Review of Literature

'A problem well stated is a problem half solved.'
2. REVIEW OF LITERATURE

2.1. SINGLE NUCLEOTIDE POLYMORPHISMS

Single Nucleotide Polymorphisms (SNPs or 'snips') are point mutations in nuclear and mitochondrial DNA. They are the commonest types of variation in the genome whereby individuals differ in only a single base position. These abundant forms of genome variations are distinguished from rare variations by a requirement for the least abundant allele to have a frequency of 1% or more (Brookes, 1999). SNPs are said to occur about once in every 1000 bases along the genome (Zhang et al., 2005a). Therefore, around 3 million SNPs are believed to be existent in the human genome occurring in genes within the regulatory regions, drug responding elements, drug metabolizing enzymes, drug transporters, drug receptors, proton pumps and non-coding regions linked to disease susceptibility. Because of their high abundance, low mutation rate and accessibility to high-throughput genotyping, SNP markers are largely preferred for population genetics and disease association studies (Syvanen et al., 1999).

**SNPs and Disease susceptibility**

SNPs are considered best markers inhabiting the genome due to their wide prevalence, easier automation for scoring, highly polymorphic & biallelic nature, co-dominance, easy reproducibility and their presence in both exonic as well as intronic regions of genes, furnishing large sets of markers near or within the locus of interest. SNPs are considered to be primarily responsible for differences between individuals that may serve as causative variations for simple and complex genetic diseases or can be indirectly associated to a diseased state by serving as linked markers for localizing a disease on the human genome map. Therefore, due to their disease association, SNPs are under close scrutiny in studies on fishing out the basis of diseases to radically improve biological understanding and to help develop more appropriate therapeutic measures (Syvanen et al., 1999). Diseases like Alzheimer’s Disease (Martin et al., 2000; Ozturk et al., 2006), Parkinson’s Disease (Martin et al., 2001; Chaudhary et al., 2005), Schizophrenia (Li et al., 2000; Verma et al., 2005), Sickle-cell anemia (Waterfall and Cobb, 2001), Beta-thalassemia (Fei et al., 1989; Kukreti et al., 2002; Chen et al., 2004), Cystic Fibrosis (Stoerker et al., 2000; Chou et al., 2005), Muscular Dystrophies (Kilimann et al., 1992; Chaturvedi et al., 2001), Phenylketonuria (DiLella et al., 1986), Type-II diabetes (Wiltshire et al., 2001; Mori et al.,
2002), coronary atherosclerosis (Su et al., 2000; Howell et al., 2005), Migraine headaches (McCarthy et al., 2001; Fumal and Schoenen, 2004) have already been found to be directly or indirectly linked to SNPs in the human genome. Many more are certain to be discovered in the near future following loads of research having been undertaken in this regard.

**SNPs and Differential Drug Response**

Not all people respond to drugs and medications in the same way leading to the emergence of Adverse Drug Reactions (ADRs) which is one of the major causes for the failure of a drug to successfully hit the market. This is why pharmaceutical companies have to test their products on large numbers of patients in order to develop a clear idea about the safety and efficacy of new medications. Research has demonstrated the presence of individual or population-specific variations like SNPs to be the reason behind differential drug responses. Examples of variable drug response due to the presence of SNPs include responses to warfarin (Rieder et al., 2005; Veenstra et al., 2005), antidiabetics (Kang et al., 2005b), Hormone replacement therapy (HRT) (Yahata et al., 2005), antihypertensives (Matayoshi et al., 2004), antiretroviral agents administered against AIDS (Saitoh et al., 2005), etanercept therapy in rheumatoid arthritis (Kang et al., 2005a) antidepressants (Binder et al., 2004), anticancer drugs (Shi et al., 2004a), bronchodilators/ antiasmatics (Choudhry et al., 2005), analgesics (Lotsch et al., 2004), anti-inflammatory drugs used in Inflammatory Bowel disease (Griga et al., 2005) and antihepatitic drugs (Naito et al., 2005). Pharmaceutical research laboratories are vastly interested in the identification of causative SNPs that are located within genes and their regulatory regions and also for indicative SNPs that may not be located within genes, but that are genetically linked/associated with genes that lead to disease and/or drug response phenotypes thus giving birth to a new branch of science called Pharmacogenetics. Pharmacogenetics is the field of research that attempts to unravel the relationship between genetic variation affecting drug metabolism (pharmacokinetic level) or drug targets (pharmacodynamic level) and interindividual differences in pharmacoresponse. Pharmacokinetic studies have shown the role of genetic variants of the cytochrome P450 (CYP) enzymes in the metabolism of drugs (Cichon et al., 2000). With the advent of lineage-independent genotyping through SNP-analysis, drug-manufacturing companies now hope to be able to utilize this new method to develop therapies for diseases in a more targeted fashion with minimum ADRs and to significantly narrow
down the number of test subjects by conducting genotyping instead (Johnson and Turner, 2005). Therefore, the major contributions of genomics to drug therapy include identification of new genes which code for vital proteins, identification of the most suitable targets for drug intervention and to help understand differential patient responses to drugs so as to target drugs more effectively during their development paving the way for individualized drug therapy (Yan, 2005).

2.2. SNP databases, examples and limitations
Deciphering the structure of human populations is very important for epidemiological studies since the rates of various phenotypes including diseases and ADRs vary in a very population specific manner owing to variations in frequencies of their genetic and non-genetic factors (Wilson et al., 2001; Risch et al., 2002). SNP databases cataloguing the exponential growth of data being generated worldwide on genotype data, allele frequencies and haplotype information for SNPs distributed across the entire human genome have facilitated resource integration and information retrieval, exchange and manipulation, thus aiding researchers to study the diversity of human genome variation and the genetic basis of disease and drug response (Isaksson et al., 2000). There are a number of public and private SNP databases to create information reservoirs of human SNPs, to confirm their existence and to measure their frequency in different populations which include National Centre for Biotechnology Information (NCBI)'s dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) (Sherry et al., 1999; Pruitt and Maglott, 2001), Ensembl (http://www.ensembl.org/) (Hubbard et al., 2002), HapMap Project (www.hapmap.org) (The International HapMap Consortium, 2003; Lin et al., 2005) and The SNP Consortium [TSC] (http://snp.cshl.org/db/map/snp) (Masood, 1999; Isaksson et al., 2000; White et al., 2001; Sachidanandam et al., 2001; Brookes et al., 2001; Thorisson and Stein, 2003) (which has been integrated into the Hapmap project, http://www.hapmap.org/cgi-perl/gbrowse/gbrowse).

The major limitation encompassing all the existing SNP databases/repositories is the fact that very low percentage of validation and confirmation have been carried out on the millions of SNPs documented in the databases (Jiang et al., 2003). The total number of samples used for each validation conducted is also a limiting factor. Moreover, prevalent technologies for assaying thousands of SNPs in patients and controls have not yet attained the desired robustness for this purpose. Not all SNPs are created equal in terms of functional implications. Therefore, it is essential to know
their effects through computational analysis before studying their involvement in diseases. 82% of the SNPs are found at a frequency of more than 10% in the global human population. But, most importantly, the microdistribution of SNPs in individual populations is not well known (Chakravarti, 1999; Chakravarti, 2001). Absence of extensive and reliable validated databases of SNPs specific for the populations subdivided into a multitude of large endogamous ethnic groups that are genetically less heterogeneous than conglomerate populations, like the population groups of the Indian subcontinent is therefore, a major concern for genomic and epidemiological studies. So far, all the enterprises undertaken globally to fish out SNPs in the human genome through creation and maintenance of SNP databases, like the Human genome project (HGP), the Hapmap project and other SNP projects including the dbSNP and TSC and smaller SNP mapping projects like the Seattle SNPs and Asian SNP project of Vita Genomics (Holden, 2002; The International HapMap Consortium, 2003; Greenhalgh, 2005; Tanaka, 2005; Shih-Hsin et al., 2007) have conspicuously ignored the huge diversity of the Indian population which envelops the largest human genetic diversity among comparable global regions, second only to Africa (Majumder, 1998; Majumder, 2001).

2.3. The Indian Population: Treasures within

The Indian population, representing 1/6th of the world population, is a potpourri of multiple cultures and evolutionary histories that provides a unique resource for complex disease analysis due to large family structures, endogamous groups, large patient pool with a majority being drug-naive. Currently, it comprises of 4693 communities with several thousands of endogamous groups, 325 functioning languages and 25 scripts (Indian Genome Variation Consortium, 2005) and thereby has the largest human genetic diversity among comparable global regions, second only to Africa (Majumder, 1998; Majumder, 2001; Roychoudhury et al., 2001; Bamshad et al., 2001; Basu et al., 2003; Kivisild et al., 2003; Cordaux et al., 2004; Kashyap et al., 2006; Sahoo et al., 2006; Sengupta et al., 2006; Thanseem et al., 2006). Furthermore, distinct religious communities, hierarchical caste and sub-caste structures and isolated tribal groups have contributed to the observed endogamy in subpopulations across the country. Ethnicity (tribal/non-tribal) and language are the major determinants of genetic affinities between the Indian populations, rather than geography (Piazza et al., 1980; Indian Genome Variation Consortium, 2008). The major spoken languages in India are grouped into Austro-Asiatic, Tibeto-Burman,
Indo-European and Dravidian language families (Grierson, 1927) while the people of India are classified into Caucasoid, Mongoloid and Australoid morphological groups with a small number of Negrito populations (Malhotra, 1978), as shown in Figure 1.1. Genetically isolated populations are vital for understanding the basis of complex diseases and identifying underlying genes (Indian Genome Variation Consortium, 2008). Therefore elucidation of the genetic diversity of these localized populations is a primary prerequisite for understanding the genetic basis of several critical community-specific complex disorders.

Population Diversity of India

India occupies only 2.4% of the world's land area while it supports over 17.5% of the world's population. India's population of approximately 1,028.7 million (Census of India, 2001; estimated 124,000 in areas of Manipur could not be covered in the enumeration) comprises approximately one-sixth of the world's population. India is expected to overtake China by 2030 and will then be the most populated country in the world. India has more than two thousand ethnic groups, all major religions and the four major families of languages (Indo-European, Dravidian, Austro-Asiatic and Tibeto-Burman languages) as well as a language isolate, the Nihali language (Summer Institute of Linguistics [SIL] International. Ethnologue report for Language Isolate. Retrieved on 2007-10-11 from the website http://www.sil.org/) spoken in parts of Maharashtra). Three millennia of language contact has led to significant mutual influence among the four language families in India and South Asia. Two contact languages have played an important role in the history of India: Persian and English (Bhatia and Ritchie, 2006). There are 1,652 languages and dialects in total (Mother Tongues of India). The languages of India primarily belong to two major linguistic families, Indo-European (whose branch Indo-Aryan is spoken by about 70% of the population) and Dravidian (spoken by about 22%). Minor linguistic families include the Austro-Asiatic and Tibeto-Burman families (with around 10 and 6 million speakers, respectively). Other languages spoken in India come mainly from the Dardic language family and also include minority languages such as Persian, Portuguese, French and English as lingua franca as well as a few language isolates. Kashmiri, considered a Dardic language, has 4.6 million speakers in India. There is also a language isolate, the Nihali language. According to the Census of India, 2001, 29 languages have more than a million native speakers, 60 have more than 100,000 and 122 have more than 10,000 native speakers. The largest of these is Hindi with 337
million while the second largest being Bengali with 207 million speakers. The Constitution of India recognises 22 languages of the 8th Schedule, spoken in different parts the country, the status of official language, viz. Assamese, Bengali, Bodo, Dogri, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Maithili, Malayalam, Santali, Marathi, Nepali, Oriya, Punjabi, Sanskrit, Santhali, Sindhi, Tamil, Telugu and Urdu.

Geographically, there are 29 states and 6 Union territories (UT) in the country with 604 districts and an estimated 256,000 Gram Panchayats. From the largest to the smallest, each State/UT of India has a unique demography, history and culture, dress, festivals, languages, etc. As per the 2001 census, 72.22% of the people live in more than 550,000 villages, and the remainder in more than 2000 towns and cities [http://www.censusindia.net/results/rudist.html]. Further complexity is lent by the great variation that occurs across this population on social parameters such as income and education.

This enormous diversity in population structure and geographical distribution points towards a possible existence of wealth of Human genetic diversity (within a variety of vital genes crucial for the maintenance of cellular homoeostasis including genes involved in DNA repair) among and between the Indian subpopulations waiting to be unearthed so as to uncover the actual genetic composition of the world population in general and the Indian scenario in particular.

2.4. DNA REPAIR

The veracity of the human genome is under continuous threat from exogenous and endogenous factors which include UV radiation, ionizing radiation, xenobiotics, water environment and cellular metabolites like activated oxygen and methyl group carriers which may attack and inflict genotoxic damages through the generation of various DNA lesions (von Sonntag, 2006; Friedberg et al., 2006). Therefore, DNA repair systems, an ensemble of several multienzyme, multistep processes, are fundamental to the maintenance of genomic stability and integrity in the face of such replication errors, environmental insults and the cumulative effects of ageing (Cheng et al., 1998).

There are several different DNA repair pathways in mammalian cells which may be summarized into the following two major categories which comprises of six roughly defined sub-categories viz.

1. Direct reversal of DNA lesions without the involvement of excision and resynthesis using photolyase or O-6-methyl-DNA-alkyltransferase.
2. Single and multi-step excision mechanisms which involve degradation of at least one damaged nucleotide followed by subsequent DNA re-synthesis which include (a) nucleotide excision repair pathway (NER) involving removal of bulky helix-distorting lesions, (b) mismatch repair (MMR) that corrects replication errors, (c) non-homologous end joining (NHEJ) and (d) recombination repair that commonly participates in the rectification of double-stranded breaks and (e) base excision repair (BER) that repairs the most wide-spread type of DNA damage viz. small non-bulky, mostly nucleo-base lesions like oxidative base damages, alkylation, base deaminations, base losses, single-stranded breaks, etc (Friedberg et al., 2006).

2.5. Base Excision Repair (BER)

The base excision repair (BER) pathway is the most pro-active DNA repair pathway that corrects DNA modifications which arise either spontaneously or from attack by reactive chemicals i.e. removal of DNA damage involving structurally non-distorting and non-bulky lesions like oxidized and ring-saturated bases, alkylated and deaminated bases, apurinic/apyrimidinic (AP) sites and certain type of mismatches (Lindahl et al., 1997). It involves the concerted effort of several repair proteins that recognize and excise specific DNA damages, eventually replacing the damaged moiety with normal nucleotide(s), restoring the DNA back to its original state.

The BER pathway comprises of removal of the damaged/inappropriate base(s) by a DNA glycosylase (e.g. hOGG1) that binds specifically to a target base and hydrolyzes the N-glycosylic bond, thereby catalyzing the excision of the base to generate a noncoding apurinic/apyrimidinic (AP/abasic) site (Krokan et al., 1997). The abasic site is subsequently recognized by 5'-acting AP endonucleases [viz. HAP1 or Ape1 (APEX1)] which incises the deoxyribose phosphate in the phosphodiester backbone immediately 5' to the lesion leaving behind a strand break with a normal 3'-hydroxyl group and an abnormal 5'-abasic terminus (Demple B and Harrison L, 1994). The resultant 5' abasic site is then replaced with the correct nucleotide(s) by two alternative processes, the Short patch/single nucleotide BER (SP BER) for single nucleotide gap filling by DNA polymerase β or Long patch/multinucleotide BER (LP BER) for the replacement of more than a single nucleotide (~7 nucleotides) (Klungland and Lindahl, 1997) [Figure 1.2].
2.5.1. Single nucleotide/short-patch BER (SP BER)

The repair of an AP site via single nucleotide/short-patch replacement involves the generation of a single nucleotide gap as a reaction intermediate created by the recognition and removal of DNA damage by the DNA N-glycosylases (Krokan et al., 1997). The resultant AP-site is recognized and processed by the apurinic/apyrimidinic endonuclease/exonucleases (APEs) (Taylor and Weiss, 1982) that hydrolyze the phosphodiester DNA backbone at the 5' side of the AP-site, generating 3'-OH and 5'-terminal deoxyribose-phosphate (5'-dRP) ends flanking the gap. DNA polymerase β, with partial support from DNA polymerase λ, catalyses excision of the 5'-sugar phosphate, the 5'-dRP by its 5'-dRPase activity and inserts one nucleotide into the repair gap (Matsumoto and Kim, 1995). Finally, ATP-independent DNA LIG3α-XRCC1 complex or the ATP-dependent LIG1-XRCC1 complex seals the ends (Sleeth et al., 2004; Fortini and Dogliotti, 2007).

2.5.2. Multinucleotide/long-patch BER (LP BER)

LP BER repairs patch sizes typically 2-8 nucleotides in length. It repairs sites with oxidized or reduced sugar groups and those with fragmented bases or sugars and thus is very important in the repair of DNA damage caused by ionizing radiation (Klungland and Lindahl, 1997). This pathway utilizes AP endonuclease for 5'-incision, but the AP site is not removed by DNA polymerase β since modification of the 5'-dRP by oxidation/reduction prevents its excision by Pol β. Repair of such sites requires proliferating cell nuclear antigen (PCNA) (Matsumoto et al., 1994; Kim et al., 1998), that encircles template DNA and forms a holoenzyme complex with DNA polymerases δ or ε in conjunction with replication factor C (RF-C: which loads the PCNA sliding clamp on the double helix), the PCNA loading factor (Biade et al., 1998). The polymerase adds a few nucleotides to the 3'OH end and generates a flap containing 5'-dRP end, which is then removed by FEN1 (Lieber, 1997) along with at least one adjacent nucleotide by phosphodiester bond hydrolysis to leave a gap of two or more nucleotides (Klungland and Lindahl, 1997; Kim et al., 1998). Finally, DNA ligase I (LIG1) covalently binds the fresh piece into the backbone to complete the process (Nilsen and Krokan, 2001; Dianov et al., 2003; Slupphaug et al., 2003; Sung and Demple, 2006).

2.6. CANCER

Cancer is a genetic disease resulting from gradual accumulation of changes in the DNA that activate proto-oncogenes and inactivate tumour-suppressor genes. The
resultant genetic instability is fuelled by DNA damage and errors made by the DNA maintenance and repair machinery (Hoeijmakers, 2001). Many cancers (retinoblastoma, Li-Fraumeni syndrome, familial adenomatous polyposis, breast cancer, SCCHN, etc) are heritable due to inheritance of specific variant allele/polymorphic variant and are associated with very high risk (Li et al., 1988; Kinzler et al., 1991; Kastan et al., 1991; Szabo and King, 1997).

Cancer is regarded as one of the major non-communicable diseases (NCD) and a main cause of death accounting for 7.9 million deaths worldwide (around 13% of all deaths) in 2007. According to the statistical data obtained from the World Health Organization (WHO) Global InfoBase (http://www.who.int/infobase/report.aspx), an online data storehouse for collection, storage and display of information on chronic diseases and their risk factors for all WHO member states, cancer killed more than 826,000 people in India in the year 2005 alone (Figure 1.3). Lung cancer is the major cause of cancer-related deaths while Squamous Cell Carcinomas of the Head and Neck (SCCHN) is a close second for men, and fourth for women of the Indian subcontinent accounting for a significant proportion of mortality contributing 3.4% of all deaths reported from India. On the other hand, cervical cancer is the most common cancer among Indian women both in terms of new cancer cases and cancer-related deaths, followed closely by breast cancer, as shown in Figures 1.4 and 1.5.

Recent advancements in cancer research have provided increasing evidences that cancer acts through a combination of high-risk variants in a set of low- and medium-penetrance genes rather than a few high penetrance genes. The ability to metabolize carcinogens or pro-carcinogens, repair DNA damage and control cell signaling and the cell cycle are important examples of low- and medium-penetrance genes fundamental to homeostasis which is why most cancers, including SCCHN and Breast Cancer are based on these main factors.

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

Squamous Cell Carcinomas of the Head and Neck (SCCHN) are the commonest forms of cancers of the head and neck that start in the cells forming the lining of the mouth, nose, throat and ear or the surface covering the tongue. The major head and neck sites include the oral cavity, the pharynx (nasopharynx, oropharynx and hypopharynx), the tongue (anterior 2/3rd and posterior 1/3rd or base of tongue), the larynx and the paranasal sinuses (Figure 1.6).
SCCHN is the 5th most common cancer worldwide (Parkin et al., 1993) with high mortality ratios among all malignancies accounting for 12% of all cancers in men and 8% of all cancers among women (Blot et al., 1996). It is primarily a disease of older men but sometimes also occurs in young adults of less than 40 years with a long history of tobacco and/or alcohol use. Hence, it is a major worldwide problem, especially in developing countries since tobacco usage rates is following an upward trend (Liu et al., 1998). Lifestyle factors like cigarette smoke contains large amounts of carcinogens like Polycyclic aromatic hydrocarbons (PAHs) such as benzo(a)pyrene, which damage DNA by covalent binding or oxidation following activation in vitro into benz(a)pyrene epoxide (Hecht, 1999). SCCHN, therefore is a result of multipolar molecular events induced by the effects of various carcinogens from habits/lifestyle factors like tobacco and alcohol use (Brennan et al., 1995; Cheng et al., 1998), influenced by environmental factors against a background of heritable genetic resistance or susceptibility. The resultant genetic damages consequently cause cell dysregulation with disruption in cell growth cycle, and/or mechanisms to repair cell damage or eliminate dysfunctional cells through apoptosis, carcinogen metabolism and cell-cycle control. Accumulation of such genetic changes over a period of time lead to cellular malfunctioning to such an extent that growth becomes autonomous and invasive mechanisms develop leading to carcinoma often via clinically evident malignant lesions (Scully et al., 2000).

According to WHO estimates, about two-thirds of the estimated 405,318 global number of new SCCHN cases arise in developing countries among which the highest rates are reported in South Asian countries including India and Sri Lanka. The Indian sub-continent accounts for one-third of the world burden. Being largely a lifestyle disease, the social habits like tobacco and alcohol use of the Indian population lie at the core of this national health problem. Betel leaf, betel nut, and various forms of tobacco are completely unregulated in India and used (inhaled or chewed) by almost 250,000,000 people. Tobacco chewing is common practice in these parts of the world, which have some of the highest incidence rates of oral and pharyngeal cancer. A mixture called pan (betel, nut and leaf, lime, catechu, tobacco and other additives) is very popular in India. It is chewed into a quid and buccal mucosa cancer usually develops at the site where the quid is kept and has most frequent contact (Borish et al., 1985; Parkin and Muir, 1992; Blot et al., 1996; Ries, 2008). Statistics show only 6% of head and neck cancer recurrence in patients who stop smoking in contrast to 37% of
head and neck cancer patients who continue smoking developing a second cancer. Those who both smoke and drink have a 15 times greater risk of developing mouth cancer than others. Alcohol drinkers are 6 times more likely than nondrinkers to develop mouth cancer. Further evidence of the importance of type of exposure and site of cancer is the high incidence of cancer of the hard palate in populations that practice reverse smoking, i.e., where the burning end of the cigarette is kept in the mouth during smoking, in parts of India, Sardinia, Venezuela and Panama (Reddy, 1974; Stoykewych et al., 1992). Therefore, SCCHN is a major problem in developing countries like India since tobacco and alcohol usage rates are rising strikingly (Liu et al., 1998). A significant fraction of the population in India is illiterate and unaware of hygiene of mouth. Therefore, poor hygiene of oral cavity (an important risk factor for SCCHN) is very commonplace rendering the population vulnerable to floor of mouth cancer, tongue and alveolar ridge neoplasms.

2.6.2. Breast Cancer

Breast cancer (Figure 1.7) is the primary subtype of cancer leading to death among women in developing countries. 13% out of the 58 million deaths worldwide in the year 2005 were caused due to cancer which included 502,000 deaths per year due to breast cancer. Well-established risk factors ascribed to breast cancer include early menarche, late menopause, age of first child’s birth, nulliparity and family history (FH) (Dumitrescu and Cotarla, 2005). The major susceptibility genes in high-risk families, namely Breast Cancer 1, early onset (BRCA1) and Breast Cancer 2, early onset (BRCA2) genes account for only a minority of the overall risk of breast cancer (Dapic et al., 2005). Furthermore, approximately only 10% of all breast cancer cases exhibit a familial pattern of incidence (Ford et al., 1998). Hence, the identification of genetic susceptibility factors that account from low to moderate breast cancer risk is an important step in the definition of individual risk to this malignancy (Costa et al., 2007a). DNA repair capacity determines cellular susceptibility to endogenous and exogenous substances. Some studies have demonstrated a strong association of higher levels of DNA damage and lower DNA repair capacity in breast cancer patients and in healthy women with a positive FH of breast cancer (Helzlsouer et al., 1996; Parshad et al., 1996; Jyothish et al., 1998; Dumitrescu and Cotarla, 2005; Garcia-Closas et al., 2006; Bau et al., 2007).

The natural history of breast cancer can be influenced by several factors. Under the influence of genetic polymorphisms, chronic exposure to higher levels of several
endogenous (e.g. oestrogens) and exogenous breast carcinogens resulting in consequent higher accumulation of DNA damage during an individual’s lifetime, may alter the waiting-time-to onset of the disease (Ladiges et al., 2003).

2.7. Defective DNA repair and cancer

DNA repair plays a very pivotal role in protection against mutagenic and cytotoxic effects of DNA damage. Lack of DNA repair activity has been shown to result in various combinations of defective embryogenesis, tissue-specific dysfunction, hypersensitivity to DNA-damaging agents, senescence, genetic instability and elevated cancer risk (Friedberg et al., 1997). DNA repair is therefore considered to play a central role in cancer biology whereby some individuals are at very high risk of cancer due to variations in genes of DNA repair and metabolism (Ma et al., 1995; Tlsty et al., 1995; Radman et al., 1995; Bohr, 1995; Mohrenweiser and Jones, 1998). Generally these damages are repaired by base excision repair (BER), which is a multistep process that involves the sequential activity of several proteins to cope with chromosomal damages that arise as spontaneous decomposition products or from reactions with metabolically or environmentally derived reactive chemicals like oxygen free radicals and alkylating agents. BER repairs sites with oxidized or reduced sugar groups and those with fragmented bases or sugars. Therefore, it is very important in the repair of DNA damage caused by ionizing radiation. Repair-deficient animals defective in BER have shown disruptions leading to lethality during embryogenesis (Xanthoudakis et al., 1996). Research on polymorphisms of DNA repair genes involved in BER activity have been shown leading to various abnormalities as is evident from study on animal models deficient in recognized BER components. For example, mutations in APEX1 (Xanthoudakis et al., 1996) and Polβ (Gu et al., 1994) have led to embryonic lethality. DNA glycosylase knock out embryonic stem cells show hypersensitivity to a variety of alkylating agents and mitomycin C (Engelward et al., 1997), etc.

Gene-environment interactions at the genotypic level can be examined through the identification and analysis of polymorphisms in DNA repair genes (Shen et al., 1998). Genetic differences in DNA repair capacity resulting from genetic polymorphisms influence risk of environmental carcinogenesis (Zheng et al., 2001; Holley et al., 2001; Zheng et al., 2002; Shen et al., 2002a; Shen et al., 2002b). Inherited polymorphisms in the genes controlling the cell cycle or functioning in the DNA
repair mechanisms may impair their function and contribute to genetic susceptibility. Inactivation or defect in DNA repair genes may be associated with increased cancer risk (Cheng et al., 1998). Genetic polymorphisms in DNA repair genes are very common events (Shen et al., 1998; Kuschel et al., 2002; Mohrenweiser et al., 2002), and some studies have shown a significant effect of some of these polymorphisms in DNA repair capacity (Matullo et al., 2001b; Clarkson and Wood, 2005; Pachkowski et al., 2006). Evidence of inherited abnormalities in DNA repair genes and genes controlling carcinogen metabolism has been found to underline increase in risk of cancers (Butkiewicz et al., 2001). Moreover, it has been suggested that DNA repair genes are associated to age related diseases which suggests that defects in DNA repair gene activity may also alter the waiting-time-to onset of diseases like cancer (Ladiges et al., 2003). Therefore DNA repair is considered to play a key role in cancer susceptibility whereby some individuals are at very high risk of cancer due to SNPs in crucial DNA repair genes (Ma et al., 1995; Tlsty et al., 1995; Radman et al., 1995; Bohr, 1995).

2.8. Epidemiological studies conducted so far on BER genes in SCCHN and Breast cancer

2.8.1. Global status

BER has been shown to be instrumental in protection from various diseases including carcinogenesis, aging, etc (Wilson, III and Bohr, 2007; Frosina, 2007). Therefore, imbalances in this compactly regulated pathway in the form of polymorphic variations or SNPs in vital DNA repair genes may result in insufficient DNA repair and increase in DNA breaks thus rendering the human system vulnerable to the debilitatory effects of grave diseases like cancers (Memisoglu and Samson, 2000; Mohrenweiser et al., 2003). This, exactly, has been shown to be the consequence in various studies conducted worldwide concerning the determination of susceptibility of polymorphisms within DNA repair genes in general including BER gene polymorphisms with the risk of cancers, including Breast cancer and SCCHN (Sturgis et al., 1999a; Sturgis et al., 2000; Duell et al., 2001; Olshan et al., 2002; Goode et al., 2002; Kim et al., 2002; Shen et al., 2002a; Moullan et al., 2003; Han et al., 2003; Shu et al., 2003; Varzim et al., 2003; Spitz et al., 2003; Smith et al., 2003a; Smith et al., 2003b; Tae et al., 2004; Deligezer and Dalay, 2004; Figueiredo et al., 2004; Benhamou et al., 2004; Rybicki et al., 2004; Hao et al., 2004; Chacko et al., 2005; Hung et al., 2005; Huang et al., 2005; Demokan et al., 2005; Geisler et al., 2005; Auranen et al., 2005;...
CJ(eview of Literature

2005; Gal et al., 2005; Hu et al., 2005; Shen et al., 2005a; Zhang et al., 2005b; Lee et al., 2005c; Ramachandran et al., 2006; Zhang et al., 2006; Kietthubthew et al., 2006; Shen et al., 2006; Han et al., 2006; Cao et al., 2006; Breast Cancer Association Consortium, 2006; Pachkowski et al., 2006; Li et al., 2006; Li et al., 2007; De et al., 2007; Berndt et al., 2007; Huang et al., 2007), although the huge diversity of Indian subpopulations have been widely overlooked in most of these literatures.

SCCHN is an excellent example of tumor due to defective gene-environment interaction contributed jointly by genetic susceptibility factors and two most important risk factors, viz. Tobacco and alcohol consumption (Schantz et al., 1988; Foulkes et al., 1996; Landis et al., 1998). Study on inter-individual differences in repair capacity have shown reduced DNA repair capacity among patients with SCCHN in comparison with control subjects (Spitz et al., 1989; Cheng et al., 1998). Polymorphisms in the genes involved in BER activity have been shown to influence the risk of several cancers including SCCHN (Sturgis et al., 1999a; Olshan et al., 2002; Tae et al., 2004).

On the other hand, epidemiological studies on human breast cancer have revealed relationship between impaired BER function due to polymorphisms in genes encoding BER core proteins and predisposition to breast cancer susceptibility among various world populations (Lunn et al., 1999; Abdel-Rahman and El-Zein, 2000; Duell et al., 2000; Duell et al., 2001; Hu et al., 2001; Goode et al., 2002; Kim et al., 2002; Wang et al., 2003; Moullan et al., 2003; Han et al., 2003; Shu et al., 2003; Smith et al., 2003a; Smith et al., 2003b; Deligezer and Dalay, 2004; Figueiredo et al., 2004; Dufloth et al., 2005; Thyagarajan et al., 2006; Zhang et al., 2006; Zhai et al., 2006).

2.8.2. National Status

Unfortunately, very few studies have been conducted concerning the association of DNA repair genes and susceptibility towards SCCHN and Breast cancer among Indian subpopulations inspite of the distinct and diverse composition of the Indian population make-up as explained earlier. Polymorphisms in the gene XRCC1 have been shown to be associated with childhood acute lymphoblastic leukemia (ALL) (Joseph et al., 2005). Recently, the distribution of some single nucleotide polymorphisms (SNPs) in the DNA repair gene XRCC1 were examined for correlations with Head and neck cancer in 110 oral carcinoma cases, 84 leukoplakia and 110 controls belonging to the South Indian population. The results suggested presence of the polymorphic variant of XRCC1 was associated with increased risk of
oral cancer compared to the wild type genotype with significant increase in risk among smokers and betel quid chewers (Ramachandran et al., 2006). In the only study conducted on Breast cancer association in the Indian population (south Indian subpopulations), Chacko et al. (Chacko et al., 2005) reported positive association of mutant genotypes of the gene XRCC1 with breast cancer risk.

2.9. Major genes selected for the study, their widely investigated SNPs and relevance of selected SNPs in cancer association studies

The Cytogenetic and structural details of the genes selected for the current study are provided in Table 1.1.

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Gene Symbol</th>
<th>OMIM ID</th>
<th>Chromosomal location</th>
<th>Strand</th>
<th>Gene Name</th>
<th>Contig ID</th>
<th>Region in Contig</th>
<th>Gene Length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>APEX1</td>
<td>328</td>
<td>14q11.2-q12</td>
<td>PLUS</td>
<td>APEX nuclease (multifunctional DNA repair enzyme 1)</td>
<td>NT_026437.11</td>
<td>1923130..1925766</td>
<td>2637</td>
</tr>
<tr>
<td>2</td>
<td>CNP</td>
<td>1267</td>
<td>17q21</td>
<td>PLUS</td>
<td>2',3'-cyclic nucleotide 3' phosphodiesterase</td>
<td>NT_010755.15</td>
<td>3841680..3855424</td>
<td>13745</td>
</tr>
<tr>
<td>3</td>
<td>ERCC2</td>
<td>2068</td>
<td>19q13.3</td>
<td>MINUS</td>
<td>Excision repair cross-complementing rodent repair deficiency, complementation group 2</td>
<td>NT_011109.15</td>
<td>Complement (18123064..18142047)</td>
<td>18984</td>
</tr>
<tr>
<td>4</td>
<td>FEN1</td>
<td>2237</td>
<td>11q12</td>
<td>PLUS</td>
<td>Flap structure-specific endonuclease 1</td>
<td>NT_033903.7</td>
<td>6865945..6870505</td>
<td>4561</td>
</tr>
<tr>
<td>5</td>
<td>LIG1</td>
<td>3978</td>
<td>19q13.2-q13.3</td>
<td>MINUS</td>
<td>Ligase I, DNA, ATP-dependent</td>
<td>NT_011109.15</td>
<td>Complement (20886893..20941750)</td>
<td>54858</td>
</tr>
<tr>
<td>6</td>
<td>OGG1</td>
<td>4968</td>
<td>3p26.2</td>
<td>PLUS</td>
<td>8-oxoguanine DNA glycosylase</td>
<td>NT_022517.17</td>
<td>9730705..9748342</td>
<td>17638</td>
</tr>
<tr>
<td>7</td>
<td>PCNA</td>
<td>5111</td>
<td>20pter-p12</td>
<td>MINUS</td>
<td>Proliferating cell nuclear antigen</td>
<td>NT_011387.8</td>
<td>Complement (5035599..5047268)</td>
<td>11670</td>
</tr>
<tr>
<td>8</td>
<td>XRCC1</td>
<td>7515</td>
<td>19q13.2</td>
<td>MINUS</td>
<td>X-ray repair complementing defective repair in Chinese hamster cells 1</td>
<td>NT_011109.15</td>
<td>Complement (16315682..16347933)</td>
<td>32252</td>
</tr>
</tbody>
</table>

Table 1.1 Details of the Genes included in the current investigation

2.9.1. BER genes

Research worldwide on base excision repair pathway and cancer risk have focussed upon genes encoding the key enzymes in this repair pathway: OGG1, APE1/APEX1, and the XRCC1 protein.

8-Oxoguanine DNA glycosylase (OGG1)

The gene OGG1 is located in the chromosomal region 3p26.2 that has repeatedly shown loss of heterozygosity in more than a few human cancers (Kohno et al.,
The product of this gene is a 345 amino acid protein that is involved in the repair of one of the most mutagenic lesions among base damages, i.e. 8-oxoguanine or 8-hydroxyguanine (8-hydroxy-2'-deoxyguanine (8-OH-dG)) present in DNA in its alternative tautomeric form. 8-Oxoguanine have been shown to cause base-pairing with adenine and result in G:C/T:A transversions in repair-deficient bacteria and yeast (Shinmura and Yokota, 2001). Among the validated sequence variants (Single Nucleotide Polymorphisms) described in sequence databases, the Ser326Cys polymorphism (dbSNP ID rs1052133) located in Exon 7 has been studied most frequently showing association between OGG1 genotypes and enzyme activity in a number of in vivo or in vitro studies, although the results have been inconsistent (Weiss et al., 2005). The rs1052133 variants show differences in DNA repair activity which has been implicated in increased risk of various cancers (Sugimura et al., 1999;Xing et al., 2001;Elahi et al., 2002). Interestingly, while one study found that the rs1052133 wild type allele (C)-containing OGG1 has a sevenfold higher activity for repairing 8-oxoguanine than the rs1052133 mutant allele (G)-containing OGG1 (Kohno et al., 1998), although two other studies could observe no association between OGG1 genotypes and enzyme activity (Dherin et al., 1999;Janssen et al., 2001). Genetic association studies on the polymorphism rs1052133/Ser326Cys (OGG1) have been conducted in various cancers associated with a wide range of risk ratios (Sugimura et al., 1999;Xing et al., 2001;Xu et al., 2002;Takezaki et al., 2002;Tsukino et al., 2004;Vogel et al., 2004;Arcand et al., 2005;Monteiro et al., 2005;Nohmi et al., 2005;Hashimoto et al., 2006).

Apurinic/apyrimidinic endonuclease (APEX1)
The gene APEX1 is located on chromosome 14q11.2–q12 (Figure 1.8.2). It encodes a 317 amino acid protein that plays a very pivotal role in the processing of abasic sites left from the incision of the damaged base. APEX1 is a multifunctional rate-limiting enzyme that not only is responsible for repair of AP sites but also functions as a redox factor maintaining transcription factors in an active reduced state (Xanthoudakis et al., 1992;Kelley et al., 2001). APEX1 cleaves the DNA backbone at the 5' side to the abasic site, exposing a 3'-Hydroxyl group and a 5'-deoxyribose phosphate group flanking the nucleotide gap (Izumi et al., 2000;Lu et al., 2001;Dianov et al., 2003;Tell et al., 2005). Additionally, APEX1 also performs processing of other 3' DNA termini that obstruct further gap filling or religation thereby allowing repair to be completed.
Several sequence variants were identified in the gene APEX1, the rate-limiting enzyme in the BER pathway (Ramana et al., 1998) including Asp148Glu (Exon 5, rs3136820 or rs1130409), Gln51His (Exon3, rs1048945) and Ile64Val (Exon 3, rs2307486) among which the Asp148Glu variant, the SNP rs1130409, that leads to an amino acid change from aspartic acid to glutamic acid (Asp148Glu) has been opined to have a possible involvement in hypersensitivity to ionizing radiation (Hu et al., 2001) and therefore have been targeted in association studies with the risk of various cancers (Hu et al., 2001; Hu et al., 2002; Zhang et al., 2006).

X-ray repair cross-complementing group 1 (XRCC1)

The gene XRCC1 is located at chromosomal location 19q13.2 (Figure 1.8.3). The product of the XRCC1 gene is a 633 amino acid protein which is essential for mammalian viability as well as a requisite for the efficient repair of single strand breaks and damaged bases in DNA. The XRCC1 protein has no known enzymatic activity and is believed to play a critical role as a scaffold protein for both single-strand break repair and base excision repair activities (Lindahl and Wood, 1999; Thompson and West, 2000; Goode et al., 2002). It preferentially binds DNA containing nicks and short gaps and has been known to physically interact with various main and accessory BER proteins like DNA polymerase β, APEX1, OGG1, PCNA, etc through the mediation of its BRCT domains (a region having extensive homology to BRCA1) and other parts as well while its absence leads to a substantial decline in the levels of its partner ligase III thereby reducing the ligation efficiency in SP-BER. Furthermore, XRCC1 loss also leads to reduction of genetic stability including increased frequencies of spontaneous and/or induced chromosome translocations and deletions (Thompson et al., 1982; Carrano et al., 1986; Zdzienicka et al., 1992; Caldecott et al., 1994; Caldecott et al., 1995; Cappelli et al., 1997; Op, V et al., 1998; Taylor et al., 2000; Goode et al., 2002; Rouse and Jackson, 2002; Hang and Singer, 2003; Marsin et al., 2003; Fan et al., 2004; Mani et al., 2004; Diano et al., 2004; Fan and Wilson, III, 2005; Campalans et al., 2005; Wong and Wilson, III, 2005; Parsons et al., 2005; Nazarkina et al., 2007; Hitomi et al., 2007; Mani et al., 2007). Among the validated SNPs in XRCC1, the most widely investigated are rs1799782 (Arg194Trp) in Exon 6 and rs25487 (Arg399Gln) in Exon 10 (Callebaut and Mormon, 1997; Hung et al., 2005). The rs1799782 (XRCC1) locus lies in the linker region that separates the NH2-terminal domain from the central BRCT1 domain and supposedly binds to the amino terminus of APEX1 (Shen et al., 2005b). Existence of rs1799782
variant allele results in a non-conservative substitution in a hydrophobic region of XRCC1 which has been shown to be associated with lower *in vitro* sensitivity of bleomycin and benzo(a)pyrene diol epoxide (Xanthoudakis et al., 1992; Wang et al., 2003). The SNP rs25487 is located within the central breast cancer susceptibility gene 1 product (BRCT1, residues 315-403) at the carboxylic acid terminal side of the PARP (polyadenosine diphosphate-ribose polymerase-interacting domain) and interacts with polynucleotide kinase (Shen et al., 1998; Tuimala et al., 2002; Fan et al., 2004). In phenotypic studies, the rs25487 mutant allele (A) has been shown to be associated with higher levels of genotoxic damage including higher levels of aflatoxin B1-DNA adducts and higher bleomycin sensitivity (Lunn et al., 1999; Abdel-Rahman and El-Zein, 2000; Duell et al., 2000; Wang et al., 2003). (Lunn et al., 1999; Matullo et al., 2001b; Wang et al., 2003), although one study did not find any such association (Palli et al., 2001). Furthermore, the rs25487 homozygous mutant genotype (AA) was found to be related to elevated sister chromatid exchange (SCE) frequencies among healthy smokers compared to cells with rs25487 homozygous wild type genotype (GG) (Abdel-Rahman and El-Zein, 2000; Duell et al., 2000). Studies on the genetic association of the SNP rs1799782 with cancer have been reported in quite a few literatures around the globe with varying results (Sturgis and Miller, 1995; Sturgis et al., 1999a; Duell et al., 2001; Stern et al., 2001; Lee et al., 2001b; Ramachandran et al., 2006; Kietthubthew et al., 2006) while rs25487 has also been studied widely (Abdel-Rahman et al., 2000; Shen et al., 2000; Butkiewicz et al., 2001; Duell et al., 2001; Divine et al., 2001; Ratmasinghe et al., 2001; Matullo et al., 2001a; Lee et al., 2001b; Gal et al., 2005; Ramachandran et al., 2006; Kietthubthew et al., 2006), although literatures on the Indian population are barely minimum.

**DNA Ligase I (LIG1)**

The gene DNA Ligase I (LIG1), located in chromosome 19 (Figure 1.8.4), codes for the main mammalian DNA ligase, i.e. Lig1, that participates in DNA replication, nucleotide excision repair (NER) and LP-BER pathways. Lig1 is an ATP-dependent enzyme that utilises the energy of phosphate anhydride hydrolysis to make a phosphodiester bond completing BER by ligation of the single-stranded nick in DNA (Martin and MacNeill, 2002; Friedberg et al., 2006; Tomkinson et al., 2006; Hitomi et al., 2007). Structurally, the Lig1 protein comprises of an oligonucleotide-binding fold domain, a catalytic adenylation domain and a unique large DNA-binding domain that,
when bound to DNA, folds back to completely encircle the DNA (Pascal et al., 2004; Dore et al., 2006; Pascal et al., 2006; Nandakumar et al., 2007). Only one study has so far investigated the association of SNPs identified in the gene LIG1 with the risk of cancer (Lee et al., 2008). Much more exploration needs to be conducted in this regard with various cancers given the huge importance of this gene in the BER pathway.

**Flap Structure-Specific Endonuclease 1 (FEN1)**

The Flap endonuclease 1 (FEN1) gene is located in chromosome 11 (Figure 1.8.5). It encodes a Mg\(^{2+}\)-dependent structure specific endonuclease that has unannealed flapped 5'-end as preferred natural substrates (Friedberg et al., 2006). Flap structures are the intermediates of different processes of DNA metabolism, such as DNA recombination, Okazaki fragment maturation during replication of lagging strand, as well as strand displacement DNA synthesis in base excision repair. FEN1 also possesses 5'-exonuclease activity and gap endonuclease activity. FEN1 is known to interact physically and functionally with a number of DNA replication and repair proteins such as the proliferating cell nuclear antigen (PCNA), helicase/nuclease Dna2, Werner Syndrome (WRN) and Bloom syndrome (BLM) proteins, replication protein A, apurinic/apyrimidinic endonuclease 1 (APEX1), DNA polymerase β, poly(ADP-ribose) polymerase 1, high mobility group protein 1, integrase of human immunodeficiency virus, transcription coactivator p300, chromatin proteins, cyclin-dependent kinases (Cdk1, Cdk2, Cyclin A), etc. FEN1 activity is significant for maintaining the integrity of repeat sequences in genome. Recent data shows the correlation between the abnormality of hFEN1 activity and arising/progression of neurodegenerative and cancer diseases (Nazarkina et al., 2008). In long-patch base excision repair (LP BER), a damaged nucleotide is displaced into a flap and removed by FEN1. FEN1 is a genome stabilization factor that prevents flaps from equilibrating into structures that lead to duplications and deletions. As an endonuclease, FEN1 enters the flap from the 5' end and then tracks to cleave the flap base (Liu et al., 2004).

**Proliferating Cell Nuclear Antigen (PCNA)**

The product of the gene Proliferating cell nuclear antigen (PCNA) [Figure 1.8.6] is an accessory factor that chiefly functions as a polymerase processivity clamp involved in replication and BER processes mediated by DNA polymerase δ. It also functions as a scaffold protein that gets physically associated with the products of a number of BER genes like DNA Polymerase δ, APEX1, DNA polymerase β, FEN1, LIG1, etc and
attracts BER enzymes towards the repair lesions (Hang and Singer, 2003; Maga and Hubscher, 2003; Oyama et al., 2004; Fan and Wilson, III, 2005; Ko and Bennett, 2005; Dionne and Bell, 2005; Xia et al., 2005; Friedrich-Heineken et al., 2005; Moldovan et al., 2007; Almeida and Sobol, 2007; Mortusewicz and Leonhardt, 2007).

The only study so far on the analysis of SNPs in the genes FEN1 and PCNA has suggested the potential of some SNPs in the regulation of gene expression of these proteins (Ma et al., 2000), although no further research has so far been carried out in this regard carrying forward these vital findings, especially with respect to cancer association.

2.9.2. Other genes included in the study
Apart from the vital BER genes mentioned above, the following genes were also included in this study.

**Excision Repair Cross-Complementing Rodent Repair Deficiency, Complementation Group 2 (ERCC2)**
The gene ERCC2 is located in the chromosomal location 19q13.3 (Figure 1.8.7). It encodes the ERCC2/Xeroderma pigmentosum Type D (XPD) protein, which is one of the seven genetic complementation groups that forms an essential component of the Nucleotide excision repair (NER) pathway, a major DNA repair pathway that removes photoproducts from UV radiation and bulky adducts from a huge number of chemicals, cross-links and oxidative damage through the action of 20 proteins and several multiprotein complexes (Ma et al., 1995; Sancar, 1995). The XPD protein functions as an ATP-dependent 5'-3' helicase joint to the basal TFIIH complex and participates in the local unwinding of DNA helix to allow RNA transcription machinery to access the promoter and to permit the NER machinery to access the lesion (Egly, 2001; Friedberg, 2001). Several studies suggest that XPD protein may participate in the repair of ionizing radiation–induced oxidative damage (Satoh et al., 1993; Leadon and Cooper, 1993). XPD is a highly polymorphic gene and correlation of its polymorphisms and cancer risk have been extensively studied (Clarkson and Wood, 2005; Benhamou and Sarasin, 2005). There are two important genetic polymorphisms in ERCC2, one causing amino acid change in codon 312 (Asp to Asn) (SNP ID rs1799793) and the other in codon 751 (Lys to Gln) (SNP ID rs13181) and there is evidence that subjects homozygous for the variant genotypes of XPD have suboptimal DNA repair capacity for benzo(a)pyrene adducts and UV DNA damage
(Spitz et al., 2001; Qiao et al., 2002). The XPD polymorphism, rs13181 located in exon 23, which consists of an A to C substitution in the coding region results in a Lys751Gln substitution in the important domain of interaction between XPD protein and its helicase activator, inside the TFIIH complex (Coin et al., 1998). The polymorphism rs13181 has been found to be functionally significant since studies have showed that it is associated with alterations in DNA repair efficiency (Cheng et al., 1998; Lunn et al., 2000; Wei et al., 2000; Seker et al., 2001; Qiao et al., 2002; Au et al., 2003; Au et al., 2004; Benhamou and Sarasin, 2005; Costa et al., 2007b). Therefore, the polymorphism rs13181 has been studied for its role in various cancers as potential susceptibility factors (Sancar and Tang, 1993; Sturgis et al., 2000; Lunn et al., 2000; Chen et al., 2000; Butkiewicz et al., 2001; Spitz et al., 2001; Caggana et al., 2001; Goode et al., 2002; Tang et al., 2002; Benhamou and Sarasin, 2002; Hou et al., 2002; Xing et al., 2002a; Xing et al., 2002b; Liang et al., 2003; Mort et al., 2003; Justenhoven et al., 2004; Wrensch et al., 2005; Sobti et al., 2007), although no such report is available on the north Indian subpopulation cluster for the risks of SCCHN or Breast cancer.

**2',3'-Cyclic Nucleotide 3' Phosphodiesterase (CNP)**

Mammalian CNP (Figure 1.8.8) encodes a 400-amino acid polypeptide composed of a C-terminal catalytic fragment (CF) and a 150-amino acid N-terminal extension (Lee et al., 2001a). CNP constitutes 4% of total myelin protein in the central nervous system. CNP is expressed at lower levels outside the central nervous system, and it has been found to associate with mitochondria and cytoskeletal proteins (Bifulco et al., 2002; Lee et al., 2005a; Lee et al., 2006). Homozygous CNP-knockout mice were viable, and showed no overt abnormalities up to a certain extent (Lappe-Siefke et al., 2003). Myelin in young knockout mice was of normal ultrastructure and protein composition, except for the absence of CNP. Older knockout mice developed progressive motor deficits and died prematurely because of diffuse brain axonal swelling and neurodegeneration leading to hydrocephalus (Lappe-Siefke et al., 2003).

It is not known whether the late onset of neural disease is caused by an absence of the CNP protein (e.g., as a structural component of myelin or the cytoskeleton) or an absence of CNPase enzymatic activity. Interestingly, 2',3' cyclic nucleotides are not present in appreciable amounts in mammalian cells, implying either that the catalytic activity of CNP is a distraction or that 2',3' cyclic mononucleotides are not the physiological substrate (Braun et al., 2004).
CNP is an attractive candidate and proposed susceptibility gene for schizophrenia since the mutant allele (A) of an exonic SNP (rs2070106) within the CNP gene showed reduced expression compared with the wild type allele (G) in the brain, and was reported to be associated with schizophrenia (Peirce et al., 2006). In a recent study, the SNP rs2070106 (CNP) was found to be potentially associated with schizophrenia (Voineskos et al., 2008).

Very few studies have comprehensively investigated associations between genetic variants in BER pathway genes with SCCHN and breast cancer risk. Not many genes involved in the Long or Short Patch Base Excision Repair pathway has so far been studied as low penetrance gene candidates for correlations with susceptibility to the occurrence of SCCHN. Furthermore, it is very unfortunate that despite the important roles of genes involved in BER in environmental carcinogenesis, no extensive genetic study has been performed in the Indian sub-populations where SCCHN and breast cancer are major cancers, both in incidence frequency and mortality, and where the general population is most closely associated with the most important risk factors. Variations of functional polymorphisms might be pertinent for defining populations and monitoring community health (Bamshad et al., 2004). Therefore an extensive study on the medium and low penetrance genes contributing towards the susceptibility of this dreaded disease is a prime requisite to devise therapeutic interventions for its eradication from the north Indian society and the global scenario as a whole.

2.10. Y chromosome microdeletions and DNA ploidy in SCCHN
Deletions of Y chromosome-specific genes like Sex determining Region Y (SRY) [Yp11.32], Zinc Finger protein, Y-linked (ZFY) [Yp11.31], Blood pressure on SHR Y 1 (BPY1) [Yq11.2], Selected Mouse cDNA on Y (SMCY) [Yq11.22], RNA Binding Motif protein 1 (RBM1) [Yq11.23] and Blood pressure on SHR Y 2 (BPY2) [Yq12.1], etc have already been reported in various cancers like prostate cancer, bladder cancer, pancreatic cancer, etc (Perinchery et al., 2000; Khaled et al., 2000; Wallrapp et al., 2001; Zeng et al., 2004; Bianchi et al., 2006). Similar studies need to be conducted on SCCHN to detect any such microdeletions of Y-chromosome. Several studies on DNA content analysis in SCCHN patients have depicted aneuploid cell populations in tumour tissues, which has a high clinical relevance since it serves as a vital prognostic factor for determining the clinical outcome of the patients (Zatterstrom et al., 1991; Stell, 1991; Seoane et al., 1999; Kolotas et al., 1999). It would
thus be interesting to verify whether a similar pattern of aneuploidy exists in Peripheral Blood Mononuclear cells (PBMC) of SCCHN subjects.

In the present study, the initial endeavour was to examine the nation-wide and subpopulation-specific distribution of SNPs within the genes playing major roles in the BER pathway viz. AP endonuclease (APEX1), DNA Ligase 1 (LIG1), X-ray repair Cross Complementing in Chinese Hamster 1 (XRCC1), in addition to the gene 2',3'-Cyclic Nucleotide 3' Phosphodiesterase (CNP).

Population stratification and differences in selection and human adaptation are major confounders in association studies. Effects of population stratification in disease association studies will be negligible if cases and controls are drawn from the same cluster used in the present study even if they do not belong to the same ethnic group. Hence, studies involving sampling cases and controls from the same population cluster need no additional corrections or confounding effects of population stratification and would additionally increase the power of such association studies (Indian Genome Variation Consortium, 2008). Therefore, the present study on the identification and validation of SNPs in DNA repair genes was followed by case-control based genetic association study with salient functional polymorphisms within the BER genes AP Endonuclease (APEX1), Flap Endonuclease 1 (FEN1), Proliferating Cell Nuclear Antigen (PCNA), DNA ligase 1 (LIG1), X-ray repair Cross Complementing in Chinese Hamster 1 (XRCC1), human Oxo-Guanine Glycosylase 1 (hOGG1) in addition to the NER gene Excision Repair Cross-Complementing Rodent Repair Deficiency, Complementation Group 2 (ERCC2) and the gene 2',3'-Cyclic Nucleotide 3' Phosphodiesterase (CNP) in ethnically matched subpopulation clusters from north India to verify susceptibility status with respect to SCCHN and Breast cancer risk.

Additionally, relative expression of the genes selected for the study at transcriptional levels was investigated in SCCHN subjects with respect to normal healthy controls to assess the overall status (downregulation or upregulation) of these genes with reference to SCCHN.

Furthermore, in addition to the study pertaining to the analysis of polymorphisms in selected DNA repair genes within Indian subpopulations, the present investigation also focussed upon the exploration of Y-chromosome microdeletions and DNA ploidy status among SCCHN patients.
The outcome of this research initiative may have widespread applications for future epidemiological and public health related investigations on the Indian population.
Figure 1.2 Schematic representation of the Base excision repair (BER) pathway

Figure obtained from Kyoto Encyclopedia of Genes and Genomes (KEGG)
Figure 1.3 Leading causes of death in India

Data retrieved online from World Health Organization (WHO) Global InfoBase (http://www.who.int/infobase/report.aspx)
Figure 1.4 New cancer cases in the year 2005

Data obtained online from World Health Organization (WHO) Global InfoBase (http://www.who.int/infobase/report.aspx)
Figure 1.5 Leading causes of cancer deaths in India in the year 2005

Data obtained online from World Health Organization (WHO) Global InfoBase (http://www.who.int/infobase/report.aspx)
Figure 1.6 Major Head and Neck sites for Squamous Cell Carcinomas of the Head and Neck (SCCHN)
Figure 1.7 Breast cancer
Figure 1.8.1 Map of the gene human OGG1 showing its location in chromosome 3. An enlarged view of the gene loci shows the exons, introns and SNPs found within the gene.

Figure obtained from NCBI’s Map Viewer (www.ncbi.nlm.nih.gov/projects/mapview/)
Figure 1.8.2 Map of the gene APEX1 showing its location in chromosome 14. An enlarged view of the gene loci shows the exons, introns and SNPs found within the gene.

Figure obtained from NCBI’s Map Viewer (www.ncbi.nlm.nih.gov/projects/mapview/)
Figure 1.8.3 Map of the gene XRCC1 showing its location in chromosome 19. An enlarged view of the gene loci shows the exons, introns and SNPs found within the gene.

Figure obtained from NCBI’s Map Viewer (www.ncbi.nlm.nih.gov/projects/mapview/ )
Figure 1.8.4 Map of the gene LIG1 showing its location in chromosome 19. An enlarged view of the gene loci shows the exons, introns and SNPs found within the gene.

Figure obtained from NCBI’s Map Viewer (www.ncbi.nlm.nih.gov/projects/mapview/ )
Figure 1.8.5 Map of the gene FEN1 showing its location in chromosome 11. An enlarged view of the gene loci shows the exons, introns and SNPs found within the gene.

Figure obtained from NCBI's Map Viewer (www.ncbi.nlm.nih.gov/projects/mapview/)
Figure 1.8.6 Map of the gene PCNA showing its location in chromosome 20. An enlarged view of the gene loci shows the exons, introns and SNPs found within the gene.

Figure obtained from NCBI's Map Viewer (www.ncbi.nlm.nih.gov/projects/mapview/)
Figure 1.8.7 Map of the gene ERCC2 showing its location in chromosome 19. An enlarged view of the gene loci shows the exons, introns and SNPs found within the gene.

Figure obtained from NCBI's Map Viewer (www.ncbi.nlm.nih.gov/projects/mapview/)
Figure 1.8.8 Map of the gene CNP showing its location in chromosome 17. An enlarged view of the gene loci shows the exons, introns and SNPs found within the gene.

Figure obtained from NCBI's Map Viewer (www.ncbi.nlm.nih.gov/projects/mapview/)