Results

'I invent nothing, I rediscover.'
4. RESULTS
4.1. Results of two-stage SNP exploration study

The genotype frequencies of SNPs in all the genes among the Indian subpopulations studied in the two-stage SNP exploration study (i.e. SNP discovery using DNA resequencing followed by SNP validation using Sequenom's MassARRAY and Illumina's BeadArray technologies) are provided in Figures 3.1.1 to 3.1.7, while the corresponding allele frequencies are presented in Figures 3.2.1 to 3.2.7. The Mean heterozygosities ($H_0$) for each gene following grouping of the subpopulations on a pan-India level are provided in Table 3.1 at the end of this section.

4.1.1. APEX1

Among the SNPs reported within the gene APEX1 in the dbSNP database of NCBI during the time of commencement of this study, two (rs1048945/G878C, rs1130409/T1865G) were found present during SNP detection stage following DNA resequencing. No novel SNP was found. The mutant allele frequencies of the SNPs were 0.052 (for rs1048945C allele) and 0.292 (for rs1130409G allele).

Following SNP Validation, both the SNPs were confirmed to be present, although the SNP rs1048945 was a private polymorphism restricted to only two subpopulations (AA-C-IP5 and AA-E-IP3). rs1130409 (also referred to as rs3136820/rs17844937/rs17857957/rs17858508/rs17857670) had a widespread prevalence across all the Indian subpopulations with an overall mutant allele frequency (MAF) of 0.31. Genotype and allele frequencies of the SNP rs1130409 are provided in Figures 3.1.1 and 3.2.1, respectively. The MAFs among Indian subpopulation groups ranged between 0.11 and 0.53. The lowest MAFs were observed among IE-NE-LPl (0.11) and DR-S-LP2 (0.18) while highest MAFs were found in OG-W-IP (0.53) and TB-NE-LPl (0.50) subpopulations. Among the linguistic groups, the Tibeto-Burman linguistic cluster demonstrated highest MAF (0.42) whereas Indo-Europeans had the lowest MAF (0.28) among the four linguistic groups of India (Figure 3.3).

Haplotype analysis using the two validated SNPs (rs1048945, rs1130409) generated the GG, GT, CG and CT haplotypes with haplotype frequencies 0.318, 0.677, 0.001 and 0.005, respectively. The results obtained from validation of the SNP rs1130409 using Sequenom's MassARRAY were further confirmed using Illumina's BeadArray platform which provided similar MAFs (Figures 3.1.7 and 3.2.7). The mean heterozygosity values for the highly prevalent APEX1 polymorphism (rs1130409) ranged between 0.51 (OG-W-IP, DR-C-IP2 and TB-NE-LP1) to 0.199 (IE-NE-LP1) while the Gst, Ht and Hs values were 0.045, 0.434 and 0.414 respectively. Mutant allele frequency data downloaded from the Hapmap project
database for the SNP rs1130409 were 0.51, 0.46, 0.32 and 0.28 for the populations CEU, CHB, JPT and YRI, respectively (The International HapMap Consortium, 2003) [Figure 3.3]. Considering the IPs and LPs of the Indo-European linguistic subpopulation cluster alone, no difference was observed between the IPs (tribals) and LPs (Large populations) for the SNP rs1130409 with both having the same MAF value (0.28) [Figure 3.4].

4.1.2. LIG1

Out of the SNPs reported within the gene LIG1 in NCBI’s dbSNP database at the time of commencement of this study, thirteen SNPs viz. rs439132, rs20579, rs3730849, rs3730850, rs3730853, rs3730861, rs3730862, rs20580, rs2288880, rs392891, rs3730994, rs2386965, rs3731004 were found present in the samples of the first stage of SNP exploration following DNA resequencing with mutant allele frequencies 0.192, 0.15, 0.308, 0.33, 0.086, 0.093, 0.464, 0.48, 0.16, 0.365, 0.058, 0.292 and 0.017, respectively. No novel SNP was found. SNPs were then selected for further validation study on Sequenom Platform based on the mutant allele frequencies of the SNPs in the first stage of SNP exploration, location (intronic/exonic/regulatory regions, etc) and spacing between SNPs (>1KB). Some reported SNPs were included into the SNP validation procedure, although they were not found during resequencing of representative samples (first stage of SNP exploration), owing to their potential importance in context of the function of the gene or to fill additional gaps in the genome. The final list for validation study included ten SNPs viz. rs3730862, rs4987181, rs4987070, rs12981963, rs20580, rs11879148, rs3730933, rs3730966, rs3731003, rs11666150.

