Send Orders for Reprints to reprints@benthamscience.ae

Current Genomics, 2018, *19*, 313-326

RESEARCH ARTICLE

Role of Sequence Variations in *AhR* Gene Towards Modulating Smoking Induced Lung Cancer Susceptibility in North Indian Population: A Multiple Interaction Analysis

Sneha Budhwar[1], Charu Bahl[1], Siddharth Sharma[1,*], Navneet Singh[2] and Digambar Behera[2]

[1]*Department of Biotechnology, Thapar University, Patiala, Punjab-147002, India;* [2]*Department of Pulmonary Medicine, Post Graduate Institute of Medical Education and Research (PGIMER), Sector 14, Chandigarh, India*

ARTICLE HISTORY

Received: October 25, 2016
Revised: June 28, 2017
Accepted: June 28, 2017

DOI:
10.2174/1389202918666170915160606

Abstract: ***Background*:** *AhR,* a ubiquitously expressed ligand-activated transcription factor, upon its encounter with the foreign ligands activates the transcriptional machinery of genes encoding for biotransformation enzymes like CYP1A1 hence, mediating the metabolism of Poly aromatic hydrocarbons and nitrosamines which account for the maximally found carcinogen in cigarette smoke. Polymorphic variants of *AhR* play a significant role and are held responsible for disposing the individuals with greater chances of acquiring lung cancer.

***Objective*:** To study the role of *AhR* variants (rs2282885, rs10250822, rs7811989, rs2066853) in affecting lung cancer susceptibility.

***Methods*:** 297 cases and 320 controls have been genotyped using PCR-RFLP technique. In order to find out the association, unconditional logistic regression approach was used. To analyze high order interactions Multifactor Dimensionality Reduction and Classification and regression tree was used.

Results*:** Subjects carrying the variant genotype for *AhR* rs7811989 showed a two-fold risk (p*=0.007**) and a marginal risk was also seen in case of individuals carrying either single or double copy of susceptible allele for rs102550822 (***p*=0.02**). Whereas the variant allele for rs2066853 showcased a strong protective effect (***p*=0.003**). SQCC individuals with mutant genotype of rs2066853 also exhibited a protective effect towards lung cancer (OR=0.30, ***p*=0.0013**). The association of rs7811989 mutant genotype and rs10250822 mutant genotype was evident especially in smokers as compared to non-smokers. *AhR* rs2066853 showed a decreased risk in smokers with mutant genotype (***p*=0.002**). MDR approach gave the best interaction model of *AhR* rs2066853 and smoking (CVC=10/10, prediction error=0.42).

***Conclusion*:** *AhR* polymorphic variations can significantly contribute towards lung cancer predisposition.

Keywords: Aryl hydrocarbon receptor, Lung cancer, Polymorphism, Smoking, Predisposition, *AhR* variants.

1. INTRODUCTION

Lung cancer is the leading cause of deaths worldwide and the incidence of lung cancer is increasing in developing countries like India. Smoking is the main causative factor for lung cancer and it has been found that lung carcinogenesis might arise due to the presence of harmful chemicals which are present in cigarette smoke [1]. These chemicals may lead to carcinogenesis of the lung either due to chronic inflammation or anomalies in the DNA repair system especially in the epithelial cells of the lung airways. Tobacco smoke contains a plethora of harmful chemicals like Polycyclic Aromatic Hydrocarbons (PAHs), N-nitrosamines, *etc.* These chemical carcinogens present in tobacco smoke exert their effect by binding to the cytosolic receptor called the aryl hydrocarbon receptor (AhR) [2]. AhR is a ligand-activated transcription factor and regulates the activity of cytochrome P450 enzymes [3]. AhR also mediates the toxic effects of a variety of environmental chemicals including PAHs [4] and plays an important role in carcinogenesis [5]. The human *AhR* gene located on chromosome 7p15 region is 50kb long in size and contains 12 exons and 10 introns [6]. *AhR* is strongly expressed in liver, adipose tissue and in bronchial epithelial cells. AhR plays a significant role in the detoxification of xenobiotics and drugs that involve the induction of phase-I metabolizing enzymes like cytochrome P450 enzymes. The components present in tobacco smoke exert their effect by binding to the AhR and then get translocated to the nucleus where the AhR-ligand dimerizes with AhR Nuclear Translocator (ARNT) and then binds further to the Xenobiotic Response Elements (XRE) which in turn regulate the transcriptional activity of cytochrome P450 enzymes [7-9]. It has

*Address correspondence to this author at the Department of Biotechnology, Thapar University, Patiala, Punjab-147002, India; Tel: 0091-9501688366; E-mail: siddharthsharma.phd@thapar.edu

1389-2029/18 $58.00+.00

been shown in *in vivo* models that, AhR causes the induction of CYP1A1 and CYP1B1 in lung tissue soon after exposure to tobacco smoke. PAHs like benzo (a) pyrene (BaP) activates AhR and increase the expression of CYP1A1. Therefore, the cross talk between AhR and CYP1A1 has been found to play an important role in tobacco smoking related diseases, especially lung cancer [10]. BaP, one of the most important members of PAHs in cigarette smoke, is a major potent carcinogen implicated in the etiology of lung cancer as it leads to the formation of BaP diol epoxidation-DNA adducts and this process is mediated through the AhR which activates the Phase1 biotransformation cascade [11, 12]. There is direct evidence that BaP induced carcinogenicity is lost in the AhR-null mice [13]. Furthermore, a study in Non-Small Cell Lung Cancer (NSCLC) patients showed a positive correlation between the expression of AhR and CYP1A1, with increased levels of AhR mRNA and protein as compared to normal lung tissue [14]. Moreover, AhR has also been found to interact with many cellular signaling cascades which might lead the propensity of cells towards proliferation, cell cycle arrest or apoptosis [15].

From the above findings it might be plausible that genetic variations within the *AhR* gene may lead to differences in transcriptional activity of the AhR and hence affect the induciblity of target genes involved in carcinogen metabolism. Furthermore, genetic variations within the *AhR* gene also may lead to altered AhR protein activity that may affect lung carcinogenesis. Thus, *AhR* polymorphisms might negatively affect the affinity and sensitivity of the AhR proteins and activation of the AhR-signaling pathway [16]. The risk towards lung cancer in relation to *AhR* polymorphism has been controversial. Studies conducted in different populations like Japanese [17], French [18, 19] and Finnish [20, 21] have yielded a negative correlation among AhR, ARNT and CYP1A1 polymorphism and lung cancer. As far our knowledge is concerned no study has been conducted in Indian population to assess the relationship between *AhR* polymorphic variants and its association towards lung cancer risk. Given the fact that AhR does play a role in carcinogen metabolism by controlling the CYP1A1 activity, we hypothesize that *AhR* gene polymorphism might affect incidence of lung cancer. To test this hypothesis, 4 SNPs in the *AhR* gene from hap-map project were selected [r^2>0.8 and AF> (5%)]. Out of the four SNPs to be studied, three (rs2282885, rs10250822, rs7811989) were located in the intronic region and thus these might affect the expression or function of *AhR* gene. They may increase or decrease gene transcription and might also influence the proper splicing of RNA or yield alternatively spliced messenger RNA variants. The non-synonymous SNP (rs2066853) resulted in substitution of arginine with lysine amino acid at 554 position which might lead to change in the primary structure of the protein and influence the function of the AhR receptor [22].

Therefore, we conducted a case-control study and performed genotyping for four SNPs in *AhR* gene to find any association between these genetic variants and lung cancer susceptibility. We also evaluated the potential gene-smoking interaction to determine the impact of the *AhR* polymorphism and smoking status in modulating lung cancer risk. We also investigated the SNP to SNP interaction and SNP to environment interaction in the study using MDR and CART analysis tools. Furthermore, we retro prospectively assessed the relationship between the four SNPs of *AhR* and their role in overall survival either individually or in combinations in the North Indian patients undergoing doublet based platinum chemotherapy.

