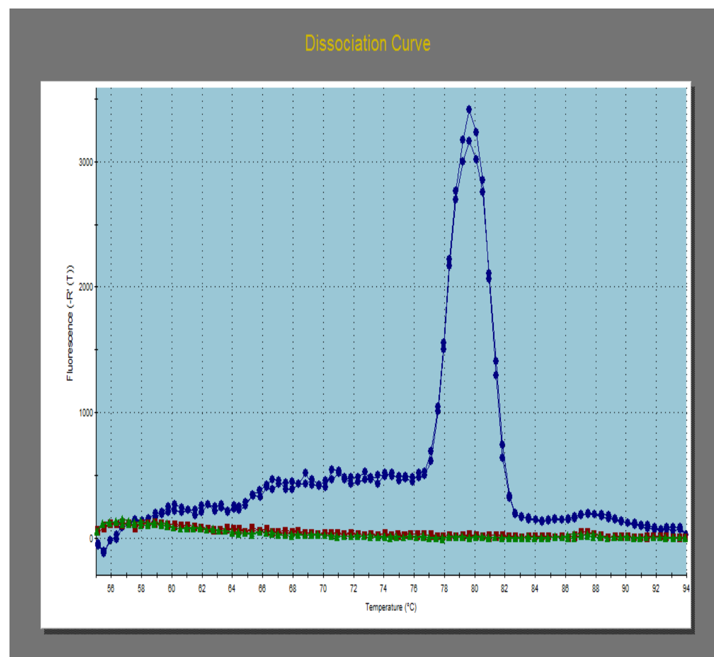

Figure 7: Dissociation Curves after $MgCl_2$ titration. Here only pure product is amplified.

The figures 6 & 7 clearly show the benefit of Primer calibration. Previously there were so many unwanted products in the NTC whereas after $MgCl_2$ titration we got clear NTC.

Also to verify that these primer sets amplify the correct product, we performed agarose gel electrophoresis using the PCR product of all markers. All of the primer sets produced correct PCR products (figure 8).

The sizes of the PCR products for the markers are as follows:

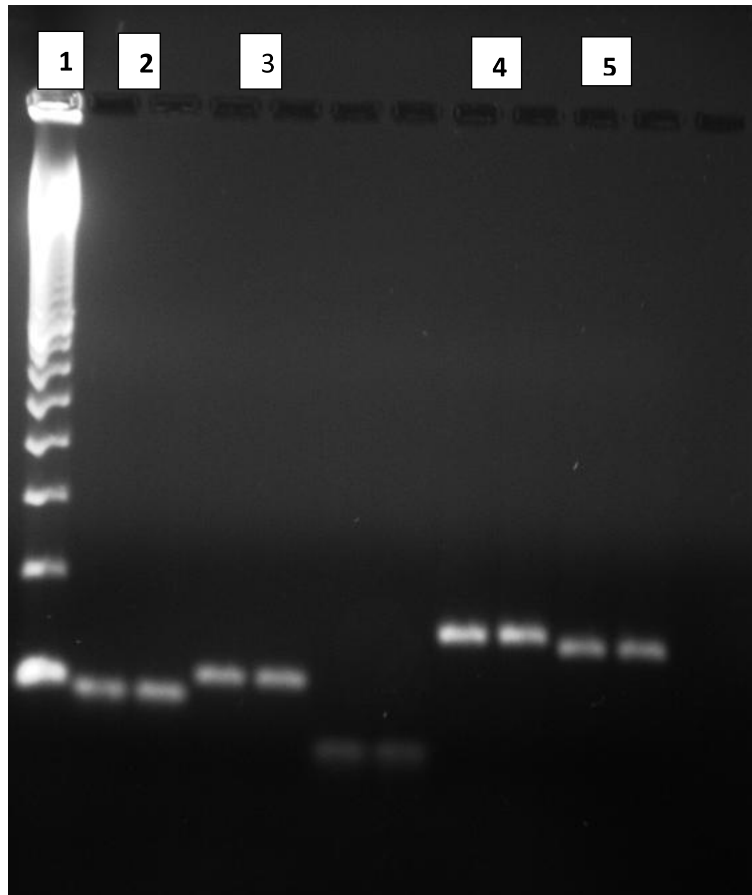

Figure 8: Agarose gel electrophoresis of PCR products of markers. The wells labelled as follows: 1: 1000bp ladder, 2: EPCAM (size=88bp), 3: PVA (size = 105bp), 4: CEACAM (size=151bp), 5: SFTPC (size=140bp).

A second verification was done for these primer sets. The PCR products for these primer sets were purified and used sequenced. The sequencing data also proved the PCR products to be the correct products of the primer sets.

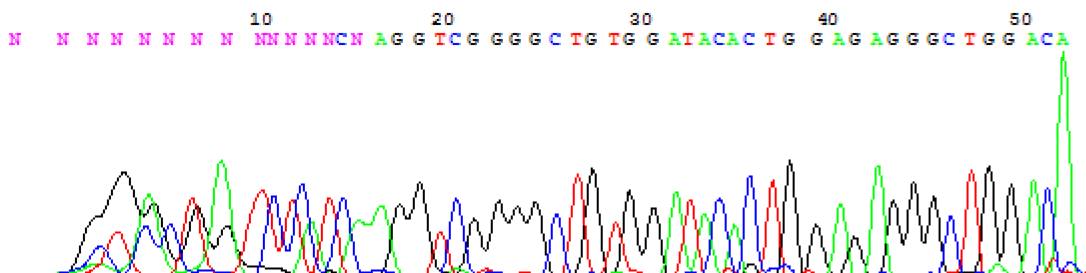

Figure 9: A chromatogram showing Sequence information for reverse primer of SFTPB

To verify the information from sequencing data, the nucleotide sequence from the chromatogram of reverse primer of SFTPB was used in blast search. The result file from the blast produced 100 % identity of the sequence with the SFTPB gene and none of the file relating SFTPB gene was available with lower identity and no other mismatch is seen.

The pairwise alignment result from the blast search is as under:

```
 Query  2     AGGTCGGGGCTGTGGATACACTGGAGAGGGCTGGACA  38
              |||||||||||||||||||||||||||||||||||||
Sbjct  1278  AGGTCGGGGCTGTGGATACACTGGAGAGGGCTGGACA  1242
```

3.1.4. Choosing a Calibrator

The purpose behind the selection of calibrator was to find a sample that can be included in all runs to be able to compare the results from different runs, and to normalize the values according to $2^{\Delta\Delta Ct}$ method. A calibrator is a cDNA molecule that has the high expression of all the markers under study and is used for calculation of R values using $2^{\Delta\Delta CT}$ method.

We tested many different cDNA as calibrator, starting with a cell line cDNA (SKMES-1 and A549) as calibrator. These two cell lines are lung cancer cell lines and they did not have good expression for all the markers under study. For some markers like EPCAM, CEACAM and CK19 they had high expression but for others the expression was very much low giving a large Ct values.

We tested some other cDNA molecules (for example Universal Reference RNA) to be used as a calibrator but most of them did not have high expression for one or the other marker. Later we quantified all the markers in another cell NCI-H441. This cell line had good expression for all the markers except SFTPC and PVA. We noticed that in some the patient tumor samples the two markers had high expression and we designed new calibrator by mixing RNA of NCI-H441 and RNA of two tumor samples that had high expression of the markers (SFTPC and PVA). The quantity in which the RNA was mixed is as under:

10 µg of NCI-H441 mRNA + 5 µg of LCa 194 T1 + 5 µg of LCa 202 T1

Where, LCa 194 T1 and LCa 202 T1 refers to the patient number whose cDNA has high expression for SFTPC and PVA.

This mixture of mRNA was reverse transcribed and used as a modified calibrator. This new calibrator has the expression for all the marker mRNAs.

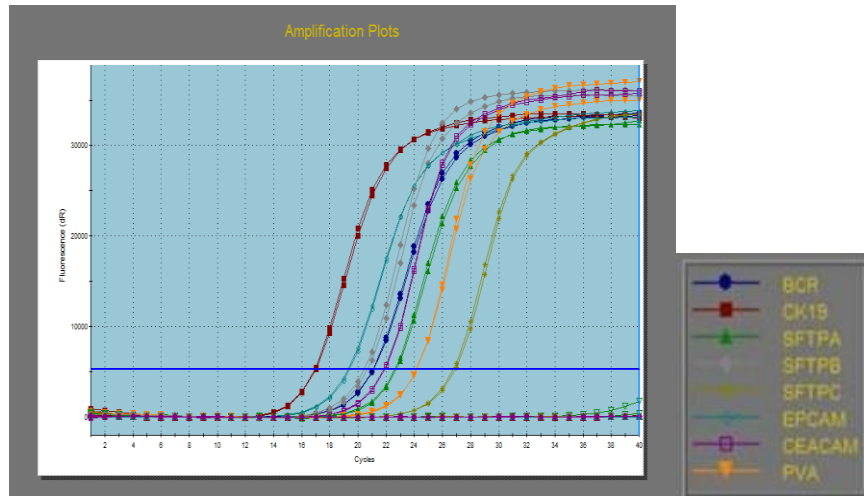

Figure 10: Amplification Curves for calibrator cDNA describes that the calibrator has good expression of all the markers.

