



Education and Culture

**Erasmus Mundus**



**UCA**

Universidad  
de Cádiz

Pharmacogenetic analysis of the Programmed cell death 6-interacting protein in  
Chinese individuals

By

Kamila Smieszkol

**Supervisors:**

Prof. Vera Ribeiro Marques (University of Algarve)

Prof. Zhao Qian Liu (Central South University)



# 1. Contents

2.	List of abbreviations .....	4
3.	Abstract: .....	7
4.	INTRODUCTION .....	8
4.1.	Pharmacogenetics.....	8
4.2.	Principles of pharmacogenetics .....	9
4.3.	Cancer .....	10
4.4.	Lung cancer .....	18
4.4.1.	Non-small cell lung carcinoma .....	19
4.4.2.	Stages of cancer.....	21
4.4.3.	Stages, diagnose and spread of Non-Small Cell Lung Cancer.....	22
4.5.	Platinum- based chemotherapy.....	28
4.5.1.	Gene polymorphisms affecting outcome of cancer therapy (cancer pharmacogenetics).....	30
4.5.2.	Platinum Resistance .....	31
4.6.	SNP (Single Nucleotide Polymorphism) .....	33
4.7.	Gene.....	34
4.7.1.	PDCD6IP (The programmed cell death 6 interacting protein) .....	34
4.8.	Polymerase chain reaction (PCR).....	36
4.8.1.	Primers .....	39
4.8.2.	PCR-RFLP (Restriction Fragment Length Polymorphism) .....	40
4.8.3.	Real-time PCR).....	41
4.9.	DNA electrophoresis .....	41
4.10.	Data analysis .....	43
4.10.1.	The Hardy–Weinberg equilibrium.....	43
4.10.2.	Minor allele frequency .....	46
4.10.3.	SPSS- Binary Logistic Regression.....	46
4.10.4.	Stratification .....	47
5.	Objective .....	48
6.	MATERIALS AND METHODS .....	49
6.1.	Ethics approval.....	49
6.2.	Tagging SNP selection .....	53
6.3.	DNA extraction and genotyping procedure .....	54
6.3.1.	Genomic DNA extraction.....	54
6.3.2.	Procedures of genotyping by PCR-RELP: .....	55
6.3.3.	Statistical analysis.....	58
7.	RESULTS.....	59

7.1. Patient Characteristics .....	59
7.2. Prevalence of the PDCD6IP rs3183982 SNP.....	61
7.3. Relation of PDCD6IP gene polymorphism with platinum-based chemotherapy response ...	61
8. CONCLUSION .....	67
9. References:.....	69

## 2. List of abbreviations

G6PD	glucose-6-phosphate dehydrogenase
GS	growth signals
GF	growth factor
PDGF	platelet-derived growth factor
PDCD6IP	the programmed cell death 6 interacting protein
PCD	programmed cell death
PDCD6	the programmed cell death 6 gene
E2F	transcription factor
TGF- $\beta$	transforming growth factor beta
Bcl-2	B-cell lymphoma 2
Apaf-1	apoptotic peptidase activating factor 1
P53	tumour protein
VEGF	Vascular Endothelial Growth Factor
HIF	Hypoxia inducible factor
CAM	cell adhesion molecules
NSCLC	non-small cell lung cancer
SCLC	small cell lung cancer
EAC	erythema annulare centrifugum
ESCC	esophageal squamous cell carcinoma
TKI	tyrosine kinase inhibitor
SNP	single nucleotide polymorphism
MRI	magnetic resonance imaging
CT	computer tomography
CAT	computed axial tomography
PET	pulmonary function test
EUS	endoscopic ultrasound
AUC	area under the concentration
CTR1	copper transporter receptor

ATP7 A/B	P-type adenosine triphosphate
GSH	glutathione
HDAC	histone deacetylases inhibitor
SAHA	pan-histone deacetylase inhibitor
ESCRT	endosomal sorting complex required for transport
TF	transcript factor
NF-kB	nuclear factor-kappaB
HCC	hepatocellular carcinoma
MVB	multivesicular body
EGFR	epidermal growth factor receptor
RTK	receptor tyrosine kinase
PCR	polymerase chain reaction
FISH	in situ hybridization
GFP	green fluorescent protein
dNTPs	deoxynucleoside triphosphates
PCR	polymerase chain reaction
PT-PCR	real-time polymerase chain reaction
PFGE	pulsed field gel electrophoresis
DGGE	Denaturing Gradient Gel Electrophoresis
TGGE	temperature Gradient Gel Electrophoresis
SDS	sodium dodecyl sulphate
PAGE	polyacrylamide gel electrophoresis
RFLP	Restriction Fragment Length Polymorphism
MNP	multiplex nested
V	voltage
I	current
R	resistance
MAF	minor allele frequency
ECOG	Eastern Cooperative Oncology Group
ECOG PS	Eastern Cooperative Oncology Group Performance Status

RECIST	Response Evaluation Criteria in Solid Tumours
CR	complete response
PR	partial response
SD	stable disease
PD	progressive disease
cTNM	Clinical diagnostic staging
UICC	Union Internationale contre le Cancer
TAE	Tris-acetate-EDTA
Tris-HCL	tris(hydroxyethyl)amino methane
EDTA	Ethylenediaminetetraacetic acid
HWE	Hardy - Weinberg Equilibrium
CI	confidential interval
OR	odd ratio
HWE P	p value for Hardy-Weinberg test

### **3. Abstract:**

#### **Introduction:**

Pharmacogenetics is the study of germline genetic variation resulting in altered absorption, distribution, metabolism and excretion associated with the efficacy or toxicity of a drug, or affecting drug interaction with the target protein. Cancer is one of the top ten leading causes of death globally. Moreover lung cancer is one of the most commonly diagnosed cancer, most patients with NSCLC are being diagnose in a late stage and lost the opportunity of surgery. Platinum –based regimens in overall survival and quality of life shown meagre improvement. Study on single nucleotide polymorphism (SNP) showing significant improvement on detecting risk of lung cancer and overall survival by study and by measure and evaluate molecularly defined biomarkers.

#### **Results:**

A total of 335 lung cancer patients, the number of responders (CR or PR) were 100 and non-responders (PD or SD) were 235. All patients received platinum-based chemotherapy, cisplatin-based was 153(45.7%) and carboplatin-based was 182(44.3%). age (P-value 0.037), histology (0.022) and ECOG (0.033) may be considered statistically significant risk factors on platinum-based chemotherapy efficacy in all population. The rs31839825 SNP showed a significant association with chemotherapy response when considering the additive model (OR=0.67, P=0.033). The relationship of the SNP with response to chemotherapy showed significant associations with chemotherapy response: for stage I-II in additive (OR=1.89, P=0.017) and dominant (OR=2.1, P=0.014) models, for EAC in additive model (OR=0.83, P=0.041), for non-smoking (OR=0.47, P=0.048), and for cisplatin (OR=0.51, P=0.046) in recessive model.

#### **Conclusion:**

Results prove that there is a correlation between response and resistance to platinum based chemotherapy. It also suggests that there is a connection between resistance to a drug and the histology and association between the gene polymorphism and lung cancer risk or chemo sensitivity. More studies need to be performed to investigate the potential linking rs3183982 polymorphism in PDCD6IP encoding protein required for apoptosis with cancer progression and prognosis

## 4. INTRODUCTION

### 4.1. Pharmacogenetics

The term pharmacogenetics was officially published by the German physician Friedrich Vogel in porphyria [1].

In October 1957, an article was published by *Journal of the American Medical Association* which showed the confluence of several scientific developments applied to medicine including the emergence of human biochemical genetics, which explained inborn errors of metabolism to be caused by enzyme malfunctions due to gene mutations. Two different adverse drug reactions shown to be caused by a specific genetic enzyme variant affecting the drug's metabolism [2].

The identification of primaquine-induced haemolytic anaemia among African-Americans (later shown to be due to glucose-6-phosphate dehydrogenase [G6PD] variant alleles) is another relevant scientific discovery in the 1950s [3].

One of the most influential discoveries for pharmacogenetics and its potential clinical utility was the identification in 1977 of the hepatic cytochrome P450 oxidase that controls debrisoquine and sparteine metabolism [4]. Subsequent population and family studies identified specific drug metabolism phenotypes and suggested 1959, but unusual drug reactions based on biochemical individuality were observed in 1930s.

This was in results to earlier observations of interindividual variability in phenylthiocarbamide taste perception and isolated cases of drug-induced that the “poor metabolism” trait was inherited in an autosomal recessive Mendelian fashion. The responsible enzyme, CYP2D6, was purified, the corresponding cDNA cloned, and the gene extensively sequenced and is now believed to be directly involved in the metabolism of ~25% of all commonly used drugs. More than 80 variant CYP2D6 alleles have since been discovered worldwide, many of which encode deficient enzyme activity, and these are carefully catalogued by the Human Cytochrome P450 (CYP) Allele Nomenclature Committee [5].

In the late 1800s, the concept of “personalized medicine” was anticipated by the Canadian physician Sir William Osler who noted “the great variability among individuals”, however the

final definition has evolved that not only patient's clinical assessment and family history were involved in guiding medical management but also personal genomic information [6].

Facilitating more effective drug therapy by using pharmacogenetic biomarkers as well as studying how genes and the environment interact to cause human disease are major areas of research involved in the identification of the genetic basis of common diseases.

Pharmacogenetics is the study of germline genetic variation resulting in altered absorption, distribution, metabolism and excretion associated with the efficacy or toxicity of a drug, or affecting drug interaction with the target protein. Generally, pharmacogenetics has focused predominantly on reducing adverse drug reactions, but recently several applications on the area of efficacy have been reported. Correct dosage, adherence to medication and treatment schedules, avoidance of drug-drug interactions and correct drug choice all have an important role to complement pharmacogenetics.

## 4.2. Principles of pharmacogenetics

Genetic polymorphism can occur when a single genetic trait is expressed in two or more different ways in the population. Each one of us have our own unique genetic coding for our body, every person reacts differently to a particular dose of a drug. Pharmacogenetics tries to study and generalize how people react based on their genetic composition. Use of properties of pharmacodynamics and pharmacokinetics in order to study these effects

- A single gene encodes a protein whose function may be relevant to several drugs e.g. a drug-metabolising enzyme or transporter.
- Different variants in the same gene may cause different functional effects, which can result in reduced or increased activity, altered binding capacity or absence of the protein [7].
- A lot of compounds are being identified with extremely potent *in vitro* activity but are found to be inactive *in vivo*. They can possess the optimal configuration and conformation needed to interact with their target receptor or enzyme, but they do not always have physicochemical properties and the best molecular form needed for their delivery to the site of action. Here helps, prodrugs that can offer many advantages over parent drugs such as increased solubility, enhanced stability, improved bioavailability, reduced side effects,

and better selectivity. Many prodrugs have been used successfully in the clinic; examples include capecitabine in cancer therapy. Many other factors apart from functional variants in genes including patient adherence, co-morbidities, and co-medications can influence the metabolism of a drug. [8,9]

### **4.3. Cancer**

Cancer is the most important unsolved medico-biological problem in the century. There are many cancer research teams investigating numerous genetic factors and complex molecular mechanisms underlying in the pathogenesis of neoplasia. Harahan and Weinberg believe that most (and probably all) type of cancer share a small number of biochemical, molecular and cellular traits called acquired capability. They created a logical science where the complexities of the disease will become understandable example rules that govern the transformation of normal human cells into malignant cancers. It is a simplification from the teaching of cell biology that virtually all mammalian carry similar molecular machinery regulating their proliferation, differentiation and death. There are more than 200 different types of cancer so this complexity provoke many questions [10]. The authors suggested that cancer cell genotypes is a manifestation of six alterations relevant for malignancy:

- Self-sufficiency in growth signals (cell proliferation, differentiation and death).

Growth signals (GS) known also as stimulatory signals are required by normal cells to activate proliferative state. Tumore cells mimicking normal GS and generate many of their own. One type of cells creates mitogenetic growth factors (GFs) in order to stimulate proliferation while cancer cells are able to synthesize GF to which they are responsive creating a positive feedback signaling loop often termed autocrine stimulation. Selfstimulation tumor cell growth is possible due to the fact that they are provided with appropriate surface receptors composed of the extracellular portion and the intracellular tyrosine kinase that is divided into two parts catalyst. Overexpression platelet- derived growth factor (PDGF) and its receptor occurs in human small cell lung cancer [11].

- Insensitivity to growth- inhibitory (antigrowth) signals.

Antigrowth signals can block proliferation by two distinct mechanisms either cell forced out for active proliferative cycle into quiescent stage  $G_0$  (they might change in the future) or cell induced

to permanently abandon their proliferative potential by entering into post-mitotic states (associated with mature cells that have differentiated). The cancer cell must therefore avoid these signals if it is to continue dividing uncontrollably. At the molecular level, nearly all antigrowth signals are funneled through a protein known as the retinoblastoma protein (classified as a tumor suppressor protein). A primary function of retinoblastoma is to bind to and inactivate E2F transcription factors. These are extremely important proteins that bind to DNA and activate genes which control the cell cycle and DNA replication. TGF- $\beta$  - transforming growth factor beta has many different mechanisms for preventing the phosphorylation of retinoblastoma (i.e. for preventing the disengagement of the brakes). Therefore, the presence of TGF- $\beta$  blocks the advancement of the cell cycle. Unfortunately many cancers can disrupt this pathway either by stopping the response to TGF-beta altogether, by producing less TGF- $\beta$  receptors on their cell surfaces or producing mutated receptors that do not respond to the presence of TGF- $\beta$

- Evasion of apoptosis.

The ability of tumor cell population to expand in number is determined not only by the rate of cell proliferation but also by the rate of cell attrition. There are two categories of apoptotic machinery regulators and effectors. The regulators are responsible for monitoring the interior and exterior environment of the cell for conditions of abnormality (DNA damage, signaling imbalance caused by the activation of cancer causing genes (oncogenes), lack of an oxygen supply or insufficient growth factors. in order to decide whether that cell should live or die. Apoptosis can occur either through an intrinsic pathway, in which signals from within the cell activate the process, or through an extrinsic pathway where death signals from outside the cell are received and processed by the cell to activate apoptosis. The primary regulators of apoptosis are proteins from the Bcl-2 family group and can either be pro-apoptotic (trigger apoptosis when activated) or anti-apoptotic (inhibit apoptosis). The anti-apoptotic proteins bind to and inactivate and the pro-apoptotic proteins in a healthy cell that does not need to die. The end result is that antigrowth signals, funnelled through Retinoblastoma protein into the cell cycle, the absence of the Retinoblastoma protein permits persistent cell division (cancer). Mitochondria in response to pro-apoptotic signals released cytochrome c and bind to a protein known as Apaf-1, resulting in the formation of the apoptosome. The apoptosome activate a group of proteins known as caspases (cellular executioners). Initiator caspase launch the other caspases in a cascade of irreversible cellular protein degradation. P53 is an extremely important protein, it is responsible for detecting DNA damage, chromosome abnormalities and arresting the cell cycle to initiate repair; if repair is not possible then apoptosis is induced. The most common method is the loss of the apoptosis

gatekeeper, the protein P53. More than 50% of all types of human cancers have a mutated or missing gene for p53, resulting in a damaged or missing P53 protein. Retinoblastoma and P53, that control both cell division and cell death; the result is repeated uncontrolled cell division that manifests itself in warts, with strong associations with the development of cancer. Cancer cells can also produce excessive amounts of anti-apoptotic proteins or less of the pro-apoptotic.

- Limitless replicative potential.

Most cells in our body can only undergo a limited number of successive cell growth-and-division cycles (Hayflick Limit) After undergoing between 40 and 60 divisions, cell growth slows down and eventually stops altogether. This state is known as senescence, and it is irreversible; although the cell does not grow or divide, it remains alive. Some cells are able to make it past the senescence barrier and continue dividing; however these cells then undergo a second phenomenon known as crisis, during which the ends of their chromosomes fuse with each other, and the cells all die on a massive scale via apoptosis. With every replication of a cell, about 50-100 nucleotides of telomeric DNA is lost. This progressive loss eventually causes the telomeres to lose their ability to protect the ends of chromosomal DNA. Left unprotected, these exposed ends become damaged. The DNA damage response is activated, leading to growth arrest; senescence. When chromosome ends fuse with each other, this irreversible damage results in the activation of apoptosis; the cell enters crisis, and dies. Cancer cells have therefore not only uncoupled their growth program from the signals in their environment, they have also breached the in-built replication limit hard wired into the cell by maintain their telomeres. 90% of them do so by increasing the production of a telomerase (by adding telomeric DNA to the ends of chromosomes). Most normal cells do not divide frequently, these cells show low telomerase activity levels. Many cancer causing proteins (oncoproteins) are able to activate the production of telomerase, while many cancer preventing proteins (tumor suppressors) such as P53 produce factors that inhibit the production of telomerase. The defining feature of a cancer cell is its ability to divide endlessly, without exhaustion, generation after generation. They achieve this by destroying the cellular timekeeper, the telomere. Immortality comes at a price; the accumulation of damaging mutations only increases with time, which is why cancer is primarily a disease of an aging population. The immortalization of cancer cells by telomere maintenance therefore represents an essential step in tumour progression.

- Sustained angiogenesis.