2. MATERIALS AND METHODS

2.1. Sample Collection

The current study enrolled a total of 320 controls having no history of lung cancer and 297 lung cancer patients recruited from the Department of Pulmonary Medicine, Post Graduate Institute of Medical Education and Research (PGIMER) Chandigarh, India. This study has been ethically cleared by the Institute ethics committee of PGIMER. Informed consent was obtained from all enrolled patients or their representatives. All the lung cancer cases were histopathologically diagnosed as having ADCC (Adenocarcinoma), SQCC (Squamous Cell Carcinoma) and SCLC (Small Cell Lung cancer). There were no age, gender, smoking, histological or TNM stage restrictions. All controls were matched for age (±10) years, sex and smoking parameters in order to avoid any sampling bias. The questionnaire required information on demographic and smoking characteristics like tobacco habits such as smoking of beedi's/cigarette (a native cigarette like stick of coarse tobacco hand-rolled in a dry tembuhurni leaf) *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 follow-up data was obtained by contacting the patients using their contact details mentioned in the medical records.

2.2. DNA Extraction and AhR Genotyping

5 ml of blood was collected in EDTA coated vacutainers from each individual enrolled in the study. Genomic DNA was isolated from that blood according to the protocol of Sodhi *et al*. [23] and stored at -20°C. The genotyping of four *AhR* SNPs was done by PCR-RFLP using specific primer sequences and restriction enzymes as described previously [22]. The detail of the primer sequences, annealing temperature, restriction enzyme, and digestion pattern of the amplified PCR are given in Table **S1**.

2.3. Statistical Analysis

Statistical analysis was conducted to check whether the four SNP'S were in Hardy-Weinberg Equilibrium (HWE) by using the formula (p2+2pq+q2=1). Pearson's χ2-test was used to determine whether there was any significant difference in allelic and genotypic frequencies between cases and controls and independent t-test was used for continuous variable like age and pack-years. To evaluate the risk of lung cancer and *AhR* polymorphism, logistic regression analysis was conducted that gives Odds-Ratio (OR), 95% Confidence Interval (CI) and p-value. The odds ratio was adjusted for sex, smoking and age. A p-value less than 0.05 is considered as highly significant [23]. SHEsis software was used for linkage disequilibrium and haplotype analysis for the four

SNP's in which D' and r^2 value were calculated [24]. Furthermore, we applied the Multifactor Dimensionality Reduction (MDR) method to identify interaction models. MDR is a non-parametric genetic model method for overcoming some of the limitations of logistic regression for the detection and characterization of SNP to SNP and SNP to environment interactions. Using MDR multi locus, genotypes are pooled into two groups of high and low risk, thus reducing the genotype predictors from n dimensions to one. Among the different genotype models generated, only those genotypic combinations, having the highest Cross-Validation Consistency (CVC), testing accuracy and significant permutation *p*-value were taken as the best interaction model. The combined effect of the variables was calculated using logistic regression analysis and a p-value less than 0.05 was considered to be statistically significant. MDR tests were performed using the version 0.5.1 of the open source MDR software package that is available online (http://www.epistasis.org) [25]. Lastly, we conducted a Classification And Regression Tree (CART) analysis to detect and characterize the high-order interactions employing CART software (6.0, Salford Systems, San Diego, California). CART is a tree based model that is created by binary recursive partitioning method and produces a decision tree to identify sub-groups at higher risk, which are found to be less visible when using logistic regression methods. The analysis is conducted in such a manner where the most significant predictor is used to split the sample into subgroups and continues until the differences are not significant. Finally, there results are obtained as classification or decision, trees having a node and sub-nodes. The risk of all genotypes sets was estimated by considering the node with low case rate as the reference to calculate the ORs and 95% CIs [26]. Univariate and multivariate analysis were evaluated by using Kaplan-Meier survival analysis and Cox proportional hazardous ratio. Kaplan-Meir was used to obtain median OS time and log rank *p*-value. Med Cal (version 16.8) software was used to compute genotypic frequencies, logistic regression analysis and survival analysis (Med calc software, Ostend, Belgium) [23].

3. RESULTS

3.1. Study Characteristics

The demographic characteristics of the study population are shown in Table **1**. The study population comprised of a total of 297 cases and 320 controls. The mean age of cases was 57.6 ± 10.81 whereas the mean age of all controls was 53.00 ± 10.42. The current study included 254 (85.52%) males and 43 (14.47%) females in cases while 265 (82.81%) males and 55 (17.18%) females made up the control group. The study was also analysed for the difference in the smoking status between the cases and controls. The study comprised of 233 (78.45%) smokers and 64 (21.54%) non-smokers in the cases group whereas 221 (69.06%) smokers and 99 (30.93) non-smokers were there in the controls.

3.2. *AhR* Variants and Lung Cancer Susceptibility

The allelic and genotypic frequencies were calculated for all the four SNPs of the AhR gene which were studied. The genotypic frequencies for *AhR* rs7811989 A>G polymorphism in both the control (n=320, $\chi2$=3.36, df=2; p-value=0.06) and case groups (n=297, $\chi2$=3.05, df=2; p=0.08) were in accordance with the HWE. For *AhR* rs7811989 as shown in Table **2**, the frequency of the mutant genotype (GG) was found to be overrepresented in cases as compared to controls and the difference in the frequency distributions was found to be significant ($\chi2$=9.43,df=2; p=0.008). The Minor Allele Frequency (MAF) distribution was more in cases as compared to controls (0.35 *vs*. 0.29). When the wild type genotype (AA) was taken as reference, it was observed that patients carrying both the mutant alleles (GG) had a two-fold risk for lung cancer which was found to be highly significant. When stratified according to histological subtypes, it was observed that SCLC subjects with the mutant genotype (GG) for the rs7811989 polymorphism had a four-fold (OR=4.24, 95%CI=1.70-10.56, ***p*=0.001**) risk for developing SCLC as compared to those subjects with both the wild alleles.

In case of *AhR* rs10250822 T>C, the genotypic frequencies in both the cases ($\chi2$=2.09, df=2; *p*=0.14) and controls ($\chi2$=0.40, df=2; *p*=0.52) were in accordance with HWE and there were no deviations. In lung cancer subjects, the frequency of the heterozygous (TC) was found to be at higher frequency as compared to the controls (45.7 *vs*. 37.5%), whereas the mutant genotype was also at slightly higher frequency in cases as compared to controls (7.5% *vs*. 4%). The MAF in cases and controls was 0.29 and 0.24 respectively. The genotypic frequencies between cases and controls were found to be significant ($\chi2$=7.37,df=2; *p*=0.02). Taking TT genotype as reference, patients having joint-genotype (TC+CC) had a marginal risk for acquiring lung cancer.

In case of *AhR* rs2282885 T>C polymorphic site, both the control ($\chi2$=0.36, df=2; *p*=0.54) and cases group ($\chi2$=2.95, df=2; *p*=0.08) were in accordance with HWE, however there was no significant difference in the distribution of the variant alleles between the cases and controls ($\chi2$=0.86, df=2; *p*=0.65). The MAF was 0.21 and 0.17 for the cases and controls, respectively. No association was also observed when we stratified the cases on the basis of histological sub-types except in case of ADCC subjects having the combined variant genotype (TC+CC). They showed a ten-fold risk of developing lung cancer.

In case of *AhR* rs2066853G>A SNP, the frequency of the wild type genotype was found to be higher in the controls as compared to cases (9.06 *vs*. 4.37%). The MAF for the controls and cases was 0.20 and 0.04 respectively. A significant association was seen in the genotypes between cases and controls ($\chi2$=6.48, df=2; *p*=0.03). As shown in Table **2**, with reference to the wild type genotype (GG) it was observed that lung cancer subjects carrying both alleles for mutant genotype (AA) exhibited a significant protective effect towards lung cancer (OR=0.34, 95%CI=0.17-0.70, *p*=**0.003**). When stratified on the basis of histology, it was observed that SQCC individuals with mutant genotype also exhibited a protective effect towards lung cancer.

