Results and Discussion

Age-related cataract (ARCs) is a complex trait, that generally develops after fourth decade of life leading to blindness worldwide. This multi-factorial disorder involves several genetic and environmental risk factors which may be a component cause, or part of a sufficient cause. So far several candidate genes have been implicated in the development of cataract and a few genes which are linked to inherited cataracts are also associated with the development of age-related cataracts. Besides genetic factors ageing, gender, familial history, habits like smoking and alcoholism, exposure to UV, diabetes act as major risk factors for ARCs. During the past decade, focus on these factors has progressed and provided insights to certain extent about the mechanisms of cataractogenesis. Yet, aetiopathogenesis of age-related cataracts still remain incompletely understood. Therefore there is a need for integrated multidisciplinary research to explore the pathophysiology of this complex, multi-factorial disorder.

In the present study an attempt was made to evaluate the contribution of epidemiological factors like sex, age, age-at-onset, obesity, habit of smoking, alcohol consumption and family history to the development of ARCs. A total of 680 age-related cataract cases including four different types and 210 healthy control subjects were investigated for the association with various epidemiological parameters.

5.1 DESCRIPTION OF THE DATA

Of different types of cataracts studied, nuclear cataract cases (NC: 29.3%) were more frequent in occurrence as compared to cortical (CC: 25.7%), posterior subcapsular (PSC: 25.6%) and mixed types (MT: 19.4%). Sex-wise distribution of different types of cataracts showed high frequency of NC (30.5%) in males followed by CC (27.7%), PSC (23.2%) and MT (18.6%) types while in females the frequency of NC cases was 28.2% followed by PSC (27.6%), CC (24.1%) and MT (20.1%) cataracts (Table-1a; Graph-1). This shows that the prevalence of NC and CC types of cataracts were low in females while that of PSC was high as compared to males, suggesting higher risk for the development of PSC in females as compared to other types. This indicates that females are more prone to PSC while males are more prone to NC and CC types of cataracts.
Based on the ages of the patients and controls recorded at the time of investigation, 4 categories were formed like young (<30yrs), middle (30-49yrs), elder (50-69yrs) and older (≥70yrs) groups (Table-1b). As per this grouping, it was observed that 67.1% of the cataract patients belonged to elder age group (50-69 years) followed by older (18.5%), middle (13.2%) and younger (1.2%) age groups. This trend was found to be similar among patients with different types of cataracts except for PSC cases where the occurrence was more in middle age group (18.4%) than the older age group (14.4%). It was also observed that in different types of cataracts the frequency of patients was high when the ages were ≥50 (NC-93.4%; CC-84.5%; PSC-79.3% and MT-83.3%) as compared to those with <50 years of age (PSC-20.7%; MT-16.7%; CC -15.4%; NC -6.6%) which suggests that cataract in general is an age-related condition. It was also observed that high proportion of cases occurring after the age of ≥50 years was seen in cases of NC (93.4%) as compared to other cataracts (CC-84.5%; MT-83.3%; PSC-79.3%) while young and middle age patients (age <50 yrs) were seen with high frequency among cases of PSC (20.7%) as compared to other types (MT-16.7%; CC -15.4%; NC -6.6%). This indicates that onset of cataract is early in PSC while it is late in NC types (Table 1b; Graph-2).

The frequency distribution of age at onset of ARCs with 5 classes each with 10 years interval, showed 39.9% of cases with an onset of 50-59yrs followed by 60-69yrs (31.2%), 40-49yrs (13.7%), ≥70yrs (8.7%) and 30-39 years (6.6%). Different types of cataracts also showed high frequency of age at onset in the age group of 50-59 years (NC: 38.7%; CC: 40.6%; PSC: 37.9%; MT: 43.2%) followed by other age at onset groups (Table 1c). In the age at onset occurring between 40-49yrs, the frequency of PSC (25.3%) was higher as compared to CC (13.1%), MT (10.6%) and NC (6.0%). This again supports the observation that NCs occur at later age while PSCs occur at earlier age. Based on the modal class (50-59 years), of the age at onset of all the types of cataracts, the patients were further grouped as those with early onset (<50 years) and those with late onset (≥50 years). This distribution also supports the observation that the frequency of early onset cases was more in PSC (32.8%) followed by CC (21.1%), MT (18.9%) and NC (9.5%) and incidence of NC cases were more in late onset group (90.5%) as compared to other cataracts (Table
Results and Discussion

1c; Graph-3). Significant difference was found in the frequency of different types of cataract cases with early onset compared to late onset cases ($\chi^2$: 31.14; p: 7.9e-7).

Based on the body mass indices, the subjects were classified into underweight (<18.5), normal weight (18.5-24.9), overweight (25.0-29.9) and obese (>30) as per WHO (1998) classification. As the frequency of obese were very less among the subjects both overweight and obese groups were considered together as cases with high BMI (≥25.0) and underweight and normal weight as with low BMI groups (<25.0) for further computations (Table-1d; Graph-4).

5.2 BASELINE CHARACTERISTIC FEATURES IN CATARACT PATIENTS AND CONTROLS STUDIED

Table-1(Graph-5) depicts general features found in patients and control groups. It showed high frequency of female cases (54.3%) as compared to males (45.7%). The mean±SD of age at investigation, age-at-onset of cataracts and BMI in the patients were 58.66±0.40; 57.45±0.40 and 22.5±0.11 respectively. The frequency of early onset (<50 years of age) cases was 20.3% and of late onset cases (≥50 years of age) was 79.7% indicating that nearly 1/5 of cases were found to be with early onset of cataract. The frequency of patients with low BMI <25.0 (83.9%) was high as compared to patients with high BMI ≥25.5 (16.1%) indicating higher risk for cataract development in subjects with low BMI. About 14.4% of cases were reported to have positive family history (FH) of cataract. For the habits like smoking and alcohol consumption only male subjects were considered because none of the female subjects were reported to be smokers. Of the male patients 49.2% were smokers and 36.7% were alcoholics. Considering the dietary pattern 93.1% of patients reported as non-vegetarians.

210 non-cataractous healthy normal individuals were studied as controls for comparison with disease group. Out of 210 cases 62.9% were males and 37.1% were females. As healthy normal females with an age above 50 were not willing to participate in the study, the frequency of females recorded with higher age group was low in the control group as compared to patients. The mean±SD for age at investigation and BMI of control subjects were 49.75±0.5 and 23.4±0.23 respectively. About 29.0% showed high BMI and 71.0%
were with low BMI. About 19.5% showed positive family history (FH) of cataracts and 74.8% were non-vegetarians. Among the male controls 38.6% were smokers and 47.7% were alcoholics.

Comparison of patients and controls showed that there was significantly high frequency of females among patients as compared to controls ($\chi^2$: 18.82; p: 0.00002). A significant difference was observed between the patients and controls regarding BMI values ($\chi^2$: 17.60; p: 0.00002) the low BMI values being high in frequency among patients (83.9%) as compared to controls (71.0%). The mean body mass index was also significantly low in patients as compared to controls (patients 22.5±0.11, controls 23.4±0.11; p≤0.0001). The frequency of smokers was significantly high among male cases as compared to male controls ($\chi^2$: 4.16; p: 0.04). The difference in the frequency of vegetarians and non-vegetarians was statistically significant between patients and controls ($\chi^2$: 54.03; p: 0.00).

The above observations suggest that females and subjects above 50 years of age, low BMI, non-vegetarian dietary habits were at greater risk for the development of age-related cataracts. Among the males the smokers showed higher risk of developing ARCs as compared to non smokers. These findings were also supported by several epidemiological studies. (Chatterjee, 1982; Framingham Eye Study, 1994; Heiba et al, 1995; Hammond et al, 1999; McCarty et al, 1999; AREDS Report No.5 2001; Cheng et al, 2000; Tsai et al, 2003; Krishnaiah et al, 2005; Kuang et al, 2005; Xu et al, 2006; Bhagyalaxmi, 2009; West, 2010; Richter et al, 2012; Duan et al, 2013).