In a developing embryo or a healing wound, require oxygen and nutrients, as well as a facility to remove metabolic wastes and carbon dioxide. The formation of new blood vessels, known as

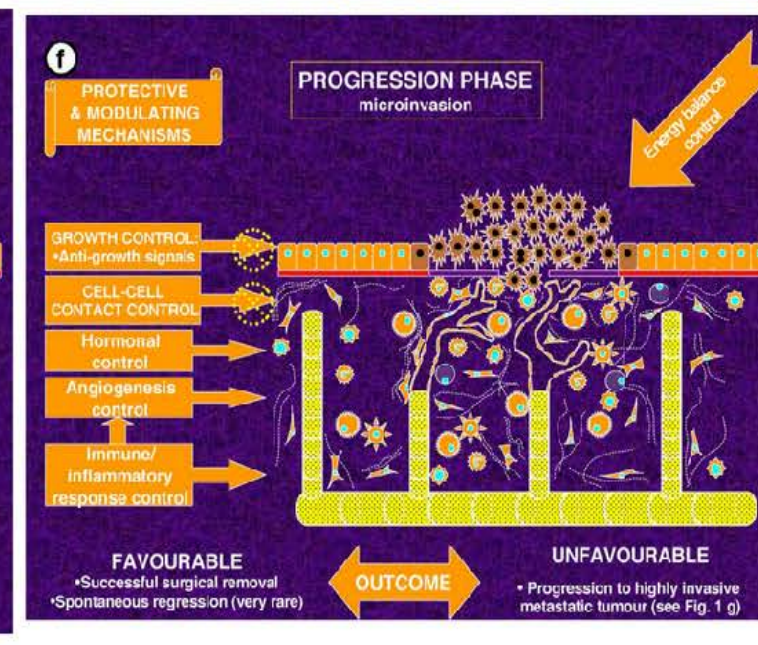
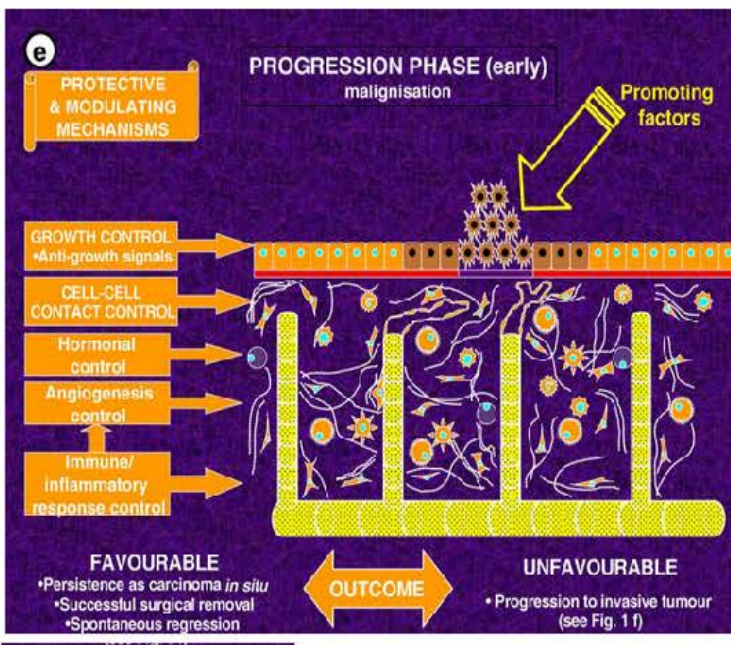
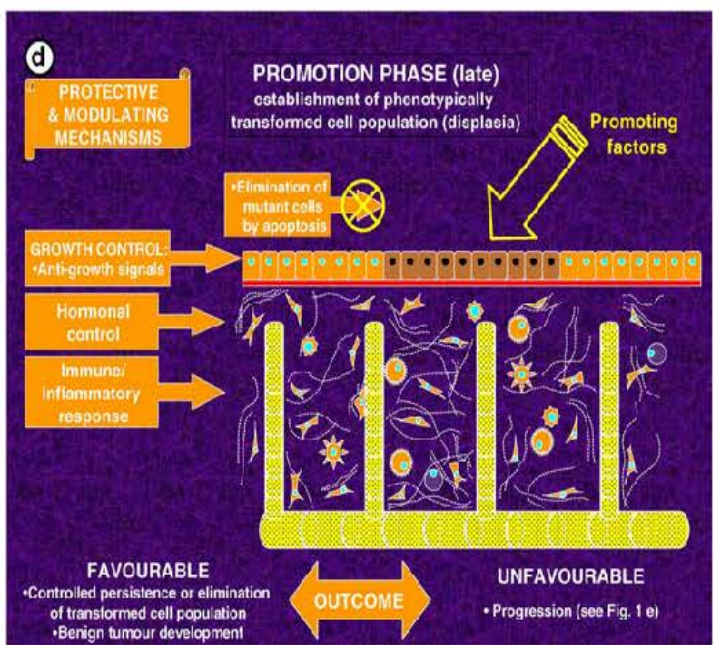
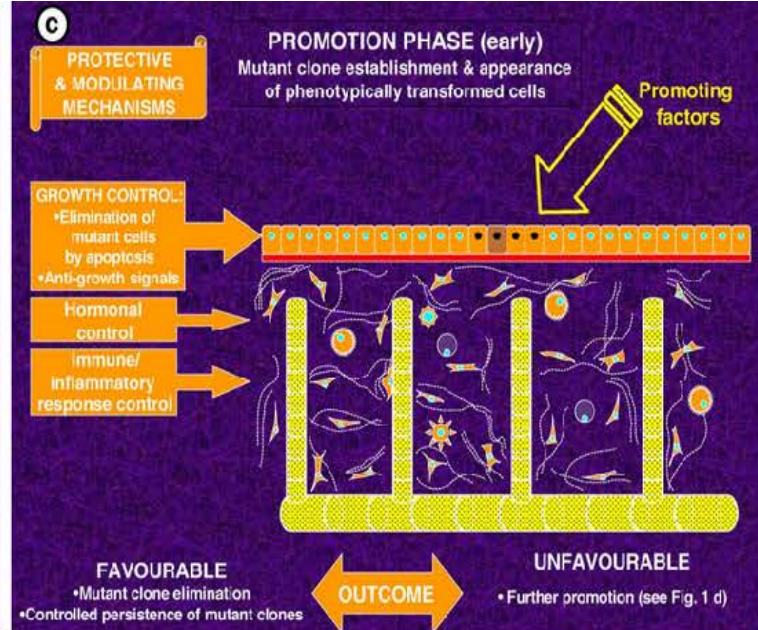
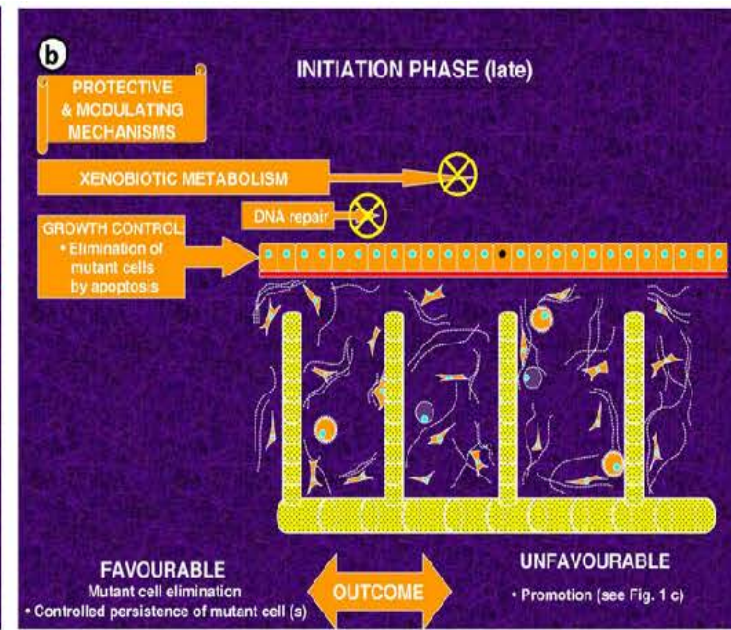
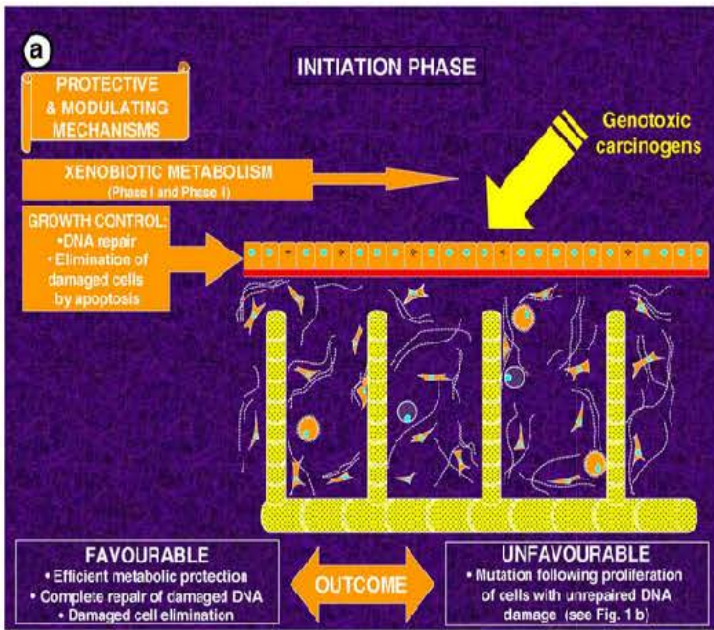
angiogenesis, satisfies these needs. In a similar manner, a growing tumour, an aggregation of cancer cells, also requires access to oxygen, nutrients and waste disposal. When the tumour begins to starve cancer cells send out signals to the cells of nearby blood vessels, inducing these innocent bystanders to grow extensions to form a supply chain and drainage channels. As a result, these rapidly growing tumours have extremely low levels of apoptosis and extremely high levels of cell division. In adults, angiogenesis is only switched on during physiological processes such as wound healing or menstruation, and then only transiently and regulated extremely carefully while in cancer cells angiogenesis process is switched on all the time. Counterbalancing positive (primary angiogenic growth factor being VEGF (Vascular Endothelial Growth Factor) causes endothelial cells to first break through existing blood vessels, then migrate towards the signal grow and replicate and form new blood vessels as directed) and negative signals (include the prototypical thrombospondin-1 protein, which inhibits the growth and migration of endothelial cells. The binding of thrombospondin-1 to endothelial cells triggers the activation of caspases, leading to apoptosis of the cell, circulate in the cellular microenvironment surrounding every cell, respectively encouraging or blocking angiogenesis. The rapid growth of a tumour results in an oxygen delivery problem; cells (hypoxic- activate the hypoxia stress response) found towards the inside of the tumour are deprived of an adequate supply of oxygen. Hypoxia inducible factor (HIF) transcription factor, initiating the production of other proteins required to mediate the effects of hypoxia. The accumulated HIF transports into the nucleus of the cell where it induces the expression of numerous target genes, including VEGF. This increase in VEGF results in shifting the balance of angiogenesis inducers and inhibitors, thereby switching on angiogenesis. By activating the angiogenic switch, tumours growth endothelial cells. Angiogenesis in a cancer is a perversion of a normal cellular process, a perversion that is an essential requirement for the development of cancer.

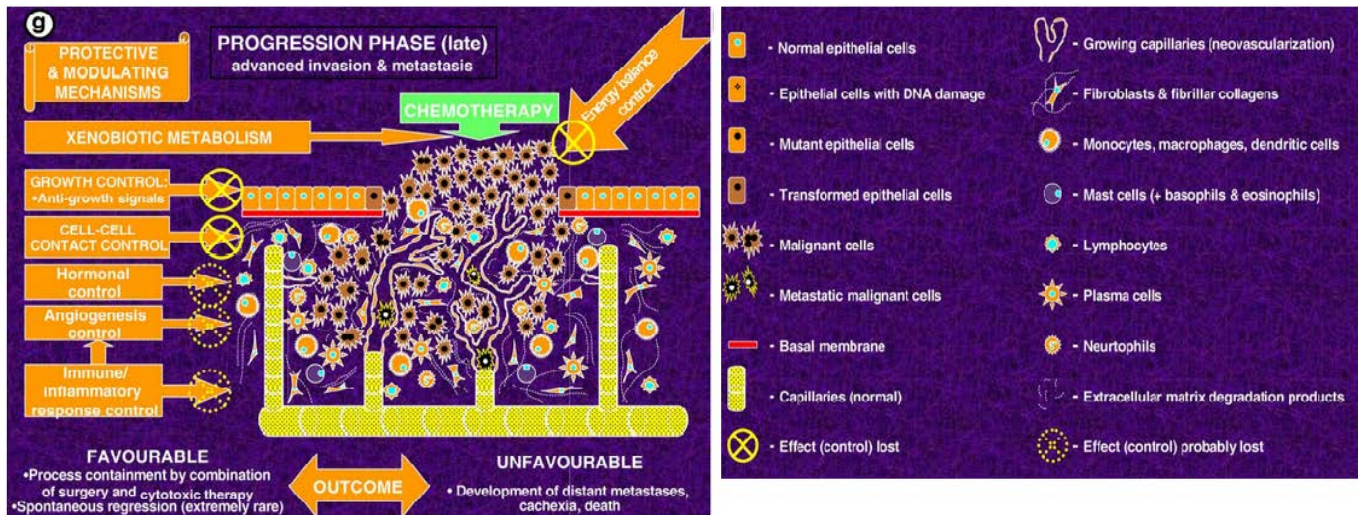
- Tissue invasion and metastasis (immune and inflammatory responses , extracellular matrix remodeling, angiogenesis and cell adhesion).

Primary tumor masses spawn pioneer cells that move out, effects adjacent tissues continue to travel to distant sites founding new colonies. Those distant settlements of tumor cells cause of 90% of human cancer deaths. There are several classes of proteins involved in invade surroundings tissue, include cell- cell adhesion molecules (CAMs). Cell-to-environment interactions in cancer involves E-cadherin expressed on epithelial cells. Adjacent cells by E-cadherin results in transmission of antygrowth via cytoplasmic contacts with  $\beta$ -catenin to intracellular signals circuits that include the Lef/ Tcf transcription factor. Several mechanisms

like mutations inactivation of E-cadherin or  $\beta$ -catenin, transcriptional repression, proteolysis of the extracellular cadherin domain cause of E-cadherin function lost and force expression of E-cadherin in cultured cancer cells. N-CAM switch in expression from a highly adhesive isoform to poorly adhesive forms in small cell lung cancer, it is an example of changes in expression of CAMs that appear to have a critical function in the processes of invasion and metastasis.

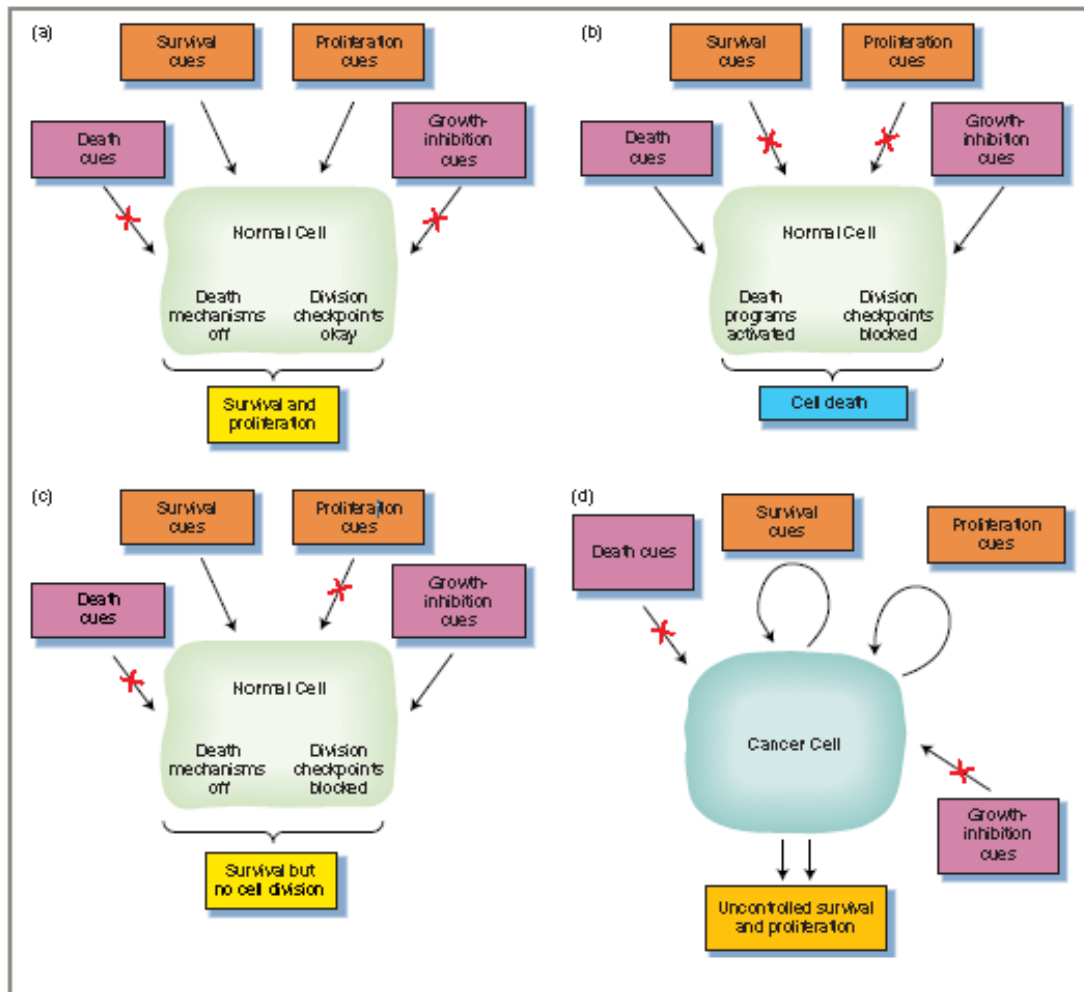
In figure 3.3.1 several types of protective/controlling mechanisms responding to tumour growth are shown. As describes above alterations in genes encoding regulatory proteins of major interconnected signalling pathways controlled by p53 that is reported for many tumours. It is very important to understand the role of individual genetic background as a factor in deregulation of cell proliferation, differentiation and death control. From the figure 3.3.1 that progression- early stage of cancer depends on these mechanism. Unfortunately tumours are never detected immediately after initiation. Most of the time diagnose is made when malignant progression continued, and invasion already started. For this reason researchers in terms of cancer risk investigate more and more polymorphisms in genes encoding factors of growth-controlling pathways rather than progression modulation.