3.3. Risk of Lung Cancer on the Basis of Smoking Status

To study the association of smoking and *AhR* polymorphism towards risk for lung cancer, the patients enrolled for the study were classified as smokers and non-smokers as shown in Table **3**. Depending upon the pack years, smokers

Table 1. Distribution of demographic characteristics of cases and controls.

Variables	Total (N)	Cases n (%)	Total (N)	Control n (%)	*p*-value
Age (years)					
Mean ± SD	297	57.87 ± 10.81	320	52.14 ± 10.42	**<0.0001**
Range		28-86		19-83	
Gender					
Male	297	254 (85.52)	320	265 (82.81)	0.932
Female		43 (14.47)		55 (17.18)	
Smoking Status					
Smokers	297	233 (78.45)	320	221 (69.06)	0.937
Non-smokers		64 (21.54)		99 (30.93)	
Pack-Years					
Mean ± SD	297	27.5 ± 34.04	320	17.61± 19.92	**<0.0001**
Histology					
ADCC	297	97 (32.65)			
SCLC		71 (23.90)			
SQCC		129 (43.43)			
Others		….			
TNM Staging					
I	273	3 (1.1)			
II		12 (4.4)			
III		138 (50.5)			
IV		120 (44)			
Overall Survival					
Dead	150	118			
Alive		32			
Performance Status					
KPS(80-100)	150	96 (64)			
KPS(60-70)		42 (28)			
KPS(<60)		12 (8)			
ECOG(0 and 1)	150	120 (80)			
ECOG(2)		29 (19.3)			
ECOG(3 and 4)		1 (0.7)			

Abbreviations: SD=Standard Deviation, n=total number of case patients or controls 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.

were categorized into heavy-smokers and light-smokers. In case of *AhR* rs7811989 A>G, we found that the smokers with the mutant genotype (GG) exhibited a three-fold risk for lung cancer as compared to non-smokers with the same genotype. Light-smokers carrying both mutant alleles showed a higher risk for lung cancer (OR=3.9, 95% CI=1.38-11.37, ***p*=0.01**) as compared to heavy smokers with similar genotype.

For *AhR* rs10250822 T>C, it was observed that the study subjects with mutant (CC) genotype were at a 2-fold risk of lung cancer (OR=2.26, 95% CI=1.10-6.20, ***p*=0.02**) as

Table 2. 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 (320) n (%)**	**Cases (297) n(%)**	**AOR (95% CI) [a]**	***p* [b]**	**Cases (97) n(%)**	**AOR (95% CI)[a]**	***p* [b]**	**Cases (129) n(%)**	**AOR (95% CI) [a]**	***p* [b]**	**Cases (71) n(%)**	**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 (1.07-2.94)	0.02	67 (51.93)	1.19 (0.15-1.88)	0.43	25 (35.21)	1.30 (0.72-2.36)	0.37
GG	20 (6.25)	33 (11.11)	2.32 (1.24-4.32)	**0.007**	9 (9.27)	1.82 (0.73-4.50)	0.91	12 (9.30)	1.89 (0.85-4.32)	0.11	13 (18.30)	4.24 (1.7-10.56)	**0.001**
GA+AG	165 (51.56)	185 (62.28)	1.51 (1.09-2.10)	**0.012**	63 (64.94)	1.77 (1.08-2.88)	0.02	79 (61.24)	1.30 (0.84-2.01)	0.23	38 (53.52)	1.61 (0.92-2.81)	0.08
A	453	376	-	-	-	-	-	-	-	-	-	-	-
G	185	218	-	-	-	-	-	-	-	-	-	-	-
MAF	-	0.35	-	-	-	-	-	-	-	-	-	-	-
-	-	**OVERALL**			**ADCC**			**SQCC**			**SCLC**		
AhR rs10250822 (T>C)	**Controls (320) n (%)**	**Cases (297) n (%)**	**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)	**0.04**	43 (44.32)	1.52 (0.92-2.50)	0.09	60 (46.51)	1.38 (0.89-2.15)	0.14	32 (45.07)	1.32 (0.75-2.32)	0.33
CC	16 (5)	22 (7.40)	1.77 (0.86-3.59)	0.11	9 (9.27)	2.48 (0.95-6.44)	0.06	7 (5.42)	1.22 (0.45-3.28)	0.68	6 (8.45)	2.85 (0.96-8.42)	0.05
TC+CC	136 (42.5)	158 (53.19)	1.44 (1.04-2.00)	**0.02**	52 (53.60)	1.60 (1.00-2.57)	0.04	67 (51.93)	1.36 (0.88-2.09)	0.15	38 (53.52)	1.45 (0.84-2.50)	0.17
T	488	414	-	-	-	-	-	-	-	-	-	-	-
C	152	180	-	-	-	-	-	-	-	-	-	-	-
MAF	0.24	0.29	-	-	-	-	-	-	-	-	-	-	-

(Table 2) contd….

-	-	OVERALL			ADCC			SQCC			SCLC		
AhR rs2282885 (T>C)	**Controls (320) n (%)**	**Cases (297) n (%)**	**AOR (95% CI) [a]**	***p* [b]**	**Cases (97) n (%)**	**AOR (95% CI) [a]**	***p* [b]**	**Cases (129) n(%)**	**AOR (95% CI) [a]**	***p* [b]**	**Cases (71) n(%)**	**AOR (95% CI) [a]**	***p* [b]**
TT	221 (69.06)	207 (69.69)	1.00 (Reference)	-	68 (70.10)	1.00 (Reference)	-	82 (63.56)	1.00 (Reference)	-	52 (73.23)	1.00 (Reference)	-
TC	88 (27.50)	83 (27.94)	1.04 (0.17-1.50)	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)	0.10	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 (4.53-21.9)	**<0.0001**	47 (36.43)	1.37 (0.87-2.16)	0.16	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 (320) n (%)**	**Cases (297) n (%)**	**AOR (95% CI) [a]**	***p* [b]**	**Cases (97) n(%)**	**AOR (95% CI) [a]**	***p* [b]**	**Cases (129) n(%)**	**AOR (95% CI) [a]**	***p* [b]**	**Cases (71) n(%)**	**AOR (95% CI) [a]**	***p* [b]**
GG	224 (70.0)	229 (77.10)	1.00 (Reference)	-	74 (76.28)	1.00 (Reference)	-	100 (77.5)	1.00 (Reference)	-	56 (78.87)	1.00 (Reference)	-
AG	67 (20.93)	55 (18.51)	0.74 (0.49-1.22)	0.15	19 (19.58)	0.83 (0.46-1.50)	0.54	23 (17.82)	0.64 (0.36-1.11)	0.11	12 (16.90)	0.59 (0.29-1.12)	0.15
AA	29 (9.06)	13 (4.37)	0.34 (0.17-0.70)	**0.003**	4 (4.12)	0.36 (0.12-1.09)	0.71	6 (4.65)	0.30 (0.11-0.78)	0.01	3 (4.22)	0.32 (0.09-1.12)	0.07
AG+GG	96 (30.0)	68 (22.89)	0.62 (0.43-0.91)	**0.01**	23 (23.71)	0.69 (0.45-1.18)	0.18	29 (22.48)	0.53 (0.32-0.88)	0.01	15 (21.12)	0.54 (0.28-1.02)	0.05
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 $\chi 2$ 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.

compared to non-smokers with similar genotype. When stratified according to pack-years, lung cancer subjects with history of lower pack years exhibited a 4-fold increased risk for lung cancer with variant alleles for the *AhR* rs10250822 SNP.

Lastly for *AhR* rs2066853 G>A, our data suggested much higher significant values in smokers as compared to non-smokers. When considering the combination of homozygous mutant and heterozygous genotype, both displayed significant value in smokers. The mutant (AA) genotype in smokers (OR=0.23, 95%CI=0.10-8.52, ***p*=0.002**) displayed significant values while non-smokers failed to show any significant value. For mutant genotype, both heavy smokers (OR=0.0.22, 95%CI=0.07-0.06, ***p*=0.02**) and light smokers (OR=0.25, 95%CI=0.08-0.081, ***p*=0.02**) showed significant value, displaying a protective effect.

