Aryl hydrocarbon Receptor gene polymorphism and its association towards Lung Cancer risk and Clinical outcomes

A Dissertation

Submitted in partial fulfilment of the requirement

For the award of degree of

Masters in Technology

Under the supervision of:

Dr. Siddharth Sharma

Assistant Professor

Submitted by:

Sneha (601404020)

Department of Biotechnology

THAPAR UNIVERSITY

PATIALA-147004

INDIA

JULY 2016

DECLARATION

I hereby declare that the work being presented in the M.Tech dissertation entitled "Aryl hydrocarbon Receptor gene polymorphism and its association towards Lung Cancer risk and Clinical outcomes" has been carried out by me during the period of July 2015 to July 2016, under the guidance of Dr. Siddharth Sharma, Assistant Professor, Department of Biotechnology, Thapar University, Patiala. Further, I declare that I have not submitted the matter embodied in this dissertation for the award of any other degree or any other qualification of any university or examining body in India/elsewhere.

Sneha
M.Tech Biotechnology
Roll. No. 601404020

Date: 15 July, 2016
Place: Patiala

CERTIFICATE

This is to certify that the dissertation entitled "**Association of *Aryl Hydrocarbon Receptor* gene polymorphism with the risk of lung cancer in North Indian population**" being submitted by Ms. Sneha in partial fulfillment for the requirement of degree of **Master of Technology** in **Biotechnology** in the **Department of Biotechnology, Thapar University, Patiala** is a bonafide work carried out under the esteemed supervision and conception of **Dr. Siddharth Sharma,** Associate Professor, Department of biotechnology, Thapar University, Patiala.

Dr. Siddharth Sharma
Associate Professor
Department of Biotechnology
Thapar University, Patiala

(countersigned by)

Dr. Dinesh Goyal
Professor & Head
Department of Biotechnology
Thapar University, Patiala

Dr. S.S Bhatia
Dean of Academic Affairs
Thapar University, Patiala

ACKNOWLEDGEMENT

This is a golden opportunity for me to convey my sincere regard for all those who enabled me to accomplish my dissertation work successfully.

My utmost gratitude goes to my esteemed advisor **Dr. Siddharth Sharma**, Assistant Professor, Department of Biotechnology, Thapar University, Patiala for his expertise guidance, kindness, motivation, patience and providing me the opportunity to complete my dissertation work. It is his confidence-imbibing attitude, splendid discussions and endless endeavours through which I have gained a lot and learned a lot building up my future and personality.

My sincere thanks to **Dr. Dinesh Goyal,** Head Department of Biotechnology for his immense concern throughout the project work and for his help and valuable advice. A special thanks to all faculty members for their constant encouragement and support throughout the project work.

I am thankful for the help rendered by all PhD scholars throughout my project period without which it would have been difficult for me to achieve my goals. My special thanks and in depth feelings of gratitude to **Ms. Charu Bahl, Ms. Amrita Singh and Ms. Shweta Lawania** for their valuable guidance, incessant encouragement, constructive criticism, expedient advice and affection during the complete course. A special thanks to Mr. Babban and Mr. Surrender our lab assistants for their help. I feel lacunae of words to express my most heartfelt and cordial thanks to all my friends especially to Shitanshu and Ankita who always stood by my side during all the tough times.

The whole credit of my achievements during the project work goes to my parents and my guide. It was their unshakeable faith in me that has always helped me to proceed further.

Date: 15 July, 2016

Place: Patiala

(Sneha)

Table of Contents

LIST OF TABLES

LIST OF FIGURES

APPENDIX 1

1. **0.5M EDTA**: Dissolved 9.306g of disodium salt of EDTA in 20ml of deionized water, and then adjusted the pH to 8.0 by 1 M sodium hydroxide. Sterilized the solution by autoclaving.
2. **10% SDS**: Dissolved 1g of SDS in 10ml of deionized water.
3. **100mM Tris-Cl** (pH 8.0): Dissolved 0.32g of Tris-Cl in 10 ml of deionized water, then adjusted the pH to 8.0 by 1M sodium hydroxide. Sterilized the solution by autoclaving.
4. **10mg/ml Proteinase K**: Dissolved 10mg Proteinase K in 1ml of double distilled water. Sterilized the solution by autoclaving.
5. **1mg/ml BSA**: Dissolved 100mg of BSA in 100ml of deionized sterile water and kept at 4 C overnight.
6. **5M Sodium chloride (NaCl)**: Dissolved 5.85g of sodium chloride in 20ml of deionized water. Sterilized the solution by autoclaving.
7. **5X TBE buffer**: Dissolved 54g of Tris base and 27.5g of boric acid in 980ml of double distilled water and then added 20ml of 0.5 EDTA. Sterilized the solution by autoclaving.
8. **Ethidium Bromide (10mg/ml)**: Dissolved 1g of ethidium bromide in 100ml of water. Mixed the solution properly.
9. **Magnesium chloride (MgCl2) (100mM)**: Dissolved 0.41gms of MgCl2 in 20ml of deionized water and sterilized by autoclaving.
10. **Sucrose (1M)**: Dissolved 3.41 g of sucrose in 10 ml of deionized water and sterilized by autoclaving.
11. **TE buffer (pH 8.0)**: Added 1ml of 100mMTris- Cl (pH 8.0) and 200µl of 0.5M EDTA solution to 8.8 ml of deionized water. Sterilized the solution.
12. **Triton X- 100 (10%)**: Took 100µl of TritonX-100 and mixed with 900µl of deionized water and mixed properly.

ABBREVATIONS

AhR	Aryl hydrocarbon receptor
SNP	Single nucleotide polymorphism
TCDD	2,3 ,7, 8-tetrachlorodibenzo-*p*-dioxin
DRE	Dioxin Responsive Elements
Bhlh	Basic helix-helix
ARNT	Aryl hydrocarbon Receptor Nuclear Translocator
HSP	Heat shock protein
CYP	Cytochrome
PCB	Polychlorobiphenyls
PAH	Polycyclic aromatic hydrocarbon
HAH	Halogenated Aromatic hydrocarbon
BaP	Benzo-*a*-pyrene
PCDF	Polychlorinated dibenzofurans
IAA	Indole-3- acetic acid
ICZ	Indole [3,2-b-] Carbazole
I3C	Indole-3-carbinol
XAP2	Hepatitis B Virus Associated Protein 2
GST	Glutathione S-transferase
UGT	Uridine diphosphate glycosytransferase
OR	Odds ratio
CI	Confidence interval
MST	Median survival time
ADCC	Adenocarcinoma
SQCC	Squamous cell carcinoma
SCLC	Small cell lung cancer
COX	Cox proportional hazard model
HR	Hazardous ratio
OS	Overall survival

LC	Lung cancer
Rs	Reference SNP
PCR	Polymerase chain reaction
PY	Pack Years
SD	Standard Deviation

ABSTRACT

The present study was focused to assess the vulnerability of an individual towards Lung cancer having a polymorphism in the exon10 (rs2066853) and two polymorphism in the intronic region (rs10250822, rs2282885) and rs7811989 of AhR. A case control study was performed on the North Indian population. Odds ratio and class interval were used to assess the potency of the association. p value were also calculated to find out the statistical significance of the analysis. Multivariate and univariate analysis were also performed to assess the survival rate.

Cumulative cigarette smoking was interrogated to be a lead player along with other parameter responsible for the susceptibility towards lung cancer. The combinatorial effect of genotype and histological forms with combined effect of polymorphism was also studied and significant associations were found to exist.

The present study provided us with wider platform to study the rs10250822, rs2066853, rs10250822, rs7811989 polymorphism. Statistically significant values were seen amongst the genetic variants of AhR and were found to have a great risk towards Lung cancer. Tobacco smoking was affirmed as congener along with other environmental contaminants and showed higher OR and therefore a strong association with lung cancer risk. Cumulative effect of polymorphism amongst the AhR variants was also shown to be highly associated with the risk towards Lung cancer. AhR rs10250822, rs2282885, rs7811989 were found to be maximally associated with the risk of Lung cancer.

CHAPTER-1

INTRODUCTION

1 .INTRODUCTION

Epidemiological studies have contributed ceaselessly to the developing familiarity with the significance of the hereditary and acquired susceptibility variables which are represented to regulate the danger connected with the encounter with the natural contaminants. Information, prevalence, circulation and understanding of these natural contaminants in the population wherein an individual's weakness, which is inclined as his genetic make-up, contrasts and records for the danger of acquiring, risks like Cancer. A commitment towards understanding the component of harmfulness so as to evaluate the danger and alleviate the mischief esteemed to happen, is dealt with under the rubric of Molecular Epidemiology.

Lung cancer is the major reason of premature deaths throughout the world, accounting one-third of cancer deaths that arising from complex associations between the genetic and environmental factors .According to the WHO reports in 2030, approximately there would be 100 million deaths caused by the consumption of tobacco resulting increase risk of lung cancer. Smoking is strongly associated with the increased risk of lung cancer. The prevalence of lung cancer can be attributed to tobacco smoke and environmental pollutants. However,80-90% lung cancer is caused due to cigarette smoking which contains formaldehyde , acroline, PAH's, nitrosamines, benzene , dioxin, isoprene, acetaldehyde which are genotoxic carcinogens and rest 10% of smokers and non-smokers develop lung cancer. Tobacco smoke interacts with the carcinogens present in the environment which increases the incidence of lung cancer. This analysis indicates that genetic differences may determine vulnerability to lung cancer. Differential susceptilities amongst individual are accounted by the complexity of exposure to carcinogen, number of multiple allele present for an enzyme to be encoded and most importantly the predisposed genes for the xenobiotic metabolism in the individual. Certain variation in the genome and metabolic pathways leads to alternation in the detoxification and metabolism of contaminants which demonstrated its role in the etiology in that disease.

Aryl Hydrocarbon Receptor (AHR) is basically a protein which is encoded by the AhR gene *.AhR*, ubiquitously present in the inactivated form in the cytoplasm, a transcription factor after its interaction with the foreign ligand it gets activated in the nucleus and activates the transcription of certain enzyme like CYP1A1, cytochromeP-450s and also helps in the metabolism of certain compounds like PAH found in cigarette smoke, which are a potent carcinogen. Additionally, AhR couple with many cell signaling transduction pathway such as cell cycle (proliferation, arrest, apoptosis), cell departure, immune response .Therefore any

alteration in the activity of protein by genetic variation or due to some other changes may influence lung carcinogenesis.

International Agency for Research on Cancer (IARC) identified fifty five known carcinogens present in cigarette smoke and some unknown carcinogens which was initially procarcinogens but after that it is subjected to enzymatic medications to form active carcinogen. Cigarette smoking produces genotoxicity which enhance the approach towards carcinogenesis which results in Lung Cancer. Hence, both genotoxicity and non-genotoxicity serves as a predominant determinant in Lung Cancer. Therefore, synergistically metabolic activation and carcinogenic detoxification is a crucial factor in development of Lung Cancer.

Polymorphic variations in the specific gene are responsible for disposing the individual with higher possibility of having endangerments like Cancer. Variations in the polymorphism of AhR lead to changes in the functionality of protein and may also affect the metabolic and detoxification pathways. This result in the formation of DNA carcinogen adducts which transverses the generation of mutation which eventually causes Lung Cancer, mostly in smokers.

In this study, to trace the role of different genetic polymorphism in the susceptibility and overall survival towards Lung cancer in North Indian population, we evaluated four SNP's (rs2282885, rs10250822, rs2066853, and rs7811989) and its combinatorial association between these SNP's towards lung cancer risk in North Indian population.

CHAPTER-2

REVIEW OF LITERATURE

2.1 *AhR: Introduction and structure*

Aryl Hydrocarbon Receptor also known as AhR is present in the cytoplasm, ligand dependent transcription factor which gets activated upon binding with ligand and control a multitude of genes.(Junchieh J.Tsay *et al*;2013).According to the behaviour of the AhR to the ligand, it quickly switches between cytoplasm and nucleus(off/on system). AhR is strongly expressed in liver, adipose tissue and strongly expressed in bronchial epithelial cells. AhR plays a significant role in the detoxification of xenobiotics and drugs that involves certain metabolizing enzymes like GSTM1 (Phase-1) to facilitate bioconversion and the enzyme cytochromeP450 CYP1A1 involves in bioactivation of aromatic hydrocarbons (Barouki *et al*., 2007). Human AhR gene is 50kb long in size .The cytogenetic location of AhR present on 'q' arm of the chromosome .The structure of AhR gene contains 12exons and 10introns (Schmidt *et al*., 1996). The activation of AhR gene results in improper modification of gene expression .Furthermore, there is alteration in the cell signaling or metabolic pathways and formation of toxic intermediates.

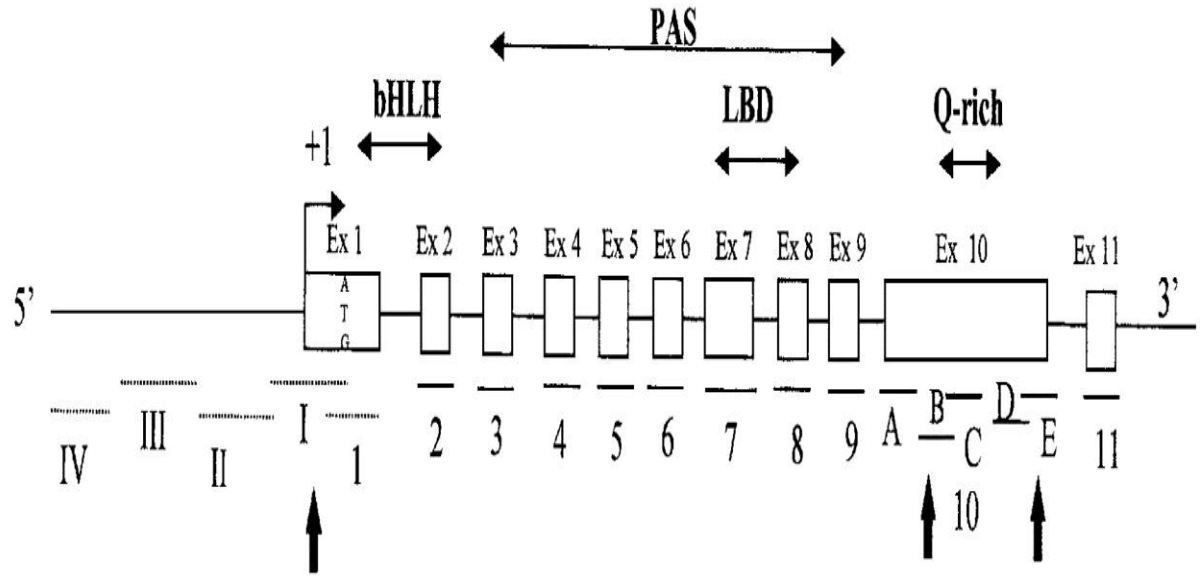

Figure 2.1: Structure of the human AhR gene **(Allan B. Okey *et al*., 2007)**

2.2: ***AhR and its Proteins***

The gene encodes a protein which participates in metabolism of carcinogens and environment contaminants and hence their polymorphism cause functional changes leading to lung cancer. AhR gene linked to a family of basic helix-loop-helix (bHLH) Period- ARNT-Single minded protein (Per-ARNT-Sim) (Wu D *et al*., 2013) AHRR and ARNT were also a part of this family found in human and drosophila. The transcription machinery gets activated from the –COOH terminus, by PAS domain which plays a predominant role in binding protein to molecule and can generally expands up to 260-310 amino acids in length. PAS (A) and PAS (B) are the two parts of PAS domain, which highly conserved regions. These regions are set apart from conserved regions. PAS (A) and PAS (B) carrying 44 amino acids repeats which can easily be distinguished by phenylalanine preceded by a chain of amino acids which ends at Histidine and Aspartic acid (Carver *et al*., 1998)

PAS domain helps in retaining cardiac tempo. The basic HLH region involved in binding of DNA to proteins which also enhance the transcription machinery specially sequence specific. The HLH region facilitates the dimerization (Meyer *et al*., 1998). It is composed of heterodimers , first one is bHLH-PAS which was present in activated form and get expressed in certain tissue while the other one is expressed all the time. Immense studies have explored that in bHLH domain to facilitate translocation, nuclear localization and some transporting signals are required (Hooper *et al*., 2011).

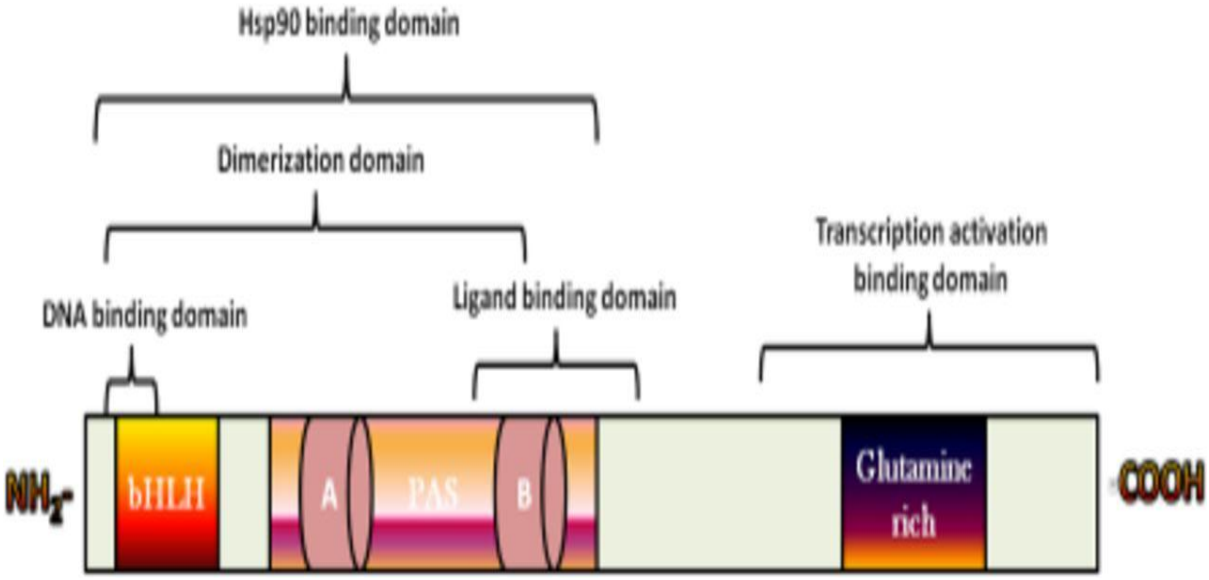

Figure 2.2: Aryl Hydrocarbon Receptor and its functional domains (Meyer et al.,1998)

2.3 ***AhR : Signalling Pathway-Cytosolic complex***

In the cytoplasm, the AhR complex was not bound to any ligand is surrounded by distinct proteins having different functions according. This complex contains two molecule of Heat Shock Protein (HSP 90), one molecule of XAP2, along with p23 (Petrulis *et al*., 2002)

AIP gene encodes a protein named XAP2.It consists of tetracopeptide repeat domain, it is helpful in assisting the interaction of HSP 90 and AhR. It also gives stability to protein after binding with p23.The transcriptional activity of AhR, XAP2 serves as both enhancer as well as repressor (Carver *et al*., 1998)

The dimer of HSP90 along with its co-chaperone p23 have a multiple role, it saves AhR receptor from proteolysis ,forcing the receptor to be in efficient conformation which is suitable for ligand binding and blocking the pre-binding of ARNT.

Hsp90 helps in efficient binding of ligand. It first binds to AhR and make it easy to orient in perfect confirmation and conserve the complex in unligated form in cytosol (Perdeww *et al*., 1998).

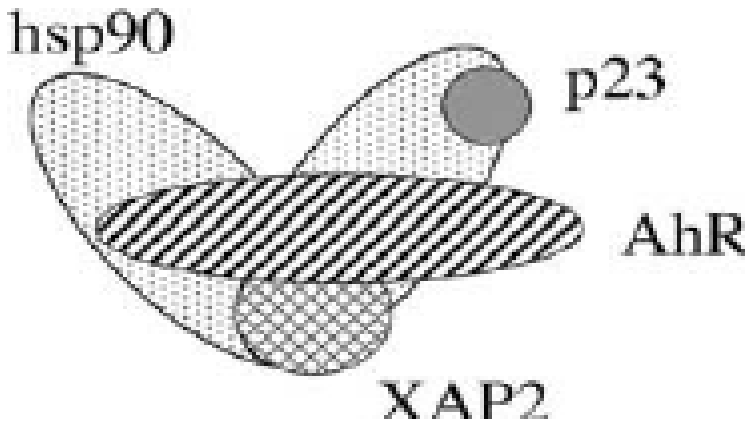

Figure 2.3*: AhR: Unligated Cytosolic Complex* **(Ramadoss *et al.*,2005)**

AhR ,ligand activated receptor resides in the cytosolic form of a complex comprising of XAP2, two molecules of Hsp90 along with co-chaperone p23.

2.4: ***Receptor Activation and DNA binding***

Initially the AhR lies in inactive state in cytoplasm (non-ligated) and shelled by heterogenous proteins and it covers the receptors until it identify the foreign ligand. After its binding with the ligand, it goes to cell by the technique of diffusion being ligand is hydrophobic in character directing to variation in the conformation. There is a strong affinity between AhR and ligand, leads to reduction in the dissociation rate (Timothy. V. Bejschlag *et al*.,2008).