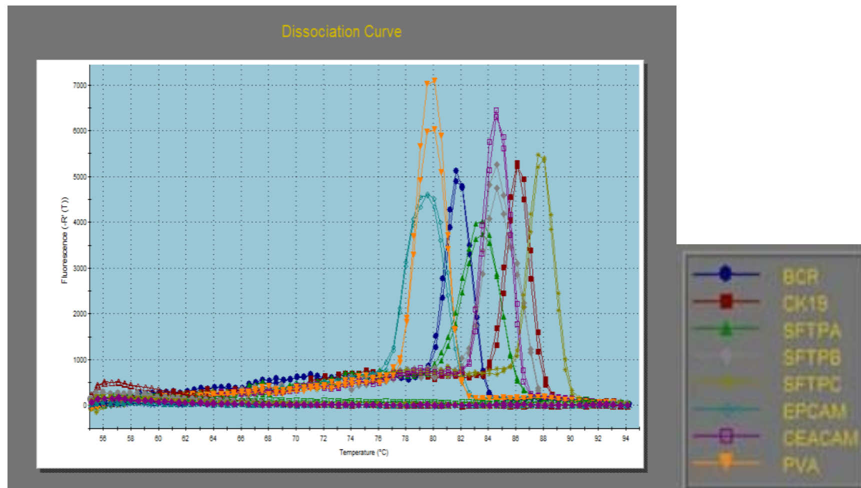

Figure 11: Dissociation Curves of all the markers in the Calibrator cDNA. All the markers have different melting points.

3.2. Marker Validation in Lung Tumors, Normal Lymph Nodes and Normal Blood

The 7 markers CK19, SFTPA, SFTPB, SFTPC, EPCAM, CEACAM and PVA, were quantified in 16 tumor samples, 12 normal blood samples and 16 normal lymph nodes samples in order to decide the best marker combination to continue with.

This project is done in collaboration with the research group of Brustgun/Helland at Norwegian Radium Hospital. The tumor samples used here were sent from Oslo and they belong to a study cohort recruited among patients undergoing surgery for NSCLC at the Norwegian Radium Hospital, Oslo. We got total RNA from normal lymph nodes from Stavanger University Hospital (SUS). The normal lymph nodes tested are not from the lung but instead they are from the colon, but it is expected that the expression of these markers does not differ much between the two. It was reverse transcribed in the lab and then mRNA markers were quantified in these using the qPCR program. The normal blood samples were taken from healthy control persons recruited at the Norwegian Radium Hospital.

The mRNA quantification data obtained for each marker from each sample was registered in a MySQL database. For each of the markers we generated stripcharts using the values from mRNA quantification. Before this many rounds of quality control were performed on the data available from the study. The quality control was established on the data to check that Ct values of the Calibrator for the markers does not differ more than 2 Ct values in all the successive runs. All the plates, where such kind of difference is observed, have been reanalyzed.

We generated the strip charts for each marker, showing the expression level in tumors, normal lymph nodes and normal blood. An example plot is shown in the figure 12.

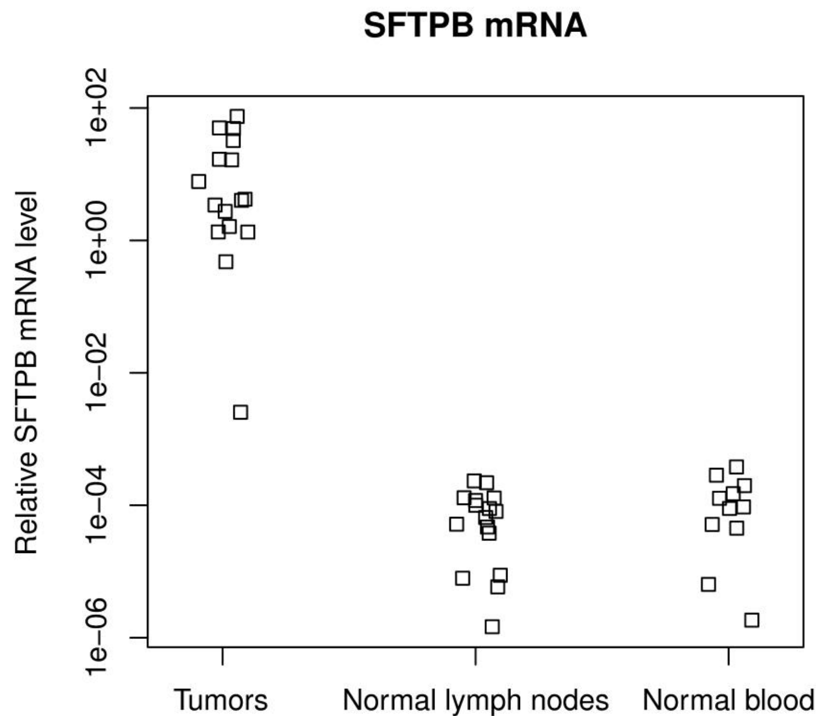

Figure 12: An example figure showing expression level of SFTPB in tumors, normal lymph nodes and normal blood respectively.

We can see that there is great difference in expression level of the marker between tumors and normal samples.

To decide which marker combination was optimal we evaluated the markers according to following criteria.

- High specificity Index
- Complementary Primary tumor expression level

Specificity Index

Specificity indexes are generated according to Ohlson et al., by computing median tumor level divided by highest normal level (either LNs or blood). The figure 13 shows the specificity indexes for all the markers.

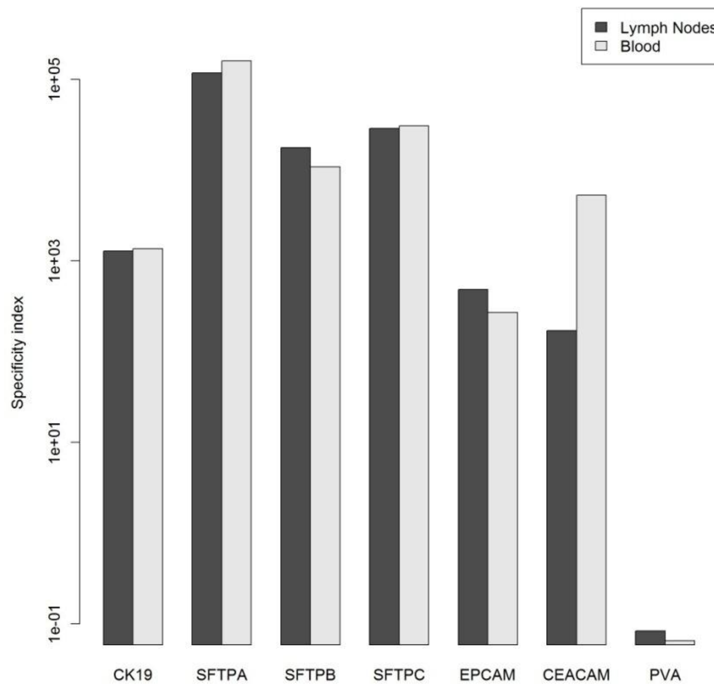

Figure 13: Specificity Indexes for all markers

The specificity indexes of all marker tells us which markers are best for the lymph nodes and which markers are best for the blood, like by observing the above picture we can see that for blood CK19,

SFTPA, SFTPB, SFTPC and CEACAM scores best where as for lymph nodes CK19, SFTPA, SFTPB and SFTPC scores best. PVA has very poor specificity index for both blood and lymph nodes.

Principal Component Analysis

We also performed Principal Component analysis to select the best candidate markers because PCA is a very powerful way to identify correlations in a data material.

Principal component analysis (PCA) involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. The data is analyzed is using PCA when we many variables (for example 7 Variables in this study). The figure 14 shows a biplot of the first two principal components.