During validation, four of the LIG1 SNPs (rs3730933, rs3730966, rs11879148, rs12981963) were found to be monomorphic while four (rs11666150/A54472C, rs3731003/C42344T, rs4987070/A8945G, rs4987181/C8885T) were private polymorphisms restricted to less than 3 subpopulations of India. Mutant alleles of rs11666150 were found in the subpopulations IE-LP2 and OG-W-IP, rs3731003 was found in OG-W-IP, rs4987070 was found in AA-C-IP5, while the SNP rs4987181 was observed in only one sample from the subpopulation DR-S-LP3. Consequently, only 2 SNPs (rs20580/C19008A, rs3730862/C8804T) have been found to have widespread prevalence (i.e. >3 subpopulations) across the country. Genotype and allele frequencies of the SNP rs20580 are provided in Figures 3.1.2 and 3.2.2, respectively while Figures 3.1.3 and 3.2.3 depicts genotype and allele frequencies, respectively of the SNP rs3730862. The overall MAFs of the SNPs rs20580 (A) and rs3730862 (T) in the Indian population were 0.48 and 0.39, respectively. Considerable variations in MAFs were noted across the four linguistic lineages and geographical zones of India, ranging between 0.24 in
IE-N-LP9 to 0.75 in IE-NE-IP1 for rs20580 (A) and from as low as 0.11 in IE-W-LP3 to 0.77 in IE-NE-IP1 for the SNP rs3730862 (T). Among the linguistic groups, the highest MAF for rs20580 (A) was observed in the Austro-Asiatic and Tibeto-Burman populations (0.63) whereas, the highest MAF observed for rs3730862 (T) was in the Austro-Asiatic (0.58) subpopulation. The Dravidian linguistic cluster exhibited lowest MAF for both the polymorphisms, 0.43 and 0.37 for rs20580 (A) and rs3730862 (T), respectively.

When compared to the world population data on the HapMap database (Figure 3.3), the overall MAF of rs20580 (A) observed in the present study was found to be in close agreement with the Central European [CEU] (0.46) population but very different from the Japanese [JPT] (0.65), Han Chinese [CHB] (0.58) and Yoruba [YRI] (0.53) populations provided in the HapMap database. Consistently, the MAF of rs3730862 (T) (0.37) was also found to be closer to that of the Central European [CEU] (0.33) and very different from the MAFs observed in Japanese [JPT] (0.58), Han Chinese [CHB] (0.51) and Yoruba [YRI] (0.01) populations (The International HapMap Consortium, 2003). Further analysis of the Indo-European linguistic subpopulation cluster alone demonstrated comparable MAFs for both rs20580A and rs3730862T among Large populations (LPs) while the Isolated populations (IPs/Tribals) showed strikingly high MAFs (Figure 3.4). Haplotype analysis using the SNPs (rs3730862, rs20580) revealed the CC haplotype (0.493) as the most frequent followed by the TA haplotype (0.363). The mean heterozygosity values, considering both the LIG1 polymorphisms together, ranged from 0.296 (IE-E-LP4) to 0.512 (TB-NE-LP1) [Table 3.1]. Gst, Ht and Hs values for the LIG1 SNPs taken together were 0.122, 0.491 and 0.431, respectively while the Gst, Ht and Hs values for individual SNPs were 0.102, 0.497 and 0.446, respectively for rs3730862 and 0.142, 0.486 and 0.417, respectively for rs20580.

### 4.1.3. XRCC1

Amongst the SNPs reported in the dbSNP database, two SNPs (rs17655329/C20824T, rs25487/G23990A) were found in the gene XRCC1 following DNA resequencing with mutant allele frequencies 0.014 and 0.091, respectively. Based on the selection criteria of inclusion of SNPs for further validation study, the SNPs rs2307186, rs25495, rs2228487, rs1799782, rs25487 and rs2307166 were chosen for validation using Sequenom's MassARRAY and Illumina's BeadArray technologies. SNP validation studies showed that three SNPs (rs25495, rs2307186 and rs2308313) were monomorphic among Indian subpopulations while the SNP rs2228487/G20824A (also referred to as rs17655329) was found to be a private polymorphism restricted to only 3 subpopulations (AA-C-IP5, AA-E-IP3 and DR-S-LP2). The SNP rs1799782/C22142T had a widespread prevalence throughout
the Indian population with an overall mutant allele frequency of 0.16. Genotype and allele frequencies of the SNP rs1799782 are provided in Figures 3.1.4 and 3.2.4, respectively. The MAFs of rs1799782 varied widely among the subpopulation groups and ranged between 0.04 in DR-S-IP4 to 0.41 in IE-NE-IP1. Within linguistic clusters, MAF was the lowest (0.10) among Dravidians and highest (0.28) among the Tibeto-Burmans (Figure 3.3). No HapMap population data was available on the allele or genotype frequencies of the XRCC1 SNP, rs1799782. Haplotype frequencies, upon consideration of the SNPs rs2228487 and rs1799782 were 0.833, 0.160, 0.005 and 0.002 for the haplotypes GG, GA, AG and AA, respectively. Average heterozygosity values of the most prevalent XRCC1 SNP rs1799782 ranged from as low as 0.084 (IE-E-LP2, DR-S-IP4) to 0.496 (IE-NE-IP1) [Table 3.1]. Gšt, Ht and Hs values were 0.068, 0.274 and 0.256, respectively.