Table 3. Genotypic frequency distribution of AhR variants among patients and controls on the basis of Smoking status and its susceptibility towards Lung cancer.

AhR rs7811989(A>G)	CASES (SMOKERS) N= 233(%)	CONTROLS (SMOKERS) N=221(%)	AOR(95% CI)[b]	*p-value*[a]	CASES (NON-SMOKERS) N=64 (%)	CONTROLS (NON-SMOKERS) N=99(%)	AOR(95% CI)[b]	*p-value*[a]
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(A>G)	**CASES N=104 (%) (Light smokers; PY≤25)**	**CONTROLS N=124(%) (Light smokers; PY≤25)**	**AOR(95% CI)[b]**	***P***	**CASES N= 129(%) (Heavy smokers;PY>25)**	**CONTROLS N=97(%) (Heavy smokers;PY>25)**	**AOR(95% CI)[b]**	***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(T>C)	**CASES (SMOKERS) N=233 (%)**	**CONTROLS (SMOKERS) N=221(%)**	**AOR(95% CI)[b]**	***P***	**CASES (NON-SMOKERS) N= 64(%)**	**CONTROLS (NON-SMOKERS) N=99(%)**	**AOR(95% CI)[b]**	***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(T>C)	**CASES N=104 (%) (Light smokers; PY≤25)**	**CONTROLS N=124(%) (Light smokers; PY≤25)**	**AOR(95% CI)[b]**	***P***	**CASES N=129 (%) (Heavy smokers;PY>25)**	**CONTROLS N=97(%) (Heavy smokers;PY>25)**	**AOR(95% CI)[b]**	***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)	0.10	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
AhR rs2282885(T>C)	**CASES (SMOKERS) N=233 (%)**	**CONTROLS (SMOKERS) N=221(%)**	**AOR(95% CI)[b]**	***P***	**CASES (NON-SMOKERS) N=64 (%)**	**CONTROLS (NON-SMOKERS) N=99(%)**	**AOR(95% CI)[b]**	***P***
0	152(65.23)	157(71.04)	Ref(1.00)	Ref.	48(75.00)	64(64.64)	Ref(1.00)	Ref.
1	69(29.61)	58(26.24)	1.22(0.79-1.87)	0.35	13(20.31)	30(30.30)	0.59(0.27-1.28)	0.18
2	12(5.15)	6(2.71)	2.17(0.77-6.13)	0.14	3(4.68)	5(5.05)	0.83(0.179-3.87)	0.81
3	81(34.76)	64(28.95)	1.30(0.86-1.96)	0.19	16(25)	35(35.35)	0.61(0.29-1.27)	0.19

(Table 3) contd….

AhR rs2282885(T>C)	CASES N=104 (%) (Light smokers; PY≤25)	CONTROLS N=124(%) (Light smokers; PY≤25)	AOR(95% CI)[b]	*P*	CASES N=129 (%) (Heavy smokers;PY>25)	CONTROLS N=97(%) (Heavy smokers;PY>25)	AOR(95% CI)[b]	*P*
0	68(65.38)	90(72.58)	Ref(1.00)	Ref.	84(65.11)	67(69.07)	Ref(1.00)	Ref.
1	29(27.88)	34(27.41)	1.10(0.60-2.01)	0.74	40(31.00)	24(24.74)	1.43(0.76-2.68)	0.26
2	7(6.73)	0(0.00)	….	…..	5(3.87)	6(6.18)	0.86(0.220-3.42)	0.83
3	36(34.61)	34(27.41)	1.35(0.76-2.41)	0.29	45(34.88)	30(30.92)	1.32(0.732-2.40)	0.350
AhR rs2066853(G>A)	**CASES (SMOKERS) N= 233(%)**	**CONTROLS (SMOKERS) N=221(%)**	**AOR(95% CI)[b]**	***P***	**CASES (NON-SMOKERS) N=64 (%)**	**CONTROLS (NON-SMOKERS) N=99(%)**	**AOR(95% CI)[b]**	***P***
0	185(79.39)	142(64.25)	Ref(1.00)	Ref.	44(68.75)	82(82.82)	Ref(1.00)	Ref.
1	39(16.73)	51(23.07)	0.53(0.32-0.87)	**0.004**	16(25.00)	16(16.16)	1.73(0.77-3.88)	0.18
2	9(3.86)	28(12.66)	0.23(0.10-8.52)	**0.002**	4(6.25)	1(1.01)	7.67(0.75-77.4)	0.08
3	48(20.60)	79(35.74)	0.43(0.27-0.67)	**0.002**	20(31.25)	17(17.17)	2.05(0.95-4.41)	0.06
AhR rs2066853(G>A)	**CASES N=104 (%) (Light smokers; PY≤25)**	**CONTROLS N=124(%) (Light smokers; PY≤25)**	**AOR(95% CI)[b]**	***P***	**CASES N=129 (%) (Heavy smokers;PY>25)**	**CONTROLS N=97(%) (Heavy smokers;PY>25)**	**AOR(95% CI)[b]**	***P***
0	79(75.96)	78(62.90)	Ref(1.00)	Ref.	106(82.17)	64(65.97)	Ref(1.00)	Ref.
1	21(20.19)	30(24.19)	0.61(0.31-1.20)	0.15	18(13.95)	21(21.64)	0.53(0.25-1.114)	0.09
2	4(3.84)	16(12.90)	0.25(0.08-0.081)	**0.02**	5(3.87)	12(12.37)	0.22(0.07-0.681)	**0.02**
3	25(24.03)	46(37.09)	0.49(0.27-0.91)	0.62	23(17.82)	33(34.02)	0.41(0.21-0.78)	0.62

[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 $\chi 2$ 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.

3.4. Combinatorial Risk Assessment of Five *AhR* SNPs *(AhR* rs7811989, rs10250822, rs2282885, rs2066853)

We further assessed the combined effect of the four SNP's of *AhR* gene in different combinations as shown in Table **S2**. It was observed that when interaction between two SNPs was evaluated, subjects who were carrying the combined variant genotype (TC+CC+AG+GG) for both rs2282885 & rs7811989 polymorphic sites had a significant association towards risk for lung cancer (OR=1.68, 95% CI=1.05-2.71, ***p*=0.03**). Individuals who had either a single or double copy of variant allele for the SNP's namely rs10250822 (T>C), rs2282885 (T>C) and rs7811989 (A>G) exhibited a 2.6-fold risk for lung cancer which was found to be significant.

3.5. Association with Haplotypes and Linkage Disequilibrium in *AhR* Variants

Haplotype frequencies and linkage disequilibrium were obtained for the four SNP's using SHEsis software. Haplotype frequencies were classified and are shown in Table **S3a**. Only those haplotype blocks were evaluated where the case and control frequencies were more than 0.03; whereas those blocks whose haplotype frequency were less than 0.03 were omitted. Correlation of general haplotype profile uncovered a critical contrast between the cases and controls. Global test for the comparison of haplotypes in cases and controls gave χ^2 =33.46, df=7, p= 2.26e^{-005}. Three haplotype blocks namely Hap1, Hap4 and Hap8 showed a marginally increased risk in patients possessing these respective haplotypes. Hap1 comprised of variant allele of *AhR* rs10250822 T/C and all others were wild type alleles. Similarly, Hap8 also contained only a single variant allele of *AhR* rs7811989 A/G. On the contrary, Hap4 consisted of variant alleles of two variants namely *AhR* rs2282885 T/C and rs7811989 A/G. Other three Hap Blocks including Hap3, Hap6 and Hap7 were found to confer a strong protective effect in patients as the subjects with these haplotypes were found to be at a lower frequency among cases as compared to controls.

Table **S3b** summarizes the result of *D'* values and r^2 values for Linkage Disequilibrium (LD) between the cases and controls together. *AhR* rs10250822 and rs2282885 illustrated a linkage disequilibrium. *AhR* rs7811989 and rs2282885 also illustrated a linkage disequilibrium *D'=0.108, r^2=0.05*) which is a strong disequilibrium. The pairwise linkage dise-

quilibrium is also illustrated by the block diagram in Fig. (**S1**).