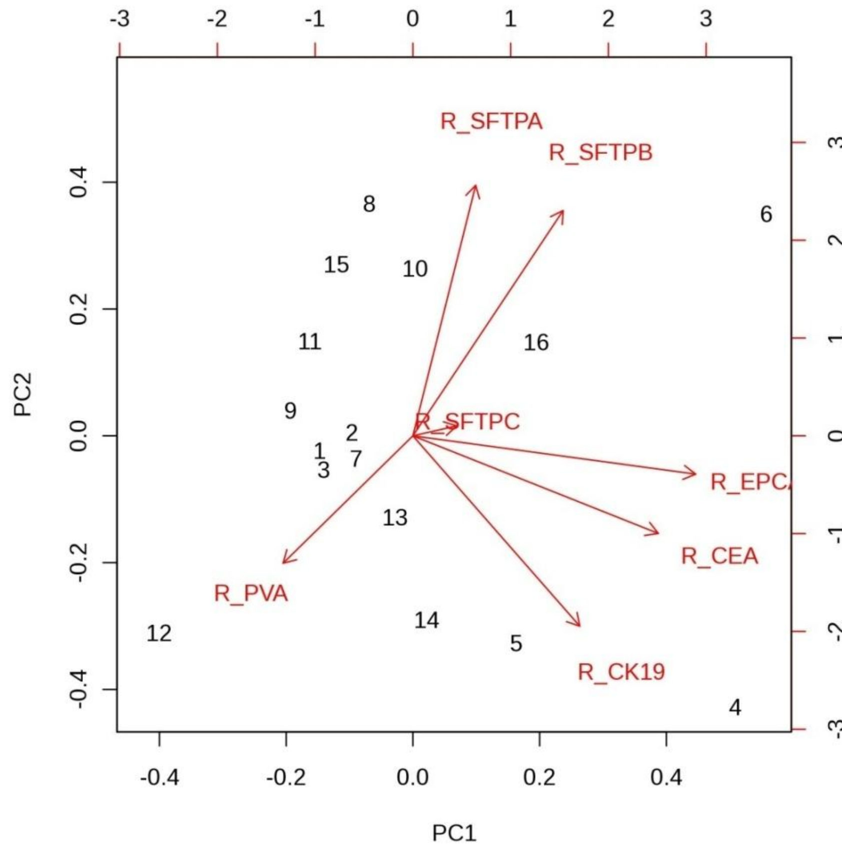

Figure 14: Principle Component Analysis

The picture above reveals that SFTPA and SFTPB are very well correlated, whereas CK19, EPCAM AND CEA have also somewhat relatedness. SFTPC also has good expression but it expresses independent of rest of the markers. We decided to use one marker from each group in a multimarker assay.

The two strategies adopted above solved the problem of marker choice. By combining the results of Specificity indexes ad PCA we finalized CK19, SFTPA, SFTPC, CEACAM and PVA.

PVA has very poor specificity index, but we wanted to evaluate it further despite this, because of its independent expression pattern in primary tumors. All the calculations and data handling procedures were performed using R (a programming language).

3.3. Marker Validation in Tumors, Patient Lymph Nodes and Patient Blood

This phase constitute the final phase of the experiment, where the motive was to quantify the selected mRNA markers in large number of tumors, patient lymph nodes and patient blood. In this phase we performed our analysis over a cohort of 55 patients that includes tumor samples, lymph nodes and blood, all three kinds of samples from the same patients. Relative mRNA quantification of each marker in each patient was calculated and the data was stored in the database. Quality checks were again applied to new data obtained.

Stripcharts were generated for all the markers. We placed a cut value for the samples from patient lymph nodes and patient blood. For the patient LNs cut off values adopted is as above the maximum value of expression in normal LNs and for patient blood the cut off is placed at the maximum value of expression in normal blood.

Cutoffs are established to define a sample as LN+/LN- and CTC+/CTC-. The other reason was to establish multimarker panel consisting of several markers that were good in combination and can reveal the presence of micrometastases. For patient LNs, those having expression level above the cut-off for at least one of the markers (CK19, CEA, and PVA) were declared as LN+, i.e. positive for lymph node metastasis, and others were considered as LN-, i.e. negative for lymph node metastasis.

Similarly for blood, patients with expression level above the cut-off for at least one of the markers were declared CT+, i.e. positive for circulating tumor cells, and others are declared CT-, i.e. without circulating tumor cells.

For SFTPA and SFTPC we cannot adopt the same cutoff value for LNs as we adopted for other markers; the reason being expression level in patient lymph nodes was generally much higher than normal lymph nodes from lung. So the normal lymph node level did not seem to be representative for normal lung lymph nodes for these markers. Also because of high back ground levels of these markers in patients LNs these markers could not be the part of classification on the basis of LN status.

Below are the figures showing relative mRNA quantification for all markers.

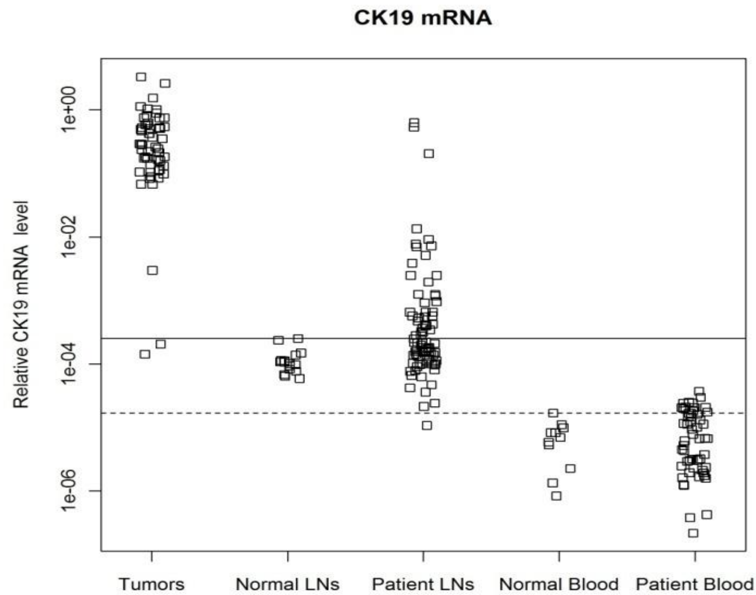

Figure 15: CK19: The picture shows that there are many patients that have expression higher than cutoff for lymph nodes where as very few for the blood. Here solid line defines cutoff for the lymph nodes and dashed line define cutoff for blood samples.

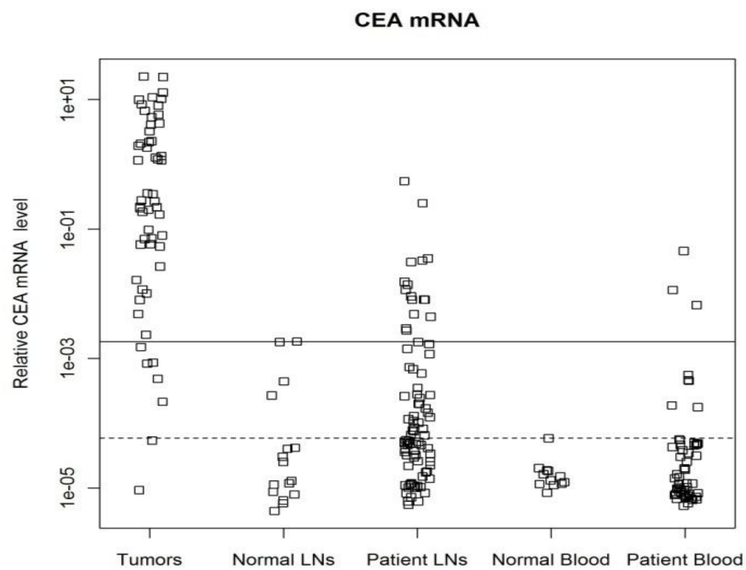

Figure 16: CEACAM5: In this case also, we can see that there are some patients that have mRNA expression higher than cutoff for both LNs and blood, but as compared to CK19 the LN+ seems few than the former.

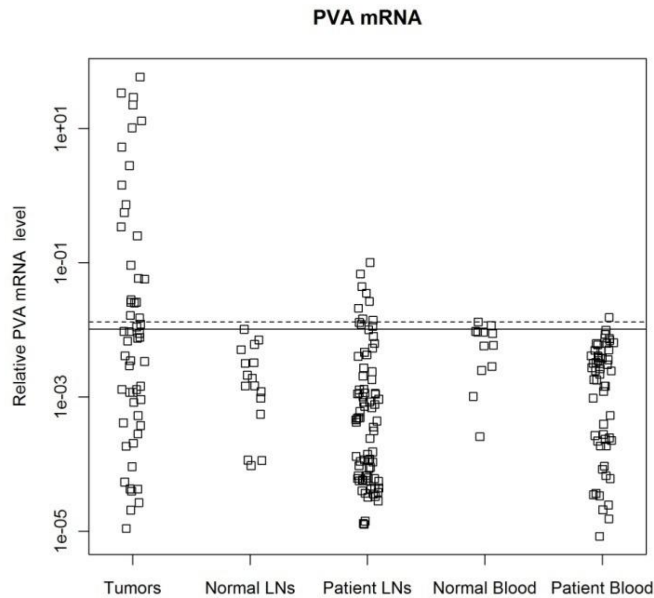

Figure 17: PVA: It does not seem to show some interesting results. We can see that there are very few patients that have mRNA expression higher than cutoff for LNs and for blood none of the patients seems CTC+. As mentioned earlier that PVA has its own independent expression than other markers. So the picture also explains the same.