5.3 EPIDEMIOLOGICAL RISK FACTORS ASSOCIATED WITH ARCs

The data on epidemiological factors were stratified into different groups to study the correlation between them and different types of ARCs. The results found are presented below.

a) Sex

An attempt has been made to evaluate gender specific differences in patients and controls regarding the cohorts planned to assess whether sex as such and/or along with confounding factors act as risk factor for ARC (Table-2).
The distribution of four age groups between the two sexes showed high frequency of patients (males: 66.2%; females: 67.8%) in the elder age group as compared to other age groups. In females the frequency of middle (15.4%) and younger age groups (1.4%) were high as compared to male patients (middle: 10.6%; young: 1.0%) while the frequency of older age group (15.4%) was lesser in female patients as compared to males (22.2%) indicating that cataracts may affect females earlier than males. Among controls there was a high frequency of subjects with middle age group (males: 53.8%; females: 51.3%) followed by elder (males: 44.7%; females: 46.2%) and older age groups (males: 1.5%; females: 2.6%) in both the sexes. The differences in the frequencies of middle and elder age groups between patients and controls were statistically significant ($\chi^2$: 14.82; p: 0.0001 and $\chi^2$: 9.02; p: 0.002 respectively). Significant differences were found in the frequencies of different age groups when male ($\chi^2$: 102.27; p: 0.00) and female patients ($\chi^2$: 48.1; p: 0.00) were compared with their corresponding controls (Table-2). This indicates that cataracts develop earlier among females as compared to males.

Considering the frequency of early onset and late onset cases in the two sexes, there was a high frequency of female patients with early onset of cataract (23.8%) as compared to males (16.1%) with corresponding decrease in late onset cases (females-76.2%; males-83.9%). This gender-wise difference found in the frequency of early onset and late onset patients was statistically significant ($\chi^2$: 6.30; p: 0.01; Table-2). This once again indicates early onset of cataract among females as compared to males.

Considering BMI the frequency of the 4 groups with underweight, normal weight, over weight and obese didn’t differ significantly in both males and female patients. When patients with underweight and normal weight were compared with their respective controls, it showed significant difference ($\chi^2$:4.41; p: 0.03 and $\chi^2$: 17.02; p: 0.00). Significant results were found when female patients were compared to female controls ($\chi^2$: 14.8; p: 0.001) regarding BMI groups.

When patients and controls were grouped into 2 broader categories as those with low BMI (< 25.0 that includes under weight and normal weight subjects) and high BMI (≥ 25.0 that includes overweight and obese subjects), a highly significant difference was found between
cases and controls with low BMI ($\chi^2$: 20.38; p: 0.00). This was mainly because of high frequency of females (85.1%) with low BMI as compared to male patients (82.6%) and also female controls (65.4%; Table-2) indicating high risk for low BMI group for developing ARCs.

There was no difference in the incidence of family history of cataracts between male and female patients.

There was a high frequency of male smokers (49.2%) as compared to male controls (38.6%) and the difference was found to be statistically significant ($\chi^2$: 4.15; p: 0.04; Table-2), suggesting smoking as a risk factor for the development of cataracts.

Among cases there was a high frequency of non-vegetarians among both male (92.6%) and female patients (93.5%) as compared to vegetarians (males: 7.4%; females: 6.5%) and non-vegetarian controls (male:73.5%; female: 76.9%). The gender based difference was highly significant when patients with non-vegetarian food ($\chi^2$: 13.35; p<0.00) habits were compared to corresponding controls. Similarly, male ($\chi^2$: 29.7; p<0.00) and female ($\chi^2$: 20.7; p<0.00) patients with vegetarian and non-vegetarian food habits showed significant difference when compared to corresponding controls (Table-2).

In conclusion the major points that can be focused from gender wise comparison were high risk of cataract development in females in general and females below 50 years of age (early onset), patients with low BMI (<25.0) and non-vegetarian food habits. In contrast, males were at risk when they were above 50 years of age with the habit of smoking and non-vegetarian food habit. The high risk for females found in the present study is also supported by the observations made by Leske et al, (1999); Cheng et al, (2000); Tsai et al, (2003); Krishnaiah et al, (2005); Bhagyalaxmi, (2009); Duan et al, (2013).

It is quite possible that higher prevalence of cataract in women (specially in high/elder age group) in the present study may be related to gender-based differences in the exposure to the environmental conditions and/or due to hormonal influences (estrogen levels) associated with menopause (Worzala et al, 2001). Another explanation given is the use of cheap cooking fuels (e.g., dried wood, twigs and sticks, leaves, cow dung), by most women
in rural India, which produces lot of smoke (particularly in ill-ventilated spaces) may be acting as an additional and cumulative source of oxidative damage affecting the eye vision. This explanation may not hold solely in modern India where there is remarkable development in rural areas at various levels especially in the use of cooking gas instead of fire wood. As described earlier, most women in Andhra Pradesh, (particularly in rural areas) are anaemic and have subnormal nutritional status which may be a confounding factor causing increased risk for cataract development in women (Shalini et al, 1994). Unexplained reasons may still be operating for the high incidence of cataracts in females. Since the proportion of females with cataracts are seen more in the age group of 50-69 years which correlates with the period soon after menopause, hormonal variations may be considered as a cause for high incidence of cataracts in females.

b) Age at Investigation

As described earlier, in the present study the subjects were grouped into 4 different age groups i.e. <30 years of age designated as young, 30-49 years as middle, 50-69 years as elder and ≥70 years as older age. Majority of the patients (67.1%) were found to be in the elderly age group of 50-69 years as compared to controls (45.2%; Table-1b) indicating the association of higher age with ARCs. The results on the association of age in concurrence with other epidemiological parameters in causing ARC are presented below (Table-3).

The frequency distribution of both sexes in different age groups showed a high frequency of male patients in older age group (54.8%) followed by elder (45.2%), young (37.5%) and middle (36.7%), while in controls high frequency of males was observed with middle age group (64.0%) followed by elder (62.1%) and older age groups (50.0%). The difference in the frequencies of age groups between the male patients and male controls was statistically significant ($\chi^2$: 102.2; p: 0.000; Table-3). Among females there was a high frequency of both middle (63.3%) and young (62.5%) patients followed by elder (54.8%) and older (45.2%) age groups, whereas in control females high frequency was observed in the older subjects (50.0%) followed by elder (37.9%) and middle (36.0%) age group. The difference in the distribution of age groups between female patients and female controls was also found to be statistically significant ($\chi^2$: 48.1; p: 0.000; Table-3).
Considering the distribution of different BMI levels, there was a high frequency of the patients and controls with normal body weight in all age groups as compared to underweight, overweight and obese patients. The BMI-based differences in the frequencies of different age groups were statistically significant when patients with underweight ($\chi^2$: 12.19; p≤0.006) normal weight ($\chi^2$: 100.7; p: 0.000) and over weight ($\chi^2$: 30.65; p: 0.000) were compared to corresponding controls (Table-3). Comparison of patients and controls with low (< 25.0) and high BMI (≥25.0) in different age groups showed significant difference in their frequencies (low BMI: $\chi^2$: 116.5; p: 0.00 and high BMI: $\chi^2$: 38.23; p: 0.00; Table-3). This difference might be due to the absence of controls in the younger age group and low frequency in older age group.

The frequency of smokers among the cases was higher in the elder and older age group (>50 years) as compared to controls which was statistically significant ($\chi^2$35.69; p: 2e-8).

Considering the frequency distribution of non-vegetarians in different age groups, significant difference was observed between the cases and controls with frequency being higher in the patient group ($\chi^2$:115.62; p:0.00).