3.6. Multifactor Dimensionality Reduction (MDR) Analysis

Tables **S4a** and **S4b** summarize the average Cross Validation Consistency (CVC) and average prediction error obtained from MDR analysis of the data set of subjects with or without lung cancer. The best interaction model is the one having maximum CVC and minimum prediction error. In Table **4a**, the best interaction model has three *AhR* variants **(*AhR* rs10250822, *AhR* rs2066853, *AhR* rs7811989)** because it had maximum 10 /10 CVC, minimum prediction rate (0.47) and permutation $p<0.001$ among all. This model also acted as the best-one for providing an insight about lung cancer risk.

The second part of the analysis identified the complex interaction among different genotypes and smoking as an environmental parameter. This analysis demonstrated that the best interaction model was the two factor model including *AhR* variant **(rs2066853)** and **Smoking**. This is the best model because it has a maximum CVC value (10/10), minimum prediction rate (0.42) and permutation *p*-value<0.0001 among all other interaction models.

The entropy dendrogram in Fig. (**S2**) demonstrates the interactions of these SNPs and smoking and their contribution in lung cancer predisposition. Also, the shorter the length connecting the two, the strength of synergy increases hence the *AhR* rs2066853 and smoking synergistically contribute to the maximum in modulating lung cancer susceptibility.

3.7. CART Analysis

CART analysis was carried out in this study to analyse the high-order non parametric interactions between the *AhR* variants. This method utilized binary recursive partitioning approach. Fig. (**1**) depicts the decision tree obtained from this data mining method. A total of seven terminal nodes were found. The terminal node having the lowest case rate (36.07) was taken as the reference to calculate the odds ratio and 95% C.I. for the other terminal node. The terminal node7 having the genotype $AhR_{rs10250822}$ (W)/$AhR_{rs7811989}$ (W) is taken as reference. The data in Table **S5** shows that subjects having the genotype $AhR_{rs2066853}$(W)/$AhR_{rs7811989}$(M) harboured 2.2 fold increased risk of developing lung cancer (OR=2.26, 95% C.I.:1.50-3.41, ***p*=0.00009**). Another terminal node 5 having the genotypic combination of $AhR_{rs2282885}$(W)/$AhR_{rs10250822}$(M)/ $AhR_{rs7811989}$(M) also exhibited a two-fold increased risk of lung cancer (OR=2.10, 95% C.I.:1.22-3.63, ***p*=0.0073**).

3.8. Association of *AhR* Polymorphism and Overall Survival in Lung Cancer Patients and also on Histological Sub-types

In the current study we also analysed the role of the *AhR* SNPs and its relation with OS of lung cancer patients as shown in Table **4**. The survival analysis was carried out for 150 lung cancer patients and the survival time was estimated by accounting the number of days from being diagnosed till follow up which was for a duration of three years. Our data showed that 118 patients were dead amid follow up period and 32 were alive. In the univariate analysis it was shown that patients with wild genotype (GG) for *AhR* rs2066853 polymorphic site had a MST of 7.3 months and those harbouring the mutant genotype (AA) had the least MST of 3.53

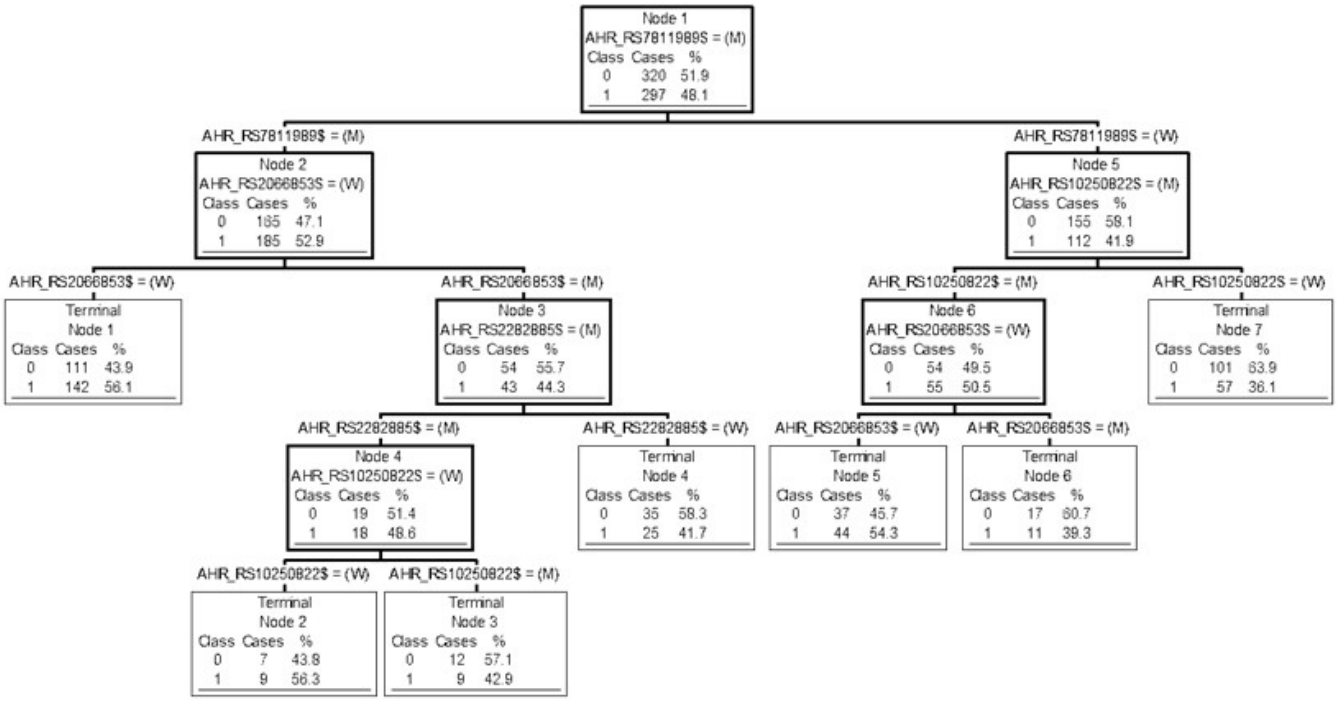

Fig. (1). CART analysis for AhR variants. W=homozygous wild type genotype, M=heterozygous + homozygous variant genotype; 0:controls, 1:cases.

Table 4. Univariate and multivariate analysis for four AhR variants.

AhR Variants	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 T>C	-							
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 T>C	-							
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 G>A	-							
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		2.56	1.68(1.09-2.59)	**0.017**
AhR rs7811989 A>G	-							
AA	57(38.0)	45(38.13)	12(37.5)	7.56	0.92	1	1	-
AG	78(52.0)	61(51.69)	17(53.1)	6.73		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.

months (H.R=2.56; Log rank p=0.013), as shown in Fig. (**2A**). However, in the multivariate analysis using Cox regression model after adjusting for different confounding predictors like histology, age, gender, smoking status, ECOG and KPS. It was observed that lung cancer patients with mutant genotype had a poor prognosis (HR=1.68, 95%CI=1.09-2.59; *p*=0.017).

Furthermore, we also stratified the OS of subjects based upon histological subtype as shown in Table **S6**. It was reported that for *AhR* rs2066853 ADCC patients with heterozygous (GA) genotype were found to have a higher MST of 10.1 months. On the contrary, lung cancer subjects carrying both the variant alleles for rs2066853 had the least MST of 8.3 months. Similarly, it was observed that SQCC patients with a mutant genotype (AA) had a lowest MST of 2.70 (HR=0.24, Log rank ***p*=0.001**) months in comparison to wild type genotype (GG) (MST=10.1) suggesting a highly significant protective effect in SQCC patients with mutant genotype as shown in Fig. (**2B**). However, multivariate Cox proportional hazards regression analysis when performed revealed an increase in death rate (HR=2.08, 95%CI=1.20-3.61; ***p*=0.008**) in SQCC subjects with mutant genotype.