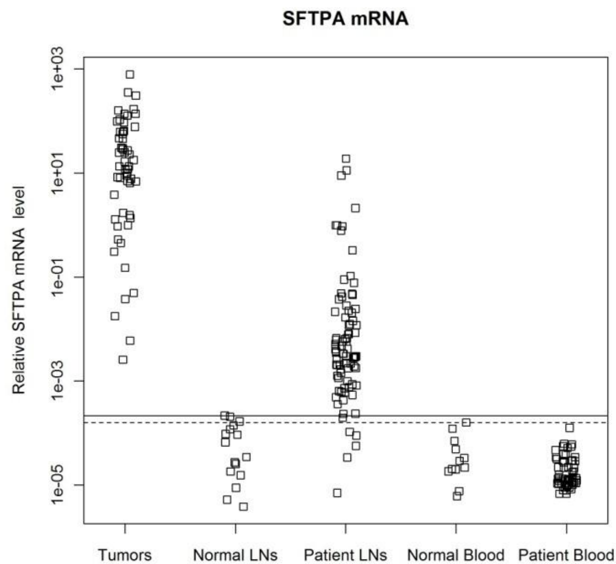

Figure 18: Stripchart illustrating the relative quantification of SFTPA in the Patient Samples.

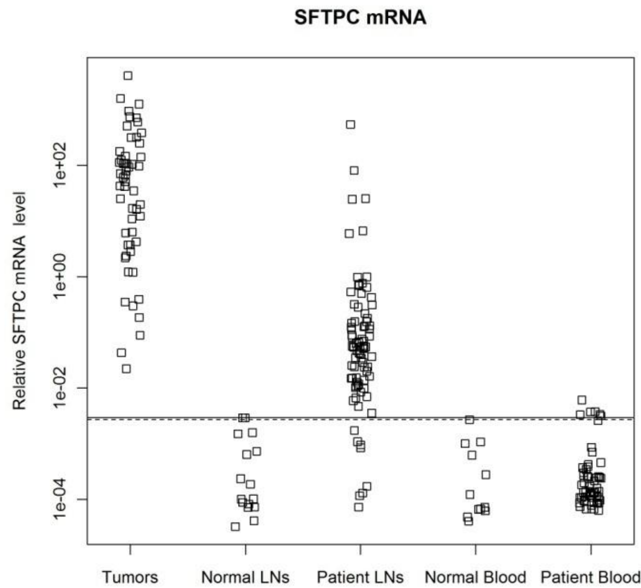

Figure 19: Stripchart showing expression of SFTPC in patient samples.

On the basis of criteria adopted for stating a sample as LN+ or LN- and CTC+ or CTC-, the patients are classified. The data obtained is mentioned in the table 11.

Table 11: Summary of LN and CTC status

Variables	Positive	Negative
Lymph Node Metastases	31	24
Blood CTCs	25	30

We have obtained significant number of patients that are positive for LN metastasis and blood CTCs. We examined whether there was any association between the detection of lymph node metastases and blood CTCs, but got no significant findings (table 12).

Table 12: Association among LN metastases and CTCs

		BLOOD	
		Positive	Negative
Lymph Nodes	Positive	11	13
	Negative	12	19

3.4. Comparison with Clinicopathological Data

Here the data obtained from the relative mRNA quantification for CK19, CEACAM and PVA were compared against the clinicopathological data (obtained from our collaborators in Oslo) for the patients studied.

Table 13: Prevalance of LN+ and CTC+ level according to clinicopathological variables

Variable	All Patients(N=55)	CTC+ (N=25)	CTC- (N=30)	LN+ (N=31)	LN-(N=24)
Sex----no (%)					
Male	23	11 (48)	12 (52)	15 (65)	8 (35)
Female	32	14 (44)	18 (56)	16 (50)	16 (50)
Histology- no (%)					
Adenocarcinoma	34	15 (44)	19 (56)	23 (67)	11 (33)
Sq Cell Carcinoma	13	7 (54)	6 (46)	4 (31)	9 (69)
Others	8	3 (37)	5 (63)	3 (37)	5 (63)
C-Stage					
Ia	16	7 (44)	9 (56)	8 (50)	8 (50)
IIa	1	0	1 (100)	0	1 (100)
Ib	24	13 (54)	11 (46)	13 (54)	11 (46)
IIb	9	2 (22)	7 (78)	6 (66)	3 (34)
IIIb	5	3 (60)	2 (40)	4 (80)	1 (20)
pT-Grade					
1	13	5 (38)	8 (62)	5 (38)	8 (62)
2	32	18 (56)	14 (44)	20 (62)	12 (38)
3	7	0	7 (100)	4 (57)	3 (43)
4	3	2 (67)	1 (33)	2 (67)	1 (33)
pN-Grade					
0	43	20 (46)	23 (54)	23 (53)	20 (47)
1	12	5 (42)	7 (58)	8 (67)	4 (33)

As we can see (Table 13), the frequency of LN+ and CTC+ vary on the basis of certain characteristic features of patients. We tested for associations between any of the clinicopathological parameters and the CTC and LN status of the patients by Fisher's exact test. The only significant (borderline) association we found was that more patients with adenocarcinoma had positive molecular LN status compared to the patients with squamous cell carcinoma (P=0.06).

23 (67 %) patients out of 34 adenocarcinoma patients, were found positive for LN+ but the number is much less for the patients with squamous cell carcinoma (only 36%). More over future survival data for the patients will reveal the clinical relevance of our findings.

If we see the results on the basis of the cancer stage, it reveals that the frequency increases with the stage of cancer, like 80 % of the patients having cancer stage IIIb have developed lymph node metastasis and 60 % of patients are positive for CTCs. Also the patients with low cancer stages have shown some tendency (50%) of LN+ that makes the findings non significant.

By looking at the T- grade, it is seen that with the increase in tumor area, frequency of LN+ and CTC+ increases. The %age is less among other patients having small tumor area.

The most interesting result in the table is seen for N-stage. For N0 patients 53 % patients are found positive for lymph node metastases in our study. Similarly 46 % of patients are found positive for CTCs. These N0 patients were declared as LN- and CTC- by other methods.

Among node positive (N1) patients in the table, we found 8 (67 %) patients out of 12 patients as LN+.

4. Discussion

In this study we established a multimarker panel consisting of CK19, CEA and PVA to classify the patients according to LN metastases and CTCs. The markers were quantified in a cohort of 55 patients and the findings were compared against the clinicopathological data. The study was designed to detect the presence of tumor cells in regional lymph nodes and circulating blood.

The presence of metastatic tumor spread to lymph nodes is recognized in many patients with different tumor types. But in some other patients this spread of tumor to lymph node and blood remain unnoticed and it keeps the disease alive in patients in spite of their surgery. There are certain reasons behind the hidden metastases;

1. Insensitivity for detection of isolated tumor cells
2. Limited amount of LN sampling from patients.

These limitations can be overcome by use of the more advanced methods that analyze the LNs at multiple sections and are sensitive enough to detect even small amount of tumors cells in the LNs. These days many modern molecular techniques are being used for this (like RT-PCR).

There are many such reports in the literature, vast majority of which have used RT-PCR to detect tumor cells in histologically negative lymph nodes and in cases where clinical data is available, the studies have shown prognosis significance of RT-PCR positivity [24, 25].

The recurrence of disease in patients of non-small cell lung cancer (NSCLC) may also occur due to the presence of circulating tumor cells (CTCs). The micrometastases may be initiated by the presence of CTCs in the peripheral blood. Different methods have been used to detect CTCs from the blood, for e.g. RT-PCR, in some previous studies [32, 36, and 37].

4.1. Relevance of RT-PCR and Molecular Markers

Reverse transcription quantitative PCR (RT-qPCR) is a combination of excellent sensitivity and specificity, low contamination risk, and high speed PCR technology [24] and provides the opportunity to establish sensitive and specific ways to analyze regional LNs [2]. This technology has been adopted in many studies related to detection of occult metastases in non-small cell lung cancer (NSCLC) patients [25, 26, 27, and 28].