In conclusion, among the four different age groups of patients, the risk in general was high for both sexes in elder age group (50-69 years of age), while it was high for female patients in middle age (≤ 49 years of age) and male patients in older age groups (≥ 70 years of age). The observation from the present data was in accordance with the study of Ughade et al, (1998) from Nagpur, India. The prevalence of cataract surgery among the patients in the age group of 40 years and above, were reported to be 13.7% in Hyderabad population from India (Dandona et al, 1999) indicating this age group conferring significant risk to ARC when compared to Melbourne and Australia population for the same age group. A survey from three districts in southern India conducted by “Aravind Comprehensive Eye Study” also showed increase in the prevalence of age related cataract significantly from the age of 40-49yrs to ≥70 years (Nirmalan et al, 2004). Data surveyed by Bhagyalaxmi, (2009) also supported the present results that showed more than 50% of the people developing cataract after the age of 50 years.
c) Early Onset and Late Onset of Cataracts

Age-related cataract as a multi-factorial condition is expected to have specific age-at-onset in the affected members of a given family. Hence information on the age-at-onset is expected to add to the understanding of genetic etiology underlying ARCs.

Based on the distribution of age-at-onset of cataracts in the patients covering 5 classes, each with class interval of 10 years, two groups were formed as cases with early onset (<50 years of age) and late onset (≥50 years of age; Table-1c). In general there was a high frequency of late onset cases (79.7%) as compared to early onset cases (20.3%; Table-1). The frequency of early onset cases appears to be substantial. The results on the association of different epidemiological parameters with early and late onset cases are presented below (Table-4).

The gender based differences between early and late onset cases was found to be statistically significant ($\chi^2 = 6.30; p≤0.01$), where there was a significantly high frequency of female patients with early onset (63.8%) as compared to male patients (36.2%) suggesting 1.8 times high risk in females for developing cataracts below 50 years. In contrast among late onset cases the difference was not significant between the frequencies among females (51.8%) and males (48.2%).

Considering other cohorts like BMI, family history and habits like smoking, alcohol consumption and non vegetarian dietary habits no significant association were found between the early and late onset groups.

In conclusion, females showed high risk of developing cataracts at an age of <50 years (early onset), while late onset patients in both the sexes with the age of ≥ 50 years showed risk for developing cataracts. This observation is not highlighted in studies conducted on ARCs so far.

d) Body Mass Index (BMI)

As described earlier, the subjects were classified into underweight (<18.5), healthy or normal weight (18.5-24.9), overweight (25.0-29.9) and obese (≥30) as per WHO (1998)
classification (Table-1d). It was further grouped into subjects with low BMI (< 25.0) and high BMI (≥ 25.0), as the frequency of obese and that of below normal BMI were very less. The results on the association of ARC with BMI in conjunction with other epidemiological parameters are presented below (Table-5).

There was a high frequency of female cases (55.0%) with low BMI (< 25.0) as compared to male cases (45.0%) and controls (34.2%). This suggests that females may be at greater risk than the males when BMI is low (< 25.0). When male and female patients were compared with respective controls a significant difference was observed (males: $\chi^2$: 4.10; p: 0.04 and females: $\chi^2$: 16.7; p: 0.00). There was a significant difference when cases with low BMI were compared to corresponding controls ($\chi^2$: 20.38; p: 0.000; Table-5).

The distribution of age groups differed significantly between the patients and controls with low (< 25.0: $\chi^2$: 38.22; p: 0.00) and high BMI (≥ 25.0: $\chi^2$: 116.52; p: 0.00). It was observed that patients in elder age group and with low BMI were in high frequency (68.7%) as compared to those with high BMI (58.7%). This was in contrast to what was observed in controls in the elder age group (low BMI: 42.3%; high BMI: 52.5%).

There was significant difference between the frequencies of non-vegetarians in patients with low and high BMI as compared to controls ($\chi^2$: 13.37; p: 0.00). This was because of high frequency of non-vegetarians in patients with low BMI (94.2%) as compared to non-vegetarian controls with low BMI (76.5%; Table-5).

In conclusion, the present results suggest that patients who were with BMI <25.0 in females, patients above 50 years of age and high intake of non-vegetarian diet may be more susceptible to develop age-related cataracts.

The present findings were also supported by two hospital-based case-control studies which showed association of low BMI with the development of nuclear opacity (Mohan et al, 1989; Leske et al, 1991). A study in Punjab, India, found that a higher prevalence of cataract was associated with short height, low weight and low weight-height ratio (Chatterjee, 1982). In contrast a study by Lu et al, (2012) showed risk for developing ARC in case of males who are both overweight and obese.
e) Familial History of Cataracts

In the present study an attempt was made to observe the influence of family history (FH) in conjuncture with epidemiological factors on ARCs. Family history of ARCs was reported in 14.4% of patients and 19.5% of controls (Table-1). The results on the association of ARCs with Family history in conjuncture of other epidemiological parameters are presented below (Table-6).

From Table-6 it may be observed that patients with positive family history (55.1%) were high in frequency among females as compared to male patients (34.9%) and also control females (46.3%; $\chi^2$: 4.45; p: 0.03; Table-6).

The distribution of BMI though not significant showed variations between familial and non-familial groups of patients and controls (Table-6). The frequency of low BMI and high BMI cases among patients and controls with family history showed significant difference ($\chi^2$: 4.18; p: 0.04). This was due to significantly high frequency of low BMI cases among familial patients (90.8%) as compared to controls (78.0%).

The frequency of smokers and non-smokers among males with positive family history differed significantly between the cases and controls ($\chi^2$: 6.81; p: 0.00), which was because of the high frequency of smokers (61.4%) with positive family history among patients as compared to controls (27.3%).

There was a significantly high frequency of non-vegetarian patients (96.9%) with positive family history as compared to non-vegetarian controls (70.7%). The frequency of patients with positive family history differed significantly among vegetarian and non-vegetarian groups as compared to controls ($\chi^2$: 17.99; p: 0.00; Table-6).

From the above observations it may be concluded that the patients with positive family history (FH), increases the risk for developing cataract if they are females, patients with low BMI, males with the habit of smoking and those with high intake of non-vegetarian diet. In most of the studies on ARCs familial segregation has been reported indicating possible genetic basis for the condition (Leske et al, 1983; The Italian-American cataract study, 1991; Framingham Eye Study, 1994; Heiba et al, 1995; Hammond et al, 2000 and
2001; McCarty et al, 2000) while a study by Ughade et al, (1998) could not support the observation. So far no reports seem to be quoted in literature correlating the family history of cataracts with other epidemiological risk factors.

f) Smoking

In the present study only males reported about their smoking habit both in patient and control groups. Therefore the data on smoker males alone was considered for analysis of results and information given by females about smoking was not reliable. An attempt was made to evaluate the association of smoking habit with other epidemiological factors influencing onset of age-related cataracts. A high frequency of male patients was observed to be smokers (49.2%) as compared to controls (38.6%; Table-1). The results on the association of smoking with other epidemiological parameters are presented below (Table-7).

There was no significant variation in the frequency of early and late onset cases among smoker and non-smoker patients ($\chi^2$: 0.24; p: 0.62).

Considering different BMI groups, there was a high frequency of smoker patients with normal weight (77.8%) followed by over weight (13.1%), under weight (5.9%) and obese (3.3%) group. Similarly in controls there was a high frequency of smokers who were of normal weight (66.7%) followed by over weight (19.6%), under weight (9.8%) and obese (3.9%) group. There was no significant difference in the frequencies of smoker patients with different BMI groups as compared to controls (Table-7).

There was a high frequency of familial cases (17.6%) with the habit of smoking as compared to controls (11.8%). There was a significant difference found between smoker patients and controls with positive family history ($\chi^2$: 6.81; p: 0.00; Table-7).

Alcoholic patients who were smokers showed lesser frequency (51.0%) as compared to controls (72.5%). There was a high frequency of non-vegetarian patients (96.1%) with the habit of smoking as compared to controls (82.4%) which showed significant difference ($\chi^2$: 10.58; p: 0.00; Table-7).
In conclusion, male subjects who are smokers and non-vegetarians can be predicted to be at risk for cataract development. Tobacco smoke is associated with a wide range of diseases, such as vascular disease, lung cancer and chronic obstructive pulmonary disease (Hecht, 2002; Shah and Cole, 2010; Taylor, 2010). Tobacco in cigarettes contains several noxious factors like nicotine, free radicals and carbon monoxide, which are reported to play an important role in the pathogenesis of ARC (krishnaiah et al, 2005; Truscott, 2005; Bhagyalaxmi, 2009; Beebe et al, 2010). Further, smoking was reported to increase the risk of developing NC and PSC types of cataracts (Ye et al, 2012).

g) Alcohol Consumption

In the present study the frequency of male patients who consumed alcohol was 36.7% and in controls it was 47.7% (Table-1). An attempt was made to evaluate contribution of alcohol consumption in concurrence with other epidemiological factors and the results found are presented below (Table-8).