4.1.4. CNP

First stage of SNP exploration study using DNA resequencing showed the presence of two SNPs (rs8078650/T2010G and rs2070106/G7103A) with mutant allele frequencies 0.345 and 0.391, respectively. SNP validation studies among 55 Indian subpopulations using Sequenom’s MassARRAY showed that both the SNPs had widespread prevalence (>3 subpopulations). Figures 3.1.5 and 3.2.5 represent genotype and allele frequencies of the SNP rs8078650 while Figures 3.1.6 and 3.2.6 provide details on the genotype and allele frequencies of the polymorphism rs2070106. Among the Indian subpopulations, lowest MAFs for the SNP rs8078650 were observed in the subpopulations DR-C-IP2 (0.02) and DR-S-IP1 (0.04) while AA-C-IP1 exhibited highest MAF (0.44) among the Indian subpopulations. On the other hand lowest MAFs for rs2070106 were observed among the subpopulations IE-S-IP1 (0.222) and IE-N-SP5 (0.23) whilst the highest MAF was found among DR-S-IP4 (0.59) and AA-NE-IP1 (0.59). The MAFs among world populations observed in the Hapmap database for the CNP SNP rs8078650 were 0.18, 0.22, 0.13 and 0.41 while that for rs2070106 were 0.41, 0.32, 0.32 and 0.05 for the populations CEU, CHB, JPT and YRI, respectively (Figure 3.3). When the Indo-European linguistic subpopulation cluster was considered alone, no appreciable difference was observed between the IPs (tribals) and the LPs (Large populations) for the SNPs rs2070106 and rs8078650. The MAFs for rs2070106 were 0.40 and 0.41 while the MAFs for rs8078650 were 0.19 and 0.16 among IPs and LPs, respectively (Figure 3.4).
4.2. Results of Genetic association study on human cancers

4.2.1. Breast Cancer

Results were successfully obtained among 225 female controls each for the SNPs rs1130409 (APEX1), rs25487 (XRCC1) and rs4989588 (FEN1) and 215 each for rs13181 (ERCC2) and rs20580 (LIG1). On the other hand, genotype results were acquired successfully for 150 breast cancer cases for rs1130409 (APEX1), rs25487 (XRCC1) and rs4989588 (FEN1), 157 cases for rs1799782 (XRCC1), 155 cases for rs13181 (ERCC2) and 142 cases for rs20580 (LIG1).

The polymorphisms rs1050525 (PCNA), rs4989586 (FEN1) and rs4989587 (FEN1) were not found in any of the cancer or control subjects.