The ability of real time RT-PCR to detect metastatic NSCLC has been demonstrated in a study relating EUS-FNA procurement [5]. In a reference study, 5 different markers (*lunx*, *KS1/4*, *CEA*, *CK19*, and *muc1*)

have been used to test with RT-PCR in order to detect over expression of any one of these genes in the patient DNA and to provide patients with NSCLC staging. In this study the method has been tested over 9 mediastinal LNs containing metastatic NSCLC (5 adenocarcinomas, one large cell carcinoma, one squamous cell carcinoma and two uncharacterized carcinomas) and 30 cervical lymph nodes were used as a negative control.

In many studies relating NSCLC metastases, RT-PCR has been used with molecular markers (genes) where the program is trained to look for the presence of marker genes in the cDNA molecule. In different studies many different genes has been used as molecular markers. We also used 7 markers in our primary study including, CK19, SFTPA, SFTPB, SFTPC, EPCAM, CEA and PVA. Most of these markers have already been used for detection of occult disease but fewer of them (SFTPA and SFTPC) are new and very little information is available regarding them in the previous literature.

In another study by Salerno and his colleagues [30], RT-PCR assay has been used to detect occult tumor cells in lymph nodes of 28 patients with NSCLC. They examined 88 N0 nodes for the expression of mRNA transcripts for mucin-1(MUC1) and detected the presence of the gene in 37.5% of N0 nodes.

In our study we used reverse transcription quantitative PCR (RT-qPCR) to detect the presence of tumor cells in lymph nodes and blood samples. We selected specific markers and designed primer sets specific to the marker genes. The primers for all markers were validated and PCR products were verified using gel electrophoresis and DNA sequencing. The validation of primer sets was a crucial step as we wanted to ensure that only correct product is amplified during PCR cycle while minimizing the amplification of unwanted products. At various times during the validation of primer sets we had to design new primer sets for many markers. Incorrect primer sets may also enhance the amplification of unwanted products. Sometimes unwanted products may also have the same threshold value of fluorescence (Ct Value) as the correct product and may give error prone results. In order to minimize this risk, PCR products were verified prior to the quantification of markers in the patient samples. The optimization of marker assays had been a great obstacle in this study.

4.2. Lymph Node Metastases

It is seen that LN metastases is quite common in non-small cell lung cancer (NSCLC) patients and its staging varies according to certain characteristic features of the lung cancer. LN metastases is often hidden and remained unnoticed by regular clinical investigations. Many scientific studies are being carried on this in order to find solution to this.

After establishment of reaction setup for quantification of mRNA markers, while optimizing primers sets by varying $MgCl_2$ and primer concentration, non-specific PCR products were no longer detected as side products during amplification of cDNA. Agarose gel electrophoresis confirmed the accuracy of the PCR products from the size of fragments generated. As the unwanted PCR products are no longer found, patient samples were used as template for quantification of mRNA markers, which was the biggest concern for the reaction scheme accepted and used.

In our study we included 7 markers in our primary study where the quantification was done using 16 tumors, 16 normal lymph nodes and 12 normal blood samples.

During secondary phase we quantified 5 (CK19, SFTPA, SFTPC, CEA and PVA) out of 7 markers over the cohort of 55 patient samples including lymph nodes and blood. In our findings we got some interesting results; more than 50 % of N0 patients have shown node positivity based on the higher expression of CK19 (figure 15). CEACAM mRNA expression was found higher in few patients. Combining the expression data of three markers (CK19, CEA and PVA), lymph node positivity is detected in 23 (53%) out of 43 node negative (N0) patients.

In the similar study by Liqiang and his colleagues [28] they have analyzed combination of molecular markers to detect lymph node (LN) metastases in NSCLC patients and found 100% accuracy with CK19 and TACSTD1 regardless of primary tumor histology. In their study CEA was declared as the best marker next to TACSTD1, showing 95 % sensitivity and according to their findings CK19 seems less likely to stand up as independent marker. These results are contradictory to our findings. In our study we found that CK19 is the best marker, with high expression level in most patients as compared to other candidate markers, for the detection of lymph node (LN) metastasis. They found PVA and SCCA1/2 as the best markers for squamous cell carcinoma and recommended PVA for use in a combination of marker assays, despite of its independent expression in some of the primary tumors where other markers are not able to detect metastases.

In another study, different detection methods (histopathology, Immunohistochemistry and RT-PCR) were compared to detect lymph node metastases in 254 mediastinal lymph nodes of 49 patients having non-small cell lung cancer (NSCLC) [31]. Out of 225 node negative (N0) patients based on histopathological screening, 32 have been found positive for CK19 mRNA and 16 patients were found to be upstaged for this marker. Similar to our findings this study also found CK19 as a good marker for detection of micrometastases in NSCLC patients.

By comparing our findings with clinicopathological data (Table 13) it is seen that 23 (53%) out of 40 node negative (N0) patients are found positive for lymph node metastases in our study. For 12 node positive (N1) patients in the table 15, we found lymph node positivity in 8 (67 %) patients. The other 4 (33 %) node positive (N1) patients did not showed positivity in our study because we tested only 1 or 2 lymph nodes from a patient, so it might be possible that during clinical investigation they had tested some other lymph nodes. The other reason could be that the patients might have obtained further chemotherapy treatments and thus the frequency of lymph node metastases had decreased. The future clinical follow up data will give relevant facts about our findings.

CK19

CK19 showed good potential as a marker with 10^3 fold higher relative concentration higher in tumors as compared against the highest concentration of CK19 mRNA in normal lymph nodes (figure 15). We observed elevated levels of CK19 mRNA in some patient lymph nodes that indicate the presence of

metastases. One can speculate that level variations may be related to the size of metastases, although this cannot be verified. In 3 tumor samples, level of CK19 expression are quite low and are easily distinguishable from the other samples. These 3 samples were considered as outliers with no Ct values and there for omitted from analysis.

CEACAM

The expression level of CEACAM in tumor samples lie between 10^{-4} and 10^{1}, and is 10^{5} times higher than normal lymph nodes (figure 16). The expression of CEACAM varies among tumor samples and is distributed evenly in the range. The variation of expression can depend upon the histology of cancer, stage of cancer and tumor size. Also in few tumor samples the expression level is low than the highest level in normal blood. These samples were considered as outliers and were not included in the statistics.

In some of the patient LNs the expression of CEACAM mRNA is much higher than the normal LNs and equals the expression level in tumors. We can assume that these patients have occult metastases.

PVA

Through previous studies, PVA has not been described as good independent marker for lymph node metastasis detection by RT-PCR, but it is suggested of having good significance in a combination of markers [28]. It is also stated that PVA is an excellent marker in Squamous Cell lung Carcinoma with 100% sensitivity [28]. In our study we found that PVA is higher in 6 LNs (figure 17).

SFTPA and SFTPC

The marker mRNA for the two genes has been found to have quite high level of expression in the lymph nodes of non-small cell lung cancer (NSCLC) patients. As shown in the figure 18 most of the patient lymph nodes have high expression levels of these marker mRNAs. It is unlikely to have lymph node metastases in so many patients. Actually we have quantified these markers in normal lymph nodes from colon sections and it did not seem any correlation for the expression of this marker with lymph nodes from lung cancer patients. To be of clinical relevance normal lymph nodes from lung are needed to be analyzed for the expression of SFTPA and SFTPC.

However it is still believed that SFTPA and SFTPC would be the good markers, based on their high values in tumor samples and lymph nodes from lung cancer patients. Since it could not be detected in normal lymph nodes from lung, SFTPA and SFTPC could not set any limit for the detection of micrometastases.

There appears to be good correlation between SFTPA mRNA level and SFTPC mRNA level in patient LNs (results not shown). This strengthens the assumption that higher SFTPA and SFTPC mRNA levels reflected the presence of tumor cells in lymph nodes.

4.3. Circulating Tumor Cells

Like lymph node metastases, RT-PCR has also a significant role in the detection of CTCs (circulating tumor cells) and accurate results have been achieved using the study. In most of the studies related to CTC detection CEA has been used as potential candidate marker to detect the tumor cells in the blood of patients.

In this study we found that, inspite of being a good marker in multimarker panel for detection of lymph node metastases, CEACAM is also the best marker for the detection of CTCs as well. It was found high levels of expression (figure 16) of this marker in 7 out of 55 patients (12%).

In other relevant studies CEA has been extensively examined as a useful marker to distinguish lung cancer, especially adenocarcinoma of the lung [33, 34]. In the study by Tanaka K and his colleagues, CEA and SCC antigen has been used to classify patients according to CTC status. The examined the samples taken from patients before surgery and after surgery. They performed analysis over 244 patients out of 34 (13.9%) patients were found CTC positive before surgery and 41 (16.8%) patients after the thoracic procedure [35]. The findings seem quite relative to our study.

Our findings seem significant in relation to the study by Nakashima S et al, where CEA has been detected in the blood of 31 out of 54 patients (57.4%). The found that incidence of total recurrence and blood borne recurrence was greater in the patients that had high levels of CEA mRNA in the blood than other patients with relatively low levels of CEA mRNA [36].