There was no significant variation in the frequency of early and late onset cases among alcoholics and non-alcoholic patients ($\chi^2$: 0.24; p: 0.62), indicating absence of any influence of alcohol consumption in decreasing the age at onset of ARCs.

Considering the distribution of BMI groups there was a high frequency of patients who consumed alcohol with normal weight (77.2%) and over weight (12.3%), followed by underweight (7.9%) and obese (2.6%) groups. Similarly in controls also there was a high frequency of subjects who consumed alcohol with normal weight (65.1%) followed by over weight (23.8%), under weight (9.5%) and obesity (1.6%). The differences in the frequencies of different BMI groups among alcoholics and non-alcoholic patients were statistically insignificant (Table-8).

There was a significant difference ($\chi^2$: 7.23; p: 0.007) found between the patients and controls who were smokers. This was due to a high frequency of smokers (68.4%) with the habit of alcohol consumption as compared to respective controls (58.7%). Within cases, patients who were smokers, with and without the habit of alcohol intake were observed to be statistically significant ($\chi^2$: 26.61; p: 0.000; Table-8). This suggests that alcoholic
patients with the habit of smoking may have a synergistic effect for the development of cataract early in life.

Significant difference was found in the frequencies of non-vegetarian and vegetarian patients who were alcoholics compared to respective controls at 9% level ($\chi^2$: 2.75; p: 0.09), this was because of high frequency of patients with non-vegetarian dietary habit (96.5%) as compared to controls (90.5%). This suggests that alcoholic subjects with non-vegetarian food habit could be at high risk of cataract development (Table-8).

In conclusion, male patients who were alcoholics and with the habit of smoking and non-vegetarian dietary habit may have increased risk for developing cataract. Among many other factors, alcohol consumption is considered to be an important risk factor for the development of cataracts though the actual role of alcohol consumption in cataract formation is not clear. A case-control study on the evaluation of risk factors for cataracts from Nagpur, India (Ughade et al, 1998) has shown a significant association of alcohol consumption and cataract formation. Studies by Lindblad et al, (2007) and Tarwadi and Agte, (2011) showed that the risk for developing cataract increases with the amount of alcohol consumption. Epidemiological study by Bhagyalaxmi, (2009) also supported the present results showing the association of alcohol consumption with ARCs. In contrast three large case-control studies conducted by Leske et al, (1983), The Italian-American Cataract Study Group, (1991) and Vitale et al, (1993) did not find any association between alcohol consumption and cataract formation.

h) Diet

In the present study an attempt has been made to observe the influence of vegetarian and non-vegetarian food habits in conjunction with other epidemiological factors on age-related cataracts (ARCs). Among cases 93.1% were found to be non-vegetarians and 6.9% vegetarians and in controls about 74.8% were non-vegetarians and 25.2% were vegetarians (Table-1). The results obtained on the association of dietary habits with other epidemiological parameters are presented below

There was a high frequency of female patients (54.5%) with non-vegetarian dietary habit as compared to male non-vegetarians (45.5%) and also control non-vegetarian females
Results and Discussion

(38.2%) which was statistically significant ($\chi^2$: 13.35; p: 0.00 and $\chi^2$: 20.7; p: 0.00; Table-9) indicating risk of cataract development in females who were non-vegetarians.

Considering the distribution of different BMI groups, there was a high frequency of patients with non-vegetarian dietary habit in the normal weight group (78.4%) as compared to other three BMI groups (under weight: 6.6%; over weight: 13.1% and obese: 1.9%) and also controls (non-veg.-63.1%; $\chi^2$: 42.6; p: 0.00). As compared to controls, patients with non-vegetarian food habit, in different BMI groups was statistically significant ($\chi^2$: 16.4; p≤0.0009; Table-9), suggesting subjects with low BMI, and high intake of non vegetarian food could be at high risk for cataract development.

There was a high frequency of patients with non-vegetarian dietary habit with BMI <25.0 (85.0%) as compared to controls (72.6%; $\chi^2$: 43.33; p: 0.000), suggesting that non-vegetarian patients with low BMI have high risk for cataract development (Table-9). There was high frequency of smokers with non-vegetarian diet among the cases as compared to controls which was statistically significant ($\chi^2$:10.57; p: 0.001).

In conclusion, it was observed that, for non-vegetarians the risk of developing cataract increases if they are females and patients with low BMI. Among the many risk factors, diet (nutrition) is reported to play an important role in the cataract formation. This study is in accordance with several other studies (Mohan et al, 1989; Ughade et al, 1998). While numerous studies have shown that nutritional deficiencies can lead to cataract formation, the role of nutritional status in human cataractogenesis is unclear (Jacques et al, 1998).

In order to study and validate the effect of diet on the age-related cataracts, a well-planned long-term follow-up study with well designed diet plan should be pursued. As follow up of subjects was not planned in the present study, the information recorded was analyzed only based on the consumption of vegetarian and non-vegetarians diet as reported by the patients and controls.

i) Multiple Logistic Regression Analysis

Logistic regression is a mathematical approach that can be used to describe the relationship of several independent variables to a dichotomous dependent variable. For the present
study, a stepwise regression was carried out for epidemiological and molecular markers together to identify the risk contributed significantly by variables in step wise manner to the development of age-related cataracts. The results obtained from multiple logistic regression analysis are presented below (Table-10).

The independent predictor variables identified by logistic regression analysis were a) sex (p< 0.05); b) age (p< 0.05); c) low BMI (p< 0.05); d) alcohol consumption (p< 0.05); and dietary habit (p< 0.05). Among the variables entered, the strong predictors that were identified are in the order of age (6.944) diet (4.350) Low BMI (2.208) sex (1.762) and alcohol consumption (0.602).

Regression Coefficient analysis was used to determine the impact of variable on the development of cataract. As it is known and understood that positive regression coefficient means – the variable increases the probability of cataract development, and negative regression coefficient means – the variable decreases the probability of cataract development.

The variables like age, sex, low BMI (<25.0), diet and alcohol consumption were found to be statistically significant for cataract development. Age has shown positive coefficient, indicating that as the age (OR: 6.944) increases by a unit value, the probability of incidence of cataracts also increases. The coefficient of sex was found to be positive indicating high risk for females (OR: 1.762) for cataract development. BMI indicating normal and underweight (Low BMI: <25.5) showed positive coefficient indicating that lower the BMI value higher is the likelihood of incidence (OR: 2.208) of cataract development, predicting non-obese individuals to be at higher risk for cataract development. Non-Vegetarian dietary habit also showed positive coefficient indicating higher risk for non-vegetarians (OR: 4.350). Alcohol consumption and age were observed to show negative regression coefficient on the incidence of cataracts (Table-10).

From the results obtained by stepwise multiple regression analysis, it may be inferred that cataract is an age-related and sex influenced process and both age and sex are the best independent variables that can predict the onset of the condition. Apart from this, other
variables that are significantly associated with the age-related cataract are low BMI and non-vegetarian dietary habit. Most of these variables are already established by reports in the literature as the risk factors for cataract development. The results from the present study also confirm the same.

The results suggest that the set of variables that are identified as independent predictors in the present study can be used in the prediction of onset of disease prior to its manifestation and also in estimating risk figures for our population and thus in determining the genetic predisposition in an individual. Based on these results it is desirable to evaluate the contribution of the above risk factors to cataractogenesis using large population based studies from India, specifically taking ethnic background into consideration. These figures can then be used as a reference risk figures while managing the disorder.