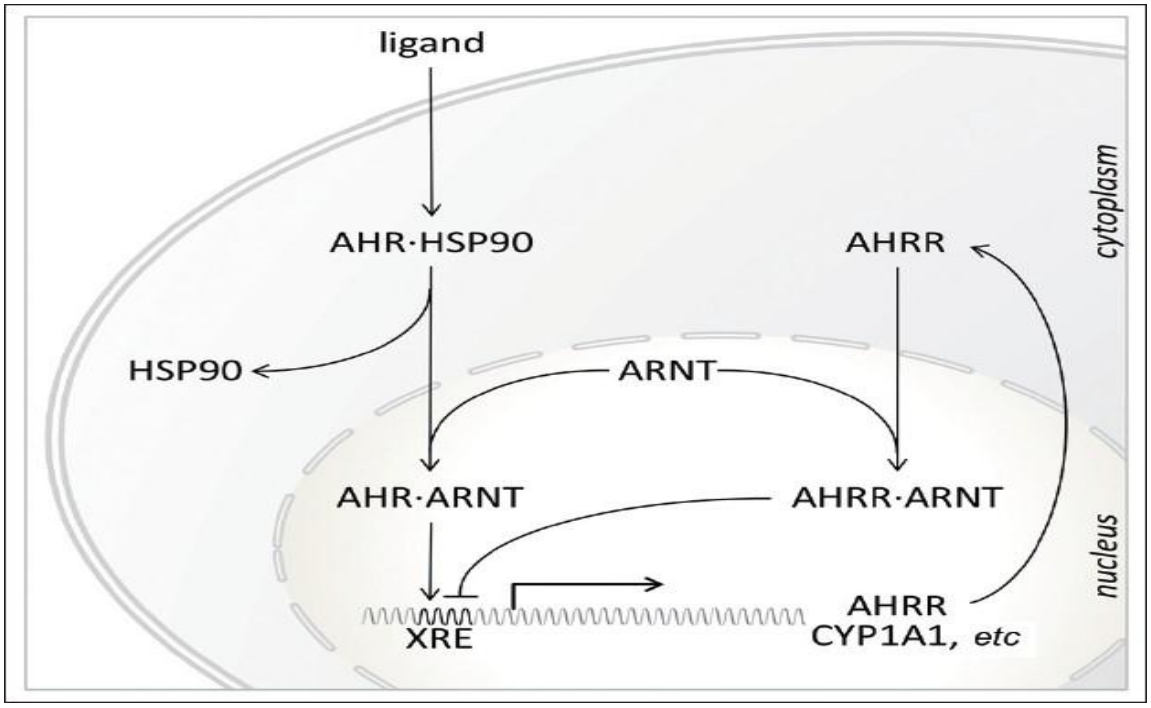

Figure 2.4: ***Schematic representation of ligand-activated AhR signaling.***

Once the AhR-ligand complex gets activated, then it goes to nucleus. It is then associate with ARNT to form AhR- ARNT heterodimer. Now this heterodimer additionally identify a consensus sequence (5'-T/GNGCGTGA/CG/CA-3') from the principal sequence 5'-GCGTG-3' called as DRE's or XRE's or AHRE. After that they bind to DRE's, these are the particular sequence on element which respond to a stimulus and then regulates the transcription process by certain enzyme(CYP1A1,CYP1A2,CYP1B1,UGT1A1,GST) which aids in detoxification of drugs and metabolism of xenobiotic. Following its binding, DNA along with ARNT-AhR engaged several proteins for transcription. This overlay the approach towards activation of gene which participates in metabolism.

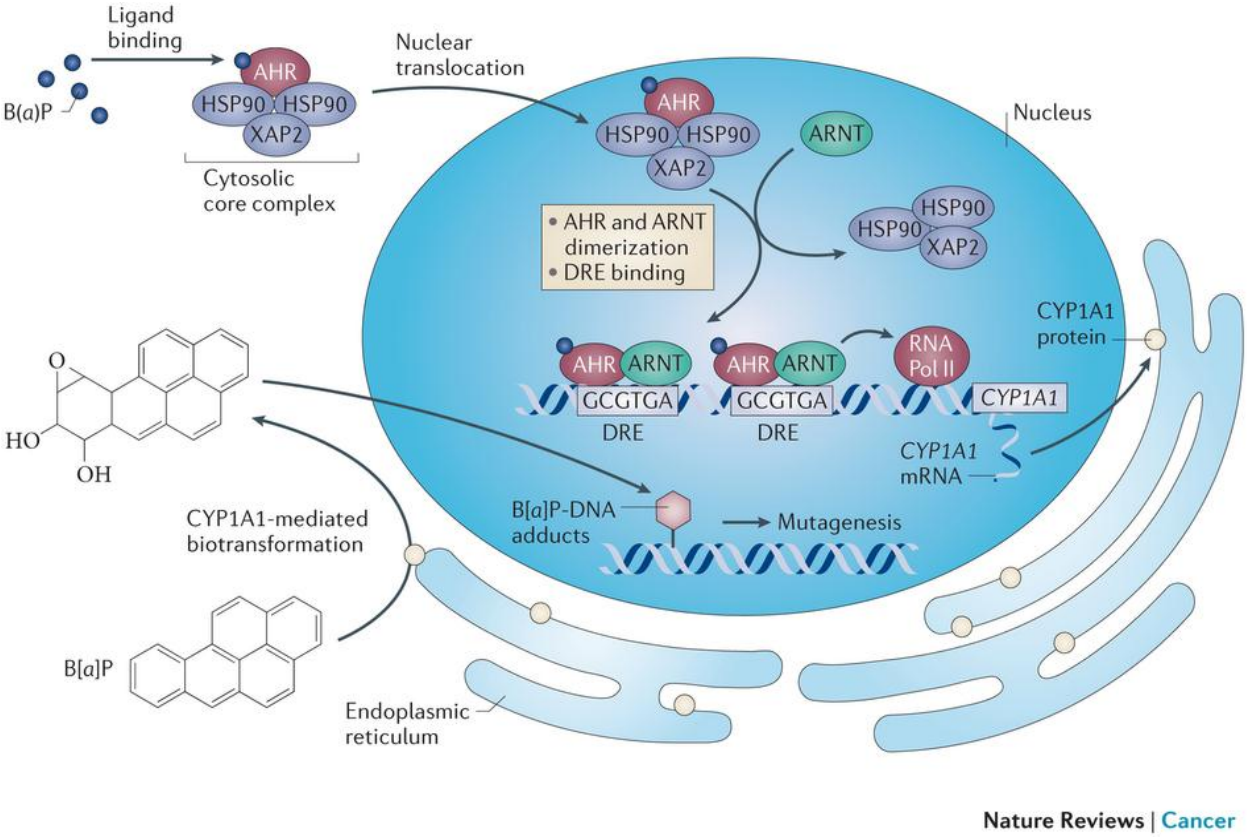

Figure 2.5: ***AhR Signaling Pathways (*****Iain M. Murray *et al*., 2014)**

2.5: ***Significance of AhR***

In humans, AhR has been viewed to be articulate intensely in numerous tissues and organs. Research have imitated that AhR shows a favorable role in retaining the cells in synchronization with the environment being. Studies endure and maintain the crucial role of AhR in numerous organ systems in the human body.

For the detoxification of drugs and environment contaminants, AhR is responsible for modifying various metabolic pathways for the regulation of Phase-1 and Phase-2 enzymes which helps in bioconversion (Tan *et al*., 2004)

Researchers show that AhR perform to restrain the process of transcription by combining enhancers, repressor, transcription factors, specific ligand binding or by modifying cell signaling pathways.

Studies stipulate interesting novel insights into the physiological function by expressing that it is vital for proper functioning of immune system. The dietary component maintains the

intestinal immunity through AhR. The compound such as Indole3carbazol (I3C) found in vegetable like cabbage, broccoli, cauliflower, from these AhR gets activated. Once AhR is activated in the cytoplasm and then it make a complex with ARNT resulting in dimerization which leads to formation of heterodimers and finally it switches to the nucleus, ultimately proceeding to formation of cancer (Hooper *et al*., 2008).

2.6: ***Ligand Independent activation***-

Another part of AHR conduct to consider is the likelihood that there is genuine ligand free initiation. In backing of this idea is the evident capacity of lifted cAMP levels to actuate the AHR. Moreover, it is conceivable that in cells that have moderately high AHR levels, for example, numerous tumor cell lines, the AHR experiences dynamic nucleocytoplasmic transporting, which could prompt AHR and ARNT heterodimerization without ligand. Support for this idea can be found after looking at the level of held atomic AHR in a human head and neck squamous cell carcinoma cell line, NH30, contrasted with typical human keratinocytes. In any case, the chances that AHR ligands are available in these cells can't be removed. Shear stress in endothelial cells can likewise prompt AHR actuation, despite the fact that the mechanism of initiation is not known. . Co-articulation of a mutant AHR not able to tie ligand (AhR-A375I) and ARNT expanded AHR-intervened transcriptional action recommends that the AHR have ability to heterodimerize with ARNT without ligand, in spite of the fact that heterodimerization potential have all the earmarks of being entirely inefficient(Iain A. Murray *et al*.,2014)

2.7: ***Role of AhR in Immune Response***

The AhR performs multiple tasks in immune system. In mice, the thymus increases in size and there is also increase in the activity until maturation, after that begins to atrophy termed as thymus atrophy which also weakens the immune system by the intake of xenobiotic compounds (Silkworth *et al*., 1995)

The maturation of B-lymphocytes shows very lethal effects in living organism. Studies reported that TCDD cease the expression of B-lymphocytes producing maturation protein. These proteins impede and halt the process of differentiation of B-lymphocytes.

TCDD, a known congener perceive to disturb the interleukin-5 production.TH2 cell was helpful in mediating the immune response. TCDD, a carcinogen activates the AhR and have

the proficiency of HSC which acts on stimuli which was there in environment (Thurmond *et al*., 2009)

The immune system of the body affects the adaptive immunity after interaction with carcinogens like TCDD. The AhR expression stimulated by leukocytes which helps in translocation from cytoplasm to nucleus (Fujimaki *et al*., 2002)

2.8: ***Role of AhR in Reproductive System***

Studies indicate that individual which exposed to trichlorophenoxy-acetic acid develops sexual abnormality results in lowered libido (Moses *et al*., 1998) It was also found that exposure of TCDD, showed change in the rate of offspring ratio. Studies have shown that exposure of dioxin in animals, results in altering the route of sexual differentiation (Mocarelli *et al*., 2000). Sometimes, it also hampers the neuro endocrine function. The coke-oven workers, showed unstable spermatogenesis causing infertility in males (Izawa *et al*., 2007).

2.9: ***AhR and Organ Development***

AhR plays an predominant role in organogenesis involved in development of organ like liver,kidney,heart and skin.Non-uniform organogenesis was seen in knockout mice. It was also helpful in cardiac vascular functioning and also in the devepment of new blood vessels.Improper renal and non fuctioning cardiovascular development,after activation of AhR by TCDD. It is widely expressed in liver (Thacckaberry *et al*., 2002)

The AhR performs diverse functions such as cardiac vascular flow, metabolism,angiogenesis, haematopoisis, organogenesis .

AhR is also found in various pathways involving oestrogen and androgen receptor.It's interaction with stimulus decreases the cellular component or stops the cellular differntiation facilitated through androgen (Bunger *et al*., 2008)

2.10: ***AhR Ligands***

AhR has been assigned as the receptor whose expression is completely ligand dependent.AhR showing a promising site for binding and hence it binds to distinct ligands .

Due to difference in structure and binding position, ligands are divided into two groups – Synthetic ligands and Endobiotic/Dietary ligands.

2.10.1: ***Synthetic Ligands***

It involves a diverse array of carcinogens like PCB, Benzo-a-pyrene ,TCDD. It involves both halogenated and non-halogenated ligands (Denison *et al*., 2003)

HAH (Halogenated Aromatic Hydrocarbon) is the most important class of ligand which triggers its activity in minimal amount.They get accrued in the microenvironment because of their chemical stability and hydrophobic nature.

TCDD (2,3,7,8-Tetrachlorodibenzo-p-dioxin) , a contaminant present in the environment formed during the organic synthesis and also produced during combustion of natural materials like wood ,oil, fossil fuels. Workers exposed to TCDD by inhalation are strongly associated to lung cancer. IARC assigned TCDD as carcinogen.

TCDD and HAH makes the compound environmentally and metabolic stable.

Polycyclic Aromatic Hydrocarbon (PAH) , a potent carcinogen in humans and found in higher amount in tabacco smoke.It also produced during coking process.They have a adverse effect on lungs,skin and causes inflammation.Epidemiologic studies suggested people exposed to PAH reported Asthma.

Substances in tabacco smoke are not reactive ,but after its metabolism they get converted into ROS which produces carcinigenic effects.

Benzo[a]pyrene (B[a]P) is the most important PAH activates by AhR,which get metabolized into toxic substances .These lethal intermediates damage the DNA resulting in mutation which progresses tumor formation in lungs (Shimizu *et al*., 2000)

Pentachlorobiphenyls (PCB's) is organochlorine mixture and a toxic contaminant present in environment .IARC classed PCB as human carcinogen.Interestingly, the major source of this is fish.Studies reported that PCB causes serious health hazards ,causes neurological disorder (Boucher *et al*., 2009).

2. 10. 2: ***Endobiotic / Dietary Ligands***

The Endobiotic ligands differ from the synthetic one because they are synthesized endogenously. Mostly these ligands have less binding affinity for AhR. It activates the transcription of gene which encodes several enzyme which helps in detoxification.

Bilirubin is the breakdown product of hemoglobin in human body, showing less affinity for AhR and not ample for AhR activation. Finding revealed bilirubin show strong affinity in yeast system but the similar approach fails in mammals. In mammalian system, TCDD shows maximum affinity (Perdew *et al*., 2005)

Examples of endobiotic /dietary ligands – Indole comprising compounds like 2-(1'H-Indoe-3'- carbonyl)-thiazole-4-carboxylic acid methyl ester (ITE), tryptophan photo products such as 6-formylindole [3,2-b] carbazole [FICZ], bilirubin and metabolites of arachidonic acid.

Tryptophan based metabolites in plants like Indigo serves as strong activators of AhR (Dension *et al*., 2003)

Crucifers like broccoli, cabbage strengthen the intestinal immune system. These vegetables contain I3C which undergoes condensation to form diindolylmethane (DIM) and indolocarbazole after exposure to gastric acid. These products have high affinity for ligand (Hooper *et al*., 2011).

The activation of cytochrome450 (CYP1A1) enzymes by AhR, formulate these ligand to facilitate biotransformation and then elimination (Bjeldanes *et al*., 1991)

2.11: ***Atypical AhR Ligands***

This class of ligands includes some uncommon compounds which seems like ligands but do not functions as ligand to AhR. But it does not bind to receptor, yet it handles to provoke the gene activation which encodes the enzyme involved in multiple roles.

Omeprazole, Myriscitin and the benezimidazole derivatives are the example of atypical ligand. Omeprazole is used for the treatment of ulcers and burns and functions as proton pump inhibitor. Literature shows induced activity of AhR, in case of failure of transduction pathway (Quattrochi *et al*., 1993).

2.12: ***AhR Serves As A Regulator***

2.12.1 ***Phase 1 Enzymes*:**

In the first phase of biotransformation, the introduction of polar functional groups increases the polarity of the drug so that it becomes more water-soluble. It involves multiple reaction including oxidation, reduction, hydrolysis. Pharmacologically, it converts pro drug which is in inactive form to the active one and then goes to the second phase of detoxification.

The target gene of AhR involves cytochromeP450 protein which encodes enzyme which helps in metabolism. Mostly, it was located in endoplasmic reticulum. Till now there are 57 cytochromeP450 identified in humans, which participate in drug metabolism.CYP1A1, CYP1A2, CYP1B1 are highly prominent amongst all .Literature shows ,CYP1A1 is the only enzyme which require AhR for its activation. It is expressed in extrahepatic tissue .Its role is to metabolize xenobiotics and biotransforms carcinogens.

Among all, CYP1A2 is directly regulated by the AhR. Studies revealed that this enzyme have able to metabolise acrylamines and heterocyclins and shows enhanced AhR expression on exposure to tobacco smoke.CYPIA2 reduces caffeine intermediates and serves as biomarker for measuring its activity (Nakajma *et al*., 1999)

2.12.2 ***Phase 11 Enzymes:***

In the second phase of biotransformation, this is the last step of detoxification involving glucorinidation, sulfation, methylation reaction. It involves addition of glucuronic acid, having methyl group to make water soluble products which then excretes out from the body.

Phase 11 enzymes include UGT specially UGT1A1, glutathione-S-transferase, ADH.

In humans, GST is the most effective enzyme involved in metabolizing xenobiotic compounds. Few cytotoxic drugs which cause severe toxicity and various side-effects , it get metabolized by GST family of enzymes (Daly *et al*., 2003).

TCDD and PCBs activates the expression of uridine diphosphate glucuronosyl transferase (UGT1A1) .Only the enhancer region of UGT1A1 is responsible for stimulating AhR ,being ligand-independent keeps DRE in its enhancer region. Rise in the level of bilirubin corresponds to strong binding with AhR, resulting increased UGT1A1 level.

Aldehyde dehydrogenase (ALDH) is basically used for the oxidation of aldehyde, lipid and carbohydrates to carboxylic acid. Findings shows that DRE facilitates the induced and constitutive expression of ALDH (Boesch *et al*., 1998).

2.13 ***Role of AhR Polymorphism in Cancer***

Polymorphism arises due to genetic variation resulting in inter- individual differences among the same population. It is seen in 1% of the population. These polymorphisms may alter the response of individual towards mediators of damaged DNA. Genetic variation leads to lung cancer susceptibility. It is due to presence of low or medium penetrant gene rather than highly

penetrating gene. Any alteration in the single nucleotide in the genome termed as Single nucleotide polymorphism. Numerous SNPs have being confirmed which effects the health-outcome triggered by environmental exposure to contaminants, which serves as a major factor in lung cancer susceptibility. In the NCBI database, there are 120 SNPs are present in coding region of AhR gene while intronic region identified within 5' and 3' UTR (untranslated regions).The molecular basis of polymorphism is due to presence of any these factors-

(1) Change in nucleotides or there is amino acid substitution in the exonic region of gene which alters the activity of enzyme.
(2) Sometimes, deletion occurs in the exonic region which leads to enzyme inactivation resulting in no protein synthesis.
(3) Changes in the polyadenylation site resulting post transcriptional repression which alters the transcripts formed.

For studying polymorphism in a specific gene, different type methods and methods are used. Here we select the Candidate gene method, in which gene and nucleotide polymorphism are taken into account. We also followed this approach for studying polymorphism in North Indian population.

In spite of higher variation in affinity towards xenobiotic exhibited by AhR, extensive amount of amino acid conservation was seen. This shows a vigorous and pro-vital function of AhR in human physiology and homeostasis (Boffeta *et al*., 2000).

Most polymorphism occurs around exon10, which encodes the transcription domain of receptor, which is important for regulation of other region of gene. It includes the basic HLH region and PAS domain. In human AhR, studies shows at codon678, polymorphism was not observed. While at codon554 and 570 polymorphism affects the functioning of proteins.

Polymorphism mainly observed in the transcription domain are R554K,P517S and V570I.Among all variants of AhR , R554K is the most commonly found polymorphism present at exon10.This region is responsible for the transcription of other gene. In women, higher risk of breast cancer occurs due to Arg554Lys polymorphism. Studies reported this SNP is not responsible for the risk of lung cancer (Kajawajiri *et al*., 1995).

Polymorphism has been widely observed at Lys554Leu, Val570Ile and Pro571Ser in humans. But there is no significant relation among polymorphism to the phenotype of TCDD toxicity.

Studies reported an predominant association when consideration is the combination of these gene, for example Lys554Leu, Val570Ile, Pro591Ser.

When in combination, the mutation in Lys554Leu and Val570Ile were seen to cancel the induction of CYP1A1.Further studies shows that a single variation in the gene shows correlation when considering with other polymorphism in the same subject. Epidemiological studies showed decreased susceptibility of individual towards lung cancer when used in combination.

Polymorphism at codon517 was prevailing in African-specific population. Studies reflected that there is no association between AhR polymorphism with ligand to receptor binding dimerization, translocation occurrence in signaling of AhR, nuclear localization, nuclear export, stability of protein along with co-activators (Harper *et al*., 2002)

Studies reported a close association of AhR with mammary gland tumor genesis. Polymorphism in AhR gene leads to higher risk of acquiring lung, breast, pancreatic cancer (Kim *et al*., 2007).

In order to assists the science fraternity in exploring the facts, extending our boundaries of wisdom and unravel the associative powers possessed by polymorphism towards risk of lung cancer.

Here, we study the four SNPs in the AhR gene-

12.13.1 AhR 2282885-

In this AhR variant, the polymorphism is located in the intronic region where nucleotide substitution take place T to C. rs2282885 variants is strongly linked with 1-OHP level among side oven workers, which are exposed to PAH.

12.13.2 **AhR** *Arg554Lys rs2066853-*

This is one of the most important polymorphism studied in AhR gene. In this, amino acid substitution occurs from arginine to lysine in AhR protein. It is located on exon10.It seems to be in the conserved region .The sequence that codes for the transactivation domain present in exon10, resulting playing a major role in upregulation and downregulation in transcription activity of AhR gene. Previous data seems ,this substitution does not affect the function of AhR to control the transcription activity, which is induced by CYP1A1 and CYP1B1.This genotype display a significant risk of lung cancer in smokers.

AhR-ARNT complex is linked with host response by which it safeguards it from the adverse effects of carcinogenesis. Any substitution or any change in nucleotide sequence disturbs the machinery of transcription by the genes which codes for enzymes. The chances of carcinogen formation increases if there is overexpression of gene or when it is not fully expressed leading transformation of procarcinogen to carcinogen.

AhR involves other receptors by showing number of sites available for binding to receptor which includes the xenobiotic compounds and environmental contaminants. After that dimerization occurs which help in translocation from nucleus to cytoplasm and finally approaches to tumorigenesis.

Smoking is considered as causation factor but not serves as a biomarker for lung cancer.

12.13.3 **AhR *10250822-*** In this polymorphism, substitution of amino acids occurs from T to C in the intronic region. Epidemiological studies shows this AhR variant is strongly linked with urinary 1-OHP level in the PAH exposed individuals, which is a known carcinogen. Furthermore, rs2282885 also show association with 1-OHP level among PAH-exposed workers.

However, polymorphism of AhR effect the level of urinary 1-OHP in individuals exposed to PAH.

12.13.4 **AhR *rs7811989-***

In this polymorphism, amino acid substitution occurs from G to A in the intronic region of AhR gene. This leads to alteration in the RNA splicing pathway which results in loss of transcription activity (Ping Bing *et al*., 2008).

CHAPTER-3

AIMS AND OBJECTIVE

AIM OF THE STUDY

1. To assess the four single nucleotide polymorphism (rs2282885, rs2066853, rs10250822, rs7811989) in AhR gene and its association towards lung cancer and specific histology.

