CHAPTER 3

Role of Quad-SNP in Cancer
3.1 Introduction

Genetics of disease etiology is an essential component behind every research objective. The human genome is a constellation of tightly regulated phenomenon but alterations in these can lead to mayhem. One of the important components of this constellation is the DNA structure. Normal DNA is double stranded and takes the shape of Watson-Crick double helix. But certain regions of the genome have transient or stable secondary structure arising as a function of sequence, interaction between two sequences or superhelicity. Formation of structures like Z-DNA, cruciform, H-DNA, triplet repeats, tetraplex, slipped structures etc have all been linked to genome instability (1) and human diseases (2). Expansion of triplet repeats lead to diseases. Expansion refers to an excessive number of repeat number of a particular sequence occurs thereby affecting the gene expression. For e.g. till 28 repeats Huntington disease doesn’t occur whereas at 38 repeats the disease sets in (3). More than 60 neurological diseases are known to occur due to triplet expansion (4). H–DNA sequences have been shown to be mutagenic and prone to double strand breaks (DSB) (5). Z DNA have been associated with large deletions and DSB (6). Germline deletions and recombination are known to be caused by quadruplex occurrences (7,8). Since genomic instability is the hallmark of cancer therefore all DNA secondary structures will be making the genome susceptible to cancer (9). G-quadruplexes are one such class of DNA structures that have been shown to be associated with various diseases especially cancer (10). Phenomenon like replication stalling and formation of single stranded overhangs can initiate instability via these motifs.

Cancer is a state when cells divide in an uncontrolled manner and thereby lead to various genomic malfunctions. Cells are usually destined to undergo a programmed cell death. When cells evade this pathway and divide indefinitely, body homeostasis is disrupted. Dynamic changes in the genome lead to cancer. Random mutations are the major contributors to the disease. Major changes include gain of function for oncogenes and loss of function for tumour suppressor genes. Six major hallmarks of cancer have been listed out; (i) self sufficiency for growth signal (ii) evasion of apoptosis (iii) sustained angiogenesis and (iv) limitless replicative potential (v) tissue invasion and metastasis (vi) insensitivity to anti-growth signals (11). Cancer can be called a combination of various diseases as numerous genes and pathways get de-regulated thereby leading to multiple anomalies. Additional hallmarks of cancer have been repeatedly proposed but the major emphasis has been laid
upon genomic instability. Genomic instability in cancers can be of many types, viz. chromosomal instability, microsatellite instability and increase in per base mutation rate (12;13). These can be manifested to translocations, deletions, amplifications and whole chromosome aneuploidy. Molecular basis of these instabilities is under scanner (14;15) and role for secondary structures has been indicated (9) (as mentioned above).

G-quadruplexes were initially discovered in the telomeric DNA (16). The telomeres occur at ends of chromosomes and help in maintaining chromosomal integrity. The chromosome spans a double stranded form throughout till the extreme 3′ end. This single stranded DNA at ends of chromosome is lost at each replication cycle. Telomere forms a ‘G’ rich cap (TTAGGG) at these ends which thereby conserves important DNA (17;18). Telomerase reverse transcriptase (TERT) and other associated proteins like TERF1, TERF2, TINF2, TERF2IP, ACD and POT1 together protects the chromosome ends from fusing together. TERT uses the underlying RNA template of single stranded overhang to add new bases to the chromosome ends. This continues till the telomere length is shortened to a critical limit. Onset of aging takes on when the telomeres are shortened and finally the cells are destined to undergo apoptosis. Usually the telomerase enzyme is found in small amount within normal cells. The cancerous cells express unusually high levels of telomerase and thus chromosome ends keep growing indefinitely. It was seen that if G-quadruplexes were formed and stabilised in the telomeres, then telomerase failed to act (19). Though this was shown in-vitro but it opened avenues for various clinical and molecular research in telomere biology. TMPyP4, PIPER, anthroquinone analogues are some of the molecules that have been targeted to stabilise G-quadruplex as a cure for cancer. Quarfloxin is a drug that has gone into clinical trials(20).