<table>
<thead>
<tr>
<th>Subpopulations</th>
<th>rs1130409 (APEX1)</th>
<th>LIG1(rs3730862,rs20580)</th>
<th>XRCC1 (rs1799782)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA-C-IP5</td>
<td>0.489</td>
<td>0.499</td>
<td>0.345</td>
</tr>
<tr>
<td>TB-N-SP1</td>
<td>0.487</td>
<td>0.489</td>
<td>0.415</td>
</tr>
<tr>
<td>IE-E-IP1</td>
<td>0.359</td>
<td>0.497</td>
<td>0.231</td>
</tr>
<tr>
<td>IE-E-LP2</td>
<td>0.372</td>
<td>0.451</td>
<td>0.084</td>
</tr>
<tr>
<td>IE-N-LP1</td>
<td>0.371</td>
<td>0.497</td>
<td>0.394</td>
</tr>
<tr>
<td>IE-W-LP1</td>
<td>0.358</td>
<td>0.495</td>
<td>0.169</td>
</tr>
<tr>
<td>DR-C-IP2</td>
<td>0.508</td>
<td>0.431</td>
<td>0.199</td>
</tr>
<tr>
<td>IE-NE-IP1</td>
<td>0.474</td>
<td>0.499</td>
<td>0.496</td>
</tr>
<tr>
<td>AA-E-IP3</td>
<td>0.478</td>
<td>0.324</td>
<td>0.199</td>
</tr>
<tr>
<td>DR-S-LP3</td>
<td>0.415</td>
<td>0.450</td>
<td>0.294</td>
</tr>
<tr>
<td>IE-W-LP2</td>
<td>0.426</td>
<td>0.342</td>
<td>0.169</td>
</tr>
<tr>
<td>IE-N-LP5</td>
<td>0.485</td>
<td>0.453</td>
<td>0.207</td>
</tr>
<tr>
<td>DR-S-IP4</td>
<td>0.467</td>
<td>0.389</td>
<td>0.084</td>
</tr>
<tr>
<td>TB-NE-LP1</td>
<td>0.512</td>
<td>0.512</td>
<td>0.433</td>
</tr>
<tr>
<td>DR-S-LP2</td>
<td>0.305</td>
<td>0.326</td>
<td>0.169</td>
</tr>
<tr>
<td>IE-NE-LP1</td>
<td>0.199</td>
<td>0.378</td>
<td>0.322</td>
</tr>
<tr>
<td>IE-E-LP4</td>
<td>0.487</td>
<td>0.296</td>
<td>0.264</td>
</tr>
<tr>
<td>IE-W-LP3</td>
<td>0.394</td>
<td>0.486</td>
<td>0.124</td>
</tr>
<tr>
<td>IE-W-LP4</td>
<td>0.477</td>
<td>0.444</td>
<td>0.124</td>
</tr>
<tr>
<td>IE-N-LP9</td>
<td>0.394</td>
<td>0.464</td>
<td>0.264</td>
</tr>
<tr>
<td>OG-W-IP</td>
<td>0.514</td>
<td>0.485</td>
<td>0.169</td>
</tr>
<tr>
<td>IE-N-SP4</td>
<td>0.359</td>
<td>0.481</td>
<td>0.294</td>
</tr>
<tr>
<td>TB-N-IP1</td>
<td>0.474</td>
<td>0.473</td>
<td>0.384</td>
</tr>
<tr>
<td>IE-N-IP2</td>
<td>0.394</td>
<td>0.452</td>
<td>0.449</td>
</tr>
</tbody>
</table>

Table 3.1 Average heterozygosity ($H_0$) values for each gene among Indian subpopulations. [SNPs considered for each gene are provided in parentheses]
Chisquare_{HWE} for genotype distributions were 0.0000, 0.2488, 16.7029, 15.8667, 0.8907 and 0.0028 among controls for the loci rs1130409 (APEX1), rs13181 (ERCC2), rs4989588 (FEN1), rs20580 (LIG1), rs1799782 (XRCC1) and rs25487 (XRCC1), respectively.

The genotype and allele frequencies of the selected SNPs among Breast cancer cases and normal healthy female controls have been provided in Figures 3.5.1 to 3.5.6.

Allele frequencies of mutant alleles were 20%, 38.1%, 6.9%, 44.4%, 15% and 39.3% in control group and 33.7%, 57.1%, 5%, 34.5%, 7% and 53.3% in breast cancer group for the polymorphisms rs1130409 (APEX1), rs13181 (ERCC2), rs4989588 (FEN1), rs20580 (LIG1), rs1799782 (XRCC1) and rs25487 (XRCC1), respectively.

3x2 contingency Chisquare values were 19.2 (P<0.0001), 24.39 (P<0.0001), 1.397 (P=0.4972), 12.06 (P=0.0024), 12.64 (P=0.0018) and 20.89 (P<0.0001) for the genotypes of rs1130409 (APEX1), rs13181 (ERCC2), rs4989588 (FEN1), rs20580 (LIG1), rs1799782 (XRCC1) and rs25487 (XRCC1), respectively which suggested an overall significant association between breast cancer incidences and genotypes for the loci rs1130409 (APEX1), rs13181 (ERCC2), rs20580 (LIG1), rs1799782 (XRCC1) and rs25487 (XRCC1).

Subsequent analysis concerned assessment of risks associated with individual mutant genotypes, WM (homozygous), MM (homozygous mutant) and WM + MM (combined mutant) with the risk of breast cancer for every polymorphism under study based on Odds ratio (OR), 95% Confidence Intervals (CI) and corresponding P values.