2. To study the effect of smoking and its association with AhR gene polymorphism towards lung cancer risk.

3. To estimate the combinatorial effect of different genotypes of AhR and its association towards lung cancer susceptibility

4. To access the overall survival of lung cancer patients in relation with the four polymorphic variants of the AhR gene.

CHAPTER-4

MATERIALS AND METHODS

4. Material and Methods

4.1: Sample Collection

The current study enrolled a total of 320 controls and 297 lung cancer patients from the Department of Pulmonary Medicine, Post Graduate Institute of Medical Education and Research (PGIMER) Chandigarh, India. This study has been revised and accepted by the Institute ethics committee of PGIMER. Informed written agreement was obtained from all enrolled patients or their representatives. All the enlisted patients were histopathologically diagnosed as having NSCLC or SCLC. Patients under observation having a previous history of cancer were excluded from the study. There was no age, gender, smoking, histological or TNM stage restrictions. The control group of the study consisted of 320 individual having no lung cancer history at the time of blood collection; they entered the hospital for health check-ups.

Around, 3-6 ml of blood was collected in vacutainers from each individual enrolled in that study. All controls were sub-grouped as sex, age and smoking parameters in order to avoid any sampling bias. Each participant filed up the detailed questionnaire with the help of trained interviewer. The questionnaire compromised information on demographic and smoking characteristics like tobacco habits such as smoking of beedi/cigarette etc. Individuals who smoked regularly were classified as smokers. They were further classified as light and heavy smokers on the basis of pack years (PY) that were calculated by the formula:

[(cigarettes or beedis per day/20)*years smoked]

PY less than or equal to 25 were considered as light smokers and PY greater than 25 were considered as heavy smokers. The medical information of cases such as histology, TNM classification, clinical staging, primary tumor size, involvement of lymph node and metastasis were obtained from medical records of the hospital.

4.2: DNA extraction

The genomic DNA was isolated using standard Protein K digestion, phenol/chloroform extraction and ethanol precipitation method from blood samples of both cases and controls (Field *et al* 1999).

Requirements:

1. Washing buffer
2. Lysis buffer
3. Phenol: Chloroform: Isoamylalcohol (25:24:1)
4. Chloroform: Isoamylalcohol(24:1)
5. Isopropanol
6. TE buffer

Preparation of Buffers:

Washing buffer, Lysis buffer and TE buffer were prepared as shown in tables below.

Table 1: Preparation of washing buffer

Stock concentration	**Working concentration**
1M sucrose	320mM sucrose
100% Triton X-100	1% Triton X-100
100mM Magnesium Chloride	5mM magnesium Chloride
100mM Tris-HCl pH (8.0)	10mM Tris-HCl pH (8.0)

Table 2: Preparation of lysis buffer

Stock concentration	**Working concentration**
1M Tris HCl pH (8.00)	400mM Tris HCl pH (8.00)
10% SDS	1% SDS
0.5M EDTA	60mM EDTA
5M NaCl	150mM NaCl
10mg/ml Proteinase-K	100µg/ml proteinase-K

Procedure of DNA Isolation

1. Took equal amount of blood and Washing Buffer (5 ml) and then it was added and mixed thoroughly.

2. After mixing, centrifuged it at 3500rpm for 5 minutes.

3.Discard the upper aqueous layer(supernatant) and add 5ml of Washing buffer (1.6ml 1M Sucrose,0.5 ml Triton X-100, 0.25ml $MgCl_2$, 0.5 ml 100mM Tris HCl and 0.26ml of water) to the pellet.

4. Then resuspend the pellet in the buffer and centrifuged again.

5. Repeat this step three times.

6.Dissolved the pellet in 5ml of Lysis buffer (1 M Tris HCl 2ml,10% SDS 0.5ml ,0.5 M EDTA 0.6ml, 5M NaCl 0.15ml,10mg/ml Proteinase-K 0.05ml and water 1.7ml) and mixed thoroughly.

7. Then it is subjected to overnight incubation at 44°C .

8. Added an equal volume of Phenol: Chloroform: Isoamyl alcohol (PCI) 25:24:1(25ml-Phenol, 2. 4 ml- Chloroform and 0.1ml- isoamyl alcohol) and mixed all the contents slowly.

9. Centrifuged at 8000rpm for 10minutes at 4°C.

10. Then slowly took the upper aqueous layer and again add PCI mix and then centrifuged at 8000rpm for 10minutes.

11. Took the upper aqueous layer and added equal amount of Chloroform: Isoamyl alcohol (24:1).

12. Centrifuged it at 6500 rpm for 5 minutes and took the upper layer.

13. To the aqueous upper layer, add equal volume of chilled Isopropanol or 2.5 times volume of absolute Ethanol and mixed it slowly.

14. Store it at -20°C for 1-2 hours.

15. Centrifuged it at 12,000 rpm for 10 min at 4°C.

16. The supernatant was discarded and the pellet of DNA was washed with chilled 70% Ethanol twice at 10,000 rpm for 5 minutes.

17. Decant ethanol and air dry the pellet.

18. Dissolved the pellet in 50µl-150µl Tris-EDTA buffer depending on the size of DNA pellet (Bartlett & White, 2003).

4.3: Quantitative and qualitative estimation of DNA template

For quantification of DNA, Thermo Scientific Nanodrop Spectrophotometer is used. It uses only 1µl of sample. It was done by using UV spectrophotometer and the absorbance of sample was noted at two wavelengths A_{260} nm and A_{280} nm. To calculate the purity of DNA, ratio of absorbance at 260nm and 280 nm is used.

If A_{260} nm/ A_{280} ratio is approximately 1.8 then the DNA sample is pure and free from contaminations of RNA and proteins.

If A_{260} nm/ A_{280} ratio is approximately 2.0 or >1.8 it shows RNA contamination in the sample, whereas a ratio less than 1.8 shows protein contamination in the sample.

The concentration of pure double stranded DNA is 50µg/ml and ratio A_{260} nm/ A_{280} is equals to 1.0

The DNA concentration in a solution was determined by using this formula given below:

Concentration (µg/ml) = A_{260} X 50 µg/ml X Dilution factor

Procedure:

1. Add 1µl of deionized water onto the lower optical surface of Nanodrop Spectrometer for cleaning.
2. After that, open the Nanodrop software displaying on the screen
3. Then select Nucleic acid Module.
4. First take a blank measurement by loading 1µl of TE and selecting blank from the options.
5. Measured the amount of nucleic acid sample by loading 1µl of DNA sample and then select the option- measure.

The Nanodrop Spectrometer automatically calculates the concentration of DNA and its purity.

4.4: Resolution of DNA Fragments on Agarose Gels

Requirements

- Electrophoresis buffer (TAE or TBE)
- Ethidium bromide solution
- Electrophoresis-grade agarose

- 6X loading dye
- DNA molecular weight markers
- Horizontal gel electrophoresis apparatus
- Gel casting platform
- Gel combs (slot formers)
- DC power supply

Preparing 5X TBE (1000ml)

- Tris base - 54 g
- Boric Acid - 27.5g
- EDTA (0.5M) - 20ml
- Make up final volume with water

Preparing 6X Loading Dye (20ml)

- 0.25% Bromophenol blue - 0.05gm
- 0.25% Xylene Cyanol - 0.05gm
- 40% Sucrose - 8gm
- Make up final volume with TE buffer

Preparation of the Agarose Gel for Electrophoresis

- Prepared a sufficient quantity of electrophoresis buffer.
- Added the required amount of Agarose to a volume of Electrophoresis buffer which is sufficient for preparing the gel. For example, for genomic DNA 0.7% gel (0.7g agarose in 100ml 0.5X TBE) was prepared while for the PCR products 1.7% gel (1.7g agarose in 100ml 0.5X TBE buffer) was prepared.
- Before pouring onto the casting plate, melted agarose was cooled to 55°C in a water bath to prevent deforming of the gel apparatus.
- Before gel casting, Ethidium bromide solution was added to the melted agarose gel to a final concentration of 0.3μg/ml to enable visualization of DNA when seen under UV Transilluminator.
- Poured the melted Agarose onto the gel casting apparatus between 0.5 and 1 cm thick and a comb is inserted after pouring gel, in such a way so that no bubbles are trapped

underneath the combs and if bubble were there on the surface of the agarose ,it was removed before the setting of the gel.

Loading and running the gel

- The gel comb was removed slowly with proper care without disrupting the sample wells, after the gel got solidified
- Placed the gel casting tray containing the set gel in the Electrophoresis tank.
- Add sufficient amount of Electrophoresis buffer to cover the gel, till the tops of the wells are submerged and make sure no air pockets were trapped within the wells.
- DNA samples were prepared by mixing 5µl DNA with 2µl of 6X loading dye and 2ul water in case of genomic DNA or by mixing 5µl DNA with 2µl of 6X loading dye in case of PCR product. Then samples were loaded into the wells with micropipette. Care was taken to prevent mixing of the samples between wells.
- A ladder or DNA molecular weight marker is loaded in the first well in case of PCR and digestion products
- Connected the electrodes to a power supply and allowed the Electrophoresis unit to run at 60 V until the dyes migrated the required distance.
- Turned off the power supply, disconnected the leads, and remove the electrophoresis buffer from the electrophoresis tank.
- DNA was visualized by placing the gel on a UV transilluminator and then photographed using Gel Documentation

4.5 Restriction Digestion of AhR SNP's

Restriction digestion utilizes an enzyme which cleaves the DNA or specific sequence .Here is the digestion of 4 SNP's-

1. BsmA1- A restriction enzyme isolated from the microorganism Bacillus Stearothermophillus Abb4, basically a E.coli strain. It recognizes GTCTC (1/5)^ sequence and cut perfect at 37^0 C in NEB buffer (10X).This buffer is used because it has maximum efficiency with BsmA1 enzyme and also helps in controlling the homeostasis condition in the reaction mixture of digestion. This is used for digesting amplicons produces from the amplification product of AhR **rs10250822.**

5'...G T C T C (N) 1^...3'
3'...C A G A G (N) 5^...5'

2. BspH1- The genomic source from which it is isolated was Bacillus species H. This restriction enzyme was used for digesting amplicons produced from the PCR-product of the AhR gene **rs2282855**.It recognizes T^CATGA site and cuts good at 37^0 C in a specific buffer.

5' . . . T ^C A T G A . . . 3'
3' . . . A G T A C^T . . . 5'

3. Dde1- The genomic source from which it is isolated is Desulfovibrio desulfuricans, an E.coli strain. It is used for digesting the amplicons of **Arg^{554}Lys rs2066853** produced after the PCR. This is used with Cut Smart buffer (10X). It recognizes the C^TNAG sequence which produces 5' protruding ends.

5'…C ^ T N A G…3'
3'…G A N T ^C…5'

4. BbV1- It is isolated from Bacillus brevis, an E.coli strain. It recognizes the GCAGG (8/12) ^ sequence and cuts good at 37^0 C. This enzyme is used for digesting amplicons of the PCR-product of **rs7811989**.

5'…G C A G C ^…3'
3'…C G T C G ^…3'

4.6 Polymerase Chain Reaction (PCR)

PCR is a very powerful technique used for the amplification of DNA fragment which generates thousand million copies of DNA with the help of DNA polymerase enzyme (Taq polymerase),

Concept-

PCR uses a small fragment of DNA sequence called as primer, here which the nucleotide bases are attached to which long DNA molecule which works as a template for producing multiple copies. The amplification of DNA is achieved with the help of polymerases, which binds to a DNA to form long strands .Then the polymerase adds nucleotides to the primer which serves as template for producing several copies of DNA. It usually amplifies a product in approximately 2 hours.

Steps-

1. Denaturation

2. Annealing

3. Extension

In the denaturation step, dsDNA denatures at high temperature and converted into ssDNA. After that the temperature is again lowered to 50-60^{0}C, which allows primers to bind to the DNA and the polymerase starts copying the DNA. In this step, primer anneals to the template resulting in primer extension. The purpose of extension is to make complimentary copy of DNA strand, hence DNA make multiple number of copies. This process is repeated again and again which doubles the DNA, then run the amplified product on agarose gel and then visualized under UV transilluminator.

<u>Requirements</u>-

- 10X PCR buffer
- BSA
- Forward Primer
- Reverse Primer
- dNTPs
- Taq DNA polymerase
- Water
- DNA sample

Table 3 - Reaction mixture of PCR- RFLP

Reagent	**Stock concentration**	**Final Reaction concentration**	**Quantity Used**
Additive 1 BSA	1000 µg/ml	100µg/ml	44µl
PCR Buffer (Mg Conc.)	10X, 25mM	1X 1.5mM total	44µl
Primer (forward)	10µM	0.5µl	22 µl
Primer (reverse)	10µM	0.5µl	22 µl
Taq Polymerase	2.0U	1.5µl	22 µl
dNTPs	10mM each	0.2	8.8 µl
PCR Grade Water			189.2 µl
DNA Template	100ng/µl	300ng	3 µl

Table 4: Cycling profile of PCR for AhR variants

Steps	**Temperature**	**Time**
Initial Denaturation	94^0 C	5 min
Denaturation	94^0 C	45 seconds
Annealing	***65°C (AhR rs2282885)***	30 seconds
	61°C (AhR rs10250822)	30 seconds
	54°C (AhR rs7811989)	30 seconds
	54°C (AhR Arg554 Lys rs2066853)	30 seconds
Polymerization	72^0 C	30 seconds
Final Extension	72^0 C	5 min

The role of PCR primers is to provide 3'-OH group, this is the region where DNA polymerases can add DNTP's .So it needs to be choosen selectively. There should not be

annealing between primer present in the mixture or any copies of same or the reverse direction if present, which results in primer-dimer formation. And also there should be no annealing of primer to them, as it forms internal hairpin and loop structure which prevent the annealing of primer to DNA template.

Table 5: List of primers and restriction enzymes used					
Gene, SNP ID	**Exons/Intron, (Nucleotide/AminoAcid susbstitution)**	**Primer sequence**	**Product size (bp)**	**Restriction enzyme**	**Restriction pattern**
AHR rs7811989	Intron A-G	F5'-GTT TTC TTG TTA CAA AGT CTG AAC AC-3' R5'-TCC TCT TCA GAA ATA AAC ACA TAA AC -3'	250	BbVI	250bp=A 199+51bp=G
AHR rs10250822	Intron T-C	F5-TGA TGC TTG GTA TGG GGT CTG AGT G-3' R5"-CCT CCG TGG GCT GAA GAA TAT GTG T -3'	412	BsmAI	412bp= T 219+19=C
AHR rs2282885	Intron T-C	F5'-AAC TGC ACT TGA CTT GGA TTA CGC T -3' R5'-AAG ATA GCA TTC AGA CTG GCA TTG G -3'	377	BspHI	377bp=A 258+119=G
AHR rs2066853	Exon G –A (Arginine-Lysine)	F5-'CAT TGA TTT TGA AGA CCT CA -3' R5'-CTG AAG GTA TGA AGG GAG- 3'	152	DdeI	152=G 135,17=A

5: Statistical Analysis-

Statistical analysis are used to check whether the 4 SNP'S were in Hardy-Weinberg equilibrium or not (p2+2pq+q2=1); where p is the frequency of the wild-type allele and q is the frequency of the variant allele). χ2-test are used to assess both the cases and control and independent t-test has been used for continuous variable like age and smoking .This was used to determine the genotypic frequency in cases and control using chi-square test. Med Cal (version 16.4.3) software was used to compute genotypic frequency in a population (Medcalc software, Ostend, Belgium).To evaluate the risk of lung cancer and AhR polymorphism, logistic regression analysis were used that gives Odds-ratio(OR), 95% Confidence interval(CI) and p-value. By using logistic regression, cases and controls are taken as independent variable while sex, age, smoking and pack years are taken as dependent variable. A p-value which is less than 0.05 is considered as highly significant. Paired-t was also used for the survival analysis. SHEsis software was used for linkage disequilibrium and haplotype analysis of 4SNP's in which D' and r2 value were calculated. Further, for overall analysis to estimate risk of lung cancer in different subgroups were made on the histology, smoking status, KPS, ECOG .Multivariate logistic regression are used to see the overall effect by a opposite genotype which was identified from single using logistic regression analysis. Moreover, multifactor dimensionality factor (MDR) was used to evaluate the best interaction model, using the online 0.5.1version of the open-source MDR software (http://www.epistasis.org).The model having minimum prediction error and maximum cross-validation consistency(CVC) considered as best model.Univariate and multivariate analysis are evaluated by using Kaplan-Meier survival analysis and Cox proportional hazardous ratio.

CHAPTER-6

RESULTS

6.1: Genotyping

The DNA is isolated from the blood samples of cancer patients and its presence is checked by running sample on 0.7% agarose gel. After isolation, the samples are preserved as stock and store at -20^{0}C. From the stock DNA, these samples were diluted to 100ng/µl which serves as a template in PCR reaction for amplification.

Specific primers are designed which is particular to a specific gene, then subjected to PCR for amplification. To check whether the DNA is amplified or not ,the amplified product were run on 1.7% agarose gel. EtBr is added to the gel, which intercalates with the DNA and helps in visualization under UV transilluminator.

The amplified product is subjected to restriction digestion with a specific enzyme which is capable of cleaving the product at a particular restriction site which is unique for all SNP's. For checking digestion, the samples were run on 2.5% or 3% agarose gel.

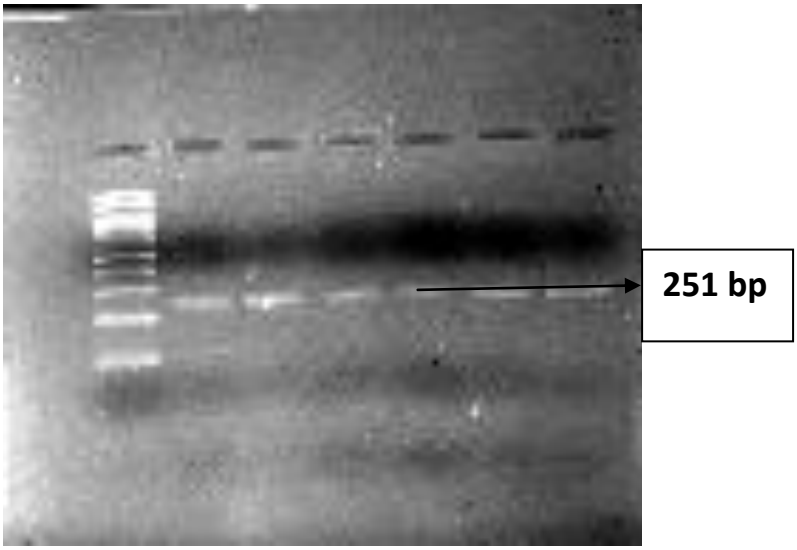

Figure 6.1.1: Representative example of PCR product of AhR rs7811989 with amplicon size 250bp

Lane 1:100bp Ladder (G Biosciences)

Lane 1, 2, 3, 4, 5, 6: PCR products

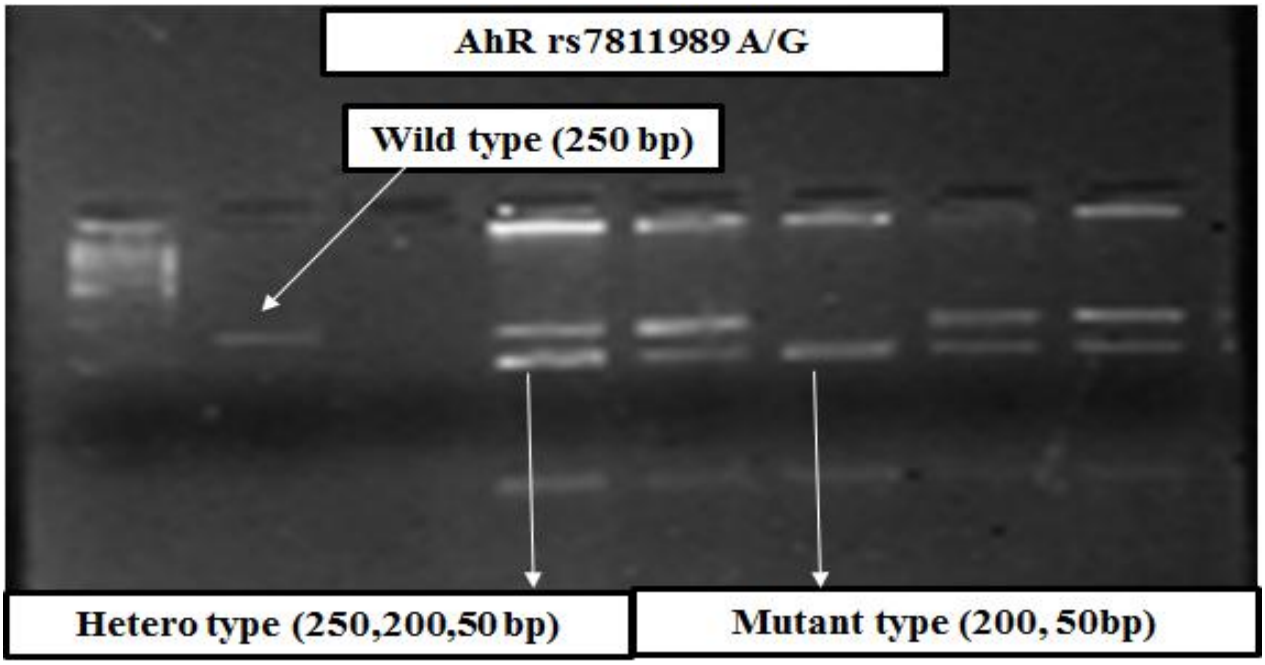

Figure 6.1.2: Representative example for the Restriction digestion of PCR product of AhR rs7811989 showing the digested products (wild: 250bp; mutant: 199 bp, 51bp; heterozygote: 250bp, 199bp, 51bp

Lane 1:100bp Ladder (G Biosciences);