In nutshell the evaluation of contribution of epidemiological factors to cataract development revealed that females were more prone to develop cataracts compared to males and showed risk for developing cataracts when they were below the age of 50 years (early onset) and also with positive family history. Males with the habit of smoking and alcohol consumption showed risk for the condition. In general, patients above the age of 50 years, with low BMI and non-vegetarian dietary habit showed risk for developing cataracts.
5.4 SCREENING FOR SEQUENCE VARIATIONS/MUTATIONS IN IDO GENE

In recent past wide range of research is being carried out to understand the molecular genetic basis of lens opacification in aged persons. As per Cat-Map database nearly 20 genes have been related to ARCs which include genes coding for crystallins, gap junction proteins, lens fibre-cell intrinsic membrane proteins, DNA repair genes, antioxidant enzymes, xenobiotic enzymes, enzymes in folate metabolism etc. (Shiels et al, 2010). However, variations identified in these genes so far are accounted for a relatively small percentage of conferring genetic risk for age-related cataract (ARC). Hence an extended research on molecular markers is expected to help to identify, test and validate contribution of different genes to the development of ARCs. In the present study, for the first time, an attempt was made to screen for variations in Indoleamine 2,3-dioxygenase (IDO) gene and its promoter region contributing to cataractogenesis.

Indoleamine 2,3-dioxygenase (IDO), is a single copy gene present on human chromosome 8. It has 10 exons, coding for a protein of 403 amino acids with a molecular weight of 45kDa. It is likely that genetic variants in the IDO gene and its promoter region may alter IDO protein and its expression. In the present study a total of 331 ARC cases (110-Nuclear cataract (NC); 110-Cortical cataract (CC) and 111-Posterior subcapsular cataract (PSC) and 210 controls were screened for sequence variations in IDO gene by SSCP analysis to identify the variations in all the 10 exons with their flanking intronic regions and 1.3kb promoter region of IDO gene.

Bioinformatic methods can help prioritize a SNP for wet lab functional studies. Hence, different complimentary bioinformatic tools like a) CLUSTAL X- to check the homology of variant in different species; b) Polymorphism Phenotyping (Polyphen)-to predict possible impact of an amino acid substitution on the structure and function of a human protein; c) Sorting Intolerant From Tolerant (SIFT)-to predict the effect of sequence changes on the protein function based on homology search and the physical properties of amino acids; d) Triton V 4.0.0 package -for protein modeling and e) Human splicing finder
(HSF) to predict the effect of variation on splicing signals were used to predict the functional characterization of variants in IDO gene identified in the present investigation.

The present study was conducted to screen for sequence variations/mutations in all the 10 exons along with intronic boundaries and also 1.3kb promoter of IDO gene using 17 sets of primers (Table-3.1).

**5.4.1 Sequence Variations/Mutations in Exons and Intrinsic Boundaries of IDO Gene**

10 sets of primers (Table-3.1) were used to screen for mutations in all the 10 exons and their flanking intronic regions of IDO gene. It revealed the presence of 4 genetic variants, two novel variants in exons and two known variants in intronic region (Figure-5.1; Table-11).

**A) Novel Variants**

SSCP analysis carried out to screen for sequence variations/mutations in exon 7 and exon 9 of IDO gene showed mobility variation. Sequencing showed the presence of two novel variations  

i) **c.596_597delinsTT** in exon 7 and ii) **C.822C>T** in exon 9.

i) **Exon 7 (c.596_597delinsTT; rs267606590)**  
Screening for variations in exon 7 and its intrinsic boundaries showed heterozygous banding pattern in two of the cases studied. Sequencing of the samples showing heterozygous banding pattern on SSCP analysis revealed the presence of a novel variation i.e. the deletion and insertion of two nucleotides in succession at position c.596 and c.597 of IDO gene by considering first nucleotide of start codon as +1 (Figure-5.2). The novel variation c.596_597delinsTT found in the present study has been registered in the NCBI SNP database as **rs267606590**. The deletion (CG) and insertion (TT) of two successive nucleotides at this position causes codon change in human IDO cDNA showing change of GCG to GTT. This leads to change in amino acid at position 199, where alanine (A) is replaced by valine (V) in variants (p.A199V). This mutation was observed in two (one with NC and one with CC) of the 331 cataract cases and in none of the controls. The two probands (with NC and CC) with this variation were males with an age at onset of 59 and 45 yrs respectively. This variation resulted in loss of restriction site for HhaI enzyme.
RFLP analysis using HhaI enzyme showed the presence of c.596_597delinsTT or p.A199V (Figure-5.3) variation in heterozygous pattern among the cases with a frequency of 0.61%. None of the samples showed homozygous mutant pattern.

![Image of RFLP analysis showing heterozygous pattern](image)

**Figure-5.2**: SSCP analysis of exon 7 and intronic boundaries of *IDO* gene (Top). Chromatograms showing (i) normal homozygote and (ii) heterozygote state for c.596_597delinsTT or p.A199V variation. Arrow shows the variant nucleotides.

![Image of SSCP analysis showing normal and heterozygote states](image)

**Figure-5.3**: Restriction analysis for genotyping p.A199V variation in exon 7 of *IDO* gene.

**Bioinformatic analysis**

a) **CLUSTAL X**

Conservation of the amino acid at this position in different species was done by using CLUSTAL X free offline software (http://www.softpedia.com/get/Science-CAD/Clustal-
Alignment of IDO amino acid sequence from several species using CLUSTAL X showed high conservation for amino acid Alanine at 199 position (Figure-5.4).

Figure-5.4: CLUSTAL X Sequence alignment of IDO protein in different species.

b) Polyphen & SIFT
Polyphen predicted the impact of p.A199V substitution in the present cases as possibly damaging. As per the properties of both amino acids, A and V are aliphatic and neutral. Analysis of this pathogenic change using SIFT and PolyPhen tools showed “probable damaging effect” with Position-Specific Independent Count (PSIC) score of 1.53 by the variant on protein function and with a significant SIFT score of 0.00.

c) Protein modeling
The mutation p.A199V was modeled by using the TRITON V 4.0.0 package. Triton uses the methodology of computational site-directed mutagenesis and the in silico testing of mutant properties. The 3D structure of p.A199V mutant (Figure-5.5) was generated by the MODELLER V9v3 program. This program uses the method of comparative protein modeling by satisfying spatial restrains to model structures of homology proteins based on the 3D structure of template protein (Figure-5.6) (typically X-ray structure from structural database) and the amino acid sequence of modeled protein.

After the homology modeling of the IDO mutant by Triton package the modeled protein structure is superimposed (Figure- 5.7) to that of wild type protein (2dot) and the resulting root mean square deviation (RMSD) value is found to be 1.19. It indicates the wide variation between the wild type and mutant protein structure.
Results and Discussion

**Figure-5.5:** 3D Structure of the p.A199V mutant model obtained

**Figure-5.6:** Location of amino acid Alanine at 199 position in wild type protein of IDO. The ribbon representation of the overall structure of Human IDO (PDB entry code 2DOTB)

**Figure-5.7:** Superimposition of wild type and mutant model.
ii) Exon 9 (C.822C>T)
Only one control sample showed heterozygous banding pattern on SSCP analysis while screening for variations in exon 9 of IDO gene. Sequencing of the variant sample showed C to T transition in heterozygous condition at c.822 of IDO gene (Figure-5.8). c.822C>T is a novel synonymous mutation detected with codon change GAC→GAT coding for aspartic acid at 274th position of the protein. The control proband was a female with 59yrs age. This change was not found in patient samples. Based on codon usage table the variant codon showed that 46% of accessibility due to codon degeneracy. So, this variation did not affect the function of the molecule. This variation showed loss of site for AatII restriction enzyme.