**Figure 3.3.1.** Stage of tumour development and involved protective and modulating mechanisms affected by common gene polymorphisms.

The regulation of somatic-cell number is a splendid example of homeostasis—the sophisticated mechanisms that maintain an organism’s physiology within normal limits. A good example of homeostasis is the regulation of somatic cell number. When too many cells of a given type are present or when there is a deficiency of cells, cell proliferation or cell death accelerates or cell death respectively (figure 3.3.2). When a mutation appears in the genes governing those homeostatic mechanisms, the consequences are dramatic: the accumulation of multiple mutations accelerating proliferation and blocking cell death in the same somatic cell is the underlying cause of cancer [12].



**Figure 3.3.2.** Overview of the regulation of cell number in normal and cancer cells. (a) Proper external cues for normal cell survival and proliferation. (b) Proper external cues for normal cell death or inhibition of proliferation. (c) Proper external cues for normal cell survival without proliferation. (d) Self-generated survival and proliferation signals in cancer cells.

Globally, cancer is one of the top ten leading causes of death. It is estimated that 7.4 million people died of cancer in 2004 and, if current trends continue, 83.2 million more will have died by 2015 [13]. The burden of cancer is increasing in economically developing countries as a result of population aging and growth as well as, increasingly, an adoption of cancer-associated lifestyle choices including smoking, physical inactivity, and “westernized” diets. In 2008 about 12.7 million cancer cases

and 7.6 million cancer deaths are estimated to have occurred in of these, 56% of the cases and 64% of the deaths occurred in the economically developing world [14, 15].

Annual report in the US estimates that there were 2 million new cases diagnosed and 750 thousand deaths caused by cancer in 2014 in the US [16].

#### **4.4. Lung cancer**

Lung is the leading cancer site in males, comprising 17% of the total new cancer cases and 23% of the total cancer cases.

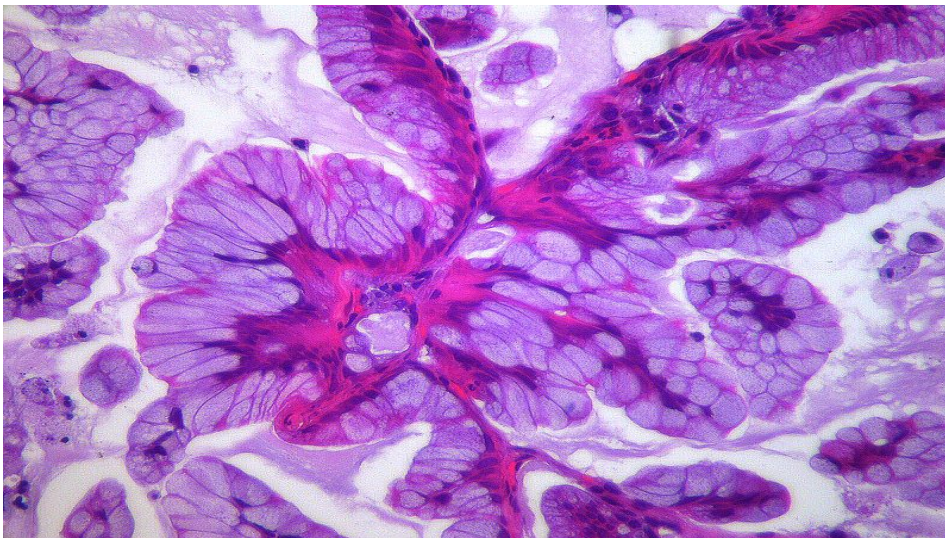
Lung cancer was the most commonly diagnosed cancer as well as the leading cause of cancer death in males in 2008 globally. Among females, it was the fourth most commonly diagnosed cancer and the second leading cause of cancer death. Lung cancer accounts for 13% (1.6 million) of the total cases and 18% (1.4 million) of the deaths in 2008. Despite their lower prevalence of smoking (less than 4% adult smokers), Chinese females have higher lung cancer rates (21.3 cases per 100,000 females) than those in certain European countries such as Germany (16.4) and Italy (11.4), with an adult smoking prevalence of about 20%. The relatively high burden of lung cancer in women is thought to reflect indoor air pollution from unventilated coal-fuelled stoves and from cooking fumes in China. Other known risk factors for lung cancer include exposure to several occupational and environmental carcinogens such as asbestos, arsenic, radon, and polycyclic aromatic hydrocarbons.

The observed variations in lung cancer rates and trends across countries or between males and females within each country largely reflect differences in the stage and degree of the tobacco epidemic. Smoking accounts for 80% of the worldwide lung cancer burden in males and at least 50% of the burden in females. The lung cancer rates are increasing in countries such as China and several other countries in Asia and Africa, where the epidemic has been established more recently and smoking prevalence continues to either increase or show signs of stability [16].

#### 4.4.1. Non-small cell lung carcinoma

Lung cancer consists of two types: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). In this work all patients recruited had NSCLC which corresponds to about 80-90% of lung cancer cases. There are 3 main subtypes of NSCLC. The cells in these subtypes differ in size, shape, and chemical make-up when looked at under a microscope. But they are grouped together because the approach to treatment and prognosis are often very similar.

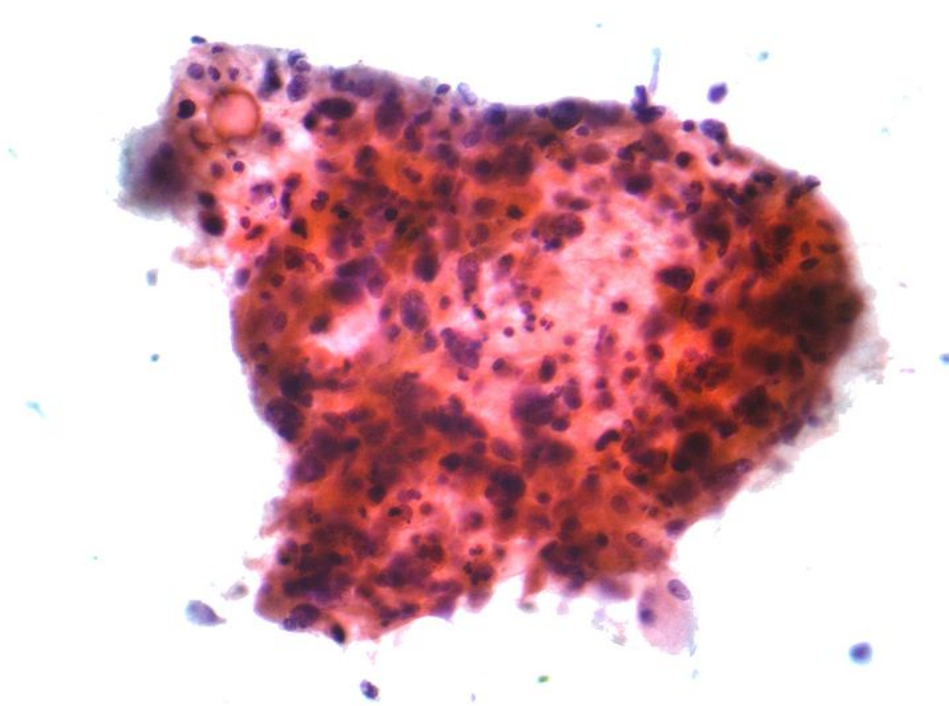
- Adenocarcinoma: About 40% of lung cancers are adenocarcinomas. These cancers tend to stain mucin positive as they are derived from the mucus producing glands of the lungs. This type of lung cancer occurs mainly in current or former smokers, but it is also the most common type of lung cancer seen in non-smokers. It is more common in women than in men, and it is more likely to occur in younger people than other types of lung cancer. Adenocarcinomas are highly heterogeneous tumours [17]. Adenocarcinomas are usually found in outer parts of the lung. They tend to grow slower than other types of lung cancer, and is more likely to be found before it has spread outside of the lung. People with a type of adenocarcinoma called adenocarcinoma in situ (previously called bronchioloalveolar carcinoma) tend to have a better prognosis than those with other types of lung cancer [18].



**Picture 3.4.1.** Squamous carcinoma lung

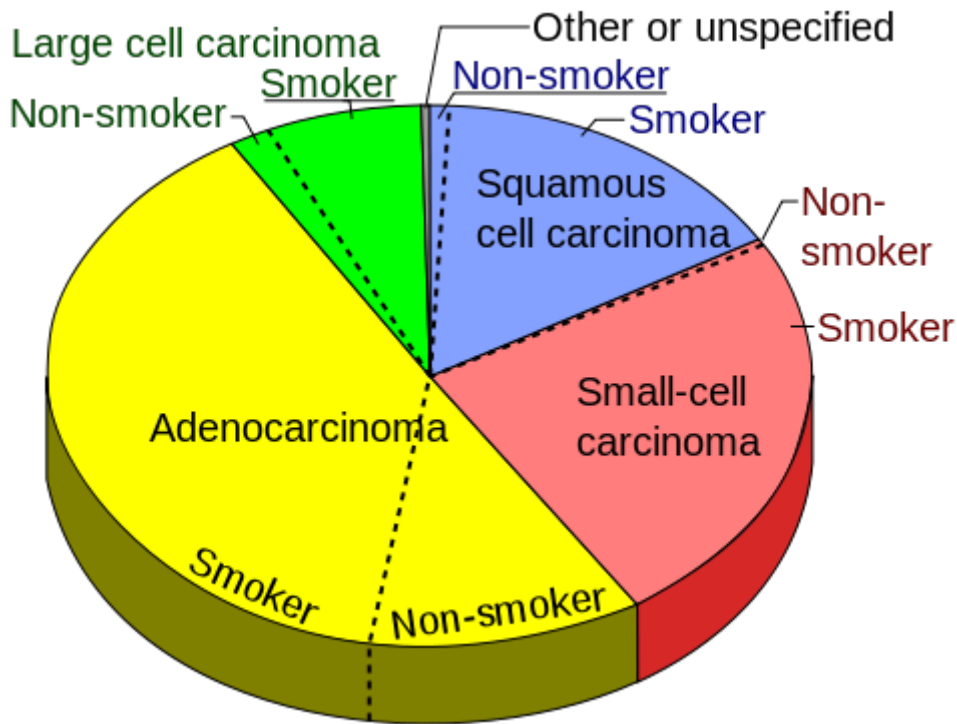
- Squamous cell (epidermoid) carcinoma: About 25% to 30% of all lung cancers are squamous cell carcinomas. These cancers start in early versions of squamous cells,

which are flat cells that line the inside of the airways in the lungs. They are often linked to a history of smoking and are more common in men than in women. It tends to be found in the middle part of the lung, it often metastasizes to loco regional lymph nodes early in its course and generally disseminates outside the thorax somewhat later than other major types of lung cancer [19].



**Figure 3.4.2** Squamous carcinoma lung- cytology.

- Large cell (undifferentiated) carcinoma: This type of cancer accounts for about 10% to 15% of lung cancers. It can appear in any part of the lung. It tends to grow and spread quickly, which can make it harder to treat. A subtype of large cell carcinoma, known as large cell neuroendocrine carcinoma, is a fast-growing cancer that is very similar to small cell lung cancer [20].
- Other subtypes: There are also a few other subtypes of non-small cell lung cancer, such as adenosquamous carcinoma and sarcomatoid carcinoma. These are much less common.



**Figure 3.4.3** Pie chart of types of lung cancer, considering the smoking status.

The main treatment methods for lung cancer are surgery, radiation, platinum-based doublet chemotherapy and tyrosine kinase inhibitor (TKI). Unfortunately most patients are being diagnose in a late stage and lost the opportunity of surgery. Platinum –based regimens in overall survival and quality of life shown meagre improvement. In advanced NSCLC median overall survival is still 10-12 months, but in some molecularly defined subgroups to have much better prognosis. Scientist constantly work on improvement in risk of lung cancer and overall survival by study single nucleotide polymorphism (SNP) and by measure and evaluate molecularly defined biomarkers [21].

#### 4.4.2. Stages of cancer

If a biopsy shows cancer, more imaging tests are done to find out the stage of the cancer. Stage means how big the tumour is and how far it has spread. NSCLC is divided into five stages:

- Stage 0** - the cancer has not spread beyond the inner lining of the lung
- Stage I** - the cancer is small and has not spread to the lymph nodes

•**Stage II** - the cancer has spread to some lymph nodes near the original tumour

•**Stage III** - the cancer has spread to nearby tissue or too far away lymph nodes

•**Stage IV** - the cancer has spread to other organs of the body, such as the other lung, brain, or liver

#### **4.4.3. Stages, diagnose and spread of Non-Small Cell Lung Cancer**

After lung cancer has been diagnosed, it is important to know the stage in order to plan treatment. Some of the tests used to diagnose non-small cell lung cancer are also used to stage the disease include: MRI (magnetic resonance imaging), CT scan (CAT scan), PET scan (to find malignant tumour cells), radionuclide bone scan, pulmonary function test (PFT), endoscopic ultrasound (EUS), mediastinoscopy (to look at the organs, tissues, and lymph nodes between the lungs), anterior mediastinotomy (to look at the organs and tissues between the lungs and between the breastbone and heart), lymph node biopsy, bone marrow aspiration and biopsy.

Unfortunately cancer can spread very fast through tissue, the lymph system, and the blood. When the primary tumour spread to another part of the body is called metastasis.

The following stages are used for non-small cell lung cancer (Figure 3.4.3.1):