Lane 2: uncut ;Lane 5: Mutant ;Lane 3,4,6,7,8,9,10,11,12,13: Wild type

Lane 14, 15: Heterozygote type

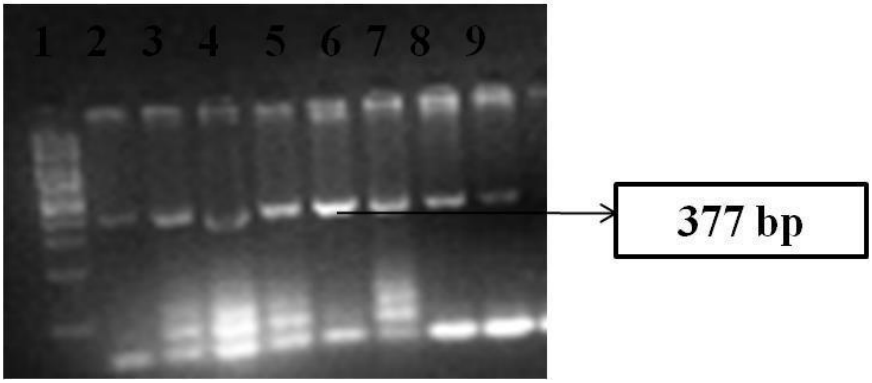

Figure 6.1.3: Representative example of the PCR product of AhR rs2282885 with amplicon size 377bp Lane 1: 100bp Ladder (G Biosciences); Lane 2, 3, 4, 5, 6 ,7, 8, 9: PCR products of amplification reaction

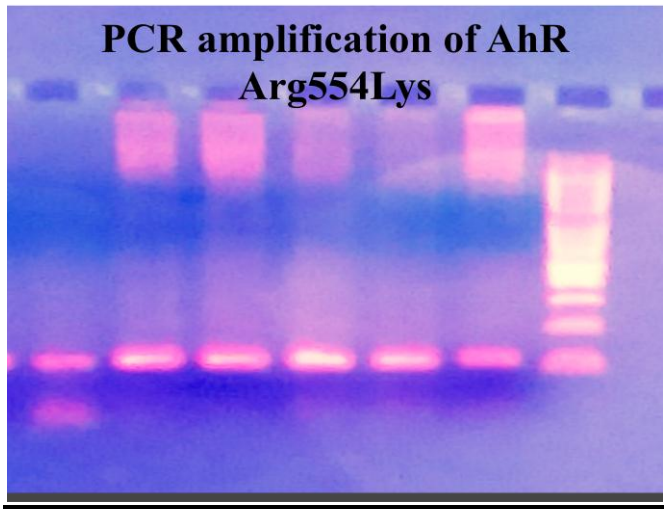

Figure 6.1.4: Representative example showing the PCR product of AhR Arg 554Lys rs2066853 (amplicon size: 152bp)

Lane 1-10: PCR Products of AhR Arg 554 Lys rs2066853

Lane 11: 50bp Ladder (G Biosciences)

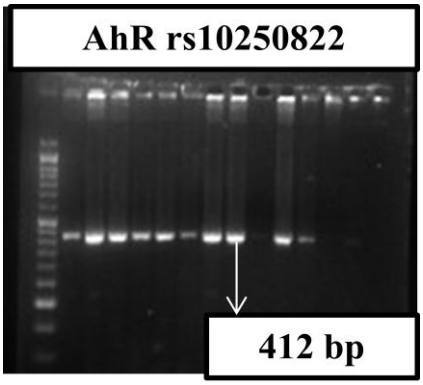

Figure 6.1.5: Representative example of PCR product of AhR rs10252882 with amplicon size 412bp

Lane 1:100bp Ladder (G Biosciences)

Lane 2, 3, 4, 5, 6, 7, 8, 9: PCR product

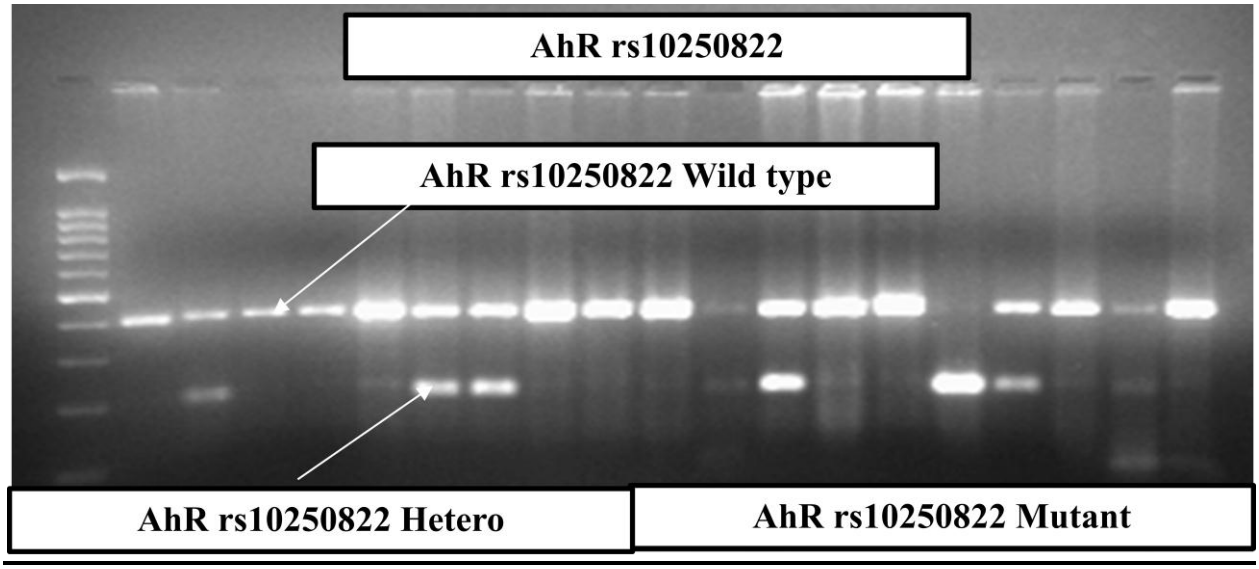

Figure 6.1.6: Representative example for the Restriction digestion of PCR product of AhR rs10252822 showing the digested products (wild: 412bp ; mutant: 213 bp;heterozygote:412bp,213bp,

Lane 1:100bp Ladder (G Biosciences);

Lane 2: uncut ;Lane 5: Mutant ;Lane 3,4,6,7,8,9,10,11,12,13: Wild type

Lane 14, 15: Heterozygote type

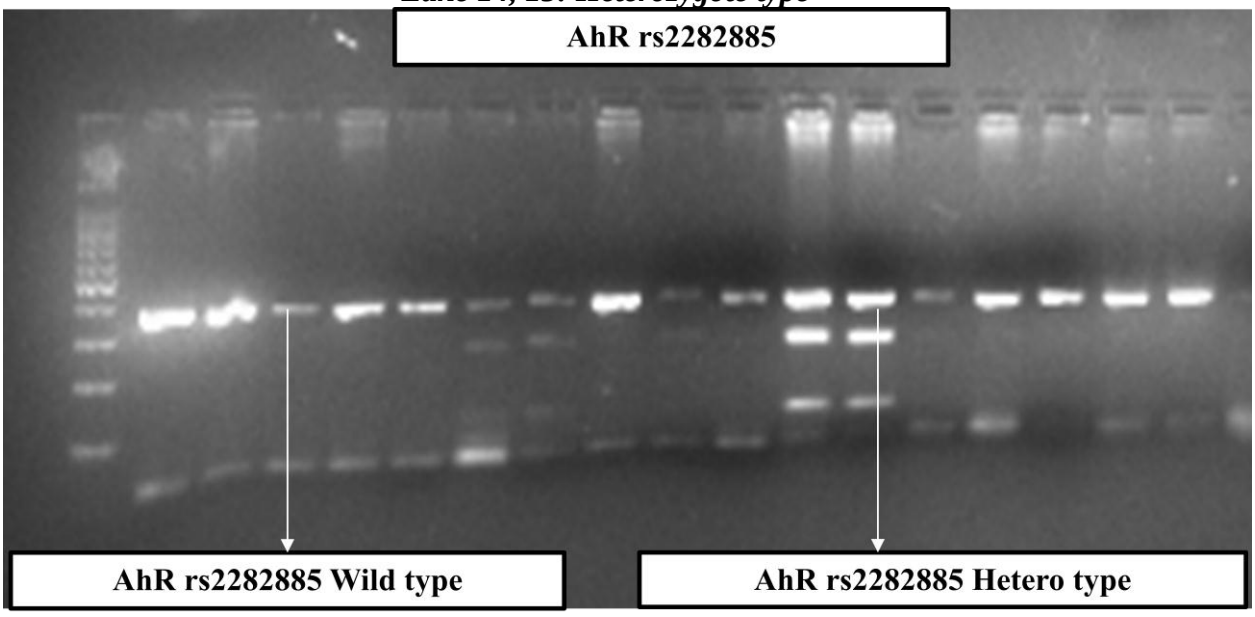

Figure 6.1.7: Representative example showing the restriction digestion of PCR product of AhR rs2282885 showing the digested products (wild type: 377bp; mutant type: 258bp, 119bp; hetero type: 377bp, 258bp, 119bp)

Lane 1:100bp Ladder (G.Biosciences); Lane 2, 3, 6, 10, 11, 12, 14, 16: Wild type

Lane 8, 13, 15: heterozygote type; Lane 9: Mutant type

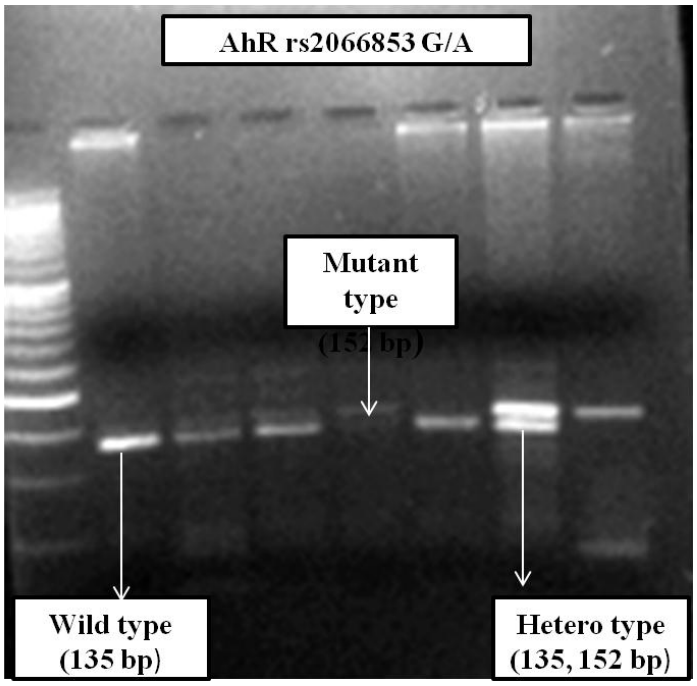

Figure 6.1.8: Representative example of restriction digestion of PCR products of AhR rs2066853 showing the digested products (wild type: 135bp, hetero type: 135bp, 152bp; mutant type: 152bp)

Lane 1: Ladder (G.biosciences); Lane 2: Uncut
Lane 3, 4, 6: Wild type; Lane 5: Mutant type; Lane 7: Heterozygote type

6.2: Epidemiology for AhR gene polymorphism

Table 6: Demographic characteristics among cases and controls			
Variable	**Cases, n (%) N=297**	**Controls, n (%) N=320**	***p* – value**
Age (years)			
Mean ± SD	57.6± 10.81	53.00 ± 10.42	**p< 0.0001**
Range			
Gender			
Male	254(85.52)	265(82.81)	0.932
Female	43(14.47)	55(17.18)	
Smoking Status			
Smokers	233(78.45)	221(69.06	0.937
Non–Smokers	64(21.54)	99(30.93)	
Pack Years			
Mean ± SD	27.5 ± 34.04	17.61± 19.92	<0.0001
Abbreviations: SD = Standard Deviation, n = total number of case patients or control subjects[a] *p*-values were derived from Pearson Chi – square test except age; Student t-test was used for age. All *p*-values are two – sided. *p*< 0.05 was considered statistically significant			

The demographic characteristics of the study population are shown in table. Demographic factors include age, gender, smoking, pack years which serves as confounding factors, were adjusted for genotypic analysis by using logistic regression. The study population encompasses a total of 297 cases and 320control. The mean age of cases was 57.6± 10.81 where as the mean age of all controls are 53.00±10.42.The current study include 254 (85.52%) males and 43(14.47%) females in cases while 265(82.81%) males and 55(17.18%) females in controls. The study was also analyzed for the difference in the smoking status of the study group. The study comprises of 233(78.45%) smokers and 64(21.54%) non-smokers in cases while 221(69.06%) smokers and 99(30.93) non-smokers. When pack years is taken in consideration, the study shows the cases had higher mean value of pack years smoked27.5±34.04 as compared to 17.61±19.92 of controls having p-value<0.0001,which is highly significant.

6.3: Genotypic distribution and association of AhR gene polymorphism among the patients with different histological types of Lung cancer

Table 7: Genotypic distribution of the AhR genetic variants and their association with risk of Lung cancer along with the stratified association analysis based on histology

		OVERALL			**ADCC**			**SQCC**			**SCLC**		
AHR rs7811989 A/G	**Controls n (%) N =320**	**Cases n(%) N=297**	**AOR (95% CI) [a]**	***p* [b]**	**Cases n(%) N =97**	**AOR (95% CI) [a]**	***p* [b]**	**Cases n(%) N =129**	**AOR (95% CI) [a]**	***p* [b]**	**Cases n(%) N=71**	**AOR (95% CI) [a]**	***p* [b]**
AA	155 (48.4)	112 (37.71)	1.00 (Reference)		34 (35.05)	1.00 (Reference)		50 (38.75)	1.00 (Reference)		33 (46.47)	1.00 (Reference)	
AG	145 (45.31)	152 (51.17)	1.40 (0.99-1.98)	0.05	54 (55.67)	1.78 (1.074-2.949)	0.025	67 (51.93)	1.19 (0.159-1.886)	0.439	25 (35.21)	1.30 (0.724-2.367)	0.372
GG	20 (6.25)	33 (11.11)	2.32 (1.24-4.32)	0.007	9 (9.27)	1.82 (0.738-4.505)	0.912	12 (9.30)	1.89 (0.851-4.32)	0.117	13 (18.30)	4.24 (1.708-10.56)	0.001
GA+AG	165 (51.56)	185 (62.28)	1.51 (1.09-2.10)	0.012	63 (64.94)	1.77 (1.089-2.881)	0.021	79 (61.24)	1.30 (0.842-2.013)	0.235	38 (53.52)	1.61 (0.929-2.819)	0.089
A	453	376											
G	185	218											
MAF		0.35											
		OVERALL			**ADCC**			**SQCC**			**SCLC**		

AHR rs10250822 T/C	**Controls n (%) N =320**	**Cases n(%) N=297**	**AOR (95% CI) [a]**	***p* [b]**	**Cases n(%) N =97**	**AOR (95% CI) [a]**	***p* [b]**	**Cases n(%) N =129**	**AOR (95% CI) [a]**	***p* [b]**	**Cases n(%) N =71**	**AOR (95% CI) [a]**	***p* [b]**
TT	184 (57.5)	139 (46.08)	1.00 (Reference)		45 (46.39)	1.00 (Reference)		62 (48.06)	1.00 (Reference)		33 (46.47)	1.00 (Reference)	
TC	120 (37.5)	136 (45.79)	1.40 (1.00-1.97)	0.046	43 (44.32)	1.52 (0.92-2.50)	0.09	60 (46.51)	1.38 (0.89-2.15)	0.144	32 (45.07)	1.32 (0.75-2.32)	0.33
CC	16 (5)	22 (7.40)	1.77 (0.867-3.590)	0.110	9 (9.27)	2.48 (0.955-6.44)	0.06	7 (5.42)	1.22 (0.455-3.28)	0.68	6 (8.45)	2.85 (0.96-8.42)	0.05
TC+CC	136 (42.5)	158 (53.19)	1.44 (1.043-2.00)	0.026	52 (53.60)	1.60 (1.00-2.57)	0.049	67 (51.93)	1.36 (0.887-2.092)	0.157	38 (53.52)	1.45 (0.84-2.50)	0.173
T	488	414											
C	152	180											
MAF	0.24	0.29											
		OVERALL			**ADCC**			**SQCC**			**SCLC**		
AHR rs2282885 T/C	**Controls n (%) N =320**	**Cases n (%) N=297**	**AOR (95% CI) [a]**	***p* [b]**	**Cases n (%) N=97**	**AOR (95% CI) [a]**	***p* [b]**	**Cases n(%) N =129**	**AOR (95% CI) [a]**	***p* [b]**	**Cases n(%) N =71**	**AOR (95% CI) [a]**	***p* [b]**
TT	221	207	1.00		68	1.00		82	1.00 (Reference)		52	1.00	

	(69.06)	(69.69)	(Reference)		(70.10)	(Reference)		(63.56)			(73.23)	(Reference)	
TC	88 (27.50)	83 (27.94)	1.04 (0.172-1.50)	0.80	23 (23.71)	0.85 (0.49-1.48)	0.57	39 (30.23)	1.27 (0.79-2.04)	0.31	18 (25.35)	0.86 (0.46-1.61)	0.65
CC	11 (3.437)	15 (5.05)	1.73 (0.75-3.96)	0.19	6 (6.18)	2.01 (0.68-5.92)	0.20	8 (6.20)	2.26 (0.83-6.12)	0.106	1 (1.40)	0.64 (0.72-5.46)	0.68
TC+CC	99 (30.93)	98 (32.99)	1.11 (0.78-1.58)	0.53	29 (29.89)	9.96 (4.53-21.90)	<0.0001	47 (36.43)	1.37 (0.879-2.162)	0.161	19 (26.76)	0.86 (0.47-1.57)	0.62
T	530	497											
C	110	133											
MAF	0.17	0.21											
		OVERALL			**ADCC**			**SQCC**			**SCLC**		
AHR rs2066853 G/A	**Controls n (%) N =320**	**Cases n (%) N=297**	**AOR (95% CI) [a]**	***p* [b]**	**Cases n(%) N =97**	**AOR (95% CI) [a]**	***p* [b]**	**Cases n(%) N =129**	**AOR (95% CI) [a]**	***p* [b]**	**Cases n(%) N =71**	**AOR (95% CI) [a]**	***p* [b]**
GG	224 (70.0)	229 (77.10)	1.00 (Reference)		74 (76.28)	1.00 (Reference)		100 (77.51)	1.00 (Reference)		56 (78.87)	1.00 (Reference)	
AG	67 (20.93)	55 (18.51)	0.741 (0.490-1.229)	0.158	19 (19.58)	0.836 (0.465-1.501)	0.549	23 (17.82)	0.643 (0.369-1.119)	0.118	12 (16.90)	0.597 (0.293-1.121)	0.15
AA	29 (9.06)	13 (4.37)	0.347 (0.172-0.702)	0.0032	4 (4.12)	0.363 (0.120-1.093)	0.717	6 (4.65)	0.305 (0.118-0.783)	0.013	3 (4.22)	0.322 (0.092-1.129)	0.076
AG+GG	96 (30.0)	68 (22.89)	0.628 (0.433-0.911)	0.014	23 (23.71)	0.694 (0.450-1.188)	0.18	29 (22.48)	0.533 (0.322-0.880)	0.0141	15 (21.12)	0.540 (0.286-1.020)	0.057

G	515	167											
A	125	27											
MAF	0.20	0.04											

[a] Adjusted Odds ratios, 95% confidence intervals and their corresponding p-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status. [b] Two- sided χ2 test for either genotype distribution or allelic frequencies between the cases and controls. Abbreviations: ADCC, Adenocarcinoma; SQCC, Squamous cell carcinoma; SCLC, Small cell lung carcinoma

AhR rs7811989-

The genotypic and allelic frequencies were calculated for all the variants of AhR which were studied. For AhR rs7811989, out of 297 cases and 320 controls were enrolled in the study. Among the 297 lung cancer patients, (51.17%) of the individual were found to have heterozygous genotype (AG), (37.71%) had homozygous wild genotype and (11.11%) individuals had mutant genotype. (GG).While in controls, (45.4%) of individual have (AA) wild genotype, (45.31%) individual had heterozygous genotype (AG) and (6.25%) had mutant genotype. In the above table, we have combine both heterozygous and mutant genotype (GA+AG) as single genotype, because of less number of mutants in both cases and controls. The genotypic frequencies for AhR rs7811989 polymorphism in the control group (n=320, $\chi 2$=3.36,p-value=0.06)were in Hardy-Weinberg equilibrium. When AA genotype taken as reference, patients having mutant genotype (OR=2.32, 95%CI=1.24-4.32,p-value=0.007) shows a highly significant value with almost 2-fold increase in the risk of acquiring lung cancer. However people having heterozygous genotype display marginal significant value (OR=1.40, 95%CI=0.99-1.98, p-value=0.05) ,decrease in the risk of lung cancer. In case of combined genotype (AG+GA), no significant value was observed when AA taken as reference.

For AhR rs7811989**,** a total of 297 cases of cancer patients were studied, out of which (32.68%) of patients diagnosed with ADCC (43.3%) people had SQCC (23.9%) had SCLC. When histology is further classified into different genotype, the percentage of reference group had ADCC (35.5%), SQCC (38.75%) and SCLC (46.47%). While (55.67%), (51.93%) and (35.21%) have heterozygous genotype. Although (9.27%), (9.30%) and (18.30%) have mutant genotype. Due to low number of mutants in cases and controls, mutant and heterozygous genotype were combined as single genotype. The total count was (64.94%), (61.24%) and (53.52%) for ADCC, SQCC and SCLC. For population study, while considering different histological forms of lung cancer, people diagnosed with ADCC having heterozygous genotype showed a statistically significant value (OR=1.78, 95% CI=1.074-2.949, p-value=0.025) towards risk for lung cancer. Similarly the combined genotype of homozygous mutant and heterozygous also show significant value (OR=1.77, 95% CI=1.089-2.881, p-value=0.021).Whereas SQCC does not show any statistically significant value with any genotype and hence not associated towards risk for lung cancer. However, SCLC display four-fold increased susceptibility for mutant genotype (OR=4.24, 95% CI=1.708-10.56, p-value=0.001) towards lung cancer. When taking wild genotype (AA) as reference, no

significant value was observed in heterozygous and combined genotype. Minor allele frequency for cases and controls are 0.30 and 0.35. For population study, while considering different histological forms of lung cancer, people diagnosed with ADCC having heterozygous genotype showed a statistically significant value (OR=1.78, 95% CI=1.074-2.949, p-value=0.025) towards risk for lung cancer. Similarly the combined genotype of homozygous mutant and heterozygous also show significant value (OR=1.77, 95% CI=1.089-2.881, p-value=0.021).Whereas SQCC does not show any statistically significant value with any genotype and hence not associated towards risk for lung cancer. However, SCLC display four-fold increased susceptibility for mutant genotype (OR=4.24, 95% CI=1.708-10.56, p-value=0.001) towards lung cancer. When taking wild genotype (AA) as reference, no significant value was observed in heterozygous and combined genotype. Minor allele frequency for cases and controls are 0.30 and 0.35.