4. DISCUSSION

Lung malignancy has developed as one of the significant causes of cancer death worldwide. It is a multifactorial disease which manifests due to environmental and genetic factors. Certain variation in the genome and metabolic pathways leads to alteration in the detoxification and metabolism of contaminants which demonstrates its role in the etiology in that disease. The process of metabolism of carcinogens present in the cigarette smoke largely governs the onset of lung tumorigenesis. The biological effect of these carcinogens is exerted by the interaction of these molecules with the receptor of the cytosol namely *AhR* (Aryl hydrocarbon receptor). AhR has been found to target various cellular processes on its activation, which includes cell division, loss of cellular adhesion, formation of DNA adducts, *etc*. These metabolic regulations are directly involved in the process of smoke induced lung carcinoma [27]. Hence, AhR does mediate the genetic and molecular abnormalities taking place during lung carcinogenesis. It is not only involved in the activation of phase I cytochrome P450 but also regulates the other pathways such as NF-κB induced inflammation which is highly expressed in lung cancer [28]. The sequence variations confer a considerable effect on the protein structure and

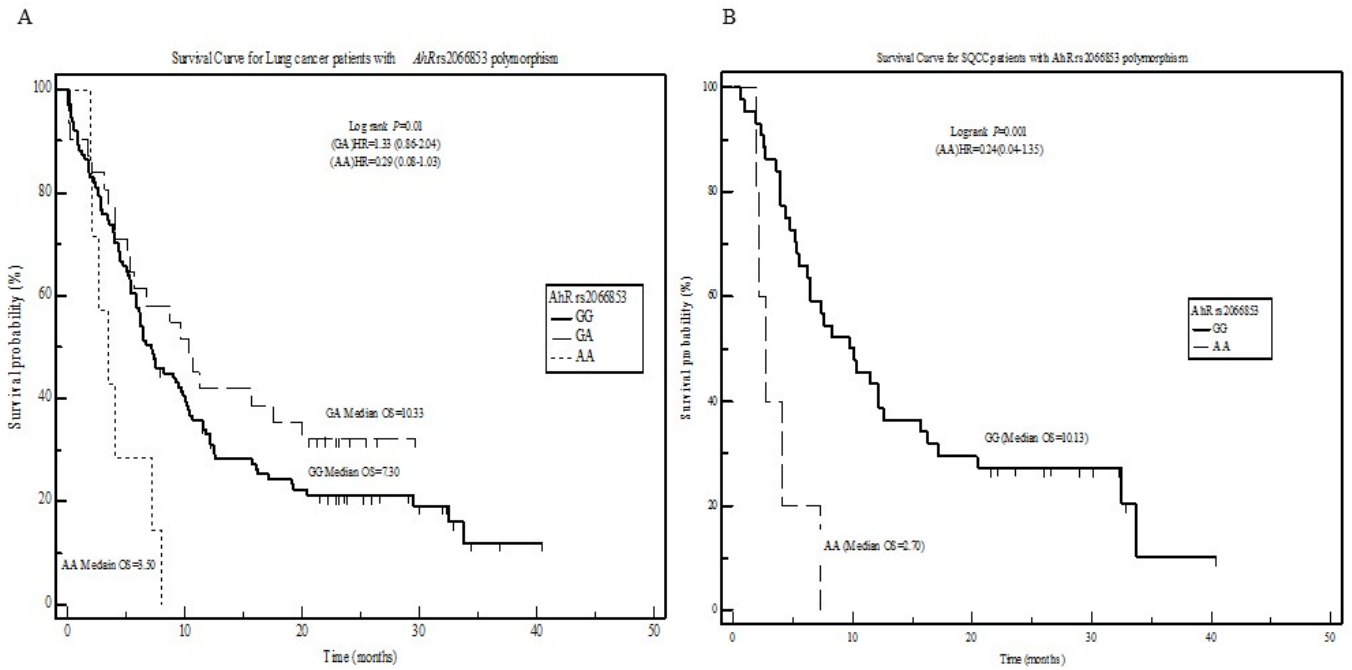

Fig. (2). Kaplan Meier Curves showing Overall survival; **A**: Survival curve for lung cancer patients with AhR rs2066853 polymorphism; **B**: Survival curve for SQCC patients with AhR rs2066853 polymorphism.

function. This is the first study which traces the role of four different genetic polymorphisms towards lung cancer susceptibility in North Indian population.

The current study revealed an increased risk of lung cancer in North Indian population in subjects carrying the variant (GG) genotype for *AhR* rs7811989 (***p*=0.007**) and also a marginal risk in case of individuals carrying either single or double copy of susceptible allele for rs102550822 (***p*=0.02**). On the other hand, the variant allele for rs2066853 showcased a strong protective effect towards lung cancer (***p*=0.003**). However, our study reported the lack of association of rs2282885 with lung cancer risk. A study evaluated the association of *AhR* polymorphism and levels of hydroxypyrene in urine and reported an increased amount of hydroxypyrene in the urine of coke exposed workers, which is a 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 related to the association of urinary 1-OHP which demonstrated that AhR signaling may partake in control of interceded PAH-metabolic activation and contribute to susceptibility to PAH exposure. In conclusion, the alteration in PAH metabolic pathway may interact with the environmental exposure and contribute towards tumorigenesis [22]. The findings in regard to rs78911989 in the present study were in accordance to the study conducted in Chinese population where SNP **rs7811989** along with **rs2158041** both residing in the intronic regions were reported to be associated with higher risk of lung cancer [29].

The present study also explored detailed dimensions of the role of these polymorphisms in regard to smoking and histology of lung tumor. The highlights of these findings were, the association of rs7811989 mutant genotype and rs10250822 mutant genotype with lung cancer especially in smokers as compared to non-smokers. However, in case of rs2066853 a decreased risk was observed in smokers with mutant genotype (p=0.002).

AhR **rs2066853** being nonsynonymous is thought to play a vital role in the area of proteins crucial to enzyme activity. Earlier studies conducted in Korean [30], Japanese [17], French [18, 19] and Finnish [20, 21] population showed where no risk was associated in regard to $Arg^{554}Lys$ polymorphism. Their results were not concordant with our analysis, in which smokers patients did not display any risk towards the disease whereas in non-smokers, protective effect was seen. Conflicting results were observed in Caucasian population study which revealed that mutant genotype displayed an increase in CYP1A1 activity in women smokers [31]. On the contrary, our findings were in sync with the study conducted in Chinese breast cancer patients where AA genotype in females conferred a protective effect towards breast cancer similar to those in our study [32]. However, a study done in non-smokers exhibited an increased CYP1A1 enzyme which was determined by ethoxyresorufin-O-deethylase assay in peripheral blood lymphocytes [33]. This study supports the findings of the present study where higher odds-ratio was observed in case of non-smokers having genotype for this SNP. As evident, there has been contradicting prediction about the functional effect of the codon 554 SNP. As this nucleotide change is a conservative replacement, therefore it might be possible that there exists no functional variability due to this polymorphism [34]. Meta - analysis study conducted recently on *AhR* rs2066853 polymorphism was also non-conclusive about the clear association of this genetic variation with different types of human cancer [35]. Previous study done in Chinese population suggests that an increase in the pack years of smokers simulta-

neously 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 [29]. Our study in the North Indian population falls very much in line with the former which also holds good for *AhR* **rs2066853** and proves a significant association of cumulative cigarette smoking on the susceptibility towards lung cancer. In our study, we observed that the individuals diagnosed with SQCC showed statistically significant values, which confirms the findings reported by the study done in Chinese population [29].