![Figure-5.8: SSCP analysis of exon 9 and intronic boundaries of IDO gene (Top). Chromatograms showing (i) normal homozygote, (ii) heterozygote for c.822 C>T variation. Arrow shows the variant nucleotide.](image)

![Figure-5.9: Restriction analysis for genotyping c.822C>T variation in exon 9 of IDO gene.](image)

RFLP (Figure-5.9) by AatII enzyme showed the presence of heterozygote (CT) and homozygous wild type (CC) with a frequency of 0.47% and 99.53% respectively among controls. All the cataract cases were of homozygous for wild type (CC: 100%) pattern.
Bioinformatic analysis

a) CLUSTAL X

Functional significance of c.822 C>T variation in exon 9 of IDO gene was predicted by using two different bioinformatic tools like CLUSTAL X and HSF.

Homology and cross species conservation analysis using CLUSTAL X showed high conservation for the amino acid aspartic acid at 274\textsuperscript{th} position of IDO protein (Figure-5.10).

b) HSF

Application of HSF for the variant c.822 C>T resulting in p.\textasciitilde asp274\textasciitilde asp was found to have a significant effect on splicing with respect to that of the wild type (WT) sequence. This mutation was predicted to result in the break of potential branch point with consensus value (CV) for the wild type and the mutant sequence to be 67.86 and 63.96 respectively. The exonic splicing enhancers (ESE) http://rulai.cshl.edu/tools/ESE2/ (Cartegni et al, 2003), finder matrices for serine/arginine rich proteins SRp40, SRp55 proteins and EIEs predicted destruction of enhancer site with a score of -100. Whereas putative exonic splicing enhancer (PESE) octomer from Zhang and Chasin predicted creation of new enhancer site.

B) Known variants

While screening for variations in exon 4 and 8 with intronic boundaries showed the presence of two variations which are already registered in NCBI database as rs4613984 (c.422+90 G>A) and rs3214412 (-/CAA) deletion (http://www.ncbi.nlm.nih.gov/snp; Table-11).
i) Intron 4 (c.422+90 G>A; rs4613984)

Screening for variations in the region of exon 4 and its intronic boundaries of *IDO* gene by SSCP analysis showed mobility shift in 6 patient samples. Sequencing of the variant samples revealed the presence of G to A substitution at position c.422+90 in the intronic region downstream to exon4 (Figure-5.11). c.422+90 G>A is a known variant registered in the NCBI SNP database with reference sequence number (rs) rs4613984. Of the 6 patients showing the variation three were with NC (one female and two males) and three were with PSC (two females and a male). None of the patients with CC or controls showed this variation. All 6 patients were above 55 years except one with PSC, who was a female of 25 years. This patient reported onset of cataract 6 months prior to this study and was not having any vision defects before that. Hence this patient was considered to have early onset of age related cataract rather than juvenile onset cataract, by the ophthalmologists. Only one case of NC showed homozygosity (AA) for the variation while the remaining five samples were heterozygous (GA). So far no clinical conditions are associated with this variation except for our study on ARCs that is published (Mamata et al, 2012). This variation resulted in loss of restriction site for HhaI enzyme.

Figure-5.12 depicts the RFLP analysis for the variation c.422G>A using HhaI enzyme. It showed the presence of a mutant homozygote (AA), and five heterozygotes (AT) and wild type homozygotes (GG) among cases occurring with a frequency of 0.30%, 1.51% and 98.18% respectively. All the controls showed wild type homozygous pattern (GG-100%).

**Bioinformatic analysis**

a) HSF

To know the functional significance of c.422G>A variation in intron 4 of *IDO* gene Human splicing finder (HSF) software was used. The variant c.422+90G>A was predicted to result in the break of two enhancer sites with score of -100 and creation of new silencer motif with a variation score of -16.23. The exon splicing enhancer finder matrice for SF2/ASF and SF2/ASF (IgM BRCA1) proteins predicted destruction of enhancer site while that for serine/ arginine rich protein was predicted to create a new site. No difference was observed between wild type and this variant with respect to creation / destruction of potential splice site/branch point but it is predicted to break enhancer sites involved in splicing.
Results and Discussion

i) Wild type (Homozygote)

<table>
<thead>
<tr>
<th>Lane</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Homozygous mutant</td>
</tr>
<tr>
<td>2, 3, 5, 6</td>
<td>Homozygous normal</td>
</tr>
<tr>
<td>4</td>
<td>Heterozygote</td>
</tr>
</tbody>
</table>

Chromatograms showing (i) homozygous wild type, (ii) heterozygote and (iii) homozygous variant for c.422G>A variation. Arrow shows the variant nucleotide.

Figure-5.11: SSCP analysis of exon4 and its intronic boundaries of IDO gene (Top).

Figure-5.12: Restriction analysis for genotyping c.422G>A gene

Lane L - 100 bp ladder

Lanes 1, 5 - Normals
Lane 2 - Variant
Lanes 3, 4 - Heterozygotes


97
ii) Intron 8 (-/CAA deletion)

Screening of exon-8 and its intronic boundaries showed (-/CAA) deletion (rs3214412) in 8th intronic region lying in between 8th and 9th exonic region of IDO gene (Figure-5.13). Two of the patients, 1 with NC and 1 with PSC were found to be homozygous (DD) for this deletion. The frequency of heterozygotes (ID) was high in PSC (10.8%) as compared to NC (7.3%), CC (5.5%) cases and controls (5.7%). This deletion is considered as polymorphic as it is found with considerable frequency among controls also. Allele frequencies were in consistent with Hardy-Weinberg equilibrium. Estimate of odds ratio showed protection for wild type allele ‘I’ (OR= 0.43; 95%CI =0.18-1.02; P=0.03) in PSC cases (Table-12) while variant allele ‘D’ showed 2 fold risk (OR= 2.28; 95%CI =1.01-5.38; P=0.03) for developing PSC.

![SSCP analysis of intron 8 and its intronic boundaries of IDO gene](image)

**Figure-5.13:** SSCP analysis of intron 8 and its intronic boundaries of IDO gene (Top). Chromatograms showing homozygous (i) wild type, (ii) heterozygote and (iii) homozygous deletion for (-/CAA) variation.
Bioinformatic analysis

HSF
Application of HSF software to know the functional significance of this deletion on splicing predicted no difference between the wildtype and mutants with respect to creation or destruction of potential splice site or branch point. Whereas exon identity elements (EIEs) and PESE (putative exonic splicing enhancer) octomer from Zhang and Chasin predicted destruction of enhancer site with a variation score of -100. Similarly use of exonic splicing regulatory sequences (ESRS) from Goren et al, (2006) predicted destruction of enhancer site.

C) Exons-1, 2-3, 5, 6 and 10
Screening for sequence variations in exons 1,2-3,5,6 and 10 didn’t show any mobility shift by SSCP analysis suggesting the absence of detectable variations in them (Figure-5.14). There is however a possibility of missing variations in these regions as SSCP technique has limitations and does not detect all the variations occurring in a given region of any gene with 100% accuracy. Though sequencing yields 100% results, it couldn’t be carried out in the present study due to the cost involved to analyze 541samples. Hence, preliminary screening for detecting variations was done by SSCP analysis which helped in identifying some of the variants in the gene.

![Exon1](image1)

![Exon2-3](image2)

![Exon5](image3)

![Exon6](image4)

![Exon10A](image5)

![Exon10B](image6)
**Figure- 5.14:** SSCP analysis of exons 1, 2-3, 5, 6, and two overlapping regions of exon10 of *IDO* gene showing absence of mobility shift.