AhR *rs10250822-*

In case of AhR rs10250822, PCR-RFLP was used to obtain the genotype of cases and controls which were recruited for population study. A total of 297 cases and 320 controls were enrolled in the study. Among the 297 lung cancer patients, (45.79%) of the individual were found to have heterozygous genotype (TC), (46.08%) had homozygous wild genotype and (7.40%) individuals had mutant genotype (CC).Whereas in controls, (57.5%) of individual have (TT) wild genotype, (37.5%) individual had heterozygous genotype (TC) and (5%) had mutant genotype. In the above table, we have combine both heterozygous and mutant genotype(TC+CC) as single a genotype, as mutants in both cases and controls were present in less number. The genotypic frequencies for AhR rs10250822 polymorphism in the control group (n=320, $\chi 2$=0.4,p-value=0.5) were deviating from Hardy-Weinberg equilibrium. In AhR 10250822, AA genotype was taken as reference, patients having joint genotype (TC+CC) comprising heterozygous and homozygous mutants (OR=1.44,95%CI=1.043-2.00,p-value=0.026) shows a marginal significant value towards risk of acquiring lung cancer. However patients with heterozygous and mutant genotype no significant value was observed when AA is taken as reference.

For AhR 10250822, a total of 297 cases of cancer patients were studied, out of which (32.68%) of people diagnosed with ADCC, (43.3%) people with SQCC, (23.9%) with SCLC. When we further categorize the genotype in terms of histology, (46.39%), (48.06%) and (46.47%) had wild genotype for ADCC, SQCC and SCLC. While (44.32%), (46.51%) and

(45.07%) have heterozygous genotype. Although (9.27%), (5.42%) and (8.45%) have mutant genotype. Due to less number of mutants in cases and controls, mutant and heterozygous genotype were combined as single a genotype. The total count was (53.60%), (51.93%) and (53.52%) for ADCC, SQCC and SCLC.

In the above table among all histological forms- of lung cancer, mutant genotype of ADCC group gave significant values and almost 2-fold increase in the advent of developing the disease (OR=2.48, 95% CI=0.955-6.44, p-value=0.06).While combined genotype (TC+CC) of same group, also display marginal significant value (OR=1.60, 95% CI=1.00-2.57,p-value=0.049)resulting chances of developing a disease. Patients having SQCC failed to show any significant value in all genotype and hence considered to be not associated with lung cancer. SCLC patients having mutant genotype (OR=2.85, 95% CI=0.96-8.42,p-value=0.05) displaying more than 2-fold increase in the risk of developing lung cancer. The Minor allele frequency for cases and controls are 0.24 and 0.29 respectively.

AhR rs2282885

In case of AhR rs2282885, PCR-RFLP was used to obtain the genotype of cases and controls which were recruited in the population study. A total of 297 cases and 320 controls were enrolled in the study. Among the 297 lung cancer patients, (27.94%) of the individual were having heterozygous genotype(TC), (69.69%) had homozygous wild genotype and (5.05%) individuals had mutant genotype (CC).While in controls, (69.06%) of individual were with (TT) wild genotype,(27.50%) individual had heterozygous genotype (TC) and (3.43%) were mutant genotype. In the given table, we have combined both heterozygous and mutant genotype (TC+CC) to form single genotype, because of less number of mutants in both cases and controls. The genotypic frequencies for AhR rs10250822 polymorphism in the control group (n=320, $\chi 2$=0.36,p-value=0.54) were deviating from Hardy-Weinberg equilibrium.

In AhR rs2282885, a total of 297 cases of cancer patients were studied, among them (32.68%) of patients were diagnosed with ADCC, (43.3%) with SQCC,(23.9%) with SCLC. When histology was further classified into different genotype, (70.10%), (63.56%) and (73.23%) had wild genotype for ADCC, SQCC and SCLC. While (23.71%), (30.23%) and (23.35%) have heterozygous genotype .Although (6.18%), (6.20%) and (1.40%) were having mutant genotype. Due to low count of mutants in cases as well as controls, mutant and

heterozygous genotype were combined together as single genotype. The total count of combined genotype was (29.89%), (36.43%) and (26.76%) for ADCC, SQCC and SCLC.

In the above table when considering different sub-groups of histology, with reference to TT genotype, it was observed that individual diagnosed with SQCC and SCLC does not show any statistically significant value. As odds ratio was nearly less than 1 for all the genotypes, hence a protective effect was observed in SQCC and SCLC patients. Whereas ADCC outstands among all, displaying around 10-fold increase in the susceptibility towards the disease. The combined (TC+CC) genotype showed highly statistically significant value (OR=9.96, 95% CI=4.53-21.90, p-value=<0.0001) in individual having ADCC, resulting elevated risk of acquiring lung cancer. The Minor allele frequency for cases and controls are 0.17 and 0.21 respectively.

AhR rs2066853

In case of AhR rs2066853, PCR-RFLP was used to obtain the genotype of cases and controls which were recruited for population study. A total of 297 cases and 320 controls were enrolled in the study. Among the 297 lung cancer patients, (18.51%) of the individual were found to have heterozygous genotype(AG), (77.10%) had homozygous wild genotype and (4.37%) individuals had mutant genotype (AA).While in controls, (70.00%) of individual were having (GG) wild genotype,(20.93%) individual had heterozygous genotype (AG) and (9.06%) had mutant genotype. In the above table, we have combined both heterozygous and mutant genotype (AG+GG) as single genotype, as number of mutants in both cases and controls were less. The genotypic frequencies for AhR rs10250822 polymorphism in the control group (n=320, $\chi 2$=35.6, p-value=0) were not following Hardy-Weinberg equilibrium. In AhR rs2066853 with reference to GG genotype, the homozygous mutant genotype displayed highly significant value (OR=0.347, 94%CI=0.172-0.702, p-value=0.0032) and also shows a highly protective effect towards the risk of acquiring lung cancer. The heterozygous genotype failed to give any significant value. While the combined effect of homozygous mutant and heterozygous also displayed significant value (OR=0.628, 95% CI=0.433-0.911, p-value=0.014) showing protective effect towards the disease. In case of AhR rs10247158 A/T, the genotypic frequencies were in Hardy Weinberg equilibrium. The percentage of individual having the mutant genotype was 3.4% of the total. The Odds ratio calculated for mutant genotype by taking AA as a reference, came out to be 1.26 (0.8-2.6) p=0.3. No significant association was seen in any of the genotypes.

In AhR rs2066853, a total of 297 cases of cancer patients were studied, among them (32.68%) of cases diagnosed with ADCC, (43.3%) people had SQCC, (23.9%) had SCLC. When histology was further classified into different genotypes, (76.28%), (77.51%) and (78.87%) had wild genotype for ADCC, SQCC and SCLC. While (19.58%), (17.82%) and (16.90%) have heterozygous genotype and total number of mutant genotype as (4.12%), (4.65%), (4.22%) respectively. Due to less number of mutants in cases and controls, mutant and heterozygous genotype are combined as single genotype. The total count was (23.71%), (22.48%) and (21.12%) for ADCC, SQCC and SCLC.

In the above table among all histological forms-of lung cancer, individual diagnosed with ADCC does not show any statistically significant value in all genotype showing no association with the risk of developing the disease. For SQCC, individual having mutant genotype (AA) (OR=0.305, 95% CI=0.118-0.783, p-value=0.013) showed a highly protective effect towards the disease. When combined it with the heterozygous genotype (AG+GG), it also displayed significant values and suggesting a protective effect towards the lung cancer. For SCLC, the combined genotype of homozygous mutant and heterozygous (OR=0.540, 95 % CI=0.286-1.020, p-value=0.057) showed a highly significant protective effect. The Minor allele frequency for cases and controls are 0.20 and 0.04 respectively.

6.4: Genotypic frequency distribution of AhR gene polymorphism among patients and control on the basis of Smoking characteristics (Pack Years) and its susceptibility towards Lung cancer

Table 8 : Genotypic frequency distribution of AhR variant among patients and controls on the basis of Smoking status and its susceptibility towards Lung cancer								
AHR rs7811989	**CASES (SMOKERS) N= 233(%)**	**CONTROLS (SMOKERS) N=221(%)**	**OR(95% CI)**	***P***	**CASES (NON-SMOKERS) N=64 (%)**	**CONTROLS (NON-SMOKERS) N=99(%)**	**OR(95% CI)**	***P***
0	87(37.33)	106(47.96)	Ref(1.00)	Ref.	25(39.06)	49(49.49)	Ref(1.00)	Ref.
1	119(51.07)	103(46.60)	1.23(0.82-1.86)	0.30	33(51.56)	42(42.42)	1.68(0.83-3.41)	0.146
2	27(11.58)	12(5.42)	2.91(1.36-6.21)	0.0056	6(9.37)	8(8.08)	1.19(0.34-4.12)	0.772
3	146(62.66)	115(52.03)	1.42(0.96-2.10)	0.07	39(60.93)	50(50.5)	1.59(0.81-3.11)	0.172
AHR rs7811989	**CASES N=104 (%) (Light smokers; PY≤25)**	**CONTROLS N=124(%) (Light smokers; PY≤25)**	**OR(95% CI)**	***P***	**CASES N= 129(%) (Heavy smokers;PY>25)**	**CONTROLS N=97(%) (Heavy smokers;PY>25)**	**OR(95% CI)**	***P***
0	37(35.57)	58(46.67)	Ref(1.00)	Ref.	50(38.75)	48(49.48)	Ref(1.00)	Ref.
1	53(50.96)	60(48.38)	1.33(0.75-2.36)	0.32	66(51.16)	43(44.32)	1.18(0.65-2.12)	0.57
2	14(13.46)	6(4.83)	3.9(1.38-11.37)	0.01	13(10.07)	6(6.18)	1.99(0.67-5.88)	0.21
3	67(64.42)	66(53.22)	1.59(0.92-2.74)	0.09	79(61.24)	49(50.51)	1.31(0.75-2.8)	0.34

AHR ***rs10250822***	**CASES (SMOKERS) N=233 (%)**	**CONTROLS (SMOKERS) N=221(%)**	**OR(95% CI)**	***P***	**CASES (NON-SMOKERS) N= 64(%)**	**CONTROLS (NON-SMOKERS) N=99(%)**	**OR(95% CI)**	***P***
0	102(43.77)	125(56.56)	Ref(1.00)	Ref.	37(57.81)	59(59.59)	Ref(1.00)	Ref.
1	113(48.49)	87(39.36)	1.49(1.00-2.21)	0.04	23(35.93)	33(33.33)	1.07(0.53-2.15)	0.83
2	18(7.72)	9(4.07)	2.26(1.10-6.20)	0.02	4(6.25)	7(7.07)	0.67(0.16-2.67)	0.57
3	131(56.22)	96(43.43)	1.59(1.08-2.33)	0.01	27(42.18)	40(40.4)	1.00(0.51-1.93)	0.99
AHR ***rs10250822***	**CASES N=104 (%) (Light smokers; PY≤25)**	**CONTROLS N=124(%) (Light smokers; PY≤25)**	**OR(95% CI)**	***P***	**CASES N=129 (%) (Heavy smokers;PY>25)**	**CONTROLS N=97(%) (Heavy smokers;PY>25)**	**OR(95% CI)**	***P***
0	44(42.30)	71(57.25)	Ref(1.00)	Ref.	58(44.96)	54(55.67)	Ref(1.00)	Ref.
1	51(49.03)	49(39.51)	1.59(0.91-2.79)	0.102	62(48.06)	38(39.17)	1.54(0.86-2.73)	0.139
2	9(8.65)	4(3.22)	3.72(1.06-13.0)	0.03	9(6.97)	5(5.15)	2.06(0.62-6.81)	0.23
3	60(57.69)	53(42.74)	1.76(1.02-3.03)	0.04	71(55.03)	43(44.32)	1.59(0.91-2.77)	0.09

To study the association of smoking with the risk of acquiring lung cancer, the patients enrolled for study were classified as smokers and non-smokers. A total of 233 lung cancer cases are smokers whereas 64 are non-smokers. In control group, 221 of them are smokers whereas 99 are non-smokers. Depending upon the smoking status, smokers were categorized into Heavy-smokers and Light-smokers. These were categorized on the basis of pack years (a number of cigarette/beedi a person smoke in a day multiplied by number of year smoked).Heavy-smokers are those which are having pack years greater than or equal to 25 (PY≤ 25) whereas light-smokers are having pack years less than 25 (PY>25).In cases,124 patients were heavy smokers while 104 are light-smokers. In control group study, heavy smokers are 97 whereas light-smokers were 124.

AhR *rs7811989*

In case of AhR *rs7811989***,** we found that smokers showed higher risk for lung cancer for those having mutant genotype (0R=2.91, 95%CI=1.36-6.21) and p-value=0.0056) which seems to be highly significant when compared with non-smokers (OR-1.19, 95% CI-0.34-4.12,p-value=0.772) . As inferred from the results, light-smokers showed a higher risk for lung cancer for those having mutant genotype (OR=3.9, 95% CI=1.38-11.37,p-value=0.01) in comparison to heavy smokers (OR=1.99, 95% CI=0.67-5.88, p-value=0.21), which was not significant. Additionally, smokers also display risk for lung cancer in homozygous mutant and heterozygote genotypes when taken in combination (OR=1.42, 95% CI=0.96-2.10,p-value=0.07) which is highly significant in comparison to non-smokers (OR=1.59,95% CI=0.81-3.11,p-value=0.172).In case of light-smokers, all genotype showed more significant values as compared to heavy-smokers. On considering the heterozygote genotype, statistically insignificant values were observed in both smokers (light or heavy smokers) which results in decreased risk for lung cancer.

AhR *rs10250822-*

For AhR *rs10250822***,** the table shows statistically significant value in case of smokers in comparison to non- smokers. With reference to wild genotype, it was observed that in case of mutant phenotype, smokers shows increased risk of lung cancer (OR=2.26, 95% CI=1.10-6.20,p-value=0.02) in comparison with non-smokers (OR=0.67, 95%CI=,0.16-2.67,p-value=0.57). If we combine both homozygous mutant and heterozygote genotype as a single genotype and compared it with wild genotype, a significant risk was observed (OR=1.59, 95% CI=1.08-2.33, p-value=0.01) in smokers whereas in non-smokers(OR=1.00,95%

CI=0.51-1.93,p-value=0.99), no such risk was observed. In AhR rs10250822, all genotype suggested more significant value in smokers as compared to non-smokers. In contrast to smokers, light smokers displayed significant values of risk for lung cancer (OR=3.72, 95% CI=1.06-13.0,p-value=0.03) when compared with heavy-smokers (OR=2.06,95% CI=0.62-6.81,p-value=0.23).Combinatorial effect of homozygous mutant and heterozygote genotype showed higher significant value in both light smokers(OR=1.76,95% CI=1.02-3.03,p-value=0.04)and heavy smokers (OR=1.59, 95% CI=0.91-2.77,p-value=0.09).

AhR *rs2282885*-

For AhR *rs2282885, i*n accordance to the reference wild genotype, the heterozygous genotype indicated a slightly higher significant value in non-smokers (OR=0.59,95%CI=0.27-1.28,p-value=0.18) in comparison with smokers of same genotype(OR=1.22,95% CI=0.79-1.87,p-value=0.35).While in case of mutant genotype, smokers show marginally significant value (OR=2.17,95%CI=0.77-6.13,p-value=0.14) towards lung cancer as compared to non-smokers having p-value=0.81 ,which was not significant. When analyzed in combination both homozygous mutant and heterozygote genotype in smokers (OR= 1.30, 95%CI=0.86-1.96,p-value=0.19),showed increase in the risk of acquiring lung cancer while in case of non-smokers(OR=0.61,95% CI=0.29-1.27,p-value=0.19) protective effect was observed**.** On sub-grouping, light-smokers and heavy-smokers failed to show any significant value, resulting no association with the lung cancer.

AhR *rs2066853*-

In case of AhR rs2066853, Tobacco smoking showed highly significant effect on the polymorphism of AhR rs2066853 towards lung cancer. As the table suggested much higher significant value in smokers was observed as compared to non-smokers. When considering the combination of homozygous mutant and heterozygous genotype, both display significant value in smokers well as non-smokers.2-fold increase in the significant value in smokers were seen. The mutant genotype in smokers (OR=0.23, 95%CI=0.10-8.52, p-value=0.002) display significant values while non-smokers failed to show any significant value. In case of heterozygous genotype, smokers have (OR=0.53, 95% CI=0.32-0.87, p-value=0.004), which is highly significant in comparison to non-smokers (OR=0.53,95% CI=0.32-0.87,p-value=0.18) which leads to increase the risk of lung cancer.

On grouping smokers into heavy and light smokers, heavy smokers (OR=0.53, 95%CI=0.25-1.114, p-value=0.09) display slightly significant protective effect to light-smokers (OR=0.61,

95% CI=0.31-1.20,p-value=0.15) with heterozygous genotype. In mutant genotype both heavy smokers (OR=0.0.22, 95% CI=0.07-0.0681, p-value=0.02) and light smokers (OR=0.25, 95%CI=0.08-0.081,p-value=0.02) shows significant value, resulting displayed protective effect.. In combination of both homozygous mutant and heterozygote genotype, both heavy-smokers and light-smokers showed slightly significant values towards risk of lung cancer. In case of AhR rs10247158 A/T, there appeared no visible stratifying difference between the two subgroups made on the basis of smoking, as the percentage of heterozygotes is very were comparable in both the cases and controls within the smokers. A similar trend was observed in case of all the non-smokers. As there was no significant difference seen within smoker group so, the further stratification into heavy and light smokers also didn't show any significant trend (data not shown).

Table 9: Genotypic Distribution based on different genotypic combinations and their association with lung cancer risk				
Genotype **AHR rs2282885, T/C & AHR rs10250822,T/C**	**Cases, n (%)** **N = 146**	**Controls, n (%)** **N = 167**	**Adjusted OR (95% CI)** [b]	***p*-value**[a]
0	99(67.8)	129(77.24)	1.00 (Reference)	
1	45(30.82)	37(22.15)	1.542(0.916-2.59)	0.102
2	2(1.369)	1(0.59)	-	-
3	47(32.19)	38(22.75)	1.667(1.022-2.719)	0.404
AHR rs2282885, T/C & AHR rs2066853 , A/G	**Cases, n (%)** **N=172**	**Controls, n (%)** **N =181**	**Adjusted OR (95% CI)** [b]	***p*-value**[a]
0	158(91.86)	159(0.87)	1.00 (Reference)	
1	12(6.97)	21(11.60)	0.538(0.251-1.154)	0.116
2	21(1.162)	1(0.55)	-	-
3	14(8.13)	22(12.15)	0.682(0.380-1.224)	0.200
AHR rs2282885, T/C& AHR rs7811989, G/A	**Cases, n (%)** **N = 133**	**Controls, n (%)** **N = 156**	**Adjusted OR (95% CI)** [b]	***p*-value**[a]
0	87(65.41)	112(71.79)	1.00 (Reference)	
1	43(32.33)	43(27.56)	1.261(0.742-2.143)	0.389
2	3(2.25)	1(0.641)	-	-
3	46(34.58)	44(28.20)	1.68(1.050-2.710)	**0.03**
AHR rs10250822, T/C & AHR rs2066853, G/A	**Cases, n (%)** **N = 131**	**Controls, n (%)** **N = 171**	**Adjusted OR (95% CI)** [b]	***p*-value**[a]
0	106(80.91)	142(83.04)	1.00 (Reference)	
1	24(18.32)	29(16.95)	0.988(0.537-1.818)	0.969
2	1(0.763)	0	-	
3	25(19.08)	29(16.95)	0.827(0.494-1.385)	0.472

AHR rs10250822, T/C & AHR rs7811989, A/G	**Cases, n (%)** **N = 138**	**Controls, n (%)** **N = 170**	**Adjusted OR (95% CI)** [b]	***p*-value**[a]
0	57(41.30)	102(60)	1.00 (Reference)	
1	78(56.52)	66(38.82)	1.035(1.0099-1.062)	0.0065
2	3(2.173)	2(1.176)	-	-
3	81(58.69)	68(40)	1.631(1.0079-1.0556)	**0.0009**
AHR rs2066853, G/A & AHR rs7811989,A/G	**Cases, n (%)** **N = 118**	**Controls, n (%)** **N = 148**	**Adjusted OR (95% CI)** [b]	***p*-value**[a]
0	87(73.72)	113(76.35)	1.00 (Reference)	
1	30(25.42)	31(20.94)	1.058(0.582-1.925)	0.851
2	1(0.84)	4(2.702)	-	-
3	31(26.27)	35(23.64)	0.93(0.56-1.539)	0.404
AHR rs10250822,T/G & AHR rs7811989,A/G & **AHR rs2066853, G/A**	**Cases, n (%)** **N =66**	**Controls, n (%)** **N = 107**	**Adjusted OR (95% CI)** [b]	***p*-value**[a]
0	43(65.15)	76(71.02)	1.00 (Reference)	
1	16(24.24)	15(14.01)	1.634(0.715-3.7308)	0.243
2	7(10.60)	16(14.95)	-	-
3	23(34.84)	31(28.97)	1.019(0.989-1.0504)	0.488
AHR rs10250822, T/C& AHR rs2282885 T/C & **AHR rs7811989 A/G**	**Cases, n (%)** **N = 82**	**Controls, n (%)** **N = 101**	**Adjusted OR (95% CI)** [b]	***p*-value**[a]
0	43(52.43)	76(75.24)	1.00 (Reference)	
1	16(19.51)	15(14.84)	2.046(0.991-4.226)	**0.052**

2	23(28.04)	10(9.90)	-	-
3	39(47.56)	25(24.75)	**2.6112(1.359-5.017)**	**0.0042**
AHR rs 2282885, T/C & AHR rs 7811989,A/G & AHR rs 2066853,G/A	**Cases, n (%) N =81**	**Controls, n (%) N = 96**	**Adjusted OR (95% CI) [b]**	***p*-value[a]**
0	63(77.77)	77(80.20)	1.00 (Reference)	
1	8(9.87)	11(11.45)	0.736(0.267-2.027)	0.553
2	10(12.34)	8(8.33)	-	-
3	18(22.22)	19(19.79)	1.079(0.512-2.275)	0.839
AHR rs2066853, G/A, & AHR rs2282885, T/C,& AHR rs10250822 , G/A	**Cases, n (%) N = 88**	**Controls, n (%) N = 110**	**Adjusted OR (95% CI) [b]**	***p*-value[a]**
0	**77(87.5)**	**94(85.45)**	1.00 (Reference)	
1	5(5.68)	7(6.36)	0.7916(0.236-2.652)	0.704
2	6(6.81)	9(8.18)	-	-
3	11(12.5)	16(14.54)	0.824(0.355-1.9115)	0.652
AHR rs7811989,A/G & AHR rs10250822,T/C & AHR rs2282885,T/C & AHR rs2066853, G/A &	**Cases, n (%) N = 37**	**Controls, n (%) N =58**	**Adjusted OR (95% CI) [b]**	***p*-value[a]**
0	33(89.18)	53(91.37)	1.00 (Reference)	
1	4(10.81)	5(8.62)	1.189(0.2904-4.8693)	0.809
2	0	0	-	-
3	4(10.81)	5(8.62)	-	-

[b]Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status. [a]Two-sided χ^2 test for either genotype distribution or allelic frequencies between the cases and controls. 0: wild genotype, 1: heterozygote genotype, 2: mutant genotype, 3: combined hetero and mutant genotype

6.5: Combinatorial risk assessment of 4 SNP's-

We further assessed the combined effect of 4 SNP's in combination of two, three and four as shown in the table above. Combinatorial analysis of the SNP's was done to estimate their effect whether they are protective or hazardous towards the lung cancer.