Statistically significant association with breast cancer susceptibility was observed for rs1130409 (APEX1) homozygous mutant (GG) (OR 3.355, 95% CI 1.363 to 8.258), heterozygous (TG) (OR 2.419, 95% CI 1.559 to 3.755) and combined mutant (GT + GG) (OR 2.523, 95% CI 1.651 to 3.856) genotypes. Mutant genotypes of the polymorphism rs13181 in the gene ERCC2 viz. homozygous mutant (CC) (OR 4.412, 95% CI 2.413 to 8.068), heterozygous (AC) (OR 2.086, 95% CI 1.246 TO 3.492) and combined mutant (AC + CC) (OR 2.672, 95% CI 1.647 to 4.334) also exhibited statistically significant association with the risk of breast cancer. Correspondingly, significant association was also observed for the rs25487 (XRCC1) homozygous mutant AA (OR 2.910, 95% CI 1.661 to 5.100) and combined mutant (AA + AG) (OR 1.408, 95% CI 0.903 to 2.195) genotypes.

On the other hand, protective association was exhibited by rs1799782 (XRCC1) homozygous mutant (TT) (OR 0.522, 95% CI 0.132 to 2.056), heterozygous (CT) (OR 0.354, 95% CI 0.194 to 0.646) and combined mutant (CT+ TT) (OR 0.373, 95% CI 0.213 to 0.654) genotypes and rs20580 (LIG1) homozygous mutant (AA) (OR 0.472, 95% CI 0.228 to 0.980),
Results

heterozygous (CA) (OR 0.444, 95% CI 0.277 to 0.713) and combined mutant (CA + AA) (OR 0.449, 95% CI 0.284 to 0.709) genotypes.

No association was observed between rs4989588 (FEN1) mutant genotypes viz. heterozygous (TA) (OR 0.618, 95% CI 0.274 to 1.390, P=0.3310, RR 0.844, 95% CI 0.657 to 1.084), homozygous mutant (AA) (OR 0.865, 95% CI 0.203 to 3.681, P=1.000, RR 0.945, 95% CI 0.548 to 1.628) and combined mutant (TA + AA) (OR 0.666, 95% CI 0.324 to 1.364, P=0.2981, RR 0.863, 95% CI 0.683 to 1.090) genotypes with breast cancer risk.

Results of association studies between the gene polymorphisms and breast cancer risk are represented graphically in terms of corresponding Odds ratios in Figure 3.6.

The association with breast cancer did not vary greatly with menopausal status. Analyses stratified by tumour grading and ER-PR status did not seem to modify the risk of breast cancer among carriers (data not shown). The four XRCC1 haplotypes CG, CA, TG, TA accounted for 53.8%, 31.2%, 8.1%, and 6.8% chromosomes, respectively in controls and 43.5%, 51%, 1.4% and 4.1% chromosomes, respectively among cases. When compared to the most common haplotype CG, the results of haplotype association studies indicated statistically significant association between the CA haplotype and risk of breast cancer (OR 2.023, 95% CI 1.475 to 2.776, P<0.0001) whereas, protective association was observed between the haplotypes TG (OR 0.2082, 95% CI 0.073 to 0.598, P=0.0011), TA (OR 0.7496, 95% CI 0.371 to 1.515) and breast cancer risk. The haplotypes TG + TA taken together also demonstrated statistically significant protective association with breast cancer susceptibility (OR 0.4542, 95% CI 0.253 to 0.817, P=0.0085).

4.2.2. Squamous Cell Carcinomas of the Head and Neck (SCCHN)

Genotype results were successfully obtained among 386 healthy normal unaffected controls for the SNP rs1130409 (APEX1), 385 each for rs13181 (ERCC2) and rs1799782, 387 for rs20580 (LIG1), 395 for rs25487 (XRCC1), 302 for rs1052133 (OGG1) and among 100 control samples for the polymorphism rs2070106 (CNP). Successful genotype results were observed among 285 SCCHN-affected cases for rs1130409 (APEX1), 280 for rs25487 (XRCC1), 275 each for rs13181 (ERCC2) and rs20580 (LIG1), 235 for rs1052133 (OGG1), 270 for rs1799782 (XRCC1) and 100 SCCHN cases for the SNP rs2070106 (CNP).

The polymorphism rs2228487 (XRCC1) was not found in any of the SCCHN cases or healthy controls included in this study.

ChisquareHWE for genotype distributions were 8.473, 0.846, 0.345, 1.373, 14.025, 8.360, 0.015 among controls for the loci rs1130409 (APEX1), rs2070106 (CNP), rs13181 (ERCC2),
rs20580 (LIG1), rs1052133 (OGG1), rs1799782 (XRCC1) and rs25487 (XRCC1), respectively.