Reports on genome-wide occurrence of G-quadruplex in the promoters took on the realisation for the regulatory potential of these motifs (21;22). Focus shifted from telomeres to promoters, pertaining to cancer biology, with the findings of a G quadruplex in c-myc promoter which is trans-activated during cancer (23). c-myc is over expressed in more than 80% of the cancers. The P1 promoter of this gene has a nuclease hypersensitive region -142bp to -115bp upstream of it which comprises of a 33-mer G rich sequence (5′-TGGGAGGGTGGGGAGGGTGGGGAAGGTGGGA-3′). This sequence is capable of forming a complex G-quadruplex (24;25). Disruption of a specific G → A residues increases the basal expression level three times and stabilization of the quadruplex by TMPyP4 leads to reduced expression.
Foldback tetraplex structures have been found in the promoters of many oncogenes like sc-abl, c-ets, c-fes/fps, c-fgr, c-fos, c-jun, c-kit, c-mos, c-rel, c-sis, c-src, c-yes, and the vav (26) proto-oncogene and characterised quadruplexes have been seen in VEGF, BCL-2, HIF1A, MYB, PDGFA, RB1 and TERT (27). A parallel G4 motif is formed in the promoter of VEGF and nucleolin (a G4 binding protein) binds to it (28;29). c-kit has two G quadruplex forming sequences in its promoter at -87 to -109 and -140 to -160bp from the transcription start site (30). Extensive structural characterisation has shown strict sequence specificity for structural integrity in these G4 motifs.

As seen previously that alterations within G-quadruplex can alter its structural integrity (31). SNPs being the most prevalent form of alteration that occurs within the genome has profound effect on the G quadruplex structure (32;33), thereby altering the G4 mediated gene regulation. SNPs have been thoroughly studied with its association to cancer. Almost all known cancers have SNP associated as a marker or as a causative change in the coding sequence or in the regulatory sequence. With advent of newer genotyping and sequencing tools the number of cancer associated SNPs increased stupendously. Novel mutations associated with cancer are discovered by sequencing whereas the frequency of an allele in the diseased population is determined by genotyping techniques like Taqman Assay, Illumina Goldengate assay, Infinium genotyping assay, GeneChip human mapping assay, iPLEX assay etc (34). The vast numbers arouse the need to orderly catalogue them in a database. The Cancer Genome project is one such initiative. It collects cancer tissues and sequences them followed by proper database update. These data are available for public use.

3.2 Aim

Cancer is not a single genic disorder but a combination of different phenomenon. It is not difficult to apprehend from the above literature that another method of cancer regulation by SNPs might be mediated by G-quadruplex motifs. Presence of SNP in G4 motifs (Quad-SNP) of oncogenes might release the repression or repress the release of polII from the gene promoter, thereby leading to aberrant gene expression. We analysed genome-wide Quad-SNPs of known oncogenes and genotyped them in head and neck cancer samples to detect cancer associated Quad-SNPs. Furthermore we showed functional correlation between the aberrant gene expression and altered G4 structure mediated by the SNP.
Chapter 3

Role of Quad-SNP in Cancer

3.3 Materials and methods

3.3.1 DNA Isolation

Sample sectioning was done by microtome (Leica GM1505) and taken to doctor for diagnosis of stage of cancer. TNE buffer was added to selected sample to make a total volume of 7ml. 0.35ml of SDS (10%) was added and mixed gently followed by incubation for 15mins at 65°C. On returning to room temperature proteinase K was added to a final concentration of 50μg/ml and kept for overnight digestion at 37°C. Following day equal volume of phenol was added, mixed and centrifuged at 3000rpm for 10mins. Supernatant was taken out and washed twice with chism and centrifuged. DNA was precipitated from the final supernatant with ethanol and KOH. Precipitated DNA was pelleted, washed and finally eluted in 500μl T.E. (pH 7.4).

3.3.2 Picogreen Quantitation

The isolated DNA sample was quantitated using PicoGreen dsDNA quantitation reagent (Molecular Probes). PicoGreen reagent a sensitive fluorescent nucleic acid stain which is specific for double stranded DNA (dsDNA). It can measure a minimum of 5ng/μl concentration of DNA. The DNA to be quantitated is measured on a scale based on absorbance measured from standard DNA.

Lambda (λ) DNA was used as a standard. Original concentration was 3μg/ml which was diluted to working stock of 1μg/ml. This was serially diluted from 1000 pg/μl to 1.9pg/μl with 0.1 T.E buffer in the following manner (Figure 3.1).
Figure 3.1  Schematic representation of λDNA standard preparation, required for Picogreen dsDNA quantitation.

25μl of standard was added to each well in a 384-well black plate. Stock DNA was diluted 200 times and 25μl was aliquoted into the plate. The PicoGreen reagent was diluted 1:200 times with 1X TE and 25μl of the diluted reagent was added to λ DNA standards and sample DNA in the 384-well black plate. The samples were mixed properly and incubated at room temperature for 5min under dark. Fluorescence was measured using a spectrofluorometer microplate reader (BMG Labtech) at 480 nm (Excitation) and 520 nm (Emission).