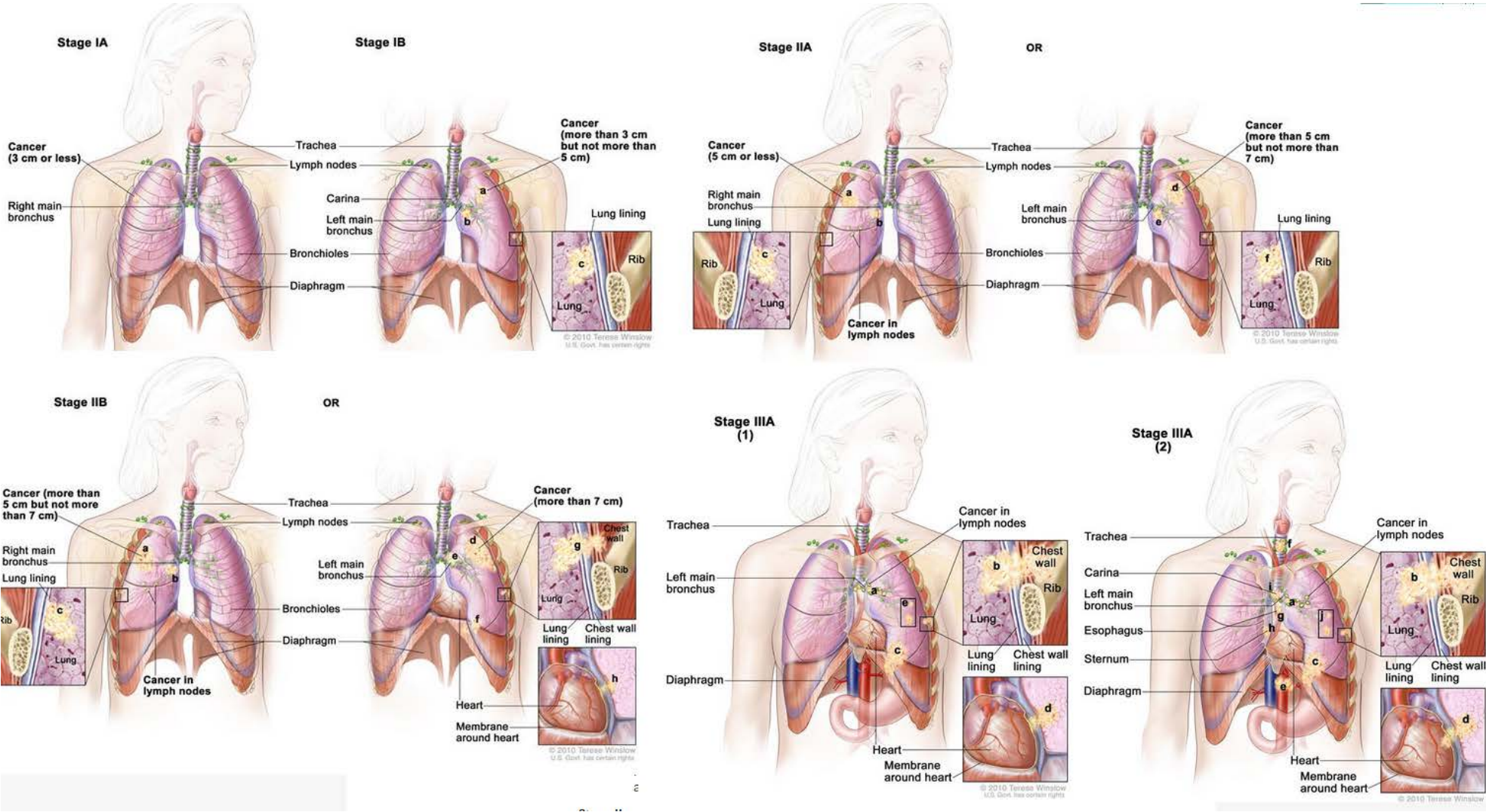
- Occult (hidden) stage

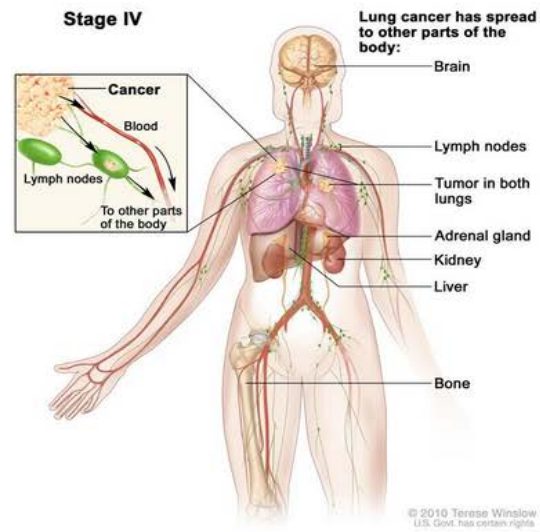
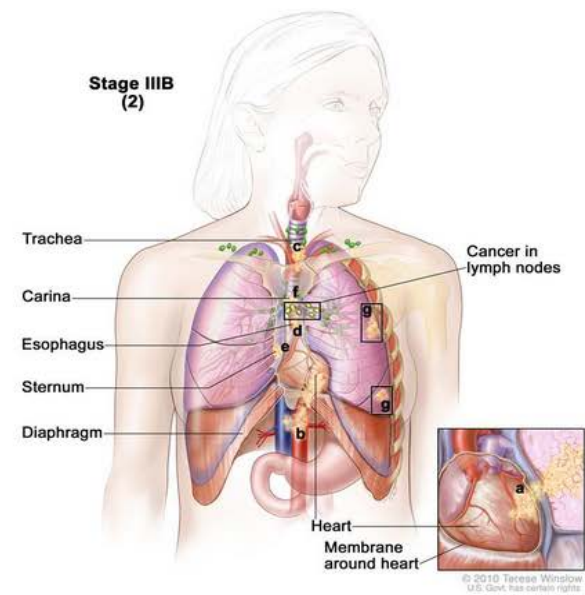
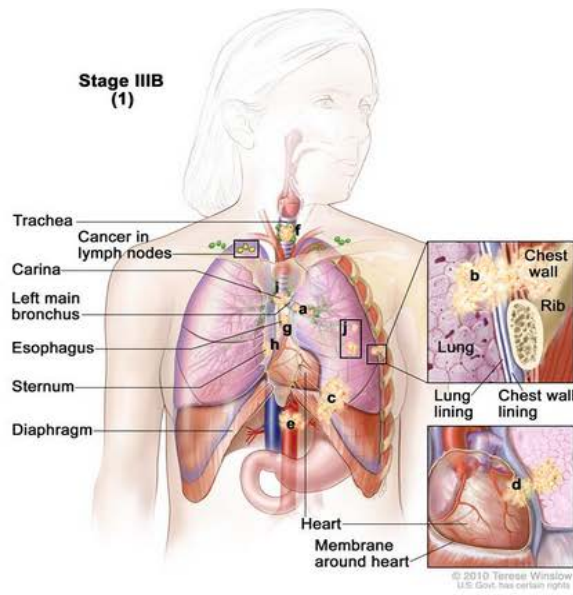
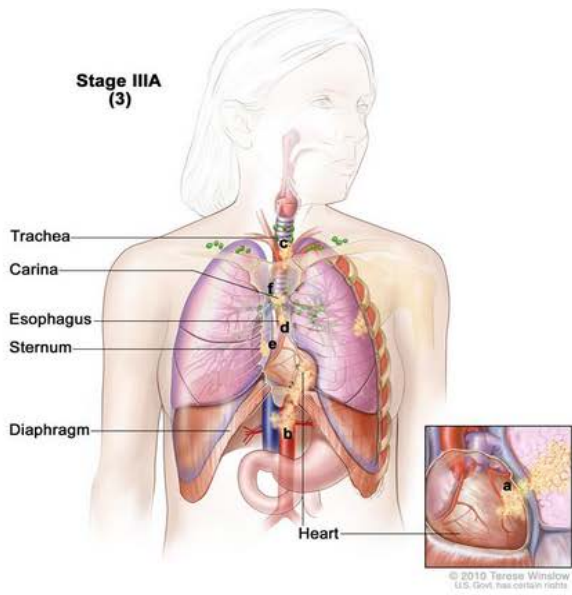
In the occult (hidden) stage, cancer cannot be seen by imaging or bronchoscopy. Cancer cells are found in sputum (mucus coughed up from the lungs) or bronchial washing (a sample of cells taken from inside the airways that lead to the lung). Cancer may have spread to other parts of the body.

- Stage 0 (carcinoma in situ)

In stage 0, abnormal cells are found in the lining of the airways. These abnormal cells may become cancer and spread into nearby normal tissue. Stage 0 is also called carcinoma in situ.

**Figure 3.4.3.1** Illustrations of stages in Non-Small Cell Lung Cancer Patients





## Stage I

- Stage IA: The tumour is in the lung only and is 3 centimetres or smaller.
- Stage IB: Cancer has not spread to the lymph nodes and there are few more possibilities:
  - a. The tumour is between 3 and 5 centimetres.
  - b. Cancer developed in main bronchus and is at least 2 centimetres below where the trachea joins the bronchus.
  - c. Cancer attached to the innermost layer of the membrane that covers the lung.
  - d. Part of the lung has collapsed or developed pneumonitis

## Stage II

- Stage IIA (1) Cancer has spread to lymph nodes on the same side of the chest as the tumour. The lymph nodes with cancer are within the lung or near the bronchus. The tumour is not larger than 5 centimetres.
- Stage IIA (2) Cancer has not spread to lymph nodes. The tumour is larger than 5 centimetres but not larger than 7 centimetres.

### Common option for IIA (1 and 2):

- a. Cancer has spread to the main bronchus and is at least 2 centimetres below where the trachea joins the bronchus.
- b. Cancer has spread to the innermost layer of the membrane that covers the lung.
- c. Part of the lung has collapsed or developed pneumonitis (inflammation of the lung) in the area where the trachea joins the bronchus

## Stage IIB

- Stage IIB (1) Cancer has spread to nearby lymph nodes on the same side of the chest as the tumour. The lymph nodes with cancer are within the lung or near the bronchus. The tumour is larger than 5 centimetres but not larger than 7 centimetres.
- Stage IIB (2) Cancer has not spread to lymph nodes. The tumour is larger than 7 centimetres.

### Common options for Stage II B (1 and 2)

- a. Cancer has spread to the main bronchus (and is less than 2 centimetres below where the trachea joins the bronchus), the chest wall, the diaphragm, or the nerve that controls the diaphragm.
- b. Cancer has spread to the membrane around the heart or lining the chest wall.
- c. The whole lung has collapsed or developed pneumonitis (inflammation of the lung).
- d. There are one or more separate tumours in the same lobe of the lung.

### Stage III

#### Stage IIIA

- IIIA (1) Cancer has spread to lymph nodes on the same side of the chest as the tumour. The lymph nodes with cancer are near the sternum (chest bone) or where the bronchus enters the lung.
- IIIA (2) Cancer has spread to lymph nodes on the same side of the chest as the tumour. The lymph nodes with cancer are within the lung or near the bronchus.
- IIIA (3) Cancer has not spread to the lymph nodes and the tumour may be any size. Cancer has spread to any of the following:
  - a. Heart.
  - b. Major blood vessels that lead to or from the heart.

#### Common option for spreading cancer in stage IIIA (1 and 2)

The tumour may be any size.

Part of the lung (where the trachea joins the bronchus) or the whole lung may have collapsed or developed pneumonitis (inflammation of the lung).

- a. There may be one or more separate tumours in the same lobe of the lung.
- b. Cancer may have spread to any of the following:
- c. Main bronchus, but not the area where the trachea joins the bronchus.
- d. Chest wall.
- e. Diaphragm and the nerve that controls it.
- f. Membrane around the lung or lining the chest wall.

- g. Membrane around the heart.

Common option for spreading cancer in stage IIIA (2 and 3)

- a. Trachea.
- b. Esophagus.
- c. Nerve that controls the larynx (voice box).
- d. Sternum (chest bone) or backbone.
- e. Carina (where the trachea joins the bronchi).

Stage IIIB

- Stage IIIB (1) Cancer has spread to lymph nodes above the collarbone or to lymph nodes on the opposite side of the chest as the tumour. Part of the lung (where the trachea joins the bronchus) or the whole lung may have collapsed or developed pneumonitis (inflammation of the lung).
- Stage IIIB (2) Cancer has spread to lymph nodes on the same side of the chest as the tumour. The lymph nodes with cancer are near the sternum (chest bone) or where the bronchus enters the lung. There may be separate tumours in different lobes of the same lung.

Common other symptoms of stage III B (1 and 2)

- a. The tumour may be any size.
- b. There may be one or more separate tumours in any of the lobes of the lung with cancer.
- c. Cancer may have spread to any of the following:
- d. Main bronchus.
- e. Chest wall.
- f. Diaphragm and the nerve that controls it.
- g. Membrane around the lung or lining the chest wall.
- h. Heart or the membrane around it.
- i. Major blood vessels that lead to or from the heart.
- j. Trachea.
- k. Esophagus.
- l. Nerve that controls the larynx (voice box).
- m. Sternum (chest bone) or backbone.

n. Carina (where the trachea joins the bronchi)

- Stage IV

In stage IV, the tumour may be any size and cancer may have spread to lymph nodes. Unfortunately there might be other options like:

- There are one or more tumors in both lungs.
- Cancer is found in fluid around the lungs or the heart.
- Cancer has spread to other parts of the body, such as the brain, liver, adrenal glands, kidneys, or bone [22].

#### **4.5. Platinum- based chemotherapy**

Surgery is the mainstay of treatment for early stage and localised disease (Stage I and II and selected IIIA). Following lung resection for NSCLC five year survival rates are I A-73% I , B-54%, II A-48%, II B-38%, IIIA-25% [6]. The majority of patients have advanced disease at the time of diagnosis and therefore are not surgical candidates, in the United Kingdom only 14% of patients diagnosed go on to have surgical resection. In those patients that are surgical candidates, more than 50% will develop a recurrence. Adjuvant chemotherapy has been used with limited success to decrease the recurrence rates but this has only yielded a survival benefit of 5%-15%. Due to advanced disease or recurrence following resection, different chemotherapy forms were proposed for patients. In the first line setting the most effective is platinum- based combination.

Unfortunately, in a recent trial of platinum combination therapies in advanced NSCLC, only 30% of patients showed objective disease response and a significant proportion suffered toxic side-effects such as neutropenia (27%), anaemia (10%), thrombocytopenia (13%), alopecia (21%) and nausea (4%). During the trial, deaths due to study drug toxicity were registered. This demonstrates the major problem of drug resistance in NSCLC to standard platinum based therapies and the associated toxicities [23].

Survival has been improved when cisplatin is given postoperatively to those with stage II or IIIA disease, with radiotherapy in stage IIIB disease, or alone in stage IV disease. Cisplatin is one of the most active agents in NSCLC with an overall response rate of about 20%. More recently it has been shown that various combinations of cisplatin or carboplatin, in particular the combination of cisplatin and paclitaxel, but also combinations with gemcitabine, docetaxel, navelbine, and irinotecan can also prolong survival in NSCLC [24].

Platinum based therapy refers to a specific class of cytotoxic chemotherapy drugs that are derived from the element platinum. The antitumor properties of platinum-containing drugs are attributable in large measure to the kinetics of their ligand displacement reactions. Their primary target is believed to be nitrogen donor atoms in the bases of DNA. The bonds formed between the metal ion and these atoms must be sufficiently long-lived to interfere with the process of cell division, or to trigger the intracellular mechanisms that recognize irreparable damage to a cell. Bonds between the base nitrogen atoms and platinum clearly fulfil this requirement.

Cisplatin (cis-diamminedichloroplatinum II) is a relatively unreactive molecule, it does not react directly with any other molecules present in biological systems that will bind to the platinum through nitrogen or oxygen donor groups,  $N^7$  position of guanine and adenine in intact DNA. Hydrolysis of cisplatin is usually considered to be a necessary prelude to its reaction with DNA. The aqua complexes also react more readily than cisplatin with molecules present in vivo other than target DNA. They therefore are involved in reactions that lead to toxicity. Cisplatin is usually given at a dosage of 50-120 mg  $m^{-2}$  per cycle.

Carboplatin is less toxic than cisplatin but also has antitumor activity on an equal-dose basis. These properties are correlated with its much lower tendency to undergo hydrolysis reactions- indeed, the rate of hydrolysis is negligibly slow under physiological conditions. Carboplatin does react directly with guanosine monophosphate, but only at slow rates. The dose is usually customised for each patient using the area under the concentration-time curve (AUC 200-350mg  $m^{-2}$ ) and renal function of the patient, because this drug undergoes extensive renal excretion. Cisplatin and carboplatin do react directly with sulphur-containing ligands (e.g., methionine, cysteine) without the need for prior hydrolysis [25]. Both agents are usually given every 3-4 weeks, for 3-6 cycles. However, because of the toxicity of cisplatin, less toxic platinum alternatives have been discovered. Two drug combination cisplatin plus paclitaxel, vinorelbin, gemcitabine or docetaxel and, carboplatin plus paclitaxel were compared with

monotherapy in terms of response rate, survival rate, efficiency and toxicity. There were not significant different between monotherapy and combination drugs between platinum agent. When compared gemcitabine (non-platinum agent) with gemcitabine plus cisplatin showed a significant advantage in response rate, survival distribution and longer time to progression [26].

#### **4.5.1. Gene polymorphisms affecting outcome of cancer therapy (cancer pharmacogenetics)**

Drug-related toxicity depends on the genotype of no tumour tissue which means that inherited polymorphisms will have a key role with respect to toxicity, a crucial dose-limiting factor in most cancer chemotherapy regimens.

Many of the variant forms of drug-metabolizing enzymes show small deviation from wild-type enzyme activity. It is possible that the combination of several polymorphisms in components of a 'biological' pathway or 'pharmacological' pathway might significantly influence therapeutic response (significant changes in the ability to metabolize drugs)

It is well understood the importance of gene variants encoding enzymes of the metabolic pathways interacting with anticancer drugs.

The main target of assessing genetic background in cancer patients (mostly those with advanced tumours) under-going cytotoxic chemotherapy or radiotherapy is to find out information allowing to select therapeutic strategies rendering maximal damage to target tumour cells with minimal side effects and eventually leading to improved survival and comfort of life. Polymorphic genes mainly being research in problem in cancer therapy: general response to anticancer drugs (toxicity and its modulation) and modulation of specific responses (sensitivity/resistance) of tumours to anticancer drugs defining the efficacy of therapeutic procedures.

Recently developed '-genomic approaches', which use microarray technology to describe gene or protein expression in its entirety will certainly be useful in refining cancer diagnoses and, in turn, predicting tumour response to specific drugs [27].

### 4.5.2. Platinum Resistance

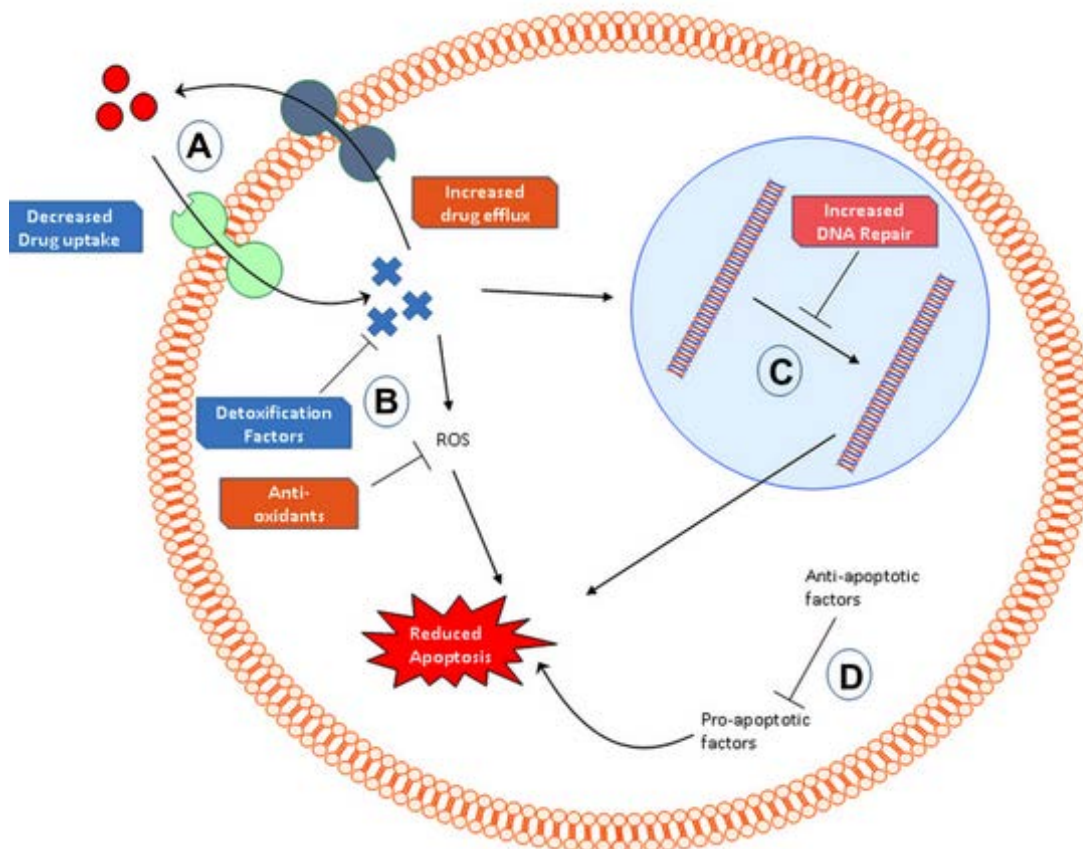
Platinum agents are used as an effective and successful chemotherapeutics, unfortunately majority of patients develop drug resistance and metastases. Recent study provide novel target for therapy. There are several factors involved in the drug resistance development, including gene mutations, genome alterations and epigenetic changes.