In other study by Wang JY, RT-PCR analysis was performed using combination of markers (CK19, CK20, hTERT, CEA). They found that healthy individuals were negative for CEA expression but patients having surgery were having significant levels of the mRNA marker. According to their findings patients with CEA expression in their blood have significantly higher risk of post-operative metastases [37].

In some of the patient blood samples, expression of CK19 mRNA (figure 15) has been found above the cutoff value but the difference is not much high than normal blood samples, so it is unlikely to consider them CTC+ and the samples should be subjected to further analysis to verify CTCs, if present any. High levels of CK19 may also be found in some patients showing CTC positivity but the marker is significant to use for detection of LN metastases.

In our study we found that out of 43 N0 patients 20 patients (46 %) were found CTC+ (Table 13) and 5 patients (42%) out of 12 N1 patients were also found CTC+ in our study. The patients found CTC+ here can be provided with further investigations and treatment therapy in order to increase the survival rate.

4.4. Future Research

In the future, the multimarker panel established in this study can be included in clinical investigations to detect micrometastases in the patients offering surgery. Also some more mRNA markers can be included in this multimarker panel. Early detection and sensitivity of the assay to detect occult disease may help in improving the survival data of patients.

The future clinical follow up data for the patients included in the study should be observed to find a clinical relevance of the findings. The markers should be quantified in more patients.

5. Conclusion

To summarize the findings of the study we can describe the most important results below:

- RT-PCR seems to have good sensitivity and specificity to detect micrometastases in non-small cell lung cancer (NSCLC) patients.
- CK19, CEA and PVA mRNA have good prospects as in multi marker panel for detection of tumor cells in regional lymph nodes from lung cancer patients. It was generally higher level of CK19 and CEA mRNA in lymph nodes from lung cancer patients than from normal lymph nodes.
- SFTPA and SFTPC mRNA might be good markers for the detection of lymph node metastases in non-small cell lung cancer (NSCLC) patients
- CEACAM mRNA is good marker in the detection of circulating tumor cells (CTCs) in the peripheral blood.
- Future clinical follow up data is important to define clinical relevance of the findings.

6. Index of Figures and Tables

FIGURE 1 A FLOW CHART DISPLAYING THE MAIN STEPS OF METHODOLOGY. IT DESCRIBES THE STEP WISE ILLUSTRATION OF ALL THE METHODS PERFORMED AT VARIOUS STAGES DURING THE PROJECT. 17
FIGURE 2: A PICTURE SHOWING VIEW OF PLATE SET UP USED FOR QPCR RUN 22
FIGURE 3: A FLOW CHART SHOWING THE DIFFERENT PHASES OF RESULT SECTION 32
FIGURE 4: AMPLIFICATION CURVES FOR DIFFERENT PRIMER SETS OF SFTPA. IT MAY BE SEEN THAT THERE ARE MANY UNWANTED PRODUCTS PRESENT IN NTC........ 35
FIGURE 5: DISSOCIATION CURVES FOR DIFFERENT PRIMER SETS OF SFTPA. IT MAY BE SEEN THAT THE UNWANTED PRODUCTS HAVE DIFFERENT MELTING POINTS THAN THE CORRECT SFTPA. 36
FIGURE 6: DISSOCIATION CURVES BEFORE $MGCL_2$ TITRATION. THERE ARE SOME UNWANTED PRODUCTS IN THE NTC. 38
FIGURE 7: DISSOCIATION CURVES AFTER $MGCL_2$ TITRATION. HERE ONLY PURE PRODUCT IS AMPLIFIED. 38
FIGURE 8: AGAROSE GEL ELECTROPHORESIS OF PCR PRODUCTS OF MARKERS. THE WELLS LABELLED AS FOLLOWS: 1: 1000BP LADDER, 2: EPCAM (SIZE=88BP), 3: PVA (SIZE = 105BP), 4: CEACAM (SIZE=151BP), 5: SFTPC (SIZE=140BP). 39
FIGURE 9: A CHROMATOGRAM SHOWING SEQUENCE INFORMATION FOR REVERSE PRIMER OF SFTPB 39
FIGURE 10: AMPLIFICATION CURVES FOR CALIBRATOR CDNA DESCRIBES THAT THE CALIBRATOR HAS GOOD EXPRESSION OF ALL THE MARKERS. 41
FIGURE 11: DISSOCIATION CURVES OF ALL THE MARKERS IN THE CALIBRATOR CDNA. ALL THE MARKERS HAVE DIFFERENT MELTING POINTS. 41
FIGURE 12: AN EXAMPLE FIGURE SHOWING EXPRESSION LEVEL OF SFTPB IN TUMORS, NORMAL LYMPH NODES AND NORMAL BLOOD RESPECTIVELY........ 42
FIGURE 13: SPECIFICITY INDEXES FOR ALL MARKERS 43
FIGURE 14: PRINCIPLE COMPONENT ANALYSIS 44
FIGURE 15: CK19: THE PICTURE SHOWS THAT THERE ARE MANY PATIENTS THAT HAVE EXPRESSION HIGHER THAN CUTOFF FOR LYMPH NODES WHERE AS VERY FEW FOR THE BLOOD. HERE SOLID LINE DEFINES CUTOFF FOR THE LYMPH NODES AND DASHED LINE DEFINE CUTOFF FOR BLOOD SAMPLES. 46
FIGURE 16: CEACAM5: IN THIS CASE ALSO, WE CAN SEE THAT THERE ARE SOME PATIENTS THAT HAVE MRNA EXPRESSION HIGHER THAN CUTOFF FOR BOTH LNS AND BLOOD, BUT AS COMPARED TO CK19 THE LN+ SEEMS FEW THAN THE FORMER. 46
FIGURE 17: PVA: IT DOES NOT SEEM TO SHOW SOME INTERESTING RESULTS. WE CAN SEE THAT THERE ARE VERY FEW PATIENTS THAT HAVE MRNA EXPRESSION HIGHER THAN CUTOFF FOR LNS AND FOR BLOOD NONE OF THE PATIENTS SEEMS CTC+. AS MENTIONED EARLIER THAT PVA HAS ITS OWN INDEPENDENT EXPRESSION THAN OTHER MARKERS. SO THE PICTURE ALSO EXPLAINS THE SAME. 47
FIGURE 18: STRIPCHART ILLUSTRATING THE RELATIVE QUANTIFICATION OF SFTPA IN THE PATIENT SAMPLES. 47
FIGURE 19: STRIPCHART SHOWING EXPRESSION OF SFTPC IN PATIENT SAMPLES. 48

TABLE 1: CLASSIFICATION OF BENIGN LUNG TUMORS [9] 9
TABLE 2: CLASSIFICATION OF PRIMARY MALIGNANT LUNG TUMORS 9
TABLE 3: CONCENTRATION OF REACTION MIXTURES FOR THE MARKERS 21
TABLE 4: VALIDATED PRIMER SETS FOR EACH MARKER 21
TABLE 5: PCR PROGRAM 22
TABLE 6: REACTION MIXTURE FOR SEQUENCING 26
TABLE 7: THERMAL CYCLER PROGRAM 26
TABLE 8: MATERIALS AND THEIR SUPPLIERS 30
TABLE 9: MARKER GENES AND THEIR ID 33

Table 10: Primer sequences of various primer sets for all markers 36
Table 11: Summary of LN and CTC status 48
Table 12: Association among LN metastases and CTCs 48
Table 13: Prevalance of LN+ and CTC+ level according to clinicopathological variables 49

7. References

1. Cancer in Norway 2008. (http://www.ncbi.nlm.nih.gov/gene/4072) Date Accessed: 12.05.2011
2. Lung Cancer. (http://www.medicinenet.com/lung_cancer/article.htm#toca) Date Accessed: 18.05.2011
3. Mitsuuchi Y , T. J. (2002). "Cytogenetics and molecular genetics of lung cancer." AMERICAN JOURNAL OF MEDICAL GENETICS 115(3): 183-188.
4. Hittelman WN, K. H., Lee JS (1996). "Detection of Chromosome instability of tissue fields at risk: insitu hybridization." JOURNAL OF CELL BIOCHEMISTRY 25: 57-62.
5. Lung cancer - Molecular Biology and Genetics (http://info.cancerresearchuk.org/cancerstats/types/lung/molecularbiologyandgenetics/) Date Accessed: 17.05.2011
6. Soria, J.-C., et al., (2003). "Chemoprevention of lung cancer." The Lancet Oncology 4(11).
7. Facts About Benign and Malignant Tumors.(http://www.mesothelioma-treatments.org/benign.html) Date Accessed: 11.06.2011
8. Benign Lung Tumors (http://emedicine.medscape.com/article/426952-overview#a0199) Date Accessed: 12.05.2011
9. The merck manuals of medical library (http://www.merckmanuals.com/professional/sec05/ch062/ch062a.html)
10. Types of lung cancer (http://www.lungcancer.org/reading/types.php) Date Accessed: 19.05.2011
11. Lung Cancer Pathology: processes and classification (<http://www.willroberts.com/lungcancer/path.html) Date Accessed: 21.05.2011
12. Lung Carcinoid Tumor (http://www.cancer.org/acs/groups/cid/documents/webcontent/003117-pdf.pdf) Date Accessed: 25.05.2011
13. http://oncolex.no/en/Lunge/Bakgrunn/Histologi.aspx Date Accessed: 17.05.2011