3.3.3 Illumina GoldenGate genotyping

Goldengate platforms allow multiplex genotyping of many SNPs in a single reaction. Maximum of 1536 polymorphisms can be studied together (35). We selected the SNPs from Cancer genome Project which catalogues genes that have been implicated in cancer. Transcription start site coordinate of these genes were downloaded from UCSC and accordingly ±1kb sequence was extracted. Our in-house PERL based programme “G4 finder” was used to search for PG4 motifs. A typical PG4 motif constituted of the general formula (G₃L₁₋₇ G₃L₁₋₇ G₃L₁₋₇G₃), where G₃ represents the run of three continuous G and L represents loop consisting of any of the bases. SNPs lying within the PG4 motifs (Quad-SNP) were mapped and only SNPs which were validated according to dbSNP were included for genotyping.

A custom panel was created by selecting and submitting the list of loci to Illumina, which is evaluated with the ADT (Illumina Assay Design Tool). The SNPScore file provided
predicted success information, validation status, and minor allele frequencies from published studies. We use the ADT SNPScore file output to refine an initial assay panel to include desired assays that are predicted to have a high likelihood of success. We selected the SNP’s with a score of 0.4 and above. The final list of were then submitted to Illumina for OPA (Oligo Pool Assay) design and synthesis.

The basic working behind the assay is as follows (Figure 3.2). For each locus three oligonucleotides (oligos) were synthesized: two allele specific oligos (ASOs) corresponding to the two alleles of the SNP, and a locus specific oligo (LSO) just downstream of the SNP, which specifies the SNPs position. The DNA is activated by biotinylation. The biotinylated DNA is then added to OPA containing all the primers. These are kept for annealing with the DNA. The annelid primers are then extended and ligated. The ASO and LSO sequences also contain target sequences for a set of universal primers (P1 and P2 for two ASOs and P3 for LSO), while each LSO also contains a particular address sequences (the illumicode) complementary to the sequences attached to beads on the sentrix array matrix (SAM). To carry further out the assay, a set of fluorescently labeled universal primers P1 and P2 (Cy3 and Cy5 labeled respectively) is added and multiplex-PCR is carried out. These are purified and eluted. The fluorescent products were then hybridized overnight with beads on the (SAM). The address sequences within the PCR amplicons hybridize to their related sequences on the beads. The SAM was then scanned for the signals from the Cy3 or Cy5 at each address sequence. These are then mapped to particular locus, and the presence of Cy3, Cy5 or both signals on a given bead type indicates AA, BB or AB genotypes.
3.3.4 SNaPshot

SNaPshot is an assay to detect the genotype at a particular locus by allele specific single base extension. The tumour and the control DNA, of the samples where significant results were seen, were amplified for the seven shortlisted loci. Primer details as in Table 3.1. The PCR products were purified by PEG purification. The purified products were subjected to single base extension by SNaPshot primers (Table 3.1). Condition for SNaPshot PCR is as follows:
Chapter 3  Role of Quad-SNP in Cancer

5X DB  0.8 µl
Reaction Mix (RM)  0.5 µl
Primer (2pm/λ)  2.0 µl
MQ  1.7 µl

PCR Condition:  94°C for 5’,
35 cycles of  96°C -10″, 50°C-5”, 60°C-30″
  4°C-10′.

Following this step CIP treatment (calf intestinal phosphate) was given at 37°C for 60 mins and denaturation at 75°C for 15 mins. Labelled samples were then run on ABI 3100 sequencer, Applied Biosystems.

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Forward Primer (5’→3’)</th>
<th>Reverse primer (5’→3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRPC5</td>
<td>CAGAAAATCCGTACCAGGGG</td>
<td>AGCAGCTCTGCTTCTTTTG</td>
</tr>
<tr>
<td>GHR</td>
<td>GTGAAACGGGAAGGAGGAGGAT</td>
<td>TTCGCTCTGCTCTCTGCAAC</td>
</tr>
<tr>
<td>CLCN2</td>
<td>AGTCCGATGTCAGACAGCTCC</td>
<td>TCCCTGCGGATGCCCCGCG</td>
</tr>
<tr>
<td>FZD 1</td>
<td>CAGGAAGGCCGGACACAGAC</td>
<td>TTGCGATTCATGAAAGCAGC</td>
</tr>
<tr>
<td>RAPGEF 5</td>
<td>AGAAGAGGCAACACAGCTCGT</td>
<td>TTTTGAGGGTGCCACCCCCCTG</td>
</tr>
<tr>
<td>CTNNB1</td>
<td>AAACCTCGGCAACAGGCCAGA</td>
<td>TACAGGAGCCGGATGAGCAG</td>
</tr>
<tr>
<td>HERC 5</td>
<td>AAAGCCAGACATCACTTAGATA</td>
<td>CTGACACACGCTTCCAGAG</td>
</tr>
</tbody>
</table>

Table 3.1  Table contains primer sequences used to amplify target regions for SNaPshot validation of Illumina genotypes.

