Summary

Cancer is a disease with its origin in both inherited and somatic genetic changes. Studies on cancer etiology are being benefited from the human genome map information which has helped in identifying the sites of losses, instability and rearrangement, acquired in the predominant chromosomal changes in cancer, to act as marker positions to unravel the presence of relevant gene(s). Breast cancer is the most common form of cancer among women. Two genes, BRCA1 and BRCA2, have been implicated in the familial form of the breast cancer which incidentally covers only approximately 10% of all the breast cancers. The etiology of the sporadic form of the breast cancer, prevalent in the majority, covering approximately 90% of the cases, is not well defined. Also, none of the known genes have been implicated in this form of cancer as yet. The information available of the human genome sequences in chromosome specific manner allows one to explore the experimentally observed regions of chromosomal loss for the probable presence of a tumor suppressor gene(s). This positioning could also pave the way for identifying and cloning the relevant genes and study their expression and involvement in the disease.

The aim of the present work was to look for a candidate region in chromosome 17 which could be explored for a presence of a previously unknown sequence or a gene relevant to sporadic breast cancer situation, and authenticate the presence of such a sequence and its association with sporadic form of breast cancer. Since involvement of more than one gene is attributed in the causation of any sporadic form of tumors, some of the relevant candidate genes and an innocuous gene for comparison purposes were included in the study to understand if their studied region could be a target for somatic mutation in sporadic breast tumors. The study was carried out with following objectives: (i) to carry out low resolution mapping of chromosome 17p13.3 and 17q21-23 regions by studying loss of heterozygosity (LOH) and microsatellite instability (MSI), to define the minimal lost region (MLR) using thinly placed, syntenic, PCR amplifiable, STS markers in the blood and tumor samples of sporadic breast cancer cases and to draw a correlation, if any, with clinical parameters; (ii) to
fine map the informative region by in-silico analysis on build 33 sequence map, carry out gene prediction with ab-initio approaches and validate the presence of a sequence or a gene partially or fully (if possible) which shows either allelic instability (LOH/MSI) or germline/somatic mutation(s); (iii) to study the somatic variation in tumor samples in selected region(s) of the following candidate genes i.e. BRCA2 (known to depict germline mutations in familial breast cancers with a role in repair and genome stability), IFN-gamma, IL-6, TGF-beta-1 (known to play a role in immune surveillance, cellular response and genome stability) and KRTHB6 (an innocuous hair specific keratin gene - a candidate for control comparison).

The study was carried out in three major parts interlinked with each other, the outcome of which is distributed in three different Chapters. The subject of CHAPTER 1 was to study alterations at 17p13.3 and 17q21-24 regions of the chromosome in breast tumors, by allelotyping with six STS ‘repeat markers’ (two on 17p13.3 and four on 17q21-23). The aim was to identify a most prevalent region of allelic loss amongst the studied regions to fine map the region at a low resolution, in these tumors. The two loci, one at 17p13.3 (D17S379) and the other at 17q21-23 region (D17S934), showed a significantly high allelic loss (LOH) of 56% and 54%, respectively, when compared to other four markers used in this study. However, the informativeness increased considerably when the frequency of LOH for all the six markers was analyzed together. It was observed that 82% of the 34 sporadic breast cancer patients showed LOH at least at one out of the six loci studied. An analysis between the LOH at 17p13+17q21-23 and MSI in the short and long arm regions together, showed a statistically significant association (p<0.009) which later guided the study to choose a specific region from the 17q arm for surveying the flanking regions for gene prediction in the second chapter (CHAPTER 2) of this presentation. The study in the breast tumors with matching lymphocytic DNA of 34 patients was examined for LOH within the short arm of chromosome 17 at band 13.3 at two polymorphic loci (D17S5 and D17S379). The D17S5 marker positioned at 17p13.3 region showed 25% loss in informative cases and a reasonably high frequency (48%) of allelic loss when the loss (LOH) for this locus and D17S379 was combined for analysis in sporadic breast tumors. These observations suggested that this region (17p13.3) could harbor important tumor suppressor gene(s), whose inactivation may contribute to tumor development or progression. The study, however, could not demonstrate any association between LOH of 17p13.3 or independent loss of each marker (D17S5 and D17S379) with tumor stage and
lymph node status, the two clinicopathological parameters studied. The tumors also showed microsatellite instability (MSI). Both the markers, D17S5 and D17S379, showed the microsatellite instability (MSI) in 12% and 6% of the studied tumors, respectively. The percentage of the MSI phenotype at 17p13.3 region was 20%. Tumors which presented with MSI in 17p13.3 region also did not show any association with any stage of the tumor, in this study. The presence of the MSI phenotype at 17p13.3 region suggested the existence of a ‘mutator’ phenotype in the studied tumors probably due to a defect in mismatch repair gene or gene expression in the studied tumors.