14. Cancer in Norway 2008. (http://www.ncbi.nlm.nih.gov/gene/4072) Date Accessed: 06.05.2011
15. Schmittgen, K. J. L. a. T. D. (2001). "Analysis of Relative Gene Expression Data Using RealTime Quantitative PCR and the $2^{\Delta\Delta Ct}$ Method " Methods 25: 402-408.
16. Rami-Porta, R. (2011). "Quantification of regional lymph node involvement in lung cancer." Thorax 66: 271-272.
17. Circulating Tumor Cells (CTCs): Emerging Technologies for Detection, Diagnosis and Treatment. (http://web.ncifcrf.gov/events/tumorcells/) Date Accessed: 25.05.2011
18. Bianca Mostert, S. S., John A Foekens (2009). "Circulating tumor cells (CTCs): detection methods and their clinical relevance in breast cancer." Cancer Treatment Reviews 35(5): 463-474.
19. Circulating Tumor Cells: Detection, Molecular Profiling & Future Prospects: CTC Detection (http://www.medscape.com/viewarticle/568351_3) Date Accessed: 06.06.2011
20. Melinda M Pettigrew, J. F. G., Yong Zhu (2007). "Respiratory symptoms among infants at risk for asthma: association with surfactant protein A haplotypes." *BMC Medical Genetics* 8(15): 1186-1471.
21. Yuxia Cao, T. V., Guetchyn Millien (2009). "Epigenetic Mechanisms Modulate Thyroid Transcription Factor 1-mediated Transcription of the Surfactant Protein B Gene." JOURNAL of biological chemsitry 285: 2152-2164.
22. http://www.ncbi.nlm.nih.gov/gene/4072 Date Accessed: 04.06.2011
http://www.ncbi.nlm.nih.gov/pubmed/7876317 Date Accessed: 04.06.2011
23. M.J. Epsy*, J. R. U., L.M. Solan, S.P.Buckwalter, M.F. Jones (2006). "Real-Time PCR in Clinical Microbiology: Applications for Routine Laboratory Testing." American Society For Microbiology **19**: 165-256.
24. Michael Mitas, D. J. C., Loretta Hoover (2003). "Real-Time Reverse Transcription-PCR Detects KS1/4 mRNA in Mediastinal LNs." Clinical Chemistry **49**: 312-315.
25. Carsten Muller-Tidow, s. D., Etmar Bulk et al (2005). "Identification of Metastases-Associated Receptor Tyrosine Kinases iCn Non-Small Cell Lung Cancer." Cancer Research **5**: 1778-1782.

26. Michael B. Wallace, M. I. B., William Gillanders, Michael Mitas (2005). "Accurate Moelcular Detection of Non-small cell lung cancer metastases in mediastinal lymph nodes sampled by endoscopic ultrasound-guided needle aspiration." Chest **127**(2): 430-437.
27. Liqiang Xi, M. C. C., Virginia R. Litle, Siva Raja (2006). "A combination of molecular markers accuratley detects lymph node metastasis in Non-Small Cell Lung Cancer patients." Clinical Cancer Research **12**(8): 2484-2490.
28. Takahi Watanabe, T. M., Yusuke Degawa, Yuna Fujita (2010). "Comparison of lung cancer cell lines representing four histopathological subtypes with gene expression profiling using quantitative RT-PCR." Cancer Cell International **10**(1).
29. Christopher T. Salerno, S. F., Gloria A. Niehans, Sam B. Ho (1998). "Detection of occult Micrometastases in Non Small Cell Lung Cancer Carcinoma by RT-PCR." Chest **113**(6): 1526-1532.
30. Tabak B, M. D., O'Day SJ, Nquyen DH, Nakayama T (2001). "The Clinical utility of multimarker RT-PCR in the detection of occult metastasis in patients with melanoma" Cancer Research **158**: 78-92.
31. Fumihiro Tanaka, Kaze Y. (2010). “Circulating Tumor Cells in Lung Cancer: current status and future perspectives” Dove Press I: 77-84.
32. Ronald A. Ghossien, Satyajit Bhattcharya, Rosai. (1995). “Molecular Detection of Micrometastases and Circulating Tumor Cells in Solid Tumors” Clinical Cancer Research
33. Okada M, Nishio W, Sakamoto T, et al. “Effect of histologic type and smoking status on interpretation of serum carcinoembryonic antigen value in non-small cell lung cancer”. Ann Thorac Surg 2004; **78**:1001–10.
34. Matsuguma H, Nakahara R, Igarashi S, et al. “Pathologic stage I non-small cell lung cancer with high levels of preoperative serum carcinoembryonic antigen: clinicopathologic characteristics and prognosis.” J Thorac Cardiovasc Surg 2008;**135**:44–9.
35. Tanaka K, Yano M. (2010) “CEA antigen and SCC antigen mRNA expression in peripheral blood predict hematogenous recurrence after resectionin patients with esophagul cancer” Annals of Surgical Oncology 10: 2796-2786.

36. Nakashima S, Natsugoe S. (2003) "Clinical significance of circulating tumor cells in blood by molecular detection and tumor markers in esophageal cancer." Surgery 133(2): 162-169.

37. Wang JY, Wu CH, Lu CY. (2006) "Molecular detection of circulating tumor in the peripheral blood of patients with colaretal cancer using RT-PCR: significance of prediction in postoperative metastases." World Journal of Surgery 30(6):1007-1013.

Appendix 1

Relative Quantification of Markers in Patient LNs and their Metastases Status

ID	R_CK19	R_SFTPA	R_SFTPC	R_CEA	R_PVA	LN Status
Patient 1	4.52E-06	2.99E-05	0.00025	3.15E-05	0.001211	FALSE
Patient 2	1.75E-05	3.18E-05	9.38E-05	0.000177	0.000394	TRUE
Patient 3	1.66E-06	3.43E-05	8.54E-05	1.04E-05	0.006992	FALSE
Patient 4	2.06E-05	7.90E-06	7.33E-05	7.84E-06	0.004086	TRUE
Patient 5	1.61E-06	6.88E-06	6.66E-05	0.006684	8.44E-06	TRUE
Patient 6	1.93E-06	9.72E-06	7.81E-05	9.46E-06	0.000187	FALSE
Patient 7	1.45E-05	8.38E-06	0.00016	8.32E-06	0.002631	FALSE
Patient 8	3.06E-06	1.02E-05	9.51E-05	1.02E-05	0.002175	FALSE
Patient 9	2.07E-05	2.14E-05	0.0003	7.29E-06	0.004962	TRUE
Patient 10	1.68E-05	6.80E-06	6.32E-05	6.76E-06	0.001455	FALSE
Patient 11	2.92E-06	1.39E-05	0.000243	7.55E-06	0.001816	FALSE
Patient 12	3.72E-05	1.32E-05	8.60E-05	4.30E-05	0.002455	TRUE
Patient 13	2.29E-06	2.71E-05	8.78E-05	3.46E-05	0.00096	FALSE
Patient 14	2.36E-06	1.56E-05	0.000107	8.47E-06	0.000246	FALSE
Patient 15	2.09E-05	1.16E-05	0.00314	4.66E-05	0.000236	TRUE
Patient 16	1.16E-05	1.40E-05	0.003773	0.000189	0.000268	TRUE
Patient 17	2.06E-05	1.13E-05	0.006108	4.53E-05	0.000222	TRUE
Patient 18	1.99E-05	1.11E-05	0.000138	8.12E-06	0.007417	TRUE
Patient 19	2.17E-07	1.25E-05	0.003365	4.99E-05	0.003786	TRUE
Patient 20	4.32E-06	8.98E-06	0.000112	6.55E-06	0.0041	FALSE
Patient 21	1.00E-05	1.23E-05	0.003319	4.92E-05	0.006302	TRUE
Patient 22	6.76E-06	9.59E-06	0.002586	3.84E-05	3.52E-05	FALSE
Patient 23	9.36E-06	5.48E-05	0.000135	0.000559	0.006434	TRUE
Patient 24	6.69E-06	3.42E-05	0.000138	8.12E-06	0.008609	FALSE
Patient 25	2.13E-06	9.04E-06	0.000116	5.64E-05	0.001791	FALSE
Patient 26	1.82E-05	1.09E-05	0.000257	7.79E-06	0.003151	TRUE
Patient 27	4.25E-07	2.14E-05	0.000132	0.000464	2.46E-05	TRUE
Patient 28	3.22E-06	1.82E-05	0.000112	0.000452	2.10E-05	TRUE
Patient 29	3.79E-07	2.22E-05	0.000424	2.55E-05	8.42E-05	FALSE
Patient 30	6.18E-06	1.08E-05	6.63E-05	5.86E-06	0.00145	FALSE
Patient 31	6.64E-06	1.36E-05	0.00012	9.72E-06	0.003594	FALSE
Patient 32	2.44E-05	2.20E-05	0.000859	1.19E-05	0.005048	TRUE
Patient 33	5.19E-06	1.42E-05	0.000128	1.63E-05	0.000277	FALSE
Patient 34	1.58E-06	0.000126	0.000383	8.49E-06	6.08E-05	FALSE
Patient 35	1.13E-05	1.33E-05	0.000136	4.67E-05	1.54E-05	FALSE