3.3.5 Biophysical experiments

JASCO-810 spectropolarimeter was used to record the CD spectra. Oligos were diluted to 3µM in sodium cacodylate buffer (100mM K+) and were taken in 1ml cuvettes with path length 1cm. Prior to experiments oligo were heated to 95°C and gradually cooled to room temperatures overnight, and stored at 4°C. CD scans were taken in a wavelength range of 220-320nm at constant temperatures (20°C) and scanning speed of 200nm/min. For each oligo three scans were taken and spectrum of the buffer was subtracted. While analyzing, the data was zero corrected at 320nm.
For calculating melting temperatures same dilution of oligo were used. Oligo were subjected to a temperature of 95°C for 10mins and then slowly cooled to 25°C at a rate of 1°C/mins. UV absorbance was noted at 295nm.

### 3.3.6 Cloning

Basic pGL3 vector, from Promega (Madison, WI, USA), was used to clone promoters FZD1. They were cloned between SacI and BglII sites upstream of the luciferase promoter, in the same direction as the gene. Promoter region was PCR amplified from normal genomic DNA, with PCR primers containing digestion site. Primers used were:

- \textit{FZD\_FP}–5′-GCGAGCTCGCCACCACCACCTACTTCCTC-3′,
- \textit{FZD\_RP}–5′-GAGATCTGGCACAAAGTTCCCAGCTC-3′

Both PCR product and vector were digested, purified and ligated. Clones were screened and sequenced for verification. Positive clones were then subjected to site-directed mutagenesis, using Quick Change Site-Directed mutagenesis Kit from Stratagene, to get desired mutation in single base within the PG4 motif.

### 3.3.7 Transfection and luciferase Assays

Primate fibrosarcoma cell line HT1080 was maintained in MEM. Breast cancer cell line MDAMB 231, and laryngeal cancer cell line HEP2 were maintained in DMEM. Prior to Transfection, 12-well plate was seeded with 2.5x10^5 cells to achieve optimum confluency. 1.5µg of plasmid was transfected per well with lipofectamine2000 (Invitrogen, Invitrogen BioServices India Pvt. Ltd, Whitefield, Bangalore), according to manufactures protocol. For Transfection control 5ng of PGL4.73 was also co-transfected. Cells were lysed after 24hrs and luciferase assay was done using the dual luciferase assay kit from Promega, according to the manufacturer’s protocol. Renilla counts were used for normalisation. All experiments were done in triplicate at room temperature. Transfection for all cell lines was done in a similar manner.
3.4 Results

3.4.1 SNP scoring and assay designing

Cancer genome project (CGP) (http://www.sanger.ac.uk/research/projects/cancergenome/), led by Prof. Mike Stratton and Andy Futreal, catalogues the genes associated with oncogenesis. It uses the high-throughput mutation detection techniques to identify acquired sequence variants in the cancer sample as compared to normal and hence tabulates genes deprecative to the development of human cancers (36). Only genetic evidence is the criteria for inclusion in the Cancer Gene Census (CGC) but it has to be reported in at least two independent studies. Cancer tissues of different origin are sequenced/genotyped for the purpose. 90% of cancer genes show somatic mutations and 20% show germline mutations.

We had two basic reasons for using this database:

(i) It provides an exhaustive list of genes causally implicated in oncogenesis. Such a list is not essential for an initial screen,

(ii) The genes have been genetically implicated in CGC but their functional relevance is not unravelled. The Quad-SNPs provided a good opportunity for functional validation of the reported data.

There were 4869 genes in the CGC catalogue where mutations associated with cancer were detected. Of these, we found genomic co-ordinates information for 4534 genes from UCSC. Flanking 1kb sequence from the transcription start site (TSS) of these genes was downloaded and in-house developed program called G4 finder was run on sequences. Of 4534, we got 709 genes with at least 1SNP within 1kb from TSS. These SNPs were then queried with respect to their presence within putative G4 (PG4) motifs. We found 1035 Quad-SNPs present in 270 genes. Amongst them 307 were validated as per dbSNP criteria.

Illumina’s Assay design tool scored the Quad-SNPs along with other SNPs in an ongoing study to give SNP scores. This indicates the feasibility of including a particular SNP in an assay as follows.

- SNP score < 0.4 Low success rate, high risk to OPA
- SNP score 0.4 - 0.6 Moderate success rate, moderate risk to OPA
- SNP score 0.6 - 1.1 High success rate, low risk to OPA
Chapter 3
Role of