### 5.4.2 Sequence Variations/mutations in Promoter of *IDO* Gene

Gene expression, and in particular transcription, in eukaryotic cells is an important process that is under complex regulation. The complex interaction of transcription factors with the transcription binding motifs could be affected by genetic variants in the promoter of a gene. Over expression of IDO is manifested in different disease conditions like cancers and several infectious diseases. But the effects of sequence variations in *IDO* gene are not reported to be with any clinical condition. Hence, screening for sequence variations in the promoter region (1.3kb) of *IDO* gene using 7 overlapping primer (Table-3.1) sets was carried out to know the association of genetic variants with ARCs

The 5’ upstream of transcriptional start site (promoter) contains several regulatory elements, which are important in IFN-γ induced IDO expression (Babcock and Carlin, 2000; Du et al, 2000). There are two distantly separated IFN-stimulated response elements (ISRE-1 and ISRE-2) and three gamma activation sequences (GAS-1, GAS-2 and GAS-3) within the 1.3 kb region upstream of the transcription start site. These response elements are shown to be essential to the transcription of IDO in response to IFN-γ (Robinson et al, 2003 & 2005; Shirey et al, 2006). Along with these, promoter also contains 3 enhancer sequences for binding of enhancer binding protein-beta (C/EBP-Beta-CCAAT; Figure-5.15)

Bioinformatic tool like JASPAR was used to identify the presence of important transcription-factor-binding sites in 1.3 kb of the 5’-flanking region of *IDO* gene. The sequence was screened at high levels of stringency for human transcription factors with a transcription factor score of 90%. The default threshold value is 80%. This software predicted a total of 118 putative transcription factor binding sites in 1.3kb promoter (Annexure- V).

Screening for sequence variations/mutations in the 1.3kb promoter region of *IDO* gene revealed the presence of 3 variants, of which 2 were novel and 1 known variation, which is already registered in NCBI data base (Figure-5.1).
**Results and Discussion**

**IDO PROMOTER SEQUENCE**

**GAS-3  C/EBP-BETA-1**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTCCTGAAG</td>
<td>ISRE-1</td>
</tr>
<tr>
<td>ATTCCCAGAG</td>
<td>GAS-1</td>
</tr>
<tr>
<td>ATTGCCTATGAATCTGATCATAAG</td>
<td>GAS-2</td>
</tr>
<tr>
<td>AACCAGTAAATTTTGAAGAATAGCGAGAGCTATTCCTGACTGTAACGAAAGCACATATGCTATACACAATTTAA</td>
<td>-979G&gt;A</td>
</tr>
<tr>
<td>TTTATTTCGTACTAAATAGCTGTACAAAACACACAGGTGCTTTCATTGGGTTTTA</td>
<td>-738A&gt;G</td>
</tr>
<tr>
<td>CTCTTGCATCTCTAGAAGGGAACTTCCCTTTCTGCTCTTCTGCTATACACAAGTTACAACCAATATGCGACACTGTTGCTT</td>
<td></td>
</tr>
<tr>
<td>CATCAGCCCCCAGTCAAGGGATATAGGCACAAAATACAGGTTGTGTTTCCG</td>
<td></td>
</tr>
<tr>
<td>GCACACCTACTATTGGAAGAATAGCGAGAGCTATTCCTGACTGTAACGAAAGCACATATGCTATACACAATTTAA</td>
<td></td>
</tr>
<tr>
<td>TTTTATTTGTGTCTCTATTCTCTTCTTCTCAGACACTGTTTGGGTTTCCG</td>
<td>-471T&gt;G</td>
</tr>
<tr>
<td>GGTGTGTGAACACATGACACTAATATTTTCAAAGAACTGTGGAAGCCTAAAAGGAAGCCATGAGAAATAACTAAATG</td>
<td></td>
</tr>
<tr>
<td>AGAGTTAGGACTGAGCCTTCTCTATTCTTATCATAATTTTTAAGGCTTCCAATGAAATGTTCTTCCGGCCA</td>
<td></td>
</tr>
<tr>
<td>CCTGTTTCTCATGCTTTGCTTTCTCAGGCTTTTCTGCTGATAATGCTGTAACCAGTGGGTTTTCCG</td>
<td></td>
</tr>
<tr>
<td>GATTACAAATTTCATATGGAATACCTTTGAGATATGCAGACACTGTTTGGGTTTCCG</td>
<td></td>
</tr>
<tr>
<td>TTTTCAATTTACAACAAAGAACACTGCTCAAGAAACATTTCTCAGACACTGTTTGGGTTTCCG</td>
<td></td>
</tr>
<tr>
<td>GTAATCTTCCTCAGACACTGTTTGGGTTTCCG</td>
<td></td>
</tr>
<tr>
<td>GAGCAGCATCAGACACTGTTTGGGTTTCCG</td>
<td></td>
</tr>
</tbody>
</table>

**Figure-5.15:** Nucleotide sequence of the *IDO* promoter showing two IFN-stimulated response elements (ISRE-1 and ISRE-2); 3 gamma activation sequences (GAS-1, GAS-2 and GAS-3) and 3 enhancer binding protein-beta (C/EBP-Beta-CCAAT) sequences in boxes. Transcription and translational start sites are highlighted in green and pink. Sequence variations identified in the present study are shown in red colour.

**A) Novel variants in the promoter sequence of *IDO* gene**

Analysis of 1.3 kb promoter of *IDO* gene revealed the presence of two novel variations (c.-979 G>A in region 2; c.-471 T>G in region 5; Table-13).

i) c.-979 G>A variant in promoter of *IDO* gene
Screening for variations in the promoter (region 2) of the \textit{IDO} gene showed heterozygous banding pattern i.e three distinct bands of which two bands were corresponding to normal homozygous banding pattern on SSCP analysis. This variation was fond in two of the cases, one with NC and another with PSC. Both the probands with NC and PSC were females of age 62 and 55yrs respectively. None among controls showed this variation. Sequencing of the variant samples showed the presence of transition G to A at position -979 from that of ATG start codon (Figure-5.16). This variant c.-979G>A found in this study is novel variation registered in NCBI SNP database with reference sequence number as \textbf{rs267606591}. This variation created a site for restriction enzyme BseMI. RFLP (Figure-5.17) using BseMI enzyme showed the presence of wild type homozygotes (GG) and heterozygotes (GA) at a frequency of 99.4% and 0.6% respectively.

![Digestion products of PCR products with BseMI showing wild type (GG) and heterozygous (GA) at a frequency of 99.4% and 0.6% respectively.](image)

**Figure-5.16:** SSCP gel and chromatograms (i-i) showing normal homozygotes and heterozygotes for the variation c.-979G>A of \textit{IDO} gene

![Digestion products of PCR products with BseMI showing wild type (GG) and heterozygous (GA) at a frequency of 99.4% and 0.6% respectively.](image)

**Figure-5.17:** Restriction analysis for genotyping c.-979G>A

### Bioinformatic analysis

Application of JASPAR software for predicting the putative transcription factors in this region showed the presence of GATA2 transcription factor binding site for wild type allele...
‘G’. This transition results in the loss of GATA2 transcription site for variant allele ‘A’ (Table-14).

**ii) c.-471T>G variant in promoter of IDO gene**

SSCP analysis to screen for variations in region 5 of the IDO promoter showed mobility shift in banding pattern for three of 331 cataract cases. Three probands with PSC showed this variation, of whom, two were females 60 and 56yrs old and one was male 47yrs old. The male proband showed early onset of cataract as compared to females. None among controls showed this mobility variation. Sequencing of the samples with mobility shift revealed the presence of T to G transversion at -471 (Figure-5.18). This novel variation c.-471T>G found in this study is registered in NCBI SNP database as rs267606592. This variation created a restriction site for AciI enzyme.

![SSCP gel and chromatograms](image)

**Figure-5.18:** SSCP gel and chromatograms (i-ii) showing normal homozygotes and heterozygotes for the variation c.-471T>G of IDO gene

![Lanes 1, 4: Normals, Lanes 2, 3, 5: Heterozygotes](image)
**Results and Discussion**

*Figure-5.19:* Restriction analysis for genotyping c.-471T>G variation in promoter of *IDO* gene

Figure-5.19 depicts the RFLP analysis for the variation c.-471T>G. It showed the presence of wild type homozygotes (TT) and heterozygotes (TG) with a frequency of 99.1% and 0.9% respectively.

**Bioinformatic analysis**

JASPAR predicted two TFBS for variant allele ‘G’ with above 95% of threshold value. Two TFBS for SPIB and ETS1 were present on sense and antisense strands respectively. No TFBS were predicted by JASPAR for normal allele ‘T’ (Table-14).