Patient 36	2.50E-05	1.21E-05	0.000134	5.52E-05	0.003484	TRUE
Patient 37	2.41E-05	1.02E-05	0.00024	7.34E-06	0.003273	TRUE
Patient 38	1.61E-05	2.93E-05	9.19E-05	1.17E-05	0.000225	FALSE
Patient 39	7.82E-06	3.90E-05	0.000342	9.46E-06	6.73E-05	FALSE
Patient 40	2.06E-05	1.77E-05	0.000109	9.59E-06	0.005962	TRUE
Patient 41	2.44E-05	1.01E-05	8.42E-05	6.69E-06	0.003195	TRUE
Patient 42	1.93E-06	5.17E-05	0.000703	1.41E-05	0.002734	FALSE
Patient 43	1.47E-05	4.71E-05	0.000261	1.17E-05	0.015251	TRUE
Patient 44	3.19E-06	6.23E-05	0.000248	2.01E-05	0.007573	FALSE
Patient 45	3.01E-06	6.06E-05	0.000196	0.011438	3.67E-05	TRUE
Patient 46	1.74E-06	5.58E-05	0.000171	1.48E-05	0.004826	FALSE
Patient 47	2.46E-06	2.88E-05	0.003721	1.91E-05	9.35E-05	TRUE
Patient 48	1.31E-05	8.95E-06	0.000213	4.82E-05	0.006237	FALSE
Patient 49	1.23E-06	1.04E-05	0.000264	2.03E-05	0.003893	FALSE
Patient 50	1.26E-06	1.21E-05	0.000107	8.70E-06	0.009855	FALSE
Patient 51	2.98E-05	7.98E-06	6.68E-05	5.30E-06	0.000187	TRUE
Patient 52	3.75E-06	2.94E-05	0.000181	0.045753	3.39E-05	TRUE
Patient 53	1.22E-05	3.93E-05	0.000457	3.02E-05	0.000531	FALSE
Patient 54	1.13E-05	1.31E-05	0.00037	7.07E-06	0.003033	FALSE
Patient 55	1.68E-05	1.04E-05	8.72E-05	6.92E-06	0.002447	FALSE

Appendix 2

Relative Quantification of Markers in Patient Blood and their CTC Status

ID	R_CK19	R_SFTPA	R_SFTPC	R_CEA	R_PVA	CTC status
Patient 1	0.006968	2.12138	81.8551	0.009131	0.001275	TRUE
Patient 2	0.007264	0.021493	0.646176	4.58E-05	0.000119	TRUE
Patient 3	0.000294	0.003331	0.011883	8.19E-05	9.00E-05	TRUE
Patient 4	0.000219	0.00099	0.019984	1.09E-05	0.044041	TRUE
Patient 5	0.0002	0.037292	0.732043	7.81E-05	0.002715	FALSE
Patient 6	0.000162	0.002677	0.089003	0.000108	0.000313	FALSE
Patient 7	0.000139	0.01243	0.534033	0.000124	0.000439	FALSE
Patient 8	9.64E-05	0.011842	0.179244	0.000146	6.82E-05	FALSE
Patient 9	0.000151	0.001145	0.040667	4.96E-05	6.12E-05	FALSE
Patient 10	0.000222	0.002781	0.010132	1.01E-05	0.012048	TRUE
Patient 11	7.65E-05	0.004158	0.068157	6.20E-06	3.20E-05	FALSE
Patient 12	0.000422	0.0041	0.06983	5.01E-05	6.23E-05	TRUE
Patient 13	0.000176	0.00285	0.006615	1.02E-05	0.004645	FALSE
Patient 14	0.00066	0.002036	0.023931	5.17E-05	6.82E-05	TRUE
Patient 15	0.005136	0.986233	0.768438	0.030927	0.00072	TRUE
Patient 16	0.000101	0.000427	0.016289	5.51E-06	2.84E-05	FALSE
Patient 17	0.009099	8.9383	0.699793	0.25	5.68E-05	TRUE
Patient 18	9.95E-05	0.00291	0.00013	1.19E-05	0.00403	FALSE
Patient 19	0.000218	0.001683	0.00173	2.19E-05	0.000691	FALSE
Patient 20	0.000316	0.024349	0.05672	0.000199	1.42E-05	TRUE
Patient 21	9.78E-05	5.66E-05	0.000116	3.97E-05	1.31E-05	FALSE
Patient 22	0.000366	0.003377	0.114229	3.23E-05	0.00011	TRUE
Patient 23	0.000571	0.002991	0.054598	0.008088	0.001141	TRUE
Patient 24	0.000137	0.002086	0.055169	0.000279	3.30E-05	FALSE
Patient 25	0.205898	0.949342	6.02099	0.013792	0.013936	TRUE
Patient 26	3.62E-05	0.005172	0.00982	0.000128	0.000774	FALSE
Patient 27	0.002489	0.077482	0.015357	0.032577	0.014478	TRUE
Patient 28	0.003879	0.327598	0.063373	0.002696	0.000457	TRUE
Patient 29	0.000531	0.01176	0.157127	0.000353	5.79E-05	TRUE
Patient 30	0.000921	0.008004	0.11344	0.001665	0.100481	TRUE
Patient 31	0.000166	0.000235	0.066064	4.81E-05	0.004245	FALSE
Patient 32	0.000153	0.001748	0.008461	1.51E-05	0.001316	FALSE
Patient 33	0.002481	0.048529	0.986233	0.000262	0.000833	TRUE

Patient 34	0.00021	0.00222	0.034435	6.52E-05	0.000421	FALSE
Patient 35	0.001946	0.989657	25.5455	5.48E-05	0.002057	TRUE
Patient 36	0.000132	0.047696	0.012824	6.59E-05	0.001103	FALSE
Patient 37	0.537747	0.041955	0.042837	0.547147	0.000492	TRUE
Patient 38	0.000213	0.020546	0.156583	8.51E-05	0.001007	FALSE
Patient 39	0.000415	0.045594	0.49827	0.00141	0.006258	TRUE
Patient 40	0.00124	0.78187	6.7039	8.38E-06	0.012913	TRUE
Patient 41	8.08E-05	0.006684	0.052374	2.60E-05	0.000117	FALSE
Patient 42	0.013462	0.089003	0.427798	0.008116	0.000141	TRUE
Patient 43	0.000406	0.016688	0.319746	0.011478	0.000152	TRUE
Patient 44	0.000143	0.006434	0.046714	1.14E-05	0.009992	FALSE
Patient 45	0.000182	0.000865	0.051296	1.11E-05	0.005355	FALSE
Patient 46	0.007732	11.2356	548.748	0.001791	0.000108	TRUE
Patient 47	0.000152	0.001258	0.023601	0.00441	0.00181	FALSE
Patient 48	0.000156	0.014378	0.130308	0.015303	0.02105	TRUE
Patient 49	0.000653	0.028262	0.99654	0.004843	0.000927	TRUE
Patient 50	0.000397	0.003734	0.076151	0.000274	9.48E-05	TRUE
Patient 51	0.000248	0.002072	0.054034	5.35E-05	0.000242	FALSE
Patient 52	0.630689	18.8959	24.847	0.035036	0.026553	TRUE
Patient 53	0.000656	0.005759	0.126745	0.00291	0.000356	TRUE
Patient 54	0.001182	0.00062	0.007017	0.000683	0.000131	TRUE
Patient 55	6.59E-05	0.000824	0.025471	0.00017	0.001157	FALSE