AhR rs2282885, T/C & AHR rs10250822, T/C-

A total of 146 cases and 167 controls studied in combination for estimating its effect towards lung cancer. In heterozygous genotype, 45(30.82%) were cases and 37(22.15%) were controls, whereas 47(32.19%) cases and 38(22.75%) controls in double variant genotype of heterozygous and homozygous mutant. The cumulative effect in mutants is not studied due to its low count in both cases and control.

From the above table the combination of these 2 SNP's when compared with the reference, heterozygous genotype (OR=1.542, 95% CI=0.916-2.59, p-value=0.102) and combined genotype of homozygous mutant and heterozygous (OR=1.667, 95% CI=1.022-2.719, p-value=0.404) give statistically insignificant value. Hence, no association between two gene was observed towards risk of lung cancer.

AhR rs2282885, T/C & AHR rs2066853, A/G-

From the above table, a total of 172 cases and 181 controls were studied in combination. In heterozygous genotype, 12(6.97%) were cases and 21(11.60%) were controls. While 14(8.13%) cases and 22(12.15%) controls in combined genotype of heterozygous and homozygous mutant. The cumulative effect in mutants was not studied due to less number of counts in both cases and control.

In reference with the wild genotype, no insignificant association was found between these 2 SNP's towards the advent of acquiring lung cancer. Both heterozygous genotype and homozygous genotype failed to give any significant value and hence their combination was not considered as a risk towards the disease.

AHR rs2282885, T/C& AHR rs7811989, G/A-

A total of 133 cases and 156 controls studied in combination for estimating its effect towards lung cancer. In heterozygous genotype, 43 (32.33%) was cases and 43(27.56%) was controls. While 46(34.58%) cases and 44(22.80%) controls in combined genotype of heterozygous and

homozygous mutant. The cumulative effect in mutants was not studied due to less number of counts in both cases and control.

The table showed a higher significance between two in AHR rs2282885 & AHR rs7811989. In reference with the wild genotype, the combined genotype shows (OR=1.68, 95% CI=1.050-2.710,p-value=0.03) a slight significant risk towards lung cancer as was observed.

AHR rs10250822, T/C & AHR rs2066853, G/A-

A total of 131 cases and 171 controls studied in combination for estimating its effect towards lung cancer. In heterozygous genotype, 24(18.32%) was cases and 29(16.95%) was controls. While 25(19.08%) cases and 29(16.95%) controls in combined genotype of heterozygous and homozygous mutant. The cumulative effect in mutants was not studied due to less number of counts in both cases and control.

Statistical analysis revealed statistically insignificant values in both homozygous mutant and heterozygous genotype towards the risk of acquiring lung cancer. Hence, no association of these two SNP's was observed towards the risk for lung cancer.

AHR rs10250822, T/C & AHR rs7811989, A/G –

A total of 138 cases and 170 controls studied in combination for estimating its effect towards lung cancer. In heterozygous genotype, 78(56.52%) were cases and 66(38.82%) were controls. While 81(58.69%) cases and 68(40.0%) controls in combined genotype of heterozygous and homozygous mutant. The cumulative effect in mutants was not studied due to less number of counts in both cases and control.

From the above table, there was a significant association observed between polymorphism in **AHR rs10250822** and **AHR rs7811989.** Heterozygous genotype failed to give any significant value when compared with reference genotype. But the individual with combined genotype involving both homozygous mutant and heterozygous genotype displayed a highly statistically significant value (OR=1.6361, 95% CI=1.0079-1.0556,p-value=0.0009) .Hence ,a significant association was found between **AHR rs10250822** and **AHR rs7811989** towards the risk of acquiring lung cancer.

AHR rs2066853, G/A & AHR rs7811989, A/G –

From the above table, a total of 118 cases and 148 controls were studied in combination. In heterozygous genotype, 30(25.42%) were cases and 31(20.94%) were controls. While

31(26.27%) cases and 35(23.64%) controls in combined genotype of heterozygous and homozygous mutant. The cumulative effect in mutants was not studied due to less number of counts in both cases and control.

Statistical analysis showed no association was observed between *AHR rs2066853* and *AHR rs7811989* towards the risk of developing the lung cancer as shown by the values, which is highly insignificant.

In the current study, we have also evaluated the triple combination for AhR polymorphism towards the lung cancer susceptibility.

AhR rs10250822, AhR rs7811989, AhR rs2066853 –

A total of 66 cases and 107 controls studied in combination for estimating its effect towards lung cancer. In heterozygous genotype, 16(24.24%) were cases were 15(14.01%) are controls. While 23(34.84%) cases and 31(28.97%) controls in combined genotype of heterozygous and homozygous mutant. The cumulative effect in mutants was not studied due to less number of counts in both cases and control.

Statistical analysis showed no association was observed between AhR rs10250822, AhR rs7811989 and AhR rs2066853.The combined genotype and heterozygous genotype failed to show any significant value towards lung cancer susceptibility.

AhR rs10250822, AhR rs2282885, AhR rs7811989 -

From the above table, a total of 82 cases and 101 controls were studied in triple combination. In heterozygous genotype, 16(19.51%) was cases and 15(14.84%) was controls. While 39(47.56%) cases and 25(24.75%) controls in combined genotype of heterozygous and homozygous mutant. The cumulative effect in mutants was not studied due to less number of counts in both cases and control.

From the above table, statistical analysis shows the presence of rs10250822 (T/C), rs2282885 (T/C) and rs7811989 (A/G) polymorphism display 2-fold increase in the significant value .In heterozygous genotype, triplet combination of these SNP's shows a significant statistical value (OR=2.046,95% CI=0.991-4.226,p-value=**0.052**) towards lung cancer susceptibility. The combinatorial effect of triplet SNP's in combined genotype involving homozygous mutant and heterozygous showed a exorbitant increase in the risk

towards lung cancer (OR=2.6112 95% CI=1.359-5.017,p-value=**0.0042**) as reflected by the values which is highly significant.

AhR rs2282855, AhR rs7811989, AhR rs2066853 –

From the above table, a total of 81 cases and 96 controls were studied in combination. In heterozygous genotype, 8(9.87%) were cases and 11(11.45%) were controls. While 18(22.22%) cases and 19(12.15%) controls in combined genotype of heterozygous and homozygous mutant. The cumulative effect in mutants is not studied due to less number of counts in both cases and control.

Statistical analysis showed no association was observed between *AHR rs2066853 and AHR rs7811989 and AhR rs2282855* towards the risk of developing lung cancer as reflected by the values, which was highly insignificant.

AhR rs2066853, AhR rs2282885, AhR rs10250822-

A total of 88 cases and 110 controls studied in combination for estimating its effect towards lung cancer. In heterozygous genotype,(30.82%) was cases and 37(22.15%) was controls. While 47(32.19%) cases and 38(22.75%) controls in combined genotype of heterozygous and homozygous mutant. The cumulative effect in mutants is not studied due to less number of counts in both cases and control.

Statistical analysis revealed statistically insignificant values in both homozygous mutant and heterozygous genotype towards the risk of acquiring lung cancer. Hence, no association of these 3 SNP's were observed towards the risk for lung cancer.

AhR rs7811989, rs10250822, rs2282885, rs2066853 –

A total of 37 cases and 58 controls studied in combination for estimating its effect towards lung cancer. In heterozygous genotype, 4(10.81%) were cases and 5(8.62%) were controls. While 4(10.81%) cases and 5(8.62%) controls in combined genotype of heterozygous and homozygous mutant. The cumulative effect in mutants was not studied due to less number of counts in both cases and control.

In reference with the wild genotype, an insignificant association was found between these 4 SNP's towards the advent of acquiring lung cancer. Both heterozygous genotypes failed to give any significant value and hence their combination is not considered as risk towards lung cancer.

6.6: Association with Haplotypes and Linkage disequilibrium in AhR variants

Table10: Analyzed Haplotype frequencies and risk towards lung cancer due to AHR Haplotypes

HAPLOTYPE	HAPLOTYPE CONSTRUCTION (rs2282885-10250822-2066853-7811989)	Case frequency, n (%)	Control frequency, n (%)	OR (95% CI) [b]	*p*-value[a]
Hap1	CTGA	31.75(0.053)	12.83(0.020)	2.720(1.406-5.262)	0.002042
Hap2	CTGG	31.39(0.053)	4.17(0.006)	8.393(3.001-23.472)	1.52e006
Hap3	TCGA	81.91(0.137)	57.22(0.089)	1.608(1.121-2.307)	0.009499
Hap4	TCGG	48.66(0.082)	41.74(0.065)	1.256(0.816-1.934)	0.299786
Hap5	TTAA	33.52(0.056)	49.41(0.077)	0.698(0.443-1.101)	0.120506
Hap6	TTAG	9.20(0.015)	25.05(0.039)	0.377(0.176-0.811)	0.009659
Hap7	TTGA	192.95(0.324)	251.06(0.391)	0.705(0.552-0.899)	0.004744
Hap8	TTGG	98.80(0.166)	76.82(0.120)	1.442(1.042-1.997)	0.026806

[b]Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by logistic regression analysis. [a]Two-sided χ^2 test for either haplotypic distribution or frequencies between the cases and controls.

Haplotype frequencies and linkage disequilibrium were ascertained for the four SNP's using SHEsis software. Haplotype frequencies were classified in table shown above. Out of every one of the all haplotype, those with cases and controls frequency less than0.03 were removed from the analysis. Correlation of general haplotype profile uncovered a critical contrast between the cases and controls. As shown in table, Hap 2(CTGG) was linked with the most noteworthy risk towards lung cancer (OR=8.393, 95% CI=3.001-23.472,p-value=1.52e006). Haplotype raising risk for lung cancer was exceedingly represented in cases in comparison with control and conveyed atleast one-heterozygous allele e.g. hap1 (OR=2.72,95% CI=1.40-5.26,p-value=0.00204) and hap 3 (OR=1.608,95% CI=1.121-2.307,p-value=0.0094).The haplotype block containing heterozygous allele for rs7811989 and wild for rest SNP showed a highly protective effect towards lung cancer as reflected by the values in hap 6 (OR=0.377,95% CI=0.176-0.811,p-value=0.00965) and hap 7 (OR=0.705,95% CI=0.552-0.899,p-value=0.0047).

Table11: D/r^2 Values for linkage disequilibrium among four AhR SNPs.				
CODON	**rs2282885**	**rs10250822**	**rs2066853**	**rs7811989**
rs2282885	-	0.049/0.001	0.041/0.002	0.168/0.009
rs10250822	-	-	0.036/0.001	0.127/0.011
rs2066853	-	-	-	0.62/0.001
rs7811989	-	-	-	-

The above table summarizes the result of *D'* values and r^2 values for linkage disequilibrium (LD) between for cases and controls together.rs10250822 and rs2282885 illustrated a linkage disequilibrium *(D'=0049, r^2=0.001).* Similarly, rs10250822 gave a linkage disequilibrium with rs2066853 *(D'=0.041, r^2=0.002*).Additionally, rs7811989 and rs2066853 also illustrated a linkage disequilibrium *(D'=0.62, r^2=0.001*) However, rs10250822 gave a weak LD with rs7811989 *(D'=0.127, r^2=0.011).*

6.7: Multifactor Dimensionality Reduction (MDR) analysis:

Table12: Multifactor Dimensionality reduction (MDR) analysis showing association of high order interactions with Lung cancer.

No. of risk factors	Best interaction model	Cross Validation Consistency (CVC)	Prediction error
1	AHR rs2282885	5/10	0.478
2	AHR rs10250822, AHR rs2066853	5/10	0.477
3	AHR rs10250822, AHR rs2066853, AHR rs7811989	**10/10**	**0.461**
4	AHR rs7811989,AHR rs10250822, AHR rs2066853, AHR rs2282885	10/10	0.472

The MDR analysis provides cross validation consistency (CVC) and average prediction error. Our analysis identifies the elite interaction among different genotype. The best interaction model are those having maximum CVC and minimum prediction error. In the above table, the best interaction model is having three AhR variant **(AHR rs10250822, AHR rs2066853, AHR rs7811989)** because it is having maximum 10 on 10 CVC and minimum prediction rate (0.461) among all. The second best interaction model is 4 AhR variant model (AHR rs10250822, AHR rs2066853, AHR rs7811989) having 10 on 10 CVC and minimum prediction error (0.472) greater than 3 AhR variant model. This model is also acted as best-one for providing lung cancer risk.

6.8: Demographic characteristics

Table 13: Patients characteristics and clinical features	
Patient characteristics	**No. of patients, n (%) N=150**
Age (years)	
Mean	58.98±10.56 (P<0.0001)
Standard deviation	
Gender	
Male	133(88.66)
Female	17(11.33)
Smoking Status	
Smokers	125(83.33)
Non–Smokers	25(16.66)
Histologic al subtypes	
ADCC	54(36)
SQCC	61(40.66)
SCLC	35(23.33)
TNM Staging	
III	84(56)
IV	64(42.66)
I ,II	2 (1.33)
ECOG performance status	
0-1	120(80)
2	29(19.33)
3-4	1(0.66)
KPS performance status	
100-80	96(64)
70-60	42(28)
50-40	12(8)
Median follow up (months)	7.283
Events ,Deaths	117

Demographic characteristics of the study population including age, smoking status, histology subtypes, TNM staging, ECOG and KPS status as shown in table above (Figure-). The present study enlisted 150 lung cancer patients with a median follow up time of 7.2 months. The present study was a hospital based case study which assessed the potential role of AhR polymorphism and its related risks for acquiring lung cancer. The mean age of lung cancer patients was 58.98±10.56 years having p-value is highly significant (P<0.0001).The study group contains 133(88.6%) males and 17(11.33%) females. Significant difference was seen in the distribution of male and females. Smoking is another factor which was further categorized into smokers and non-smokers based on smoking status of patients. Likewise, histology is further classified according to its subtypes. There are 54(36%) ADCC, 61(40.66) SQCC, 35(23.3%) patients.TNM staging was also evaluated for the cancer patients involving 84 (56%) patients having stage III, 64(42.6%) having stage IV and 2(1.33%) having stage I, II. As inferred from the table above, the frequency of occurrence of stage I and II was very low in comparison to stage III, IV. In the present study, there were 117 deaths were reported. The patients were further classified on the KPS (Karnofsky's performance scores) and ECOG (Eastern Karnofsky's performance scores performance status) performance status.

As inferred from the results, in our study there were 120 (80%) patients with ECOG PS 0-1, 29 (19.3%) were ECOG PS 2 and 1 (0.66%) were ECOG PS 3. Based upon In our study there were 127 (52.6%) patients with ECOG PS 0-1, 73 (30.2%) were ECOG PS 2 and 41 (17.0%) were ECOG PS 3.Furthermore, there were 96(54%) patients with KPS score between 100-80, 42 (28%) with KPS score of 70-60 and 12 (8%) with KPS score of 50-40.

6.9: Association of AhR polymorphism and Overall Survival (Table 14)

GENES	CASES n(%) N=150	DEATH n(%) N=118	ALIVE n(%) N=32	Univariate analysis			Multivariate analysis	
				MST (months)	Log rank p-Value	Unadjusted HR[a]	Adjusted HR[b] (95% CI)	p-Value
AhR rs2282885								
TT	98(65.3)	78(66.10)	20(62.5)	7.23	0.98	1	1	
TC	42(28.0)	32(27.11)	10(31.2)	7.33		0.99	1.22(0.79-1.88)	0.36
CC	10(6.66)	8(6.77)	2 (6.25)	6.40		0.93	1.05 (0.72-1.54)	0.78
AhR rs10250822								
TT	64(42.6)	54(45.76)	10(31.2)	7.56	0.76	1	1	
TC	73(48.66)	54(45.76)	19(59.3)	5.70		0.99	1.14(0.76-1.70)	0.50
CC	13(8.66)	10(8.47)	3(9.37)	10.13		1.27	0.83(0.57-1.21)	0.34
AhR rs2066853								
GG	112(74.66)	90(76.27)	22(68.7)	7.30	0.013	1	1	
GA	31(20.66)	21(17.79)	10(31.2)	10.3		1.33	0.70(0.43-1.16)	0.17
AA	7(4.66)	7(16.85)	0(0)	3.53		0.29	1.68(1.09-2.59)	0.017

AhR rs7811989								
AA	57(38.0)	45(38.13)	12(37.5)	7.56		1	1	
AG	78(52.0)	61(51.69)	17(53.1)	6.73	0.92	0.95	0.93(0.62-1.39)	0.72
GG	15(10.0)	12(10.16)	3(9.37)	8.03		1.12	0.76(0.52-1.12)	0.17
[a] Unadjusted Hazards ratio for Kaplan meier analysis, [b] hazards ratio adjusted for age, sex, smoking, histology, stage, KPS, ECOG								

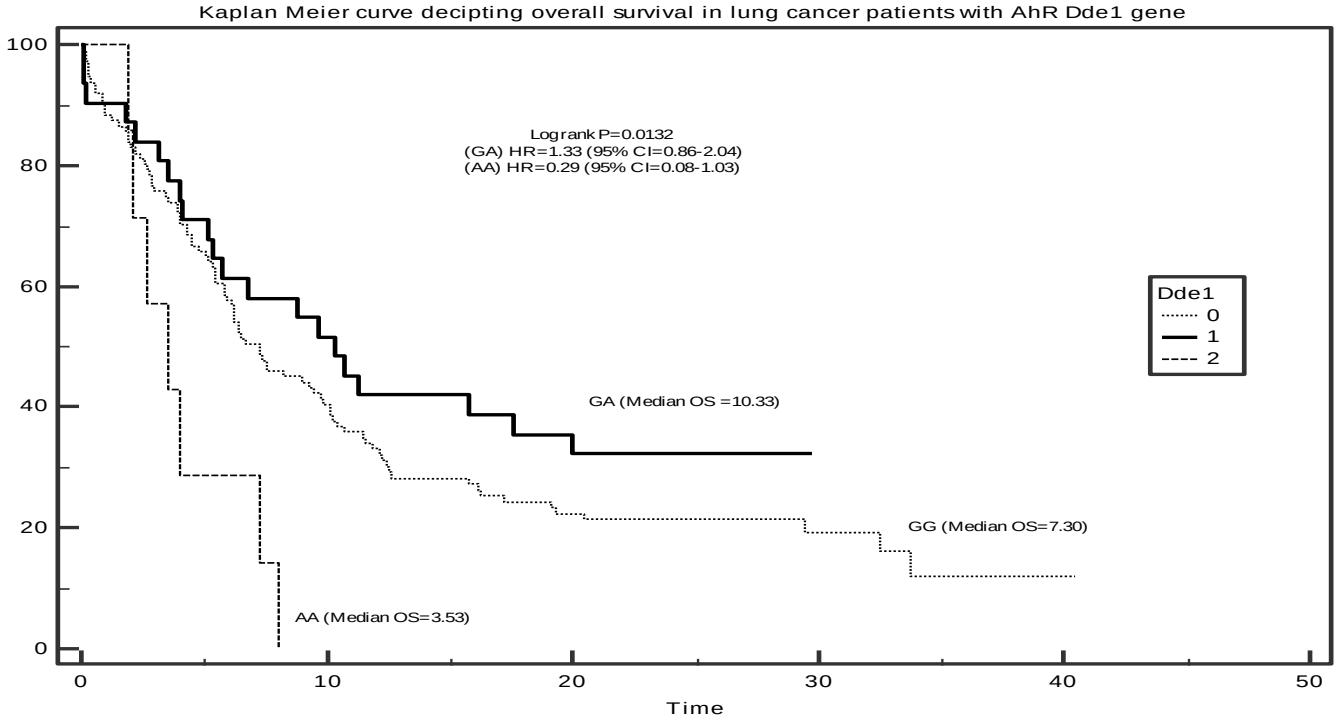

Figure 5.5: Kaplan Meier graph showing statistically significant value in AhR Dde1 gene

The above table demonstrates the 4 polymorphic variants of AhR i.e AhR rs2282885, AhR rs10250822, AhR rs2066853, AhR rs7811989 and its association with overall survival. The survival analysis was carried out after calculating the data obtained after follow-up trial of patients. Their survival time was estimated by accounting the number of days from data of their diagnosis till death. The survival curves and univariate analysis was obtained using Kaplan-Meier survival analysis, whereas multivariate analysis and hazard ratio estimation are obtained using Cox- regression proportional hazards. Our study uncovered that there were an aggregate of 150 lung cancer patients. Out of which 118 patients were dead amid follow up period and 32 were alive. If the present study, an occurrence of AhR rs2066853showed an increased in the MST (median survival time).Patients with wild genotype (GG) had a MST of 7.3 months, those with mutant genotype (AA) had a MST of 3.53 months and those with heterozygous genotype had the most exorbitant increase in the MST of 10.3 months (Log rank p=0.01) ,which is significant as shown in above table. But when Cox regression analysis were performed using different predictor variables such as histology, age, gender ,smoking status, ECOG, KPS .It was found that increased in the death rate in individual having (AA) mutant genotype was observed (OR=1.68, 95% CI=1.09-2.57, p-value=0.017) using (GG) as reference.