It has become a great challenge in response of lung cancer patients the development of innate ad acquired resistance. Cisplatin is taken up through a combination of passive diffusion and active transport via membrane bound transporters such as Na<sup>+</sup>, K<sup>(+)</sup>-ATPase1 and several copper transporters such as copper transporter receptor 1 (CTR1). It have been development there is a correlation between cisplatin resistance and expression level of these transporters. To yield an active compound of cisplatin by series of aquation reactions, with one or both chlorides being replaced by a water inside the cell.

Activated cisplatin is a highly reactive compound, capable of binding DNA, RNA, proteins and membrane phospholipids. The primary cytotoxic mechanism of cisplatin is the formation of DNA adducts. The majority of adducts formed by cisplatin cross links between adjacent purine bases represent the vast and the 1, 2-intrastrand. DNA protein cross links with consist of 1, 3-intrastrand. Formed adducts changed structure of the DNA molecule leading to steric changes in the DNA helix. Changes in DNA structure are recognised and repaired by complex network of DNA repair pathways, terminating in the destruction of the cell. DNA repair mechanisms show differences in tumour cells to normal counterparts, and some of them may also be acquired, the role of this DNA repairs in cisplatin resistance tumours is investigating. Tumour cell lines with acquired resistance to cisplatin have been shown to have an increased capacity for removing cisplatin- induced DNA lesions in comparison to their cisplatin sensitive counterparts. The tumour suppressor gene, p53 is very important in damage recognition and initiation of repair and binds with a high affinity to palatinate DNA. The studies show that loss of p53 confer cisplatin resistance in ovarian cancer, whilst restoration of wild type p53 via adenoviral transfection can restore sensitivity and promote apoptosis.

Researchers have demonstrated interactions between p53 signalling and cisplatin in lung cancer cells [28].

Cisplatin resistant phenotype observed in the different cancer types (most in lung cancer) by contribution of multiple mechanisms and pathways in tumour cells (figure 3.5.2.1). These include, decreased intracellular drug accumulation due to alterations in expression of receptors such as CTR1 and the copper-transporting P-type adenosine triphosphates, ATP7A and ATP7B (A), increased inactivation of cisplatin by intracellular thiol-containing molecules such as glutathione (GSH) and metallothionein (B), increased DNA repair due the enhanced capacity of tumour cells to remove cisplatin-induced DNA lesions (C), suppressed caspase activity and reduced apoptotic response as a result of alterations in pro-apoptotic factors such as Bax, Bak and apoptosis inhibitors such as Bcl-2 (D).



**Figure 3.5.2.1** Mechanism of cisplatin resistance in tumour cells.

Inhibitors of histone deacetylases HDAC is been recently propose for treatment NSCLC patients. Performed study was to examine the capability of the pan-histone deacetylase inhibitor SAHA and of ST35595, a novel hydroxamate-based compound, to interfere with the proliferative and invasive potential of NSCLC cells. The results confirmed an improvement in

terms of pro-aprotic, anti-invasive activity *in vitro* and antimetastatic activity *in vivo* in ST35595 over SAHA [29].

#### **4.6. SNP (Single Nucleotide Polymorphism)**

The term *polymorphism* has been defined in 1986 as a ‘Mendelian trait that exists in the population in at least two phenotypes, neither of which occurs at a frequency of less than 1%’ [30].

It is generally believed that the genomes between two randomly selected individuals contain approximately 0.1% difference or variation. This variation is called ‘polymorphism’ and it arises because of mutations. The simplest form of DNA variation among individuals is the substitution of one single nucleotide for another. This type of change is called ‘single nucleotide polymorphism’ (SNP) and it is found to be more common than other types of polymorphisms. It is estimated that SNPs occur at a frequency of approximately one in 1,000 base pairs (bp) throughout the genome and more than three million SNPs have been charted so far. These simple changes are believed to be stable and not deleterious to organisms. According to a published report, 50% of SNPs occur in noncoding regions, 25% lead to missense mutations, and the remaining 25% are silent mutations. No synonymous SNPs (altering amino acids) may often produce disease and therefore may be subjected to natural selection. SNPs can be observed between individuals in a population, and may, for example, influence promoter activity (gene expression), messenger RNA (mRNA) conformation (stability), and translational efficiency. Therefore, they may be responsible for the susceptibility of an individual to many common diseases, therapeutic drug metabolism, and genome evolution. They may also play a direct role with or without other factors in the phenotypic expression of diseases or traits such as tallness, curly hair, and individuality. In recent years, application of clinic genetic knowledge has revolutionized our ability to understand the effects of nucleotide substitutions and genetic basis of several complex and common disorders [31].

## 4.7. Gene

The man is a carrier of an average of 10-20 genes that predispose to serious illness. However, this does not mean that you have to get sick.

Nearly all cancers are due to gene, because before the disease develops, it must result in damage to a particular gene. Genetic defect that makes the cells begin to multiply uncontrollably and do not start the development of a restraining mechanism. This does not mean, however, that bad genes can be only in the genetic material inherited from the ancestors. Sometimes they are damaged by e.g. Smoking or inhalation of toxic fumes (lung cancer), or too frequent eating red meat (colorectal cancer). The risk in people under 30 years is 0.5%. Then - for 65 years - it increases to 12%, and even old age reaches 23%. The so-called lifetime risk of developing cancer in all people, irrespective of the family burden is assessed at 33%. In contrast, hereditary conditions can speak in approx. 5% of all cancers.

In inherited cancers involving oncogenes, tumour suppressor genes and the genes responsible for the repair of DNA damage. Oncogenes cause uncontrolled division and proliferation of cells. It have been described for more than 50 of such genes. They arise from mutations of proto-oncogenes, or normal genes, which have each cell. Suppressor genes supervise the growth and cell division (not dangerous in normal cell). Only they damage due to the mutation causes impairment of their function, and it promotes the formation of cancer. The genes responsible for reconstruction repair of damaged DNA, e.g. under the influence of radiation. When they are defective, mutated genes, which can lead to cancer development [32].

### 4.7.1. PDCD6IP (The programmed cell death 6 interacting protein)

Programmed cell death (PCD) is a physiologically regulated cell type-specific deletion that takes place during various developmental stages of multicellular organisms and is essential for the establishment and the maintenance of cellular homeostasis. PCD occurs by apoptosis, which refers to the morphological changes that can be observed in cells undergoing PCD. These include plasma membrane blabbing, cell shrinkage, chromatin condensation, and DNA degradation. The process eventually culminates with the fragmentation of the cells into

apoptotic bodies that, *in vivo*, are rapidly phagocytized by the surrounding cells. Programmed cell death has been recognized as an important terminal pathway for cells of multicellular organisms, and is involved in a variety of biological events that include morphogenesis, maintenance of tissue homeostasis, and elimination of harmful cells [33].

This gene encodes a protein that functions within the endosomal sorting complex required for transport ESCRT pathway in the abscission stage of cytokinesis, in intraluminal endosomal vesicle formation, and in enveloped virus budding. Previous studies have shown that overexpression of this protein can block apoptosis. Apoptosis play a very important role in cancer treatment and in carcinogenesis (describe before). Some kind of damage trigger a series of biochemical events, leading to characteristic cell morphology and death. In addition, the product of this gene binds to the product of the programmed cell death 6 (PDCD6) gene, a protein required for apoptosis, in a calcium-dependent manner. This gene product also binds to endophilins, proteins that regulate membrane shape during endocytosis [34].

Overexpression of this gene product and endophilins result in cytoplasmic vacuolization, which may be partly responsible for the protection against cell death. Several alternatively spliced transcript variants encoding different isoforms have been found for this gene [35].

Dysfunction in an important terminal pathway leads to various cancers in humans. The PDCD6IP gene is a plausible cancer susceptibility gene. The localisation of this gene is in chromosome 3p22 harbouring several sequence variations in the promoter region. mRNA levels of target genes may be affected by genetic variations in the promoter region by altering transcription factor (TF) binding sites. A polymorphism (rs28381975) was found in the promoter region of PDCD6IP, which could disrupt the binding patterns of the c-rel transcription factor. C-rel is a key nuclear factor-kappaB (NF-kB) subunit that plays an important role in cell development, proliferation, differentiation, and protection against apoptosis. In the study on rs28381975 it was hypothesized that the function of the gene might be affecting the bindings of specific TFs, thus involved in Hepatocellular Carcinoma HCC tumorigenesis [36].

In this work we hypothesized that this reactions maybe be involved in NSCLC tumorigenesis and the association between rs3183982 and the Platinum based chemotherapy in a Chinese

population. It is believed that understanding connections between SNP and response rate and drug resistant are very important for future improvement of a treatment.

The ESCRT pathway is a key mediator of multivesicular body MVB biogenesis but also plays critical roles in retroviral budding, cytokinesis abscission, receptor down-regulation, and other normal and pathological processes. The ESCRT components are conserved in all five major subgroups of eukaryotes. Studies shown growing number of links identified between ESCRT-mediated protein sorting in the MVB (multivesicular body) pathway and various human diseases. MVB deliver cargo destined for degradation to the vacuole or lysosome and plays an important role in modulating the amplitude and kinetics of the signalling pathways they activate and thus the maintenance of normal cellular homeostasis.

The endocytic down-regulation of numerous signalling receptors such as the EGFR (epidermal growth factor receptor) plays an important role in modulating the amplitude and kinetics of the signalling pathways they activate and thus the maintenance of normal cellular homeostasis. EGFR is one of the best-studied receptor tyrosine kinases RTKs, and its excessive signalling is associated with the development of a variety of human cancers, including mammary carcinomas, squamous carcinomas and glioblastomas, as well as other malignant diseases. The MVB pathway terminates receptor signalling via lysosomal degradation of the receptor, and thus plays an important role in modulating the amplitude and kinetics of various signalling pathways from activated receptors [37].

#### **4.8. Polymerase chain reaction (PCR)**

Among the molecular biology techniques most commonly used in biotechnological research aimed to analyze the state of the environment concerned, there may be mentioned:

- polymerase chain reaction (PCR),
- gel electrophoresis and its variants,
- cloning
- DNA fragments,
- DNA sequencing
- Southern blot hybridization,
- Northern blot hybridization,

- in situ hybridization, FISH,
- GFP protein,
- DNA microarrays.

DNA Polymerase Chain Reaction, allows to quickly replicate the nucleic acid segment. Although this is very sensitive and efficient, and thanks to its simplicity regarded today as the most important tool in molecular biology. Within a few hours may be amplified fragments of the starting DNA  $10^6 - 10^9$ . Before conducting the PCR reaction mixture need to be prepared consisting of: the material for amplification, which is called the matrix DNA, primer, deoxynucleotides (deoxynucleoside triphosphates dNTPs), thermostable DNA polymerase enzyme and buffer. The primers are DNA fragments that are complementary to sequences flanking the DNA segment searched. Most primers selected on the basis of the data-connected resident in the literature or are constructed based on the sequence available in the gene bank. The reaction mixture is periodically (25-40 cycles), heated to a time-dependent temperature phase of the cycle.

The main steps of the reaction are depicted in figure 3.8.1:

- Denaturation of DNA molecules
- Primer binding (annealing )
- DNA chain elongation ( extension )

These steps are repeated cyclically, resulting in exponential growth of the reaction product (reaction product of each template in the next cycle of PCR).

When denatured by the high temperature DNA is interleaved to a single strand, which in the next step hybridize with the primer pair. Then DNA polymerase on each rebuilding missing strand complementary DNA segment, starting from its 3'. Joining repeated in each reaction cycle and consequently receive the amplified fragment located between the primers. Each of the stages of the cycle has a different temperature, therefore, the PCR reaction is performed in a device called a thermal cycler, wherein the temperature change occurs tens of degrees in a few seconds. During the first cycle of each single stranded DNA segment produced another piece, so that after the first cycle, the number of fragments is doubled, in a further amount of DNA is four times greater and so on.

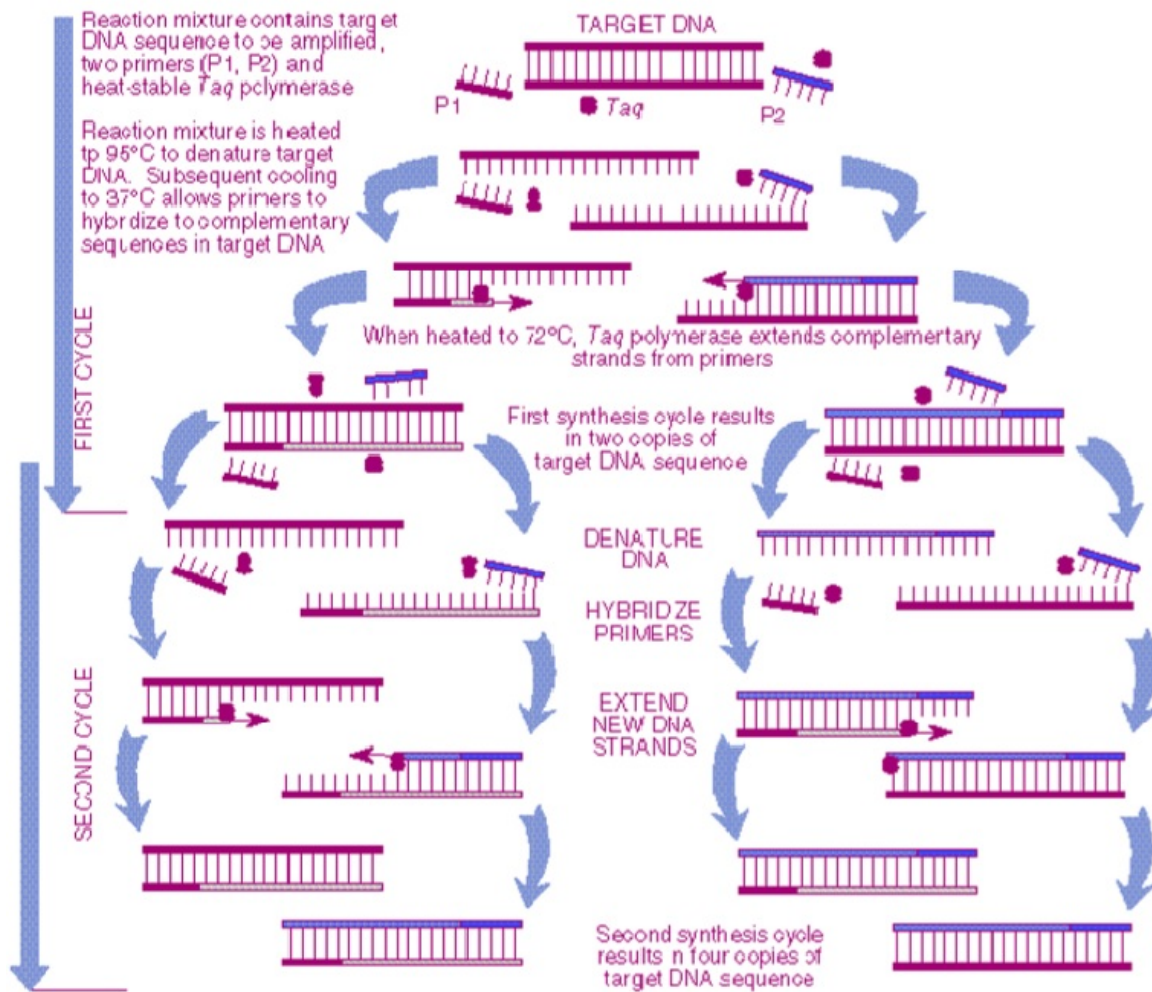
The most important advantage of PCR is the possibility of obtaining multiple copies of genes fragments with a small amount of genetic material. PCR products are the basis for further analysis and electrophoretic techniques for sequencing, which in the case of environmental samples, where we are dealing with a mixture of organisms, must be preceded by cloning.

It should also be noted that the use of PCR to environmental samples may be associated with obstacles resulting from the presence of inhibitors which may cause weakening or complete inhibition of the amplification reaction. Inhibitors can be inorganic and organic compounds of different origin (e.g. Polysaccharides, urea or humic acids) [38].