The genotype and allele frequencies of the SNPs among SCCHN cases and healthy control subjects are provided in Figures 3.7.1 to 3.7.7. Mutant allele frequencies were 28.5%, 38.5%, 34.4%, 43.9%, 36.9%, 14.4% and 40.4% among the controls and 20.5%, 39.0%, 41.1%, 35.8%, 28.1%, 6.7%, 33.0% among SCCHN cases for the SNPs rs1130409 (APEX1), rs2070106 (CNP), rs13181 (ERCC2), rs20580 (LIG1), rs1052133 (OGG1), rs1799782 (XRCC1) and rs25487 (XRCC1), respectively.

The corresponding 3x2 contingency Chisquare values were 11.35 (P=0.0034), 0.232 (P=0.8905), 7.417 (P=0.0245), 9.383 (P=0.0092), 12.61 (P=0.0018), 18.38 (P=0.0001) and 7.816 (P=0.0201) for rs1130409 (APEX1), rs2070106 (CNP), rs13181 (ERCC2), rs20580 (LIG1), rs1052133 (OGG1), rs1799782 (XRCC1) and rs25487 (XRCC1), respectively which implied an overall significant association between the prevalence of SCCHN and genotypes of the loci rs1130409 (APEX1), rs13181 (ERCC2), rs20580 (LIG1), rs1052133 (OGG1), rs1799782 (XRCC1) and rs25487 (XRCC1).

Subsequent analysis pertaining to the assessment of risks associated with individual mutant genotypes of every polymorphism covered in this investigation with regards to SCCHN risk depicted statistically significant association for rs13181 (ERCC2) homozygous mutant (CC) (OR 1.680, 95% CI 1.014 to 2.784), heterozygous (AC) (OR 1.531, 95% CI 1.092 to 2.149) and combined mutant (AC + CC) (OR 1.560, 95% CI 1.128 to 2.158) genotypes. Conversely, significant protective association with SCCHN risk was observed among the rs1130409 (APEX1) homozygous mutant (GG) (OR 0.349, 95% CI 0.182 to 0.670) and combined mutant (TG + GG) (OR 0.679, 0.496 to 0.928) genotypes. Protective association was also observed between rs20580 (LIG1) heterozygous (CA) (OR 0.661, 95% CI 0.470 to 0.930), homozygous mutant (AA) (OR 0.510, 95% CI 0.314 to 0.830) and combined mutant (CA + AA) (OR 0.623, 95% CI 0.450 to 0.862) genotypes and SCCHN susceptibility. Furthermore, rs1052133 (OGG1) heterozygous (CG) (OR 0.659, 95% CI 0.462 to 0.940), homozygous (GG) (OR 0.257, 95% CI 0.107 to 0.618) and combined mutant (CG + GG) (OR 0.606, 95% CI 0.427 to 0.859) genotypes also exhibited significant protective association with predisposition towards SCCHN. Mutant genotypes of both the XRCC1 polymorphisms studied, viz. rs25487 (XRCC1) homozygous mutant (AA) (OR 0.572, 95% CI 0.354 to 0.923), heterozygous (GA) (OR 0.677, 0.486 to 0.943) and combined mutant (GA + AA) (OR 0.650, 95% CI 0.475 to 0.888) genotypes and rs1799782 (XRCC1) heterozygous (CT) (OR 0.388, 95% CI 0.242 to 0.624) and combined mutant (CT + TT) (OR 0.391, 95% CI
Results

0.252 to 0.606) genotypes also demonstrated significant negative/protective association with SCCHN risk. On the other hand, mutant genotypes of the polymorphism rs2070106 (CNP) viz. homozygous mutant (AA) (OR 1.090, 95% CI 0.497 to 2.394), heterozygous (GA) (OR 0.908, 95% CI 0.492 to 1.675) and combined mutant (CT + TT) (OR 0.959, 95% CI 0.545 to 1.688), did not exhibit any association with the risk of SCCHN. Results of genetic association studies between the polymorphisms selected in the present study with SCCHN risk presented in terms of odds ratios of mutant genotypes is shown in Figure 3.8. Association of the selected SNPs with SCCHN risk did not vary greatly with gender or tumour grading (data not shown).

Comparison with SNP500Cancer database

Figure 3.9 shows the comparison of combined mutant genotype (WM + MM) frequencies for the SNPs studied between control samples included in the present study and the SNP500Cancer database of the Cancer Genome Anatomy Project. The mutant allele frequencies for SNPs rs1130409 (APEX1), rs13181 (ERCC2), rs20580 (LIG1), rs1052133 (OGG1), rs1799782 (XRCC1) and rs25487 (XRCC1) were 0.495, 0.211, 0.549, 0.225, 0.129, 0.353, respectively among the controls samples included in the SNP500Cancer database, whereas, the corresponding MAF values for the control samples included in the present study were 0.285, 0.344, 0.439, 0.369, 0.144 and 0.404, respectively.