Furthermore in AhR rs102508222 uncovered that an increased median survival time of10.12 months in mutant genotype (CC) in comparison to wild genotype (TT) having MST of 7.56 months and heterozygous genotype (TC) having MST of 55.7 months. But no significant difference in the MST was observed in univariate analysis. Similarly, after adjusting with different predictor variables in cox regression analysis, it does not demonstrate any significant association.

Further investigation in AhR rs2282885, no significant difference in the MST was observed in all genotypes. While, wild and heterozygous genotype showed almost similar median survival time (MST=7.23 and 7.33) having log rank p=0.98, which is not significant resulting no association was observed with overall survival. When further multivariate analysis was performed, no significant association was observed in death rate. In AhR rs7811989, the mutant genotype (GG) showed increased in the median survival time of 8.03 months in comparison to the wild and heterozygous genotype (MST=7.56 and 6.53) having log rank p=0.92, which is not significant. Further multivariate analysis were performed using Cox-proportional hazard model resulting death rate does not show any significant association in allgenotype.

6.10:Association of AhR SNP on overall survival according to tumor histology

Table 15 Association of AhR SNPs on Overall Survival according to Tumor histology																		
GENES	**ADCC**						**SQCC**						**SCLC**					
AhR rs2282885 T/C (N)	**Cases n(%) N =54**	**Univariate analysis**			**Multivariate analysis**		**Cases n (%) N=61**	**Univariate analysis**			**Multivariate analysis**		**Cases n (%) N=35**		**Univariate analysis**		**Multivariate analysis**	
		MST	**p-value**	**HR[a]**	**H.R[b]**	**p-value**		**MST**	**P-value**	**HR[a]**	**H.R[b]**	**p-value**		**MST**	**p-value**	**HR[a]**	**H.R[b]**	**p-value**
TT	35 (64.8)	6.10		1	1		39 (63.9)	10.2		1	1		24 (68.5)	7.23		1	1	
TC	13 (24.07)	19.2	**0.06**	2.01	0.692 (0.29-1.63)	0.40	19 (3.11)	6.46	0.38	0.76	1.21 (0.61-2.38)	0.57	10 (28.5)	3.53	0.19	0.60	1.72 (062-4.74)	0.29
CC	6 (11.1)	7.56	0.38	1.57	0.693 (0.23-2.08)	0.51	3 (4.9)	6.40	0.18	0.45	1.67 (0.88-3.15)	0.11	1 (2.85)	5.83	0.55	0.55	3.15 (0.34-29.1)	0.31
TC+CC	19	9.00	**0.05**	1.87	0.63	0.21	22	6.4	0.25	0.71	1.33	0.37	11	5.80	0.16	0.59	1.94	0.17

	(35.1)				(0.31-1.29)		(36.0)				(0.70-2.51)		(31.4)				(0.74-5.09)	
	ADCC						**SQCC**						**SCLC**					
AhR rs10250822 T/C (N)	**Cases n (%) N=54**	**Univariate analysis**			**Multivariate analysis**		**Cases n (%) N=61**	**Univariate analysis**			**Multivariate analysis**		**Cases n (%) N=35**	**Univariate analysis**			**Multivariate analysis**	
		MST	**p-value**	**HR[a]**	**H.R[b]**	**p-value**		**MST**	**p-value**	**HR[a]**	**H.R[b]**	**p-value**		**MST**	**p-value**	**HR[a]**	**H.R[b]**	**P-value**
TT	23 (42.5)	8.03		1	1		25 (40.9)	8.20		1	1		16 (45.7)	5.83		1	1	
TC	27 (50.0)	5.06	0.24	0.69	1.69 (0.81-3.51)	0.15	32 (52.4)	7.56	0.24	0.69	0.85 (0.45-1.62)	0.63	14 (40)	3.53	0.24	0.69	0.46 (0.12-1.71)	0.25
CC	4 (7.40)	8.80	0.65	0.78	1.47 (0.79-2.74)	0.21	4 (6.55)	5.16	0.65	0.78	-	-	5 (14.2)	12.6	0.65	0.78	0.56 (0.26-1.23)	0.15
TC+CC	31 (57.4)	5.43	0.26	0.71	1.67 (0.84-3.32)	0.14	36 (59.0)	7.56	0.26	0.71	0.88 (0.47-1.65)	0.70	19 (54.2)	6.73	0.26	0.71	0.35 (0.13-0.95)	**0.04**
	ADCC						**SQCC**						**SCLC**					
AhR rs2066853	**Cases n (%)**	**Univariate analysis**			**Multivariate analysis**		**Cases n (%)**	**Univariate analysis**			**Multivariate analysis**		**Cases n (%)**	**Univariate analysis**			**Multivariate analysis**	

G/A (N)	**N=54**						**N=61**						**N=35**					
		MST	**p-value**	**HR[a]**	**H.R[b]**	**p-value**		**MST**	**p-value**	**HR[a]**	**H.R[b]**	**p-value**		**MST**	**p-value**	**HR[a]**	**H.R[b]**	**p-value**
GG	43 (79.6)	6.10		1	1		44 (72.1)	10.1		1	1		25 (71.4)	7.23		1	1	
GA	10 (18.5)	10.3	0.19	1.69	0.38 (0.15-0.93)	**0.03**	12 (19.6)	9.66	0.48	1.33	0.80 (0.34-1.87)	0.61	9 (25.7)	4.00	0.88	1.06	0.95 (0.32-2.75)	0.92
AA	1 (1.8)	8.03	0.94	0.93	1.03 (0.27-3.95)	0.95	5 (8.1)	2.70	0.001	0.24	2.08 (1.20-3.61)	**0.008**	1 (2.8)	3.53	0.43	0.46	0.54 (0.14-2.06)	0.37
GA+AA	11 (20.3)	10.3	0.21	1.60	0.40 (0.17-0.93)	**0.03**	17 (27.8)	5.36	0.70	0.87	1.21 (0.60-2.43)	0.57	10 (28.5)	3.53	0.98	0.98	0.87 (0.31-2.47)	0.80
	ADCC						**SQCC**						**SCLC**					
AhR rs7811989 A/G (N)	**Cases n (%) N=54**	**Univariate analysis**			**Multivariate analysis**		**Cases n (%) N=61**	**Univariate analysis**			**Multivariate analysis**		**Cases n (%) N=35**	**Univariate analysis**			**Multivariate analysis**	
		MST	**p-value**	**HR[a]**	**H.R[b]**	**p-value**		**MST**	**p-value**	**HR[a]**	**H.R[b]**	**p-value**		**MST**	**p-value**	**HR[a]**	**H.R[b]**	**p-value**
AA	22 (40.7)	6.43		1	1		26 (42.6)	10.13		1	1		9 (25.7)	3.20		1	1	

AG	28 (51.8)	5.93	0.65	1.15	0.51 (0.25-1.07)	0.07	33 (54.0)	7.33	0.10	0.61	1.53 (0.79-2.98)	0.20	17 (48.5)	6.73	0.27	1.59	0.35 (0.11-1.07)	**0.06**
GG	4 (7.4)	6.76	0.66	0.79	0.80 (0.32-2.01)	0.64	2 (3.2)	6.40	0.46	2.00	0.51 (0.12-2.14)	0.36	9 (25.7)	10.1	0.12	2.09	-	-
AG+GG	32 (59.2)	6.17	0.74	1.10	0.56 (0.27-1.13)	0.10	35 (57.3)	7.33	0.19	0.67	1.34 (0.70-2.57)	0.37	26 (74.2)	7.30	0.14	1.77	0.32 (0.12-0.80)	**0.01**

[a] Unadjusted Hazards ratio for Kaplan meier analysis, [b] hazards ratio adjusted for age, sex, smoking, stage, KPS, ECOG

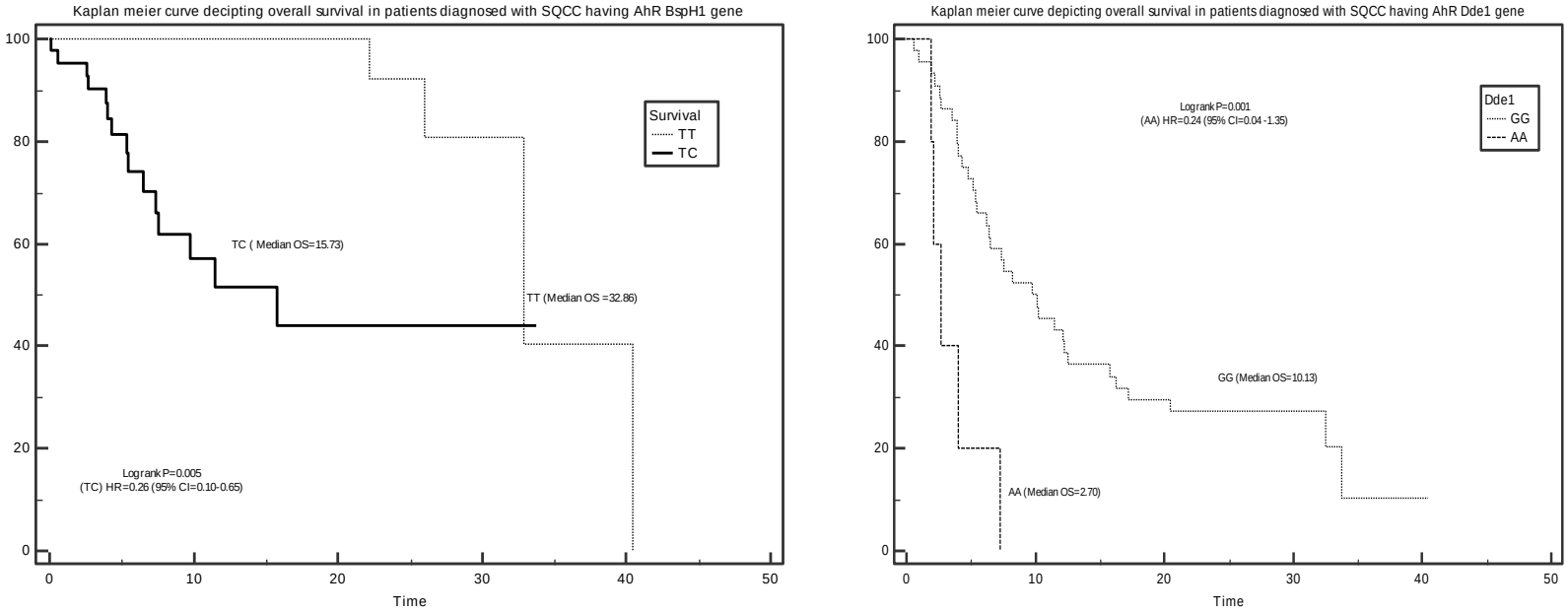

Figure 5.3- Kaplan Meier graphs showing statistically significant Log rank p-value and Hazardous ratio

Histological stratification of the patients into their different sub-types was done to find whether this polymorphism plays any prognostic role in subgroups individually.

AhR *rs2282885*

The patients recruited for the study were stratified according to histology and then univariate analysis was performed using Kaplan-Meier survival analysis. It was found that ADCC patients showed 3-fold increase in the median survival time of 19.2 months in (TC) heterozygous genotype as compared to the (TT) wild genotype (Log rank p=0.06,HR=2.01) showing an marginally significant association for overall survival. Similarly, the combinatorial effect of homozygous mutant and heterozygous(TC+CC) also showed increase in the median survival time of 9.0 months (HR=1.87,p=0.05) in comparison to wild genotype (MST=6.10 months),resulting showing significant association with overall survival .Then genotype are adjusted with different predictor variable factors, after that multivariate analysis were performed using Cox-regression analysis. But it was found that the patients having heterozygous genotype, decrease in the hazardous ratio (HR=0.69, p-value=0.40) which is not significant and does not show any association with survival rate. Similarly, the combined genotype (TC+CC) also showed no association with overall survival (HR=0.63, p-value=0.21). In SQCC patients, decrease in the median survival rate was observed in all genotype in comparison to wild genotype when taken as reference. Statistical analysis revealed no association with overall survival after adjusting with Cox regression analysis. In SCLC patients, statistical analysis showed higher median survival rate in wild genotype (TT) of 7.23 months where as carriers of single allelic variants (*TC*) and double allelic variants (*CC*) had an MST of 3.53 and 5.83 months respectively. When survival rate was calculated, it does not show any significant association with any genotype.

AhR *rs10250822-*

Amongst all the histological forms of lung cancer, in this study we saw that patients belonging to ADCC subgroup showed higher median survival time of 8.80 months in homozygous mutant genotype (CC) showing an decreased hazardous ratio(HR=1.47,p-value=0.68) in comparison to other genotypes. After adjusting with Cox regression analysis survival rate did not show any significant association with any genotype however, mutant genotype (CC) revealed a H.R of 1.47 (H.R=1.47, p=0.21).In SQCC patients, wild genotype have better survival time of 8.20 months while heterozygous(TC) and combined genotype (TC+CC) show similar median survival time of 7.56 months . Then applying the Cox

proportional hazards model after adjusting for age, sex, smoking, stage, ECOG, KPS and histology, our data showed that heterozygous genotype decreased the death rate (H.R=0.85,95% C.I=0.45-1.62, p=0.63) using wild type genotype (CC) as reference. SCLC outstands the rest by displaying higher median survival rate of 12.6 months in mutant genotype(CC) having log rank p=0.78 and p-value=0.65.accordingly applying Cox-regression analysis revealed no significant association with death rate. But in combined genotype (TC+CC), 2-fold decrease in the median survival time (MST=6.73,HR=0.71,p-value=0.26).After applying Cox proportional hazards model, statistical analysis showed decrease in the death rate (HR=0.35,p-value=0.04) resulting the combined genotype shows a significant protective effect.

AhR *rs2066853-*

Statistical analysis showed patients having ADCC were found to showcase higher median survival rate of 10.1 months in wild genotype (GG) and heterozygous(GA) genotype having median survival time of 9.6 months which was higher in comparison to mutant(GG) and combined genotype (GA+GG).However in both univariate and multivariate analysis, no significant association was observed with overall survival. On the contrary patients with a mutant genotype falling in the subgroup of SQCC having least median survival time of 2.70 months in comparison to other genotype when wild is taken as reference (MST=10.1). Univariate analysis using Kaplan-Meier method, showcase decrease in the hazardous ratio in (AA) mutant genotype (HR=0.24, Log rank p-value=0.001) resulting highly significant protective effect for mutant genotype in SQCC patients. However, multivariate Cox proportional hazards regression analysis was performed, statistical analysis revealed increase in death rate which shows highly significant association of mutant genotype(HR=2.08,p-value=0.008) in individual diagnosed with SQCC. In SCLC patients, wild genotype showed increase in median survival time of 7.23 months in comparison to other genotype. However, no such association was observed for SCLC patients in univariate and multivariate analysis.

AhR *rs7811989*

The patients recruited for survival analysis are further stratified according to histology subtypes. Among all the histological forms of lung cancer, the ADCC patients having wild genotype(AA) showed higher median survival time of 6.43 months and lowest hazardous ratio among all genotype (HR=0.79,p-value=0.66).No such association of overall survival was observed in both univariate and multivariate analysis. In SQCC patients, statistical

analysis revealed wild genotype having higher median survival time of 10.13 months. While similar median survival time of 7.3 months was observed in heterozygous (AG) and combined genotype of heterozygous and homozygous mutant (AG+GG).The mutant genotype shows 2-fold decrease in the hazardous ratio (HR=2.0,p-value=0.46) but in contrast decrease in the hazardous ratio was observed in heterozygous genotype(HR=0.06,p-value=0.10) and combined genotype (HR=0.67,p-value=0.19). After adjusting with Cox regression analysis, no significant association was observed. In accordance with wild genotype,2-fold increase in the median survival time was observed in heterozygous genotype (MST=6.73 months). Univariate analysis was performed using Kaplan-Meier curves, showing increased hazardous ratio (HR=1.58,p-value=0.27). Then applying the Cox proportional hazards model after adjusting the predictor variables , decrease in the hazardous ratio was observed (HR=0.35,p-value=0.06).As inferred from the results ,the heterozygous genotype showcase marginal protective effect and slightly associated with death rate. However, more than 2-fold increase in median survival time was observed in combined genotype (MST=7.3 months). Univariate analysis was performed. In accordance with wild genotype (AA), statistical analysis showed increase in the hazardous ratio (HR=1.77,p-value=0.14).After adjusting with Cox regression analysis, SCLC patients showcase dip in the hazardous ratio of combined genotype (HR=0.32,p-value=0.01).Highly significant protective effect was observed in SCLC patients having combined genotype(AG+GG).

Table 16: Genotypic combinations and overall survival						
		Univariate analysis			Multivariate analysis	
Genotype **AhR rs2282885, T/C & AhR rs10250822,T/C**	**Cases, n (%)** **N = 76**	**MST**	**p-value**	**H.R[a]**	**H.R.[b]**	**p-value**
0	44(57.8)	8.20		1	1	
1	25(32.8)	8.30	0.35	1.31	0.91(0.46-1.77)	0.78
2	32(42.1)	5.80	0.89	1.03	1.28(0.70-2.33)	0.42
AhR rs2282885,T/C & AhR rs2066853,G/A	**Cases, n (%)** **N =81**					
0	71(87.6)	6.73		1	1	
1	5(6.17)	11.2	0.94	1.03	1.61(0.45-5.75)	0.45
2	11(13.5)	4.03	0.62	0.62	1.08(0.48-2.42)	0.83
AhR rs2282885,T/C & AhR rs7811989,A/G	**Cases, n (%)** **N = 81**					
0	43(53.08)	8.20		1	1	
1	25(30.86)	7.33	0.66	0.88	1.03(0.57-1.87)	0.87
2	38(46.91)	8.03	0.90	0.97	1.03(0.61-1.73)	0.89
AhR rs10250822,T/C & AhR rs7811989,A/G	**Cases, n (%)** **N = 60**					
0	43(71.6)	7.33		1	1	

1	12(20.0)	5.36	0.90	1.04	1.05(0.46-2.39)	0.90
2	17(28.3)	5.36	0.86	0.94	0.99(0.49-1.99)	0.98
AhR rs 10250822,T/C &AhR rs 7811989,A/G	**Cases, n (%)** **N =85**					
0	28(32.9)	7.26		1	1	
1	42(49.4)	5.70	0.82	0.94	1.00(0.57-1.76)	0.97
2	57(67.0)	6.23	0.99	1.00	0.88(0.51-1.52)	0.66
AhR rs2066853,G/A & AhR rs7811989,A/G	**Cases, n (%)** **N = 65**					
0	42(64.6)	6.2		1	1	
1	17(26.1)	5.7	0.86	1.05	0.88(0.41-1.90)	0.75
2	23(35.3)	5.70	0.96	0.96	0.93(0.49-1.76)	0.83
[a] Unadjusted Hazards ratio for Kaplan meier analysis, [b] hazards ratio adjusted for age, sex, smoking, stage, KPS, ECOG, Histology.						

6.11 Combined effect of AhR SNPs on overall survival

In the present study, combined analysis of two SNPs was used to evaluate the effect of different genotypic combination on overall survival towards the lung cancer. We extended our analysis to explore the consolidated impact of four AhR SNPs whether they were observed to be associated with overall survival. The aggregate impact of these five SNPs was studied in combination of doubles as shown in Table 7. Our study couldn't find any significant relationship to foresee whether the distinctive two-fold combination among the four AhR SNPs have any impact on the overall survival rate of the lung cancer patients.

Table 17 : Effect of AhR SNPs and treatment outcome				
Genotype **AhR rs 2282885**	**Good response n(%) N = 61**	**Poor response** **n (%) N = 52**	**Adjusted OR (95% CI) [b]**	***p*-value[a]**
TT	41(27.3)	34(26.3)	1	
TC	18(52.0)	14(57.8)	1.14(0.44-2.91)	0.78
CC	2(20.5)	4(15.7)	0.65(0.10-3.92)	0.63
TC+CC	20(72.6)	18(73.6)	1.53(0.62-3.78)	0.35
AhR rs 10250822	**Good response n(%) N = 61**	**Poor response** **n (%) N = 52**	**Adjusted OR (95% CI) [b]**	***p*-value[a]**
TT	25(40.9)	27(51.9)	1	
TC	28(45.9)	21(40.3)	1.36(0.58-3.20)	0.47
CC	8(13.1)	4(7.69)	1.79(0.33-9.62)	0.49
TC+CC	36(59.0)	25(48.0)	1.21(0.65-2.67)	0.63
AhR rs 2066853	**Good response n(%) N = 61**	**Poor response** **n (%) N = 52**	**Adjusted OR (95% CI) [b]**	***p*-value[a]**
GG	46(75.4)	40(76.9)	1	
GA	13(21.3)	10(19.2)	0.77(0.28-2.11)	0.61
AA	2(3.27)	2(3.84)	0.28(0.03-2.06)	0.21
GA+AA	15(24.5)	12(23.0)	0.62(0.24-1.63)	0.34
AhR rs 7811989	**Good response n(%) N = 61**	**Poor response** **n (%) N = 52**	**Adjusted OR (95% CI) [b]**	***p*-value[a]**

AA	21(34.4)	24(46.15)	1	
AG	32(52.4)	21(40.38)	0.94(0.39-2.24)	0.89
GG	8(13.11)	7(13.46)	0.87(0.21-3.75)	0.87
AG+GG	40(65.5)	28(53.84)	0.90(0.38-2.15)	0.82
[b]Adjusted Odds ratios, 95% confidence intervals and their corresponding *p*-values were calculated by unconditional logistic analysis after adjusting for age, gender, smoking status, histology, stage, regimen, KPS,ECOG. [a]Two-sided χ^2 test.				