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 [3]. It was seen that if the individual was exposed to more smoke inbuilt possessed increased capacity to detoxify the inhaled carcinogen, leading to enhanced CYP1A2 activity [36]. The possible explanation of the various findings regarding the ***AhR* rs7811989** 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 [37]. For example 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 [38]. Another study reports the association of *AhR* rs2282885 with the inducibility of CYP1A2 gene, which validates the involvement of *AhR* mediated pathway and also a higher risk towards lung cancer [36]. 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 supports 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 [39]. As of now, no vivid studies on rs2282885 have 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 Tetrachlorodebenzo-p dioxin (TCDD) induced carcinogenesis. Polymorphism in **rs10250822** does not showcase an association towards the risk of acquiring infertility in males as was stated in Iranian population [40]. This SNP continues to be unexplored vividly by researchers and thus we do not have enough instances to validate our work with.

We have also analyzed the haplotype and linkage disequilibrium in this study where a strong linkage disequilbrium was observed in between rs7811989 & rs2282885 and rs7811989 & rs2066853. As it has been recently said that, there are other polymorphism within the *AhR* gene along with *AhR* Rs2066853 which have a 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 [41]. Another study done in Chinese lung cancer patients also calculated the linkage disequilibrium between the different genetic variants within *AhR* gene [29].

Considering the above mentioned facts and the effect of the interaction between various SNP's and environmental parameters such as smoking, MDR approach gave the best interaction model comprising *AhR* rs2066853 and smoking (CVC=10/10, prediction error=0.42), which contribute maximum to the arena of lung cancer predisposition in North Indians. We also evaluated the high order SNP interaction using CART where $AhR_{rs2066853}$(W)/$AhR_{rs7811989}$(M) illustrated 2.2 fold increased risk of developing lung cancer (OR=2.26, 95% C.I.:1.50-3.41, ***p*=0.00009**) which was the highest among all combinations. This is probably the first attempt wherein the interactions of the *AhR* variants and other environmental factors have been analysed using MDR and CART.

Being a major player in the detoxification process AhR protein along with other downstream genes has been recently explored for its interesting contribution in prognosis of cancer patients [42]. Taking this into account, we also analyzed the association of these polymorphisms with overall survival and prognosis of lung cancer patients. Our data suggests that the patients having mutant genotype for $Arg^{554}Lys$ showcased increase in the death rate when multivariate Cox hazardous proportional ratio was used. Similar study done on American population, demonstrated that $Arg^{554}Lys$ polymorphism elevates the CYP1A1, CYP1A2 activity which brings change in activation of gene expression. The heterozygote genotype displayed risk for soft tissue sarcoma [41]. Another study conducted in breast cancer females also evaluated the role of *AhR* polymorphism in predicting the death rate, however no correlation was observed in this case [43].

CONCLUSION

Some interesting conclusions were drawn from the current study which can help in establishing the role of *AhR* variants in modulating lung cancer predisposition in North Indians. Certain limitations of this study include a smaller sample size in various subgroups and also more detailed analysis to find out the relevance of *AhR* variants in developing cigarette smoke induced lung cancer. Further studies with larger sample size are required to validate these findings and pave a way for using *AhR* variants a predictors in lung cancer susceptibility.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study was approved by the Ethics committee of the Post Graduate Institute of Medical Education and Research (PGIMER), Chandigarh, India.

HUMAN AND ANIMAL RIGHTS

No animals were used in this study. The reported experiments/study were in accordance with the Helsinki Declaration of 1975, as revised in 2008 (http://www.wma.net/en/20activities/10ethics/10helsinki/).

CONSENT FOR PUBLICATION

Informed consent was obtained from all enrolled patients or their representatives.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

We would like to express our gratitude to all the subjects who participated in this current study and the work was supported by grant from the Indian Council of Medical Research, New Delhi, India. (Grant No. 5/13/126/2011/NCD-III).

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's website along with the published article.

REFERENCES

[1] Yokota, J.; Shiraishi, K.; Kohno, T. Genetic basis for susceptibility to lung cancer: Recent progress and future directions. *Adv. Cancer Res*., **2010**, *109*, 51-72. Available from: https://www.sciencedirect.com/science/article/pii/B978012380890500028?via%3Dihub

[2] Tsay, J.J.; Tchou-Wong, K.M.; Greenberg, A.K.; Pass, H.; Rom, W.N. Aryl hydrocarbon receptor and lung cancer. *Anticancer Res*., **2013**, *33*(4), 1247-1256.

[3] Nebert, D.W.; Dalton, T.P.; Okey, A.B.; Gonzalez, F.J. Role of Aryl hydrocarbon receptor-mediated induction of the CYP1 enzymes. *J. Biol. Chem*., **2004**, *279*(23), 23847-23850.

[4] Xue, W.; Warshawsky, D. Metabolic activation of polycyclic and heterocyclic aromatic hydrocarbons and DNA damage: A review. *Toxicol. Appl. Pharmacol*., **2005**, *206*(1), 73-93.

[5] Shimizu, Y.; Nakatsuru, Y.; Ichinose, M.; Takahashi, Y.; Kume, H.; Mimura, J.; Ishikawa, T. Benzo [a] pyrene carcinogenicity is lost in mice lacking the aryl hydrocarbon receptor. *Proc. Natl. Acad. Sci. U.S.A*., **2000**, *97*(2), 779-782.

[6] Schmidt, J.V.; Su, G.H.; Reddy, J.K.; Simon, M.C.; Bradfield, C.A. Characterization of a murine Ahr null allele: Involvement of the Ah receptor in hepatic growth and development. *Proc. Natl. Acad. Sci. U.S.A*., **1996**, *93*(13), 6731-6736.

[7] Shimada, T. Xenobiotic-metabolizing enzymes involved in activation and detoxification of carcinogenic polycyclic aromatic hydrocarbons. *Drug Metab. Pharmacokinet*., **2006**, *21*(4), 257-276.

[8] Petrulis, J.R.; Perdew, G.H. The role of chaperone proteins in the aryl hydrocarbon receptor core complex. *Chem. Biol. Interact*., **2002**, *141*(1), 25-40.

[9] Morrow, D.; Qin, C.; Smith, R.; Safe, S. Aryl hydrocarbon receptor-mediated inhibition of LNCaP prostate cancer cell growth and hormone-induced transactivation. *J. Steroid Biochem. Mol. Biol*., **2004**, *88*(1), 27-36.

[10] Dzeletovic, N.; McGuire, J.; Daujat, M.; Tholander, J.; Ema, M.; Fujii-Kuriyama, Y.; Poellinger, L. Regulation of dioxin receptor function by omeprazole. *J. Biol. Chem.,* **1997**, *272*(19), 12705-12713.

[11] Ding, X.; Kaminsky, L.S. Human extrahepatic cytochromes p450: Function in xenobiotic metabolism and tissue-selective chemical toxicity in the respiratory and gastrointestinal tracts. *Annu. Rev. Pharmacol. Toxicol.,* **2003**, *43*(1), 149-173.

[12] Sachse, C.; Brockmöller, J.; Bauer, S.; Roots, I. Functional significance of a C→ A polymorphism in intron 1 of the cytochrome P450 CYP1A2 gene tested with caffeine. *Br. J. Clin. Pharmacol*., **1999**, *47*(4), 445-449.

[13] Nakajima, M.; Yokoi, T.; Mizutani, M.; Kinoshitah, M.; Funayama, M.; Kamataki, T. Genetic polymorphism in the 5′-flanking region of human CYP1A2 gene: Effect on the CYP1A2 inducibility in humans. *J. Biochem*., **1999**, *125*(4), 803-808.

[14] Lin, P.; Chang, H.; Tsai, W.T.; Wu, M.H.; Liao, Y.S.; Chen, J.T.; Su, J.M. Overexpression of aryl hydrocarbon receptor in human lung carcinomas. *Toxicol. Pathol*., **2003**, *31*(1), 22-30.

[15] Marlowe, J.L.; Fan, Y.; Chang, X. The Aryl hydrocarbon receptor binds to E2F1 and inhibits E2F1-induced apoptosis. *Mol. Biol. Cell*, **2008**, *19*(8), 3263-3271.