**B) Known variant**

Screening for variations in region 3 of promoter detected the presence of a known variant which is registered in NCBI data base as rs118067147 (Table-13).

i) **c.-738A>G variant in promoter of *IDO* gene**

Three of 331 cataract cases (2NC and 1PSC) and one of 210 controls showed heterozygous banding pattern on SSCP analysis, which on sequencing revealed the presence of A to G transition at position -738 from that of first nucleotide of start codon (Figure-5.20). Among the two probands with NC one was female and other was male who were 62yrs and 50yrs old respectively. The third proband with PSC was a female of 58yrs age. The control subject with this variation is a male, 42 yrs old and the possibility of him developing cataract at later age cannot be ruled out. Frequency of heterozygotes found in cases and controls was 0.9% and 0.5% respectively. The c.-738A>G identified in this study is a known variation registered in NCBI SNP database with rs118067147.

**Bioinformatic analysis**

Bioinformatic analysis predicted the presence of FOXC1 and SOX10 transcription factor binding sequences for wild type allele ‘A’, with a relative score threshold value of 96%. FOXC1 and SOX10 lies on sense (+) and antisense (-) strands of DNA molecule respectively. Variant allele ‘G’ showed the presence of binding site for FOXC1 only with a
lesser threshold value of 94% and loss of SOX10 site when compared to normal allele (Table-14).

![Image of SSCP gel and chromatograms](image)

Figure-5.20: SSCP gel and chromatograms (i-ii) showing normal homozygous and heterozygotes for the variation c.-738A>G of IDO gene

C) Sequences of the regions 1, 4, 6, 7 studied in the promoter of IDO gene
Screening of 1, 4, 6 and 7 regions in the promoter of IDO gene didn’t show any variations on SSCP analysis. All 331 cases and 210 controls showed normal homozygous banding pattern. (Figure-5.21)

![Image of SSCP gel and chromatograms](image)
Results and Discussion

**Figure-5.21:** SSCP analysis of the regions (1, 4, 6, 7) in the promoter of *IDO* gene showing normal banding pattern

So far, no studies are available on the genetic variations of *IDO* gene causing cataracts except for one report (Mamata et al, 2011). Few studies describing the frequencies of the variants of *IDO* gene in some normal populations are registered in NCBI SNP database. A total of 252 human *IDO* variant sequences had been deposited in the SNP database (http://www.ncbi.nlm.nih.gov/snp). Of these 24 variants (18 non synonymous and 6 synonymous) are located in exonic regions, 219 variants in intronic region and 9 in 5’ untranslated region (UTR).

A study conducted by Arefayene, (2009) on inter individual variation of *IDO* gene among the population of 48 African American (AA) and 48 Caucasians (CA) showed the presence of 17 genetic variants of which 9 were in intronic region, 3 were in exon/intron splice junction and 4 (3 non synonymous and 9 bp deletion) in exonic region of *IDO* gene. *In vitro* functional studies for the variants Arg77His in exon 3 and 9bp deletion showed reduced IDO activity. An unpublished data by Arefayene, (2008) showed the presence of c.-738G>A variation in heterozygous state in promoter of *IDO* gene in one of the 48 Caucasian samples and its functional studies showed no significant effect of FOXC1 site on IDO regulation. Only two reports are published on the association of Indoleamine 2,3-dioxygenase polymorphisms with clinical conditions, one in pre-eclampsia patients and another in women with recurrent spontaneous abortions in Iranian women focusing negative results (Nishizawa et al,2010; Amani et al, 2011).

Only one study by Soichot et al, (2011) on the population of 41 unrelated healthy individuals has reported the presence of a VNTR polymorphism in the promoter region of *IDO*. Females with two VNTR (V2/V2) repeats showed significantly lower serum tryptophan concentrations when compared to wild (V1/V1) and heterozygous (V1/V2) condition.

In the present study screening for variants in exons and their intronic boundaries and promoter of *IDO* gene showed the presence of 7 genetic variants. Of which two were in exonic region (novel variants: c.596_597delins TT; c.822 C>T), two in intronic regions
Results and Discussion

(c.422+90G>A; (-/CAA) deletion;) and three were in the promoter region of IDO gene (c.-979 G>A; c.-738A>G; c.-471T>G). Variant c.596_597delins TT in exon 7 resulted in the change of amino acid Ala to val at position 199. Application of bioinformatic tools like SIFT, and Polyphen showed possible damaging effect on the IDO protein. Comparison of wild type and mutant protein structures by Triton package indicated wide variation with RMSD value of 1.19. Non synonymous variants present in exonic region altering the amino acid sequence of a gene product may affect the cellular phenotype at various levels. They may directly alter the stability of the native protein structure and the folding rate, resulting in a reduced concentration of the protein (Karplus et al, 2003). Polymorphisms residing in the substrate recognition site may further affect protein interactions and other biochemical activities inside the cell (Sunyaev et al, 2001). Hence *in vitro* studies were conducted for this variation to understand the variable IDO expression and its activity.

Due to degeneracy of the codons, the synonymous variation c.822C>T found in exon 9 didn’t bring any change in amino acid sequence of IDO protein.

The two known intronic variations c.422+90 G>A in intron 4 and (-/CAA) deletion in intron 8, though not directly involved in the creation or destruction of potential splice site or branch point may affect the enhancer sites associated with the process of splicing. The variant c.422+90G>A was predicted to result in the break of two enhancer sites of splicing with score of -100. NCBI (http://www.ncbi.nlm.nih.gov/snp/4613984? report=GEN) database also showed the presence of this variation in heterozygous state accounting for 0.05% in different Asian, African and Europian populations. Considering the (-/CAA) deletion in intron 8, estimates of odds ratio showed significant protection for PSC with wild type (CAA= 1). The deletion of CAA resulting in destruction of enhancer sites involved in splicing process. Destruction of enhancer sites and branch points may affect the posttranslational modifications of RNA splicing.

The novel variants c.-979G>A and c.-471T>G detected in the promoter region of IDO gene had considerable effect on the TFBs. The presence of variant allele ‘A’ at c.-979 resulted in the loss of site for GATA2 which is a zinc finger transcription factor playing a critical role in regulation of gene expression especially in hematopoietic cells. GATA2 co-
occupies chromatin sites along with Scl/TAL1 complex and either activates or represses the transcription in a context-dependent manner (Bresnick et al, 2012). The presence of other mutant allele ‘G’ at c.-471 position resulted in the creation of two sites for SPIB and ETS1. AS per NCBI database (www.ncbi.nlm.nih.gov/gene/) SPIB transcription factor belongs to the subfamily of ETS and acts as transcriptional activator by binding to the purine rich sequence on DNA i.e 5’ GAGGAA3’. ETS1 transcription factor acts as a transcriptional activator or repressor for several genes which are involved in stem cell development, cell senescence and death. It also directly regulates the expression of cytokines and chemokines in a wide range of cells. Further, it is interesting to note the presence of two variants c.-738 A>G (already reported in data base) and c.-979 G>A (novel variant found in this study) in the same proband with NC which predicted loss of GATA2 and SOX10 TFB sites. SOX10 factor acts as a transcriptional activator by forming a protein complex with other proteins. Its function is strongly influenced by its co-existing proteins in the complex. However, further in vitro studies are required to deduce the exact role of altered transcription factors observed in the present study influencing the regulation of IDO gene and development of cataract. Quantitative studies or familial segregation studies could not be done for the variants detected in promoter region because the families were not available. As all the cases in the present study were inpatients of government hospital and were from different places so we couldn’t trace the families.

Based on the observations made in the present study, it is inferred that the variations found in IDO gene including the promoter may be affecting the activity or the levels of IDO and in turn the synthesis of UV filters causing accumulation of UV filter modified proteins in the lens leading finally in cataract formation. As far as the authors are aware studies are not published on the role of genetic variants of IDO gene and its promoter linking the pathogenesis of ocular diseases including cataract. This study appears to be the first finding demonstrating the involvement of variations in IDO gene (including 10 exons and the promoter) relating to the development of cataracts.