There are at least a few modifications to the PCR, offering new opportunities for research below few listed:

- "Multiplex PCR" PCR Real-time PCR
- PT-PCR
- PFGE
- DGGE / TGGE
- Horizontal agar
- Polyacrylamide
- SDS-PAGE

## DNA Amplification Using Polymerase Chain Reaction



**Picture 3.8.1.** Illustration of all PCR steps.

### 4.8.1. Primers

The nucleotide sequence of the primers is complementary to the initial portion of the DNA strand amplified (forward) and in accordance with amplified propagated terminal fragment of the DNA strand (reverse). The primers usually have a length of 15-25 nucleotides, it is important that their nucleotide sequence is unique, i.e. not repeated in the test genome. The primers should contain 40-60 % G and C nucleotides, and should not contain inverted repeats. Melting point ( $T_m$ ) of the double stranded molecule formed by the two primers should be similar (the difference  $T_m < 5^\circ \text{C}$ ), primer terminal nucleotides must be G or C, but not more than one. The primers should not be complementary to one another

(max. 3 consecutive nucleotides can be complementary), or contain stretches of purines or pyrimidines [39].

#### **4.8.2. PCR-RFLP (Restriction Fragment Length Polymorphism)**

PCR-restriction fragment length polymorphism (RFLP)-based analysis is a popular technique for genotyping.

RELP is a fragment of DNA of predictable size resulting from digestion (cutting) of a strand of DNA by a given restriction enzyme. DNA sequence alterations (mutations) that destroy or create the sites at which a restriction enzyme cuts DNA change the size (and number) of DNA fragments resulting from digestion by a given restriction enzyme.

The technique exploits that SNPs, MNPs and microindels often are associated with the creation or abolishment of a restriction enzyme recognition site. The first step in a PCR-RFLP analysis is amplification of a fragment containing the variation. This is followed by treatment of the amplified fragment with an appropriate restriction enzyme. Since the presence or absence of the restriction enzyme recognition site results in the formation of restriction fragments of different sizes. The length of DNA fragments produced after treatment with a restriction enzyme to DNA is determined on the basis of electrophoretic separation in agarose gels in the presence of the size markers.

The main advantage is inexpensive. It is Easy to design and there are no requirements for expensive instruments and extensive training of laboratory staff. PCR-RELP is applicable to analysis of single nucleotide polymorphisms as well as microindels. With all advantages always comes disadvantages includes: relatively large amounts of hand-on-time, long time to complete the analysis and it is not suitable for high-throughput analysis. Requires that a variation generates or abolishes a restriction enzyme (which can be expensive) recognition site. Moreover, since PCR-RFLP consists of several steps including an electrophoretic separation step, it is relatively time-consuming. Exact genotyping cannot be achieved in the event that there is more than one nucleotide variation in a restriction enzyme recognition site [40, 41].

### 4.8.3. Real-time PCR)

The reaction Real-time PCR amplification reaction of DNA, the quantity of which is monitored during the course of the reaction due to the presence in the reaction mixture of dyes or fluorescent probes. Dyes connecting with the nucleic acid emit a fluorescent signal proportional to the quantity of DNA. The advantages of this technique include high sensitivity and speed of execution. Its use allows to determine the relative and absolute amount of the matrix, measuring the level of gene expression, to determine gene copy number and to distinguish alleles.

## 4.9. DNA electrophoresis

Electrophoresis is the process of moving charged molecules in solution by applying an electric field across the solution. Often classified according to the presence or absence of a solid supporting medium or matrix through which the charged molecules move in the electrophoretic system.

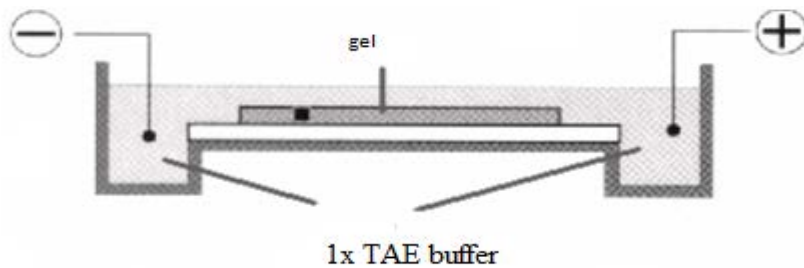
All types of electrophoresis basic are governed by the principles illustrated by Equation:

$$\text{Mobility of a molecule} = \frac{(\text{applied voltage}) \times (\text{net charge on the molecule})}{(\text{friction of the molecule})}$$

The mobility (rate of migration), of a molecule increases with increased applied voltage and increased net charge on the molecule. Most electrophoretic systems employ an equal and constant voltage on all of the cross-sectional areas of the paper strips, gels, or solutions employed in the electrophoretic separation. These electric fields are best defined in terms of volts per linear centimeter. Ohm's law ( $V = IR$ ) dictates that voltage ( $V$ ) is a function of current ( $I$ ) and resistance ( $R$ ). The nature of the electrophoresis apparatus and buffer composition dictates the resistance in the system. Current is often used to define the voltage requirements of an electrophoretic separation. The resistance of the system is important because it will determine the amount of heat generated during electrophoresis. Since electrophoretic mobility is also a function of temperature, heating of the separation matrix must be controlled.

The friction experienced by molecules during electrophoretic migration reflects both molecular size and molecular shape. If the electrophoresis is carried out in a medium that offers significant barriers to the movement of macromolecules through it (as is the case with polyacrylamide and agarose, two very commonly used systems), molecular size may prove to be the most important determinant of mobility. If the charge to mass ratio on the macromolecules being separated is approximately equal (as in several of the cases discussed below), molecular size becomes the sole determinant of electrophoretic mobility [42]. Several methods have been developed, based on the use of various physical phenomena for separation of macromolecules, due to their variable characteristics. These methods have become the most popular techniques of electrophoretic techniques. These techniques achieve unmatched levels of resolution, and combined with immunostaining techniques, also an unprecedented level of selectivity. Speed of movement of electrically charged macromolecules is dependent on its charge, size, shape, motion, and environment resistance. Based on these properties, a quick separation can be made of various macromolecules using a relatively simple apparatus and with relatively low cost. The main areas of application of electrophoresis are: biochemistry of proteins and nucleic acids, molecular biology, pharmacology, forensic medicine, veterinary medicine, medical diagnostics. It should also be noted that the use of electrophoretic techniques enabled us to obtain the complete sequence of human DNA in the recently ended genomics project. All commonly used methods in the final stage of DNA sequencing technology are based on electrophoresis. Agarose is mainly used to separate larger macromolecules such as nucleic acids whereas a polyacrylamide gel may be used to separate nucleic acids or proteins [43].

Agarose is a linear polymer composed of alternately arranged residues of D- and L-galactose linked by  $\alpha(1-3) - \alpha(1-4)$  glycosidic bonds. Approx. 800 galactose residues create a single string. It is isolated from algae. Agarose in commercial form is a white or slightly yellow powder. Dissolves very readily in boiling water and remains in a liquid state up to a temperature of about 40°C. Below this temperature solidifies in the form of a porous gel. After solidification remains in this form, even at elevated temperatures of up to several tens °C. Pore size can be adjusted using different agarose concentrations. The pore size and sieving characteristics of a gel are determined by adjusting the concentration of agarose in the gel. The higher the concentration, the smaller the pore size. Usually agarose gels are prepared at concentrations in the range of 0.4, 1, and 2% and are used in a horizontal electrophoresis apparatus [44].



**Picture 3.9.1.** Horizontal electrophoresis.

## 4.10. Data analysis

### 4.10.1. The Hardy–Weinberg equilibrium

An individual's genotype is the combination of alleles found in that individual at a given genetic locus. If there are two alleles in a population at locus A (A and a), then the possible genotypes in that population are AA, Aa, and aa. Individuals with genotypes AA and aa are homozygotes (i.e., they have two copies of the same allele). Individuals with genotype Aa are heterozygotes (i.e., they have two different alleles at the A locus). If the heterozygote is phenotypically identical to one of the homozygotes, the allele found in that homozygote is said to be dominant, and the allele found in the other homozygote is recessive.

Attendance genes (alleles) - incidence

$$P(A) = p = \frac{2*AA + Aa}{2*(AA + Aa + aa)}$$

$$P(a) = q = \frac{2*aa + Aa}{2*(AA + Aa + aa)}$$

$$P(A) + P(s) = p + q = 1$$

attendance genotypes - the incidence

$$P(AA) = AA / (AA + Aa + aa)$$

$$P(Aa) = Aa / (AA + Aa + aa)$$

$$P(aa) = aa / (AA + Aa + aa)$$

$$P(AA) + P(Aa) + P(aa) = 1$$

### LAW Hardy-Weinberg

Randomly reproduce infinitely large population (there is no genetic drift), in which there is no selection, mutation and migration, frequencies of alleles and genotypes are fixed in subsequent generations.

$$\text{generation 1 : } P(A) = p \qquad P(a) = q$$

$$\text{generation 2: } P(A) = p \qquad P(a) = q$$

$$\text{generation n : } P(A) = p \qquad P(a) = q$$

Balance genetic: In such a population there is a constant relationship between turnout alleles and attendance genotypes,

$$p + q = 1$$

$$(p + q)^2 = 1$$

$$p^2 + 2pq + q^2 = 1$$

$$P(AA) = p^2 \qquad P(Aa) = 2pq \qquad P(aa) = q^2$$

$(AA)$ ,  $P(AA)$ ,  $P(aa)$  - the value actually observed in the population (observed)

$P^2$ ,  $2pq$ ,  $q^2$  - expected values, theoretical, calculated in accordance with the Hardy - Weinberg equilibrium (expected)

Verifying the state of genetic balance in a population means determining if the studied population is in equilibrium, i.e. if the number of genotypes observed is consistent with the expected number. Since at small deviations it is difficult to conclude whether such compliance occurs, the comparison includes conducting a statistical test such as  $\chi^2$ .

H0 (null hypothesis) :

In equilibrium population Hardy - Weinberg, i.e. pairing occurs in the random population, lack of selection, mutation, migration

$$H_0 : P ( AA ) = p^2 \quad \text{and} \quad P ( Aa ) = 2pq \quad \text{and} \quad P ( aa ) = q^2$$

H1 ( alternative hypothesis ) :

population is not in equilibrium Hardy - Weinberg ,i.e. in the population there is random mating and / or there are selection , mutation , migration

$$H_1: P ( AA ) \neq p^2 \quad \text{or} \quad P ( Aa ) \neq 2pq \quad \text{or} \quad P ( aa ) \neq q^2$$

$$T (\text{test}) = \chi^2 = \sum [ ( \text{Obs} - \text{Exp} )^2 / \text{Exp} ]$$

where:

- Operation ( Observed ) - the number of observed individuals of a certain phenotype
- Exp ( expected) - the expected number of individuals of a certain phenotype
- sum (  $\Sigma$  ) will contain three components - respectively for the three genotypes at each gene locus of the test (AA , Aa , aa) .
- the calculated value T is compared with the value read from the array TMAX  $\chi^2$  test for the appropriate number of degrees of freedom and predetermined probability ( confidence level) of the rejection or acceptance hypothesis of zero).
- generally assumed probability (confidence level )  $\alpha = 0.95 ; 0.99$  or  $0,999$  corresponds to the so-called significance level  $P = 0.05 ; 0.01 ; 0,001$ , respectively.
- The number of degrees of freedom associated with the size of  $\chi^2$  is equal to the number of data classes [45, 46].

It is important to recognize that the Hardy-Weinberg equilibrium is a neutral equilibrium, which means that a population perturbed from its Hardy-Weinberg genotype frequencies will indeed reach equilibrium after a single generation of random mating (if it conforms to the other assumptions of the theorem), but it will be a new equilibrium if allele frequencies have changed. This property distinguishes a neutral equilibrium from a stable equilibrium, in which a perturbed system returns to the same equilibrium state. It makes sense that the Hardy-Weinberg equilibrium is not stable, since a change from the equilibrium genotype frequencies will generally be associated with a change in allele frequencies (p and q), which will in turn lead to new values of  $p^2$ ,  $2pq$  and  $q^2$ . Thereafter, a population that meets Hardy-Weinberg assumptions will remain at the new equilibrium until perturbed again.

#### **4.10.2. Minor allele frequency**

**Global minor allele frequency (MAF):** dbSNP is reporting the minor allele frequency for each included in a default global population. The rs number is an accession number used by researchers and databases to refer to specific SNPs. It stands for Reference SNP cluster ID. Since this is being provided to distinguish common polymorphism from rare variants, the MAF is actually the second most frequent allele value. In other words, if there are 3 alleles, with frequencies of 0.50, 0.49, and 0.01, the MAF will be reported as 0.49. The current default global population is 1000Genome phase 1 genotype data from 1094 worldwide individuals [47].

#### **4.10.3. SPSS- Binary Logistic Regression**

The software name stands for Statistical Package for the Social Sciences. (Brand, maker, etc.)

Logistic regression is used to predict a categorical (usually dichotomous) variable from a set of predictor variables. With a categorical dependent variable, discriminant function analysis is usually employed if all of the predictors are continuous and nicely distributed; logic analysis is usually employed if all of the predictors are categorical; and logistic regression is often chosen if the predictor variables are a mix of continuous and categorical variables and/or if they are not well distributed (logistic regression makes no assumptions about the distributions of the predictor variables). Logistic regression has been especially popular with medical research in which the dependent variable is whether or not a patient has a disease. For a logistic regression, the predicted dependent variable is a function of the probability that a particular subject will be in one of the categories (for example, the probability that Suzie Cue has the disease, given her set of scores on the predictor variables) [48].

#### **4.10.4. Stratification**

In statistics, stratified sampling is a method of sampling from a population. When sub-populations vary considerably, it is advantageous to sample each subpopulation (stratum) independently. Stratification is the process of grouping members of the population into relatively homogeneous subgroups before sampling. The strata should be mutually exclusive: every element in the population must be assigned to only one stratum. The strata should also be collectively exhaustive: no population element can be excluded. Then random or systematic sampling is applied within each stratum. This often improves the representativeness of the sample by reducing sampling error. It can produce a weighted mean that has less variability than the arithmetic mean of a simple random sample of the population [49].

## 5. Objective

The aim of this work was to investigate the influence of rs3183982 gene polymorphisms in the **PDCD6IP** gene on the platinum-based chemotherapy response in Chinese non-small cell lung cancer patients. The research was focused on reaction in cells that are important in NSCLC development, as well as the fundamentals resistant to platinum based chemotherapy.

Main focus:

1. Investigating the SNP in the PDCD6IP
2. Study the process of forming cancer to understand the correlation between reaction in cells and the chances by overexpression of some proteins
3. Research meaning of platinum resistance and connection to the poor response to group of patients.
4. Try to find answers about genetic markers with lung cancer risk, ways of treatment, early diagnose.

## 6. MATERIALS AND METHODS

### 6.1. Ethics approval

The clinical research was registered on the Chinese Clinical Trial Registry. The study protocol was approved by the Ethics Committee of Xiangya School of Medicine, Central South University, Changsha, China. Prior to the study, all patients provided written informed consent in compliance with the code of ethics of the World Medical Association (Declaration of Helsinki).

In total, 335 Chinese lung cancer (non-small cell lung carcinoma) patients were recruited from November 2011 to May 2013 at Xiangya Hospital of Central South University and Hunan province tumour Hospital in Changsha Hunan. The demographic and basic clinical characteristics were collected for each patient. The data included age, smoking status, histology, gender, stage, chemotherapeutic regimen, and Eastern Cooperative Oncology Group Performance Status (ECOG PS). The information is shown in the table 5.1.1.

**Table 5.1.1.** Main clinical characteristics for the lung cancer patients receiving platinum-based chemotherapy treatment.

Characteristics	Responders		Non-responders	
	N	(%)	N	(%)
Total	100	100.0	235	100.0
<u>Age (years)</u>				
≤55	52	52%	98	42%
>55	48	48%	137	58%
<u>Gender</u>				
Male	79	79%	174	74%
Female	21	21%	61	26%
<u>Histology</u>				
ESCC	39	39%	132	56%
EAC	61	61%	103	44%
<u>Stage (non-small cell lung carcinoma)</u>				
I-II	5	5%	15	6%
III-IV	95	95%	220	94%
<u>Smoking</u>				
Yes	58	58%	105	45%
No	42	42%	130	55%
<u>Polymorphism</u>				
Wild Type	14	14%	49	21%
Heterozygous	36	36%	66	29%
Homozygous	50	50%	120	51%
<u>Chemotherapeutic regimen</u>				
Cisplatin	39	39%	114	49%
Carboplatin	61	61%	121	51%
<u>ECOG</u>				
0-1	96	96%	233	99%
>1	4	4%	2	1%

The following patients were included in the study:

- Patients who were diagnosed lung cancer by histology or cytology;
- Patients who had never received any radical or biological therapy before first line chemotherapy;
- All patients who had received at least two cycles of first line chemotherapy;
- All patients who had received a follow up exam in six months.

The following patients were excluded:

- Patients who were pregnant or feeding new-borns;
- Patients who had been diagnosed with other malignancies;
- Patients who had brain metastases;
- Patients who had active infections.

Patients enrolled in this study received first-line platinum-based chemotherapy. Patients were treated with different drugs that were grouped either cisplatin or carboplatin. (Detailed in table 5.1.2).