4.3. Results of Gene expression studies

All the genes, except ERCC2, were expressed among both cases and control cDNA samples. A 200bp β-actin fragment was amplified in all the samples as internal control. Analysis of gene expression profile of the genes APEX1, FEN1, LIG1, XRCC1 and OGG1 using Semi-quantitative and Real Time PCR depicted some deviation between cases and control samples. Calculation of the relative expression of the genes, after normalization of the results to the reference gene β-actin, using ΔΔCt method demonstrated an appreciable down-regulation in expression among the genes APEX1, FEN1, LIG1, OGG1 in cases compared to the controls while gene XRCC1 showed upregulation although there was a huge inter-sample difference in gene expression values. The gene CNP showed no change in expression between cases and control samples. Figure 3.10 shows a Box plot and a histogram of the relative expression levels of the genes APEX1, CNP, FEN1, LIG1, XRCC1 and OGG1 represented in terms of ΔΔCt values. Relative mRNA expression, inferred from 2ΔΔCt values were 0.265, 0.108, 0.020, 2.061, 0.133 and 1.298 for the genes APEX1, LIG1, OGG1, XRCC1, FEN1 and CNP, respectively.
Semi-quantitative RT-PCR, corroborated with the findings from Real Time PCR, which showed an overall decrease in expression of the genes APEX1, LIG1, OGG1 and FEN1 among cases (Figures 3.11 and 3.12). Further studies may be fruitful to unravel any correlation between the gene expression patterns and genotype risk in SCCHN patients.

4.4. Analysis of DNA ploidy status

All the samples (SCCHN cases and controls) analysed by DNA flow cytometry (FC) showed only a diploid peak in more than 99% of the cells studied. No aneuploid peak was found in any of the samples included in the investigation (Figure 3.13).

4.5. Y chromosome loss microdetection

PCR conducted on blood and tissue DNA samples of SCCHN subjects using the Y-chromosome-specific genes SRY, ZFY and BPY1 showed successful amplification in all sets of primers among all the male SCCHN subjects when observed using 1.5% agarose gel. No band was observed for the Y-chromosome specific genes SRY, ZFY and BPY1 in females (negative controls). The D8S262 primer (chromosome 8p23) was amplified in all the subjects. All the normal healthy males (positive controls) showed bands for both the D8S262 primer and for the Y-chromosome specific genes SRY, ZFY and BPY1. Thus, the present study concerning PCR-based screening of Y chromosome loss could not detect any Y chromosome DNA microdeletion among SCCHN cases.
Figure 3.1.1 Graphical representation of genotype frequencies of the SNP rs1130409 in the gene APEX1 among Indian subpopulations
Figure 3.1.2 Graphical representation of genotype frequencies of the SNP rs20580 in the gene LIG1 among Indian subpopulations
Figure 3.1.3 Graphical representation of genotype frequencies of the SNP rs3730862 in the gene LIG1 among Indian subpopulations
Figure 3.1.4 Graphical representation of genotype frequencies of the SNP rs1799782 in the gene XRCC1 among Indian subpopulations following SNP validation
Indian subpopulations

Figure 3.1.5: Graphical representation of genotype frequencies of the gene CNT in the Overall Indian study.
Figure 3.1.6 Graphical representation of genotype frequencies of the SNP rs2070106 in the gene CNP among Indian subpopulations
Figure 3.1.7 Genotype frequencies of rs1130409 (APEX1) among Indian subpopulations following SNP validation using Illumina's BeadArray.
Figure 3.2.1 Mutant allele frequencies of the SNP rs1130409 in the gene APEX1 among Indian subpopulations.
Figure 3.2.2 Mutant allele frequencies of the SNP rs20580 in the gene LIG1 among Indian subpopulations
Figure 3.2.3 Mutant allele frequencies of the SNP rs3730862 in the gene LIG1 among Indian subpopulations.
Overall (India)  
TB-N-SP1  
TB-N-IP1  
TB-NE-LP1  
OG-W-IP  
IE-W-LP4  
IE-W-LP3  
IE-W-LP2  
IE-W-LP1  
IE-N-SP4  
IE-N-LP9  
IE-N-LP5  
IE-N-LP1  
IE-N-IP2  
IE-NE-LP1  
IE-NE-IP1  
IE-E-LP4  
IE-E-LP2  
IE-E-IP1  
DR-S-LP3  
DR-S-LP2  
DR-S-IP4  
DR-C-IP2  
AA-E-IP3  
AA-C-IP5  

Figure 3.2.4 Mutant allele frequencies of the SNP rs1799782 in the gene XRCC1 among Indian subpopulations
Figure 3.2.5 Mutant allele frequencies of the SNP rs8078650 in the gene CNP among Indian subpopulations
Figure 3.2.6 Mutant allele frequencies of the SNP rs2070106 in the gene CNP among Indian subpopulations
Figure 3.2.7 Mutant allele frequencies of the SNP rs1130409 (APEX1) among Indian subpopulations following SNP validation using Illumina’s BeadArray
Figure 3.3 Comparison of mutant allele frequencies within the Indian population and between the Indian linguistic subpopulation clusters and the world population data obtained from Hapmap project.