Genetic alterations were studied on 17q arm in sporadic breast tumors at four syntenic STS ‘repeat ‘markers’ (D17S855-D17S934-D17D787-D17S948), between 17q21-23 region, covering an approximately 18 Mb region on build 33 sequence map. It was observed that D17S855 marker was lost in 26% of informative sporadic breast tumors. This marker incidentally is intragenic to BRCA1 gene. Among the other studied markers from the 17q region, D17S934 marker showed a high frequency (54%) of loss (LOH) in the studied cases. The absence of microsatellite instability at D17S934 locus, in this study, with a high frequency of loss, made the somatic alteration (LOH) at D17S934 locus important to explore the locus further. The flanking markers on either side of D17S934 locus (D17S855 and D17S787, D17S948) also showed a loss (LOH) thus, suggesting the importance of the whole region (17q21-23) in sporadic breast cancer pathogenesis. The Markers on 17q21-24 regions were also evaluated for the microsatellite instability phenotype. Marker D17S855, D17S787 and D17S948 showed MSI in 10%, 35% and 9% of the studied tumors, respectively. The MSI at D17S787 locus was significantly associated with the stage II tumors in this study (p<0.009), suggesting that the MSI at D17S787 locus could have acted as an early event in the studied tumors. It was further observed in this study that the loss (LOH) of D17S948 marker was significantly (p<0.030) associated with the stage III tumors. The same marker, D17S948, also exhibited the instability (MSI), supporting further the conclusions drawn of the 17q21-23 region to be very prone to genomic loss or instability.

Although the sample size in this study was small, nevertheless, the main aim of the study was to focus on a specific chromosome region for in-silico analysis and carry out ab-initio gene prediction, based on the initial observations of high LOH and MSI in specific regions of
chromosome 17. Observation of a significant correlation between the MSI and LOH phenotypes for some of the markers studied and a correlation of these with the stage II and III of sporadic form of breast cancer was again interesting. The two minimal deleted (MLR) regions (I- D17S5-D17S379; II- D17S934-D17S787) observed in this study became the focus for in-silico mapping of the region. The availability of the high-resolution human genome sequence map made the job possible and this became the part of this work in the form of the second chapter (CHAPTER 2).

In-silico analysis of the studied marker region for positioning was carried out, as a part of the second chapter (CHAPTER 2) of this work, on build 33-sequence map (representing the integrated map of all other aligned map i.e, RNA map, disease map, sequence map, BAC map, contig map etc.). The analysis of marker regions was done for 100 Kb (window size) sequence region for each marker sequence.

The results of in-silico studies in this chapter, apart from the location of known and predicted genes, indicated towards a follow up of the exploratory studies of ab-initio gene prediction and its experimental validation from more than one region of chromosome 17. However, it was decided to focus on D17S934 marker region on 17q 21-23 which showed high loss (LOH) in sporadic breast tumors in this study. This region was subjected to ab-intio gene identification and experimental validation. Two classes of computational methods which were used for gene identification were: i) Similarity based methods such as BLAST, which compares segments of genomic DNA with known genes, proteins, or expressed sequence tags (ESTs). ii) Ab-intio gene finding programs such as GENSCAN and FGENES, which predicts gene structure on the basis of statistical models of exon-intron and splice signal composition without using sequence similarity information. It was decided to first predict the gene by computational approaches and then validate the predicted exons by experimental approach.