**Table 5.1.2.** Drugs used by patents in present study.

Number	Drug name	Shortcut	Number of patients	%	Category
1	pemetrexed disodium/ cisplatin	PP	80	24%	Cisplatin
2	Gemcitabine/ paclitaxel	DP	141	42%	Carboplatin
3	Docetaxel/ cisplatin	DP	52	15%	Cisplatin
4	Paclitaxel/ carboplatin	TP	22	7%	Carboplatin
5	Gemcitabine/ carboplatin	NP	27	8%	Carboplatin
6	vinorelbine/cisplatin	NP	13	4%	Cisplatin

Chemotherapy response was assessed after the first two cycles of chemotherapy according to the RECIST guideline (version 1.1) for solid tumours. That guideline provides the definitions of the criteria used to determine objective tumour response for target lesions. Complete Response (CR): Disappearance of all target lesions. Any pathological lymph nodes (whether target or non-target) must have a reduction in short axis to <10 mm. Partial Response (PR): At least a 30% decrease in the sum of diameters of target lesions, taking as reference the baseline sum diameters. Progressive Disease (PD): At least a 20% increase in the sum of diameters of target lesions, taking as reference the smallest sum on study (this includes the baseline sum if that is the smallest on study). In addition to the relative increase of 20%, the sum must also demonstrate an absolute increase of at least 5 mm. Stable Disease (SD): Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD, taking as reference the smallest sum diameters while on study. The patients that showed CR or PR were regarded as platinum sensitive, while, in the case of PD or SD, were regarded as platinum resistant [50].

Stage of tumour was determined by the cTNM classification of the Union Internationale contre le Cancer (UICC) [51] for non-small cell lung cancer patients. In this study, four stages of cancer were classified into two groups as I-II and III-IV.

## **6.2. Tagging SNP selection**

This study is related to the ongoing research at the Institute of Clinical Pharmacology, in which the association of gene polymorphisms with chemotherapy is investigated. The DNA samples analysed were obtained from non-small cell lung cancer (NSCLC) patients that were treated by platinum-based chemotherapy. The work was focused on a rare genetic variant in PDCD6IP. The SNP selected for this study was rs3183982.

The selected SNP satisfied the following criteria:

- (1) The SNP was chosen from the International HapMap Project Phase II database of Chinese population (<http://www.hapmap.org/>); [51]
- (2) SNP was a haplotype tagger SNP;
- (3) Minor allele frequency (MAF) was larger than 5% in Hunan Province population China;

### **6.3. DNA extraction and genotyping procedure**

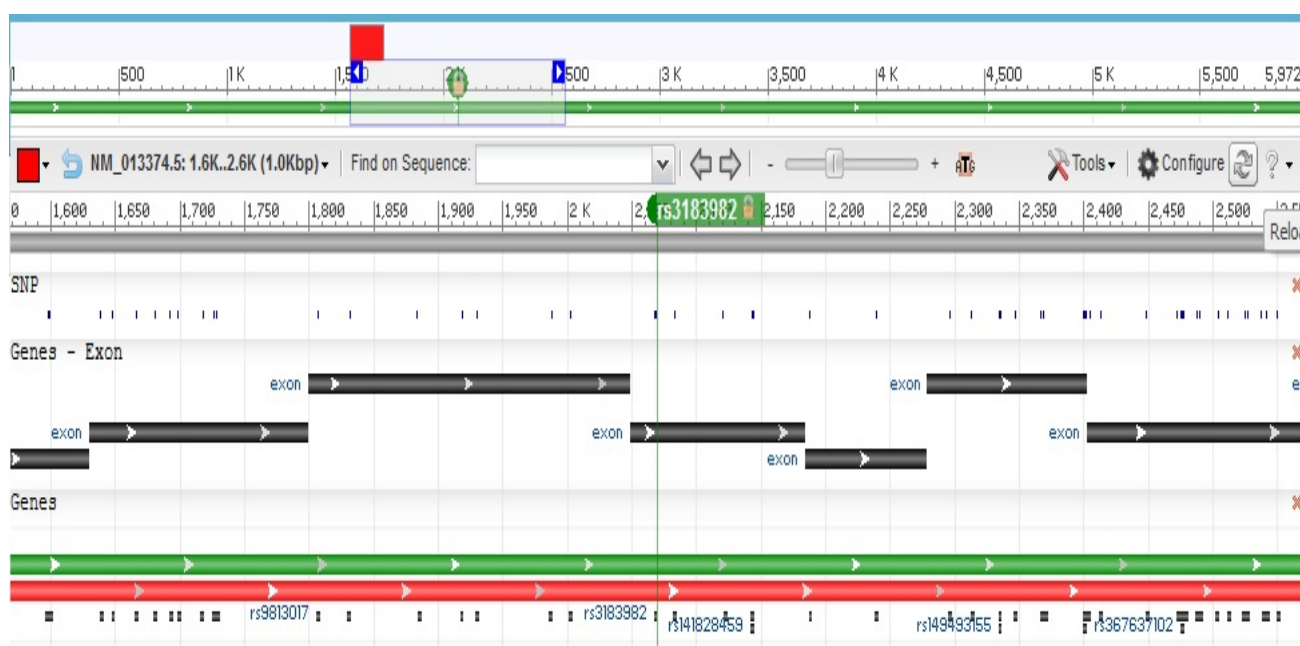
Approximately 5 ml venous blood was collected from each patient for genetic studies. Genomic DNA was extracted from whole peripheral blood using E.Z.N.A. SQ Blood Kit II (Omegabio-tek) according to the following protocols.

#### **6.3.1. Genomic DNA extraction**

- Before starting the procedure frozen blood was thawed quickly in a 37°C water bath with gentle agitation and stored on ice.
- 5 ml blood were added to a nuclease-free 50 ml centrifuge tube containing 30 ml (5 fold the volume of blood) Buffer NL (in order to lyse the plasma membrane). The mixture was centrifuged at 2000 x g for 5 minutes at room temperature.
- The supernatant was discarded and the tube was left inverted on a clean piece of absorbent paper for 2 min.
- 5 ml Buffer XL (containing chemotropic salt and proteinase K) were added to the tube containing the nuclei pellet, and the pellet was resuspended through use of a vortex agitator.
- The suspension was incubated at 65°C for 10-15 minutes in a heating block.
- 5 ml of isopropanol were added to the lysate, and mixed by inversion until the DNA precipitate become visible as threads or clumps, followed by centrifugation at 2000 x g for 5 minutes at room temperature.
- The supernatant was poured out and the tube drained on a clean absorbent paper towel. The pellet was washed by addition of 5 ml 70% ethanol and vortex agitation, followed by centrifugation at 2000 x g for 5 minutes at room temperature. The ethanol solution was carefully poured out.
- The pellet was dried for 10-15 min, with the tube inverted on a clean absorbent paper towel.
- 1 ml of DNA rehydration solution was added, followed by incubation at 65°C for 1 hour to rehydrate DNA. The purified DNA samples were stored at -20°C.

Primers were designed by using Primer3 software. Primer sequences were forward primer 5' – TGTTTTGACTCCTATTGGATTTT – 3', reverse primer 5'-CATGTA CTTTCATACCTTTGTGCC – 3'. The gene is located 3p22, the exon category.

The presence of the variant allele creates a TaqI restriction site. The PCR product (361bp) was therefore digested with Taq<sup>I</sup> an *E. coli* strain that carries a Taq<sup>I</sup> (overproducing plasmid) that has a two amino acid replacement at its amino terminus. The enzyme has a higher level of expression without interfering with its catalytic properties, resulting in three fragments: carry AA 360bp, AG 260bp + 230bp + 130bp, GG 230bp + 130bp.



**Picture 5.3.1.1.** The PDCD6IP gene with the location of the selected polymorphism

### 6.3.2. Procedures of genotyping by PCR-RELP:

a) Polymerase chain reaction (PCR) amplification

Components of the reaction mixture:

- 10.3  $\mu$  L HPLC grade Water (BioSune),
- 2  $\mu$  L 10  $\times$  PCR Buffer, (TransGEN)
- 0.3  $\mu$  L 25mM dNTP Mix, (CutSmart)

- 0.2  $\mu$  L 0.5mM Primer Mix, (BioSune)
- 0.2  $\mu$  L 5 U/  $\mu$  L HotStart Taq, (BioLabs)
- Approximately 2  $\mu$  L 10 ng/  $\mu$  L DNA in a final volume 15  $\mu$  L. Yes, professor they select randomly around 20 samples and measure DNA concentration with nanodrop and it was always very similar so they always do like this.

PCR conditions (thermocycler 6321, Eppendorf):

- denaturation at 94 °C for 15 minutes,
- 35 cycles (denaturation at 94°C for 45 seconds, annealing at 59 °C for 45 seconds and extension at 72 °C for 45 seconds),
- additional extension at 72 °C for 10min.
- samples were kept at 4 °C until analysis

b) RFLP (Restriction Length Fragment Polymorphism) analysis

Components of the reaction mixture:

- 15.5  $\mu$  L Nanopure water,
- 0.5  $\mu$  L cutSmart enzyme Taq<sup>o</sup>I
- 2  $\mu$  L 10  $\times$  NEB Buffer
- 2  $\mu$  L DNA in a final volume 20  $\mu$  L.

Samples were incubated at 65°C for 60 minutes, and then kept at 4°C until analysis.

c) DNA electrophoresis:

Materials:

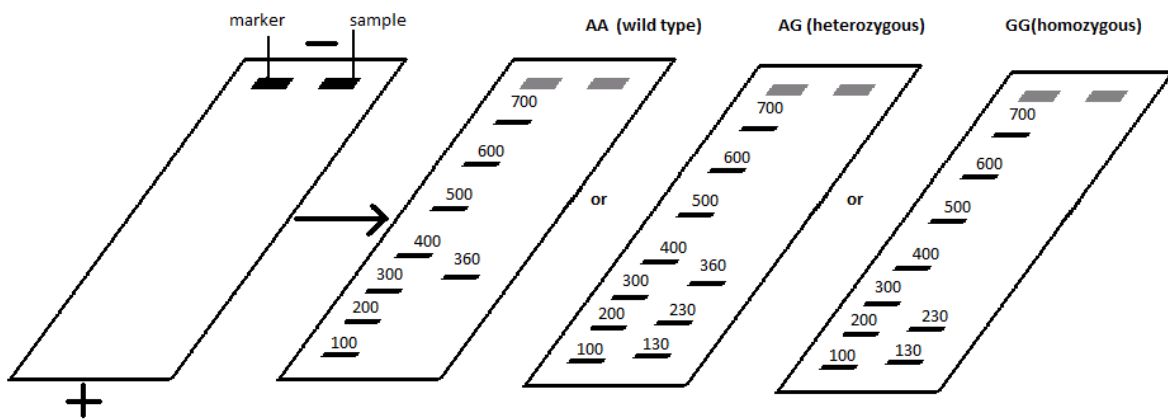
- Agarose 1.5%
- 1x TAE - ethidium bromide, (0.04 mol/L Tris-HCL and 0.001 mol/L EDTA)

- loading dye, (gel red)
- marker, (TransGEN)
- sample DNA (a digestion product)

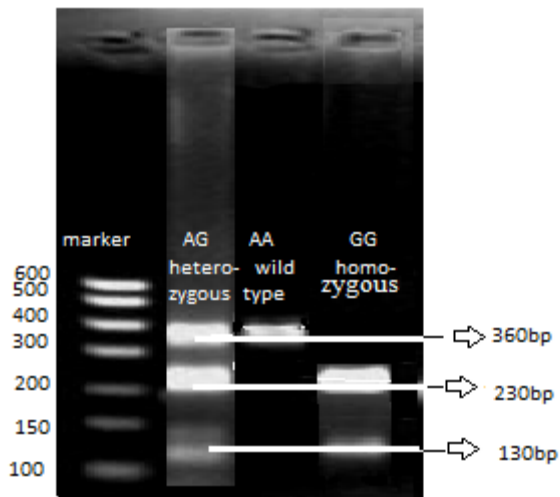
The conditions were:

- current 70 mA,
- tension 150V,
- time 25-30 minutes.

Fragment analysis; the allelic discrimination of the SNP was determined by the numbers of fragments and relative positions in the gels. Restriction patterns were analysed by using Dongsheng Bio-Imaging system. Figure 5.3.2.1. Shows the different possible genotype patterns, resulting from the possible 260bp, 230bp and 130bp. Figure 5.3.2.2. Shows actual results, representative of different samples analysed.



**Picture 5.3.2.1.** Expected results after DNA electrophoresis.



**Picture 5.3.2.2.** The results after DNA electrophoresis.

### 6.3.3. Statistical analysis

Genotypic frequencies were tested for the Hardy - Weinberg Equilibrium (HWE) by using the chi - square test. The association of genotypes and chemotherapy response was tested with logistic regression analysis. The potential covariates on chemotherapy response were selected by using binary logistic regression. Age, sex, smoking status, chemotherapeutic regimen, and classification, ECOG and histology type were considered as covariates. All the above analysis were performed using SPSS 13.0 (SPSS Inc., Chicago, Illinois, USA). Odds ratios (OR) with their 95% confidence intervals (CIs) were used to assess response to chemotherapy.  $P < 0.05$  was considered as significant. The relationship of the SNP with response of chemotherapy in different categories, was performed by the stratification analysis by SPSS 13.0.

## 7. RESULTS

### 7.1. Patient Characteristics

A total of 335 lung cancer patients were recruited in this study. All the people were Chinese with the average age of  $55.3 \pm 9.2$  range (22-76), in the work examined 150 patients (44.8%) which were  $\leq 55$  years old, and 185 (55.2%)  $> 55$  years old. The basic demographic and clinical characteristics were shown in Table 5.1.1. and 6.1.1. The non-small cell lung carcinoma patients were more likely to be non-responders than responders, the number of responders (CR or PR) were 100 and non-responders (PD or SD) were 235. The patient sample comprised 253 (75.5%) males and 82 (24.5%) females. In which concerns the differentiation in the histology 174 patients had esophageal squamous cell carcinoma (ESCC) and 161 Erythema annulare centrifugum (EAC). The majority 315 (95.0%) patients were on III-IV stage of NSCLC, when only 20 (5.0%) were stage I-II. The majority of patients were smokers 188 (56.1%). All patients received platinum-based chemotherapy, cisplatin-based was 153(45.7%) and carboplatin-based was 182(44.3%). The ratio of ECOG 0-1 was in 329 patients and was lower in responders than non-responders (96 vs 99%,  $P=0.033$ ).

There were no significant differences between responders and non-responders in gender, stage of cancer, smoking status and chemotherapy regiment, indicating that the populations were adequately matched. Covariate analysis showed that age ( $P$ -value 0.037), histology (0.022) and ECOG (0.033) may be considered statistically significant risk factors on platinum-based chemotherapy efficacy in all population.

**Table 6.1.1.** Main clinical characteristics of lung cancer patients.

Characteristics	Summary		P-Value	OR (95% C.I).
	N	(%)		
Total	335	100.0%		
<u>Age (years)</u>			0.037	1.72(1.03-2.86)
≤55	150	44.8%		
>55	185	55.2%		
<u>Gender</u>			0.321	1.46(0.69-3.11)
Male	253	75.5%		
Female	82	24.5%		
<u>Histology</u>			0.022	0.51(0.28-0.91)
ESCC	174	51.9%		
EAC	161	48.1%		
<u>Stage (non-small cell lung carcinoma)</u>			0.267	0.85(0.49-1.48)
I-II	20	5.0%		
III-IV	315	95.0%		
<u>Smoking</u>			0.482	1.27(0.69-3.11)
Yes	188	56.1%		
No	147	42.9%		
<u>Chemotherapeutic regimen</u>			0.560	0.85(0.49-1.48)
Cisplatin	153	45.7%		
Carboplatin	182	44.3%		
<u>Polimorphism</u>				
Wild Type	63	18.8%		
Heterozygous	102	30.4%		
Homozygous	170	50.8%		
<u>ECOG</u>			0.033	0.146(0.25-0.85)
0-1	329	98.2%		
>1	6	1.8%		

## 7.2. Prevalence of the PDCD6IP rs3183982 SNP

The SNP (3183982) information of PDCD6IP gene was shown in table 6.2.1. The call rate was 100%. The minor allele frequency was  $G=0.3303/1654$ . The genotype frequencies agreed with the Hardy-Weinberg equilibrium. This mean that for rs 3183982, minor allele is 'G' and has a frequency of 33.03% in the 1000Genome phase 1 population and that 'g' is observed 1654 times.

**Table 6.2.1.** The gene polymorphisms examined in this study.

Gene	dbSNP	Call Rate(%)	Polymorphism	MAF	HWE P
PDCD6IP	rs3183982	335(100)	A/G	$G=0.3303/1654$	0.60

MAF: minor allele frequency

HWE P: p value for Hardy-Weinberg test

## 7.3. Relation of PDCD6IP gene polymorphism with platinum-based chemotherapy response

The frequency distribution of the possible genotypes observed for all subjects was listed in table 6.3.1, that summarises the results of the analysis of the **rs3183982** SNP in all 335 lung cancer patients. In this study after analysis 63 patients (18.8%) were classified base on pictures obtained after electrophoresis as carriers of the wild-type genotype, 103 (30.4%) homozygous for the variant, 170 (50.8%) heterozygous.

The rs31839825 SNP showed a significant association with chemotherapy response when considering the additive model (OR=0.67, P=0.033). No statistically significant

association was found between SNP in both recessive, dominant models and chemotherapy response table 6.3.2.