AA- Austro-Asiatic, DR- Dravidian, IE- Indo-Europeans, TB- Tibeto-Burman, IND- Overall mutant allele frequency of the Indian population, CEU-CEPH (Utah residents with ancestry from northern and western Europe), YRI- Yoruba in Ibadan, Nigeria, JPT- Japanese in Tokyo, Japan, CHB- Han Chinese in Beijing, China.

No Hapmap data was available for the SNP rs1799782 (XRCC1).
Figure 3.4 Comparison of mutant allele frequencies between IPs and LPs of the Indo European (IE) linguistic subpopulation cluster
Figure 3.5.1 Representation of genotype and allele frequencies of the SNP rs1130409 in the gene APEX1 among breast cancer cases and normal females
Figure 3.5.2 Representation of genotype and allele frequencies of the SNP rs4989588 in the gene FEN1 among breast cancer cases and normal females
Figure 3.5.3 Representation of genotype and allele frequencies of the SNP rs20580 in the gene LIG1 among breast cancer cases and normal females
Figure 3.5.4 Representation of genotype and allele frequencies of the SNP rs13181 in the gene ERCC2 among breast cancer cases and normal females.
Figure 3.5.5 Representation of genotype and allele frequencies of the SNP rs1799782 in the gene XRCC1 among breast cancer cases and normal females
Figure 3.5.6 Representation of genotype and allele frequencies of the SNP rs25487 in the gene XRCC1 among breast cancer cases and normal females
Figure 3.6 Representation of genetic association of selected SNPs with the risk of Breast cancer determined in terms of odds ratios of mutant genotypes

OR>1 denotes positive association while OR<1 signifies protective/negative association with Breast cancer risk
Figure 3.7.1 Representation of genotype and allele frequencies of the SNP rs1130409 in the gene APEX1 among SCCHN cases and controls.
Figure 3.7.2 Representation of genotype and allele frequencies of the SNP rs20580 in the gene LIG1 among SCCHN cases and control samples.
Figure 3.7.3 Representation of genotype and allele frequencies of the SNP rs13181 in the gene ERCC2 among SCCHN cases and controls
Figure 3.7.4 Representation of genotype and allele frequencies of the SNP rs1052133 in the gene OGG1 among SCCHN cases and controls
Figure 3.7.5 Representation of genotype and allele frequencies of the SNP rs1799782 in the gene XRCC1 among SCCHN cases and controls.
Figure 3.7.6 Representation of genotype and allele frequencies of the SNP rs25487 in the gene XRCC1 among SCCHN cases and controls
Figure 3.7.7 Representation of genotype and allele frequencies of the SNP rs2070106 in the gene CNP among SCCHN cases and controls
Figure 3.8 Representation of genetic association of selected SNPs with the risk of SCCHN determined in terms of odds ratios of mutant genotypes

OR>1 denotes positive association, while OR<1 signifies protective/negative association with SCCHN risk
Figure 3.9 Comparison of the combined mutant genotype (WM+MM) frequencies between the present study and the data available from SNP500Cancer database of the Cancer Genome Anatomy Project.
Figure 3.10 Box plot and Histogram representing relative gene expression levels of the genes selected for study among SCCHN cases compared to normal healthy controls, inferred on the basis of corresponding ΔΔCp values. ΔΔCp > 0 denotes over expression, while ΔΔCp < 0 demonstrates decrease in gene expression among SCCHN cases compared to controls.
Figure 3.11 Results of Gene expression analysis in SCCHN cases and controls using semi-quantitative RT-PCR

Control samples in each gel are represented with a downward arrow (↓) sign while rest are all SCCHN cases.
Figure 3.12 Box plot depicting relative expression of selected DNA repair genes in SCCHN cases compared to controls obtained using semi-quantitative RT-PCR and represented in terms of Normalised relative band intensity values.
Figure 3.13 Representation of results of DNA ploidy status using flow cytometry following PI staining. Only a single peak depicting diploidy was observed in all samples which is indicative of no change in the ploidy status among SCCHN samples.