[16] Jiang, T.; Bell, D.R.; Clode, S.; Fan, M.Q.; Fernandes, A.; Foster, P.M.; Tran, L. A truncation in the aryl hydrocarbon receptor of the CRL: WI (Han) rat does not affect the developmental toxicity of TCDD. *Toxicol. Sci*., **2009**, *107*(2), 512-521.

[17] Kawajiri, K.; Watanabe, J.; Eguchi, H.; Nakachi, K.; Kiyohara, C.; Hayashi, S. Polymorphisms of human Ahr receptor gene are not involved in lung cancer. *Pharmacogenet*., **1995**, *5*(3),151-158.

[18] Cauchi, S.; Stucker, I.; Cenee, S.; Kremers, P.; Beaune, P.; Massaad-Massade, L. Structure and polymorphisms of human aryl hydrocarbon receptor repressor (AhRR) gene in a French population: Relationship with CYP1A1 inducibility and lung cancer. *Pharmacogenet*., **2003**, *13*(6), 339-347.

[19] Cauchi, S.; Stucker, I.; Solas, C.; Laurent-Puig, P.; Cenee, S.; Hemon, D.; Jacquet, M.; Kremers, P.; Beaune, P.; Massaad-Massade, L. Polymorphisms of human aryl hydrocarbon receptor (AhR) gene in a French population: Relationship with CYP1A1 inducibility and lung cancer. *Carcinogenesis*, **2001**, *22*(11), 1819-1824.

[20] Anttila, S.; Lei, X.D.; Elovaara, E.; Karjalainen, A.; Sun, W.; Vainio, H.; Hankinson, O. An uncommon phenotype of poor inducibility of CYP1A1 in human lung is not ascribable to polymorphisms in the AHR, ARNT, or CYP1A1 genes. *Pharmacogenet*., **2000**, *10*(8), 741-775.

[21] Anttila, S.; Tuominen, P.; Hirvonen, A.; Nurminen, M.; Karjalainen, A.; Hankinson, O.; Elovaara, E. CYP1A1 levels in lung tissue of tobacco smokers and polymorphisms of CYP1A1 and aromatic hydrocarbon receptor. *Pharmacogenet*., **2001**, *11*(6), 501-509.

[22] Bin, P.; Leng, S.; Cheng, J.; Dai, Y.; Huang, C.; Pan, Z.; Chen, W. Association of Aryl hydrocarbon receptor gene polymorphisms and urinary 1-hydroxypyrene in polycyclic aromatic hydrocarbon-exposed workers. *Cancer Epidemiol. Biomarkers Prev*., **2008**, *17*(7), 1702-1708.

[23] Sodhi, K.K.; Bahl, C.; Singh, N.; Behera, D.; Sharma, S. Functional genetic variants in pre-miR-146a and 196a2 genes are associated with risk of lung cancer in North Indians. *Future Oncol*., **2015**, *11*(15), 2159-2173.

[24] Shi, Y.Y.; He, L. SHEsis, a powerful software platform for analyses of linkage disequilibrium, haplotype construction, and genetic association at polymorphism loci. *Cell Res*., **2005**, *15*(2), 97-98.

[25] Ritchie, M.D.; Motsinger, A.A. Multifactor dimensionality reduction for detecting gene - gene and gene - environment interactions in pharmacogenomics studies. *Future Med*., **2005**, *6*(8), 823-834.

[26] Srivastava, A.; Sharma, K.L.; Srivastava, N.; Misra, S.; Mittal, B. Significant role of estrogen and progesterone receptor sequence variants in gallbladder cancer predisposition: A multi-analytical strategy. *PLoS One*, **2012**, *7*, e40162. Available from: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0040162

[27] Tsay, J.J.; Tchou-Wong, K.M.; Greenberg, A.K.; Pass, H.; Rom, W.N. Aryl hydrocarbon receptor and lung cancer. *Anticancer Res*., **2013**, *33*(4), 1247-1256.

[28] Hayden, M.S.; Ghosh, S. Signaling to NF-kappaB. *Genes Dev*., **2004**, *18*(18), 2195-2224.

[29] Chen, D.; Tian, T.; Wang, H.; Liu, H.; Hu, Z.; Wang, Y.; Sun, W. Association of human aryl hydrocarbon receptor gene polymorphisms with risk of lung cancer among cigarette smokers in a Chinese population. *Pharmacogenet. Genomics*, **2009**, *19*(1), 25-34.

[30] Kim, J.H.; Kim, H.; Lee, K.Y.; Kang, J.W.; Lee, K.H.; Park, S.Y. Aryl hydrocarbon receptor gene polymorphisms affect lung cancer risk. *Lung Cancer,* **2007**, *56*(1), 9-15.

[31] Harper P.A.; Wong, J.Y.; Lam, M.M. Polymorphisms in the human AH receptor. *Chem. Biol. Interact*., **2002**, *141*(1-2), 161-187.

[32] Long, J.R.; Egan, K.M.; Dunning, L.; Shu, X.O.; Cai, Q.; Cai, H.; Dai, Q.; Holtzman, J.; Gao, Y.T.; Zheng, W. Population-based case-control study of AhR (aryl hydrocarbon receptor) and CYP1A2 polymorphisms and breast cancer risk. *Pharmacogenet. Genomics*, **2006**, *16*(4), 237-243.

[33] Smart, J.; Daly, A.K. Variation in induced CYP1A1 levels: Relationship to CYP1A1, AhR receptor and GSTM1 polymorphisms. *Pharmacogenet*., **2000**, *10*(1), 11-24.

[34] Daly, A.K.; Fairbrother, K.S.; Smart, J. Recent advances in understanding the molecular basis of polymorphisms in genes encoding cytochrome P450 enzymes. *Toxicol. Lett.,* **1998**, *102-103*, 143-147. Available from: https://www.sciencedirect.com/science/article/pii/S0378427498002999

[35] Luo, C.; Zou, P.; Ji, G.; Gu, A.; Zhao, P.; Zhao, C. The aryl hydrocarbon receptor (AhR) 1661G > A polymorphism in human cancer: A meta-analysis. *Gene*, **2013,** *513*(1), 225-230.

[36] Dobrinas, M.; Cornuz, J.; Eap, C.B. Pharmacogenetics of CYP1A2 activity and inducibility in smokers and exsmokers. *Pharmacogenet. Genomics*, **2013**, *23*(5), 286-292.

[37] Hirose, Y.; Chiba, K.; Karasugi, T. A functional polymorphism in THBS2 that affects alternative splicing and MMP binding is associated with lumbar-disc herniation. *Am. J. Hum. Genet*., **2008**, *82*(5), 1122-1129.

[38] Pohjanvirta, R.; Tuomisto, J. Short-term toxicity of 2, 3, 7, 8-tetrachlorodibenzo-p-dioxin in laboratory animals: effects, mechanisms, and animal models. *Pharmacol. Rev*., **1999**, *46*(4), 483-549.

[39] Izawa, H.; Kohara, M.; Watanabe, G.; Taya, K.; Sagai, M. Effects of diesel exhaust particles on the male reproductive system in strains of mice with different aryl hydrocarbon receptor responsiveness. *J. Reprod. Dev.,* **2007**, *53*(6), 1191-1197.

[40] Safarinejad, M.R.; Shafiei, N.; Safarinejad, S. Infertility polymorphisms in Aryl hydrocarbon receptor gene are associated with idiopathic male factor. *Reprod. Sci*., **2013**, *20*(12), 1423-1432.

[41] Berwick, M.; Matullo, G.; Song, Y.S. Association between aryl hydrocarbon receptor genotype and survival in soft tissue sarcoma. *J. Clin. Oncol*., **2004**, *22*(19), 3997-4001.

[42] Swanson, H.I.; Bradfield, C.A. The AHR receptor: Genetics, structure and function. *Pharmacogenet*., **1993**, *3*(5), 213-230.

[43] Long, J.R.; Cai, Q.; Shu, X.O.; Cai, H.; Gao, Y.T.; Zheng, W. Genetic polymorphisms in estrogen-metabolizing genes and breast cancer survival. *Pharmacogenet. Genomics*, **2007**, *17*(5), 331-338.