**Table 6.3.1.** Allele frequencies of the gene polymorphism in 335 patients.

Gene	Polymorphism	Wild-type (N/%)	Heterozygous (N/%)	Homozygous(N/%)
PDCD6IP	rs3183982	53(18.8)	102(30.4)	170(50.8)

**Table 6.3.2.** Association of single nucleotide polymorphism and platinum-based chemotherapy response in all lung cancer patients.

Gene	Polymorphisms	Genotype	Responders N(%)	Non-responders N(%)	Additive		Dominant		Recessive	
					OR(95%CI)	P value	OR(95%CI)	P value	OR(95%CI)	P value
PDCD6IP	rs3183982	AA	14(14)	49(21)	0.67(0.47-0.97)	<b>0.033</b>	0.56(0.29-1.09)	0.089	0.62(0.37-1.04)	0.072
		AG	50(50)	120(51)						
		GG	36(36)	66(28)						

The values in bold were considered statistically significant.

N=Number of patients.

Esophageal squamous cell carcinoma (ESCC) and Erythema annulare centrifugum (EAC) are very different biologically [52]. Smokers and non-smokers, younger and elders, male and female, chemotherapeutic regimen (cisplatin and carboplatin) also may be different in chemotherapy response.

To further investigate the relationship of the SNP with response to chemotherapy, it was considered that there could be some associations among subgroups. To investigate this hypothesis, a stratification analysis was performed (Table 6.3.3.). The subgroups were divided into ESCC or EAC by histology, smoking or non-smoking by smoking status,  $\leq 55$  years old or  $> 55$  years old by age, male or female by gender, I-II or III-IV by stage of cancer, cisplatin or carboplatin by chemotherapeutic regimen. There were few results that showed significant associations with chemotherapy response: for stage I-II in additive (OR=1.89, P=0.017) and dominant (OR=2.1, P=0.014) models, for EAC in additive model (OR=0.83, P=0.041), for non-smoking (OR=0.47, P=0.048), and for cisplatin (OR=0.51, P=0.046) in recessive model.

In this study, the number of independent variables for PDCD6IP is 6. The SNP was analyzed in three genetic models. To take into account the problem of multiple testing, the significance threshold is therefore  $0.05 / (3 \times 6) = 0.0027$ . Using this threshold, no association remained significant.

**Table 6.3.3.** Stratification analyses of the associations of the polymorphism and chemotherapy efficacy in

- a) Dominant (A),
- b) Recessive (B),
- c) Additive(C) models.

a)

A additive Variable	Characteristics	P-Value	95% C.I. for EXP (B) OR
Histology	ESCC	0.306	0.77(0.46-1.27)
	EAC	<b>0.041</b>	0.83(0.45-1.53)
Smoking status	Smoking	0.352	0.80(0.50-1.28)
	Non-smoking	0.147	0.68(0.40-1.15)
Age (year)	≤55	0.226	0.75(0.47-2.00)
	<55	0.489	0.85(0.54-1.35)
Gender	Male	0.156	0.76(0.52-1.11)
	Female	0.293	0.63(0.27-1.49)
Stage	I-II	<b>0.017</b>	1.89(1.12-3.17)
	III-IV	0.076	0.74(0.52-1.03)
Chemotherapy regimen	Cisplatin	0.212	0.70(0.41-1.22)
	Carboplatin	0.305	0.79(0.51-1.23)

b)

B recessive Variable	Characteristics	P-Value	95% C.I. for EXP (B) OR
Histology	ESCC	0.318	0.70(0.34-1.42)
	EAC	0.312	0.69(0.33-1.42)
Smoking status	Smoking	0.949	0.98(0.49-1.96)
	Non-smoking	<b>0.048</b>	0.47(0.21-1.01)
Age (year)	≤55	0.262	0.69(0.36-1.32)
	<55	0.956	0.98(0.49-1.97)
Gender	Male	0.312	0.75(0.52-1.11)
	Female	0.306	0.53(0.16-1.79)
Stage	I-II	0.419	4.14(0.13-13.63)
	III-IV	0.211	0.72(0.44-1.20)
Chemotherapy regimen	Cisplatin	<b>0.046</b>	0.51(0.24-1.08)
	Carboplatin	0.920	0.96(0.48-1.93)

c)

C Dominant Variable	Characteristics	P-Value	95% C.I. for EXP (B) OR
Histology	ESCC	0.784	0.79(0.35-2.20)
	EAC	0.260	0.63(0.28-1.41)
Smoking status	Smoking	0.110	0.48(0.19-1.18)
	Non-smoking	0.931	1.04(0.43-2.51)
Age (year)	≤55	0.945	0.97(0.45-2.09)
	<55	0.337	0.66(0.29-1.54)
Gender	Male	0.339	0.72(0.37-1.41)
	Female	0.527	0.62(0.14-2.76)
Stage	I-II	<b>0.014</b>	2.10(1.12-3.78)
	III-IV	0.153	0.65(0.35-1.18)
Chemotherapy regimen	Cisplatin	0.711	1.20(0.45-3.17)
	Carboplatin	0.147	0.57(0.27-1.22)

## 8. CONCLUSION

The fact that chemo sensitivity differs from person to person reflects a limited understanding of the molecular mechanism of lung carcinogenesis. Existing evidence suggests that there are many differences among patient subgroups that may determine individual susceptibility to lung cancer. Nowadays individualized therapies are becoming more and more popular. There are many reasons to think that is the future for cancer therapies. Out of the entire group of patients taken into this study, 70% were resistant to the platinum based chemotherapy which means that the treatment is not effective enough to beat the cancer. Almost 95% of patients were stage III-IV, suggesting that the diagnosis of NSCLC is poor, people are being diagnosed too late and have already developed metastasis. There is a lack of adequate tests and well-characterized genetic markers for the diagnostics and characterization of chemotherapy resistance in lung cancer.

Results between responders and non-responders have shown significant differences in age, histology, and ECOG that might suggest that acquisition of resistance can be in future more deeply investigated in different age groups and histology. This can have a consequence in efficiency, response and future survival rate for patients so it is extremely important to investigate these variations. These results show that there is a correlation between response and resistance to platinum based chemotherapy. They also suggest that there is a connection between resistance to a drug and the histology. Henceforth, these findings open the way for new studies on the use of different treatment in patients with ESCC and EAC. There is a great need for new therapies and new biomarkers focusing on molecular targets helping to identify patients who will benefit from therapy

The distribution of genotypes for the SNP analysed in the present study, PDCD6IP rs3183982, shows a significant association with chemotherapy response. Our results showed that this might clarify the actual association between the gene polymorphism and lung cancer risk or chemo sensitivity. The knowledge about the association of rs3183982 SNP in PDCD6IP with chemotherapy resistance could be the basis for choosing an alternative treatment in the

patients that have that polymorphism. In the future, this SNP might be used as a marker helping to diagnose lung cancer in early stage, and maybe even changes in the cells before the cancer cells start proliferation.

However, our results require future verification and investigation. Researchers discovered that there are many proteins whose expression is associated with platinum resistant treatment [53].

The stratification analysis of the presence of rs3183982 SNP showed significant associations among subgroups: for stage I-II, for EAC, for non-smoking, and for cisplatin. This is further proof that histology and platinum based chemotherapy are being affected by SNP. Over the years, there were plenty of studies that explored the relationship of polymorphisms of candidate genes with chemotherapy response (54, 55, 56). However, the associations between chemotherapy response and rs3183982 in PDCD6IP polymorphism have not been detected in any population based on a candidate gene approach. We presented here the first case-control study evaluating the association between the rs3183982 polymorphism within the promoter of PDCD6IP and chemotherapy response.

In this study, I investigated the association of the transporter gene PDCD6IP SNP with platinum-based chemotherapy response. Our results showed that SNPs rs3183982 might be related with chemotherapy response in lung cancer patients. More studies need to be performed to investigate the potential linking of rs3183982 polymorphism in PDCD6IP encoding protein required for apoptosis with cancer progression and prognosis. Future investigations should be focused on the level of the transcriptional activity of PDCD6IP, using luciferase assays in transiently transfected NSCLC cells.

## 9. References:

- [1] Scott A. G. *Personalizing medicine with clinical pharmacogenetics* Genet Med. 2011;13(12):987-95
- [2] Motulsky AG, Gurwitz D, *Drug reactions, enzymes, and biochemical genetics:50 years later.* Pharmacogenomics 2007. 8:1479-1484.
- [3] Beutler E. et al., *Study of glucose-6-phosphate dehydrogenase: history and molecular biology.* Am J Hematol. 1993. 42:53–58.
- [4] Mahgoub A, Idle JR, et al., Polymorphic hydroxylation of debrisoquine in man. Lancet. 1977. 2:584–586.
- [5] Sim SC, Ingelman-Sundberg M. *The Human Cytochrome P450 (CYP) Allele Nomenclature Web site: a peer-reviewed database of CYP variants and their associated effects.* Hum Genomics. 2010. 4:278–281.
- [6] Issa AM. *Personalized medicine and the practice of medicine in the 21st century.* McGill J Med. 2007. 10:53–57.
- [7] Newman W. G. *Pharmacogenetics: Making Cancer Treatment Safer and More Effective inhibit the activity of a drug by facilitating its removal.* Springer Science+Business Media B.V. 2010. 39-44.
- [8] Barbara Driscoll *Lung Cancer Volume I Molecular Pathology Methods and Reviews* humana press 2003. Vol. 74
- [9] Yan-hui Yang, Herve Aloysius et al. *Enzyme-mediated hydrolytic activation of prodrugs*, Department of Medicinal Chemistry, Ernest Mario School of Pharmacy, Rutgers, The State University of New Jersey, Piscataway, 2011, 1(3):143–159
- [10] Hanahan D, Robert A. Weinberg A. *The Hallmarks of Cancer Review*, Cell Press 2000, Vol. 100, 57–70,
- [11] Goodsell D. S., *The Molecular Perspective: The ras Oncogene*, The Oncologist, 1994(3): 263-264.
- [12] Loktionov.A., *Mini-Review Common gene polymorphisms, cancer progression and prognosis*, Cancer Letters 2004, Vol. 208, 1–33
- [13] *World Health Statistics*, World Health Organisation. 2008.
- [14] Jemal A, Bray F et al. *Global cancer statistics.* CA Cancer J Clin 2011.61: 69-90.
- [15] Ferlay J, Soerjomataram I, et al., *Cancer Incidence and Mortality Worldwide: IARC CancerBase*, 2012, No. 11 Lyon.

- [16] Cancer facts and figures 2014, [www.cancer.org](http://www.cancer.org)
- [17] Mitchell, Richard Sheppard; et al. *Chapter 13, box on morphology of adenocarcinoma* Robbins Basic Pathology Philadelphia: Saunders. ISBN 1-4160-2973-7.
- [18] Travis, William D; Brambilla, et al. *Pathology and Genetics of Tumours of the Lung, Pleura, Thymus and Heart*. World Health Organization Classification of Tumours. Lyon: IARC Press. 2010, ISBN 92-832-2418-3.
- [19] Wang, LC; Wang, L; et al. *Analysis on the clinical features of 22 basaloid squamous cell carcinoma of the lung*. Journal of Cardiothoracic Surgery 2011, 6: 10.
- [20] Fernandez FG, Battafarano RJ (October 2006). *Large-cell neuroendocrine carcinoma of the lung*. Cancer Control 2006, 13 (4).
- [21] Lv H, Han T., *Genetic polymorphism of GSTP1 and ERCC1 correlated with response to platinum-based chemotherapy in non-small cell lung cancer*. Med Oncol. 2014, (8):86
- [22] *Non-Small Cell Lung Cancer Treatment (PDQ)*, National cancer Institute at the national institute of health, 2014
- [23] Paul I., Mark J., et al. *Apoptosis block as a barrier to effective therapy in non small cell lung cancer* World Journal of W J C O Clinical Oncology 2014, 10;5 (4)588-94(1)
- [24] Burczyński M. E. *An introduction to toxicogenomics* CRC PRESS, 2000. 40-59.
- [25] Kelland L. R., Farrel N. P. *Preclinical Perspectives on Platinum Resistance*, Human 2000, Volume 59, Issue 4 Supplement, pp 1-8
- [26] Cosaert J, Quoix E., *Platinum drugs in the treatment of non-small-cell lung cancer*, Br J Cancer, 2002, 7;87(8):825-33.
- [27] Ulrich C. M., Robien K. et al., *Cancer pharmacogenetics: polymorphisms, pathways and beyond*, NATURE, 2003 volume 3
- [28] O'Grady S., Martin P, et al., *The role of DNA repair pathways in cisplatin resistant lung cancer*, Anti-Tumour Treatment reviews, 2014, 40 (10), 1161-70.
- [29] Zuco V, Cassinelli G, et al., *Targeting the invasive phenotype of cisplatin-resistant Non-Small Cell Lung Cancer cells by a novel histone deacetylase inhibitor*, Biochem Pharmacol. 2015 Jan 16. pii: S0006-2952(15)00027-1.
- [30] Vogel F., Motulsky A. G. *Problems and Approaches*. Second edition. Berlin, Springer-Verlag 1996.
- [31] Komar A. A. *Single Nucleotide Polymorphisms Methods and Protocols* Human Press Second Edition 2009. 5-20.
- [32] Cooper D. *Human Gene Evolution* BIOS Scientific Publishers Limited 1999. 7-15.

- [33] Vito P. , Pellegrini, et al. *Cloning of AIP1, a Novel Protein That Associates with the Apoptosis-linked Gene ALG-2 in a Ca<sup>2+</sup>-dependent Reaction jobs*, The Journal of Biological Chemistry, 1999. 274, 1533-1540.
- [34] Zhang K, et al., *Variations in the PDCD6 gene are associated with increased uterine leiomyoma risk in the Chinese*. Genet Test Mol Biomarkers, 2013 Jul. PMID 23551056,
- [35] PubMed RefSeq, Jan 2012, <http://www.ncbi.nlm.nih.gov/>
- [36] Qiang Yu,1, Chunxiao Zhou,1 et al., *A Functional Insertion/Deletion Polymorphism in the Promoter of PDCD6IP Is Associated with the Susceptibility of Hepatocellular Carcinoma in a Chinese Population*, DNA AND CELL BIOLOGY, 2013, Vol. 32, Number 8,
- [37] Suraj Saksena, Scott D. E., *Biochemical Society Focused Meetings ESCRTs and human disease*, Biochemical Society Transactions, (2009) 37, (167–172)
- [38] Lewis B., *Genes VII*, Oxford University Press, 2000.
- [39] Yuryev A., *PCR Primer Design (Methods in Molecular Biology)* human Press.
- [40] Peter J., Russell H. P. et al., *Genetics Fourth Edition*, Collins College Publishers U.S.A 1996.
- [41] Rasmussen H. B, *Restriction Fragment Length Polymorphism Analysis of PCR-Amplified Fragments (PCR-RFLP) and Gel Electrophoresis – Valuable Tool for Genotyping and Genetic Fingerprinting, 2012*
- [42] Wilson, K., and Walker, J. M. *Principles and Techniques of Practical Biochemistry*, Chapter 9, Cambridge University Press, 1994
- [43] Hames B. D., *Gel electrophoresis of proteins* oxford university press, 1998. 15-40.
- [44] Lee PY, Costumbrado J et al., *Agarose gel electrophoresis for the separation of DNA fragments.*, J Vis Exp. 2012 Apr 20;(62).
- [45] Andrews C.A., *The Hardy-Weinberg Principle*, Nature Education Knowledge, 2010, 3(10):65.
- [46] Lynch M., Walsh B. et al., *Genetics and Analysis of Quantitative traits* Sinauer Associates, Publisher , U.S.A, 1998.
- [47] PubMed.gov MAF
- [48] Fogiel M., et al., *The genetics problem Solver*, Staff of Research and Education Association, U.S.A 1993.
- [49] Griffiths A. F., Miller J. H., *An introduction to Genetic Analysis* W. H. Freeman and Company, New York, 2000. ISBN-10: 0-7167-3520-2

- [50] Eisenhauer, E.A., et al., *New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1)*. Eur J Cancer, 2009. **45**(2): p. 228-47.
- [51] Störkel S, Eble JN., et al., *Classification of renal cell carcinoma: Workgroup No. 1. Union Internationale Contre le Cancer (UICC) and the American Joint Committee on Cancer (AJCC)*. Cancer, 1997, **(5)**: 987-9 1997.
- [52] Bandla S, Pennathur A., et al., *Comparative genomics of esophageal adenocarcinoma and squamous cell carcinoma*. Ann Thorac Surg. 2012. **94**(3):1042.
- [53] Goli Samimi<sup>2</sup>, Daniel Fink, et al., *Analysis of MLH1 and MSH2 Expression in Ovarian Cancer before and after Platinum Drug-based Chemotherapy*, Clin Cancer Res April 2000 6; 1415
- [54] Chen S, Huo X, Lin Y et. al., *Association of MDR1 and ERCC1 polymorphisms with response and toxicity to cisplatin-based chemotherapy in non-small-cell lung cancer patients*, Environ Health. 2010. Vol.213, Issue 2, 140–145
- [55] Jennifer J., James T et al., *Proteins Associated with Cisplatin Resistance in Ovarian Cancer Cells Identified by Quantitative Proteomic Technology and Integrated with mRNA Expression Levels*, Molecular & Cellular Proteomics 2006, 5:433– 443,.
- [56] Hongyu Liu, Yonggang Liang, et al., *Gene silencing of BAG-1 modulates apoptotic genes and sensitizes lung cancer cell lines to cisplatin-induced apoptosis*, Cancer Biology & Therapy 15, 20109:10, 832-840.