6.12 Effect of AhR SNPs and treatment outcome

Associations between polymorphisms and the efficacy outcome of response rate and clinical benefit were assessed logistic multiple regression analysis. Sixty-one patients experienced a good response (CR+PR) whereas fifty-two patients (SD+PD) did not respond to chemotherapy. We report that none of the four SNPS of the AhR gene were able to act as significant predictors of response rate or clinical benefit overall.

CHAPTER-7

DISCUSSIONS

Discussion

Lung malignancy has developed up as one of the significant cause of cancer death worldwide. It is a multifactorial disease which gets affected by environmental and genetic factors. Certain variation in the genome and metabolic pathways leads to alteration in the detoxification and metabolism of contaminants which demonstrated its role in the etiology in that disease. This is the first study which traces the role of four different genetic polymorphisms (rs7811989, rs2066853, rs10250822, rs2282885) towards the lung cancer susceptibility in North Indian population.

The results from our study reported that polymorphism in the two SNP's of AhR gene **(rs10250822, rs2282885)** are concordant with the study conducted in Chinese population for coke exposed workers. These workers showcased increased amount of hydroxypyrene in the urine, which is metabolite of PAH. The occurrence of detoxifying enzyme and their expression was increased in the presence of PAH. Variation in the AhR (rs10250822, rs2282885) essentially relate the association of urinary 1-OHP which demonstrated that AhR-signaling may partake in control of interceded PAH-metabolic activation and contribute to susceptibility of PAH exposure (Ping Bin *et al*., 2008).

In conclusion, the alteration in PAH metabolic pathway may interact with the environmental exposure and contribute towards tumorigenesis. Polymorphism in **rs10250822** do not showcase an association towards the risk of acquiring infertility in males as was stated in Iranian population (Berwick *et al*., 2004). This SNP continues to be unexplored vividly by researchers and thus we do not have enough instances to validate our work with.

Earlier study in Chinese population, suggests that an increase in the pack years of smokers simultaneously increased the OR and validated the hypothesis that validates the hypothesis that suggests that as the number of pack years increases the susceptibility of an individual towards acquiring lung cancer increases (OR=0.23 ,95% CI= 0.10-8.52, p-value= 0.002). It showcased a similar trend in the sub-grouping of the population in smokers and non-smokers followed by light and heavy smokers on the basis of pack years. Our study in the North Indian population falls very much in line with the former which also hold good for AhR **rs2066853** and proves a significant association of cumulative cigarette smoking on the susceptibility towards lung cancer. In our study, we seen that the individuals diagnosed with SQCC showed statistically significant values, which confirms the findings reported by the study in Chinese population (Daru et al., 2008).

Another study suggests a significant effect of AhR rs2282885 and rs2066853 polymorphism on the CYP1A2 inducibility, which confirmed the involvement of the AhR mediated pathway (Nebert *et al*., 2004). It was seen that if individual was exposed to more smoke inbuilt possessed increased capacity to detoxify the inhaled carcinogen, leading to enhanced CYP1A2 activity (Maria *et al*., 2013).

Our data suggests that the patients having mutant genotype showcased increase in the death rate when multivariate Cox hazardous proportional ratio was used. Similar study was involved in American population, which demonstrated that Arg554Lys polymorphism elevates the CYP1A1, CYP1A2 activity which bring change in activation of gene expression. The heterozygote genotype displayed risk for soft tissue sarcoma (Marianne Berwick et al., 2004).

AhR rs2066853 being nonsynonymous is thought to play a vital role in the area of proteins crucial to enzyme activity. Previous studies in Korean, Japanese and French population where no risk was observed in Arg554Lys polymorphism for lung cancer. Their variants were concordant with our analysis, in which smokers patients does not displayed any risk towards the disease whereas in non-smokers protective effect was seen. Conflicting results was observed in Caucasian population based study revealed that mutant genotype displayed increase CYP1A1 activity in women smokers, which leads to formation of DNA adducts in coke-oven workers.

However a study done in non-smokers exhibited an increased CYP1A1 enzyme which was determined by ethoxyresorufin-O-deethylase assay in peripheral blood lymphocytes. This study supports the finding of the present study where higher odds-ratio was observed in case of non-smokers having mutant genotype for this SNP.As evident, there has been contradicting prediction about the functional effect of the codon554 SNP. As this nucleotide change is a conservative replacement, therefore it might be possible that there exist no functional variability due to this polymorphism. However certain reports have recently documented that there are other polymorphism within the AhR gene which have substantial linkage disequilibrium with this polymorphism. So it might be possible that as SNP might not be functionally significant alone. But interacting with other such polymorphic variant they might produce a significant effect on the function of the AhR protein.

The possible explanation of the various finding regarding the ***AhR rs 2282885*** polymorphism is that it happens to be located in the intronic region of the *AhR* gene, wherein the gene expression is dysregulated leading to the decrease or increase in the gene transcription levels and it has been seen to influence the proper splicing of RNA leading to alternatively spliced RNA variants (Hirose *et al.,* 2008). For e.g. an intronic region in the *AhR* gene which alters RNA splicing at either 38 or 43 amino acid near the end of the carboxy terminus results in the deletion from the transactivation Domain of the Receptor, therefore these intronic mutations are accountable for differences in sensitivity to the xenobiotic induced toxicity (Pohjanbirta *et al*., 1998).

In similar study populations, SNPS ***rs2158041*** and ***rs7811989*** both residing in the intronic regions were reported to be associated with higher risk of Lung cancer (Daru *et al*., 2008). Another study reports the association of *AhR* rs 2282885 with the inducibility of CYP1A2 gene, which validates the involvement of *AhR* mediated pathway and also a higher risk towards Lung Cancer (Maria *et al*., 2013).

In an Iranian population study, it was seen that *AhR* rs2282885 SNP with a homozygous mutant genotype showed a threefold increase to acquire infertility in males. Literature backs the fact by holding the release of PAHs from the diesel exhaust responsible for the decreased sperm production due to perturbed spermatogenesis and testicular functions (Izawa *et al.,* 2007)

As of now no vivid studies on ***rs2282885*** has been reported or seen in association with the risk towards acquiring Lung cancer however this SNP shows a strong association with Idiopathic Male factor infertility which is a direct repercussion of differential sensitivity towards TCDD induced carcinogenesis (Safarinejad *et al*., 2013).

CHAPTER-8

CONCLUSION

Conclusion

In conclusion, our study showed the association between AhR polymorphism (rs7811989, , rs2282885, rs10250822) towards increased susceptibility of lung cancer in North Indian population. On the basis of histology, ADCC showed higher risk for lung cancer in three AhR variants (rs7811989, rs10250822, rs2282885).Trends changed for rs2066853 which showed protective effect in SQCC, SCLC patients. Doublet combination of two SNP's (rs10250822, rs7811989) showed the maximum risk of having lung cancer. Genotypic combination of three SNP's (rs10250822, rs2282885, rs2811989) displayed higher risk of having the disease.

Our study analyzed the effect of smoking in AhR gene polymorphism, resulting smokers are more susceptible to lung cancer. However light smokers showcased a higher risk of acquiring the disease in rs7811989, rs10250822.MDR analysis displayed best interaction model (rs10250822, rs2066853, rs7811989), having risk for cancer. Survival analysis showed that for rs20066853, a significantly higher MST was observed in patients having the carrier genotype whereas the MST significantly decreased in patients having variant genotype. Further, when adjusted Hazard ratio was calculated for the same, a significant 1.5 fold risk was observed in the case of heterozygous patients.

CHAPTER-9

REFERENCES

1. Kawajiri, K., Watanabe, J., Eguchi, H., Nakachi, K., Kiyohara, C., & Hayashi, S. I. (1995). Polymorphisms of human Ah receptor gene are not involved in lung cancer. *Pharmacogenetics and Genomics*, *5*(3), 151-158.
2. Cauchi, S., Stücker, I., Solas, C., Laurent-Puig, P., Cénée, S., Hémon, D., ... & Massaad-Massade, L. (2001). Polymorphisms of human aryl hydrocarbon receptor (AhR) gene in a French population: relationship with CYP1A1 inducibility and lung cancer. *Carcinogenesis*, *22*(11), 1819-1824.
3. Kim, J. H., Kim, H., Lee, K. Y., Kang, J. W., Lee, K. H., Park, S. Y., ... & Hong, Y. C. (2007). Aryl hydrocarbon receptor gene polymorphisms affect lung cancer risk. *Lung Cancer*, *56*(1), 9-15.
4. Harper, P. A., Wong, J. M., Lam, M. S., & Okey, A. B. (2002). Polymorphisms in the human AH receptor. *Chemico-biological interactions*,*141*(1), 161-187.
5. Bjeldanes, L. F., Kim, J. Y., Grose, K. R., Bartholomew, J. C., & Bradfield, C. A. (1991). Aromatic hydrocarbon responsiveness-receptor agonists generated from indole-3-carbinol in vitro and in vivo: comparisons with 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin. *Proceedings of the national academy of sciences*, *88*(21), 9543-9547.
6. Denison, M. S., & Nagy, S. R. (2003). Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals*.*Annual review of pharmacology and toxicology*, *43*(1), 309-334.
7. Ramadoss, P., Marcus, C., & Perdew, G. H. (2005). Role of the aryl hydrocarbon receptor in drug metabolism. *Expert opinion on drug metabolism & toxicology*, *1*(1), 9-21.
8. Shimada, T. (2006). Xenobiotic-metabolizing enzymes involved in activation and detoxification of carcinogenic polycyclic aromatic hydrocarbons. *Drug metabolism and pharmacokinetics*, *21*(4), 257-276.
9. Petrulis, J. R., & Perdew, G. H. (2002). The role of chaperone proteins in the aryl hydrocarbon receptor core complex. *Chemico-biological interactions*,*141*(1), 25-40.
10. Morrow, D., Qin, C., Smith, R., & Safe, S. (2004). Aryl hydrocarbon receptor-mediated inhibition of LNCaP prostate cancer cell growth and hormone-induced transactivation. *The Journal of steroid biochemistry and molecular biology*, *88*(1), 27-36.

11. Dzeletovic, N., McGuire, J., Daujat, M., Tholander, J., Ema, M., Fujii-Kuriyama, Y., & Poellinger, L. (1997). Regulation of dioxin receptor function by omeprazole. *Journal of Biological Chemistry*, *272*(19), 12705-12713.
12. Ding, X., & Kaminsky, L. S. (2003). HUMAN EXTRAHEPATIC CYTOCHROMES P450: Function in Xenobiotic Metabolism and Tissue-Selective Chemical Toxicity in the Respiratory and Gastrointestinal Tracts*.*Annual review of pharmacology and toxicology*, *43*(1), 149-173.
13. Sachse, C., Brockmöller, J., Bauer, S., & Roots, I. (1999). Functional significance of a C→ A polymorphism in intron 1 of the cytochrome P450 CYP1A2 gene tested with caffeine. *British journal of clinical pharmacology*,*47*(4), 445-449.
14. Nakajima, M., Yokoi, T., Mizutani, M., Kinoshitah, M., Funayama, M., & Kamataki, T. (1999). Genetic Polymorphism in the 5′-Flanking Region of HumanCYP1A2 Gene: Effect on the CYP1A2 Inducibility in Humans. *Journal of biochemistry*, *125*(4), 803-808.
15. Schlezinger, J. J., Liu, D., Farago, M., Seldin, D. C., Belguise, K., Sonenshein, G. E., & Sherr, D. H. (2006). A role for the aryl hydrocarbon receptor in mammary gland tumorigenesis. *Biological chemistry*, *387*(9), 1175-1187.
16. Kim, J. H., Kim, H., Lee, K. Y., Kang, J. W., Lee, K. H., Park, S. Y., ... & Hong, Y. C. (2007). Aryl hydrocarbon receptor gene polymorphisms affect lung cancer risk. *Lung Cancer*, *56*(1), 9-15.
17. Lin, P., Chang, H., Tsai, W. T., Wu, M. H., Liao, Y. S., Chen, J. T., & Su, J. M. (2003). Overexpression of aryl hydrocarbon receptor in human lung carcinomas. *Toxicologic pathology*, *31*(1), 22-30.
18. Rowlands, C. J., Staskal, D. F., Gollapudi, B., & Budinsky, R. (2010). The human AHR: identification of single nucleotide polymorphisms from six ethnic populations. *Pharmacogenetics and genomics*, *20*(5), 283-290.
19. Inouye, K., Pan, X., Imai, N., Ito, T., Takei, T., Tohyama, C., & Nohara, K. (2005). T cell-derived IL-5 production is a sensitive target of 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin (TCDD). *Chemosphere*, *60*(7), 907-913.
20. Inouye, K., Pan, X., Imai, N., Ito, T., Takei, T., Tohyama, C., & Nohara, K. (2005). T cell-derived IL-5 production is a sensitive target of 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin (TCDD). *Chemosphere*, *60*(7), 907-913.
21. Amazit, L., Pasini, L., Szafran, A. T., Berno, V., Wu, R. C., Mielke, M., ... & Mancini, M. A. (2007). Regulation of SRC-3 intercompartmental dynamics by

estrogen receptor and phosphorylation. *Molecular and cellular biology*,*27*(19), 6913-6932.

22. Nambu, J. R., Lewis, J. O., Wharton, K. A., & Crews, S. T. (1991). The Drosophila single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. *Cell*, *67*(6), 1157-1167.
23. Hankinson, O. (1994). The role of the aryl hydrocarbon receptor nuclear translocator protein in aryl hydrocarbon receptor action. *Trends in Endocrinology & Metabolism*, *5*(6), 240-244.
24. Skupinska, K. A. T. A. R. Z. Y. N. A., Misiewicz, I. R. E. N. A., & Kasprzycka-Guttman, T. E. R. E. S. A. (2004). Polycyclic aromatic hydrocarbons: physicochemical properties, environmental appearance and impact on living organisms. *Acta Pol Pharm*, *61*(3), 233-240.
25. Shimizu, Y., Nakatsuru, Y., Ichinose, M., Takahashi, Y., Kume, H., Mimura, J., ... & Ishikawa, T. (2000). Benzo [a] pyrene carcinogenicity is lost in mice lacking the aryl hydrocarbon receptor. *Proceedings of the National Academy of Sciences*, *97*(2), 779-782.
26. Boucher, O., Muckle, G., & Bastien, C. H. (2009). Prenatal exposure to polychlorinated biphenyls: a neuropsychologic analysis. *Environmental health perspectives*, *117*(1), 7.
27. Crawford, R. B., Holsapple, M. P., & Kaminski, N. E. (1997). Leukocyte Activation Induces Aryl Hydrocarbon Receptor Up-Regulation, DNA Binding, and Increased Cyp1a1Expression in the Absence of Exogenous Ligand.*Molecular pharmacology*, *52*(6), 921-927.
28. Quattrochi, L. C., & Tukey, R. H. (1993). Nuclear uptake of the Ah (dioxin) receptor in response to omeprazole: transcriptional activation of the human CYP1A1 gene. *Molecular pharmacology*, *43*(4), 504-508.
29. Murray, I. A., & Perdew, G. H. (2008). Omeprazole stimulates the induction of human insulin-like growth factor binding protein-1 through aryl hydrocarbon receptor activation. *Journal of Pharmacology and Experimental Therapeutics*,*324*(3), 1102-1110.
30. Carver, L. A., LaPres, J. J., Jain, S., Dunham, E. E., & Bradfield, C. A. (1998). Characterization of the Ah receptor-associated protein, ARA9.*Journal of Biological Chemistry*, *273*(50), 33580-33587.

31. Carver, L. A., & Bradfield, C. A. (1997). Ligand-dependent interaction of the aryl hydrocarbon receptor with a novel immunophilin homolog in vivo. *Journal of Biological Chemistry*, *272*(17), 11452-11456.
32. Meyer, B. K., Pray-Grant, M. G., Heuvel, J. P. V., & Perdew, G. H. (1998). Hepatitis B virus X-associated protein 2 is a subunit of the unliganded aryl hydrocarbon receptor core complex and exhibits transcriptional enhancer activity. *Molecular and cellular biology*, *18*(2), 978-988.
33. Petrulis, J. R., Kusnadi, A., Ramadoss, P., Hollingshead, B., & Perdew, G. H. (2003). The hsp90 co-chaperone XAP2 alters importin β recognition of the bipartite nuclear localization signal of the Ah receptor and represses transcriptional activity. *Journal of Biological Chemistry*, *278*(4), 2677-2685.
34. Silkworth, J. B., Lipinskas, T., & Stoner, C. R. (1995). Immunosuppressive potential of several polycyclic aromatic hydrocarbons (PAHs) found at a Superfund site: new model used to evaluate additive interactions between benzo [a] pyrene and TCDD. *Toxicology*, *105*(2), 375-386.
35. Ramadoss, P., & Perdew, G. H. (2005). The transactivation domain of the Ah receptor is a key determinant of cellular localization and ligand-independent nucleocytoplasmic shuttling properties. *Biochemistry*, *44*(33), 11148-11159.
36. Perdew, G. H., & Bradfield, C. A. (1996). Mapping the 90 kDa heat shock protein binding region of the Ah receptor. *IUBMB Life*, *39*(3), 589-593.
37. Meyer, B. K., & Perdew, G. H. (1999). Characterization of the AhR-hsp90-XAP2 core complex and the role of the immunophilin-related protein XAP2 in AhR stabilization. *Biochemistry*, *38*(28), 8907-8917.
38. Kazlauskas, A., Poellinger, L., & Pongratz, I. (1999). Evidence that the co-chaperone p23 regulates ligand responsiveness of the dioxin (Aryl hydrocarbon) receptor. *Journal of Biological Chemistry*, *274*(19), 13519-13524.
39. Kazlauskas, A., Poellinger, L., & Pongratz, I. (1999). Evidence that the co-chaperone p23 regulates ligand responsiveness of the dioxin (Aryl hydrocarbon) receptor. *Journal of Biological Chemistry*, *274*(19), 13519-13524.
40. Thurmond, T. S., Staples, J. E., Silverstone, A. E., & Gasiewicz, T. A. (2000). The aryl hydrocarbon receptor has a role in the in vivo maturation of murine bone marrow B lymphocytes and their response to 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin. *Toxicology and applied pharmacology*, *165*(3), 227-236.

41. Schneider, D., Manzan, M. A., Yoo, B. S., Crawford, R. B., & Kaminski, N. (2009). Involvement of Blimp-1 and AP-1 dysregulation in the 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin-mediated suppression of the IgM response by B cells. *Toxicological sciences*, kfp028.
42. Fujimaki, H., Nohara, K., Kobayashi, T., Suzuki, K., Eguchi-Kasai, K., Tsukumo, S., ... & Tohyama, C. (2002). Effect of a single oral dose of 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin on immune function in male NC/Nga mice.*Toxicological Sciences*, *66*(1), 117-124.
43. Mocarelli, P., Gerthoux, P. M., Ferrari, E., Patterson, D. G., Kieszak, S. M., Brambilla, P., ... & Sampson, E. J. (2000). Paternal concentrations of dioxin and sex ratio of offspring. *The Lancet*, *355*(9218), 1858-1863.
44. Moses, M., Lilis, R., Crow, K. D., Thornton, J., Fischbein, A., Anderson, H. A., & Selikoff, I. J. (1984). Health status of workers with past exposure to 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin in the manufacture of 2, 4, 5-trichlorophenoxyacetic acid: Comparison of findings with and without chloracne. *American journal of industrial medicine*, *5*(3), 161-182.
45. Thackaberry, E. A., Gabaldon, D. M., Walker, M. K., & Smith, S. M. (2002). Aryl hydrocarbon receptor null mice develop cardiac hypertrophy and increased hypoxia-inducible factor-1α in the absence of cardiac hypoxia.*Cardiovascular toxicology*, *2*(4), 263-273.
46. Bunger, M. K., Glover, E., Moran, S. M., Walisser, J. A., Lahvis, G. P., Hsu, E. L., & Bradfield, C. A. (2008). Abnormal liver development and resistance to 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin toxicity in mice carrying a mutation in the DNA-binding domain of the aryl hydrocarbon receptor. *Toxicological Sciences*, *106*(1), 83-92.
47. Boffetta, P. (2000). Molecular epidemiology. *Journal of internal medicine*,*248*(6), 447-454.
48. Jones, L. C., & Whitlock, J. P. (2001). Dioxin-inducible Transactivation in a Chromosomal Setting ANALYSIS OF THE ACIDIC DOMAIN OF THE Ah RECEPTOR. *Journal of Biological Chemistry*, *276*(27), 25037-25042.
49. Kawajiri, K., Watanabe, J., Eguchi, H., Nakachi, K., Kiyohara, C., & Hayashi, S. I. (1995). Polymorphisms of human Ah receptor gene are not involved in lung cancer. *Pharmacogenetics and Genomics*, *5*(3), 151-158.
50. Pohjanvirta, R., Wong, J. M., Li, W., Harper, P. A., Tuomisto, J., & Okey, A. B. (1998). Point mutation in intron sequence causes altered carboxyl-terminal structure

in the aryl hydrocarbon receptor of the most 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin-resistant rat strain. *Molecular pharmacology*,*54*(1), 86-93.

51. Pohjanvirta, R., Wong, J. M., Li, W., Harper, P. A., Tuomisto, J., & Okey, A. B. (1998). Point mutation in intron sequence causes altered carboxyl-terminal structure in the aryl hydrocarbon receptor of the most 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin-resistant rat strain. *Molecular pharmacology*,*54*(1), 86-93.
52. Murray, I. A., Patterson, A. D., & Perdew, G. H. (2014). Aryl hydrocarbon receptor ligands in cancer: friend and foe. *Nature Reviews Cancer*, *14*(12), 801-814.