For the computational prediction of the gene, the sequence of D17S934 marker region was derived from the BAC Clone (Accession no. AC015936.14), which included the marker sequence as well. The average gene size sequence (approximately 11.5 kb) analyzed by GENSCAN resulted in a full-length gene with a putative promoter sequence. The predicted gene showed three Exons, first being very short (the shortcoming of the software). The rest of two exons were average sized. The predicted gene sequence was analyzed for repeat sequence and a possible presence of EST in the area by using repeat masker and BLAST against dbEST.
database. The repeatmasker analysis of the predicted gene sequence showed that this sequence carried interrupted microsatellite sequences. The Exons of the predicted gene also carried the repeated sequences (GGT)n. The predicted gene sequence carried a previously reported partial transcript in intron-1 region of predicted gene. The known transcript from this region was derived from the retinal cell cDNA. The primers designed on the exon boundaries the previously unknown of predicted gene were used for experimental validation through exon specific amplification both in genomic DNA target and mRNA (cDNA) target of the predicted exons. The Exon-2 and Exon-3 of the predicted gene showed specific amplification in the genomic DNA. However, mRNA target converted in cDNA was informative only for exon 3 of the predicted gene (Third party annotation Accession no. TPA BK 000585). Further, the Exon 3 specific product of predicted gene (BK000585) from cDNA was sequenced by direct PCR product sequencing which resulted in a matching sequence data resembling the genomic Exon 3 sequence of predicted gene. The BLAST analysis of this transcript against dbEST and MGC database (Mammalian gene collection) did not result in a significant match, suggesting this transcript to be a novel one. The presence of this was tested in lymphocyte, normal breast and tumor tissue DNA and mRNA (cDNA). It was confirmed repeatedly that probably a partial expressed product (transcript) of a novel gene in this region, of which Exon-3 is a part could be a target for mutation in the studied patients with sporadic breast cancer. The partial-expressed EST observed in this study carried a repeat sequence and the BLAST (tBLASTx) analysis of the transcript(s) did not yield significant similarity with any known protein sequence in the existing databases. The BLAST analysis of the complete predicted gene (genomic) against EST databases revealed the localization of another EST in the intron 1 of the predicted gene. This suggests that this region is potentially transcribing in nature and requires construction of full-length transcripts from this region. The importance of the partial transcript positivity for mutations in the paired sample study in sporadic breast cancer patients consistently, suggests the involvement of this transcript and as yet unexplored region around D17S934 marker in sporadic breast cancer. In near future it is desired to obtain a full length of this gene unless this encoded mutant transcript turns out to be a member of a group of small nuclear RNA with a role in sporadic breast cancer. The need in future will be to explore its function as well. This study of finding novel transcript by Geneprediction coupled with RT-PCR also lays foundation to identify the novel gene (complete or partial) from those regions
which apparently are gene barren but show involvement in sporadic cancer through LOH and MSI studies.

Based on the results from the first chapter, it was further realized, to investigate the relevant and non-relevant (unrelated to cancer) candidate gene regions to understand the germline status of these genes and evaluate generalized genomic instability by somatic mutation analysis, as a part of CHAPTER 3, since most of the study samples were advanced stage tumors and showed genomic instability which correlated with the marker status mentioned in detail already. This study observed that out of the studied blood and tumor DNA samples of sporadic breast cancer patients, 2 out of 30 (6 %), 8 out of 29 (27 %), and 4 out of 25 (16 %) cases showed somatic changes in the exon 2 (5'UTR) region of BRCA2, (CA)n of IFN-gamma-Intron 1 and –174 nucleotide position of the promoter of IL-6 genes, respectively. A relationship between the somatic change in these selected regions of a ‘caretaker gene’, BRCA2 and the cytokine genes and the LOH at 17p & q regions was observed only between the nucleotide polymorphism at –174 position in IL-6 and combined LOH for D17S855+D17S787, (p<0.030), D17S855+D17S948, (p<0.007), D17S787+D17S948 (p<0.038). Further, the trend of contraction of (CA)n allele in Intron 1 of IFN-gamma gene in tumor tissues along with somatic changes in exon 2 of BRCA2 did suggest the advantage these somatic changes could provide to tumor growth and progression but not initiation of the tumor. However, the germline polymorphism in 11 out of 30 (36%) cases in exon 2 of BRCA2 gene, the presence of (CA)12 repeat in 16 out of 29 (55%) cases and –174 G/C polymorphism in 8 out of 25 (32%) in IL-6 provided an interesting piece of information with respect to the germline genetic background of these genes which again could play a role in not only tumor progression but also development and susceptibility to tumor formation. The role of genetic background as a reason for susceptibility, at least of some of the studied genes/regions of genes, if not all the studied genes/gene regions, was supported by the observation in another critical gene implicated in breast cancer in recent years, the TGF-beta1 gene, where in Case-Control studies, the prevalence of a specific genotype showed a 4 fold more risk in sporadic breast cancer cases possessing this genotype (-1340G>A) for TGF-beta1 promoter region (II·). A study carried out in out of cases of sporadic breast cancer, in an innocuous, hair specific keratin gene, KRTHB6, not involved in breast cancer, did not show both somatic instability and germline genotype susceptibility. This suggested that involvement of
polymorphism in BRCA2 in the 5'UTR region in the exon2, a repeat size of (CA)n known to influence the expression of IFN-gamma, -174 promoter polymorphism in IL-6 and of TGF-beta1 provide a susceptible genotypes for sporadic breast cancer cases and somatic changes in some such cases could again be helping tumor progression aggressively, the observations which need to be followed up in future for further information.

To conclude it was very encouraging in this work to first narrow down the region of loss in tumor samples from sporadic breast cancer patients (though limited in number) and correlate the loss or instability to stages of cancer for better prognosis. However, exploring one of the minimum lost region (MLR) in 17q21-23 around a marker, D17S934, for assigning unannotated genes by prediction and further validating the presence of a coding exon (a partial gene with its transcript Accession no.) with mutations in tumor samples, has set the future tone to carry out studies to completely fish out apparently a meaningful gene involved in sporadic form of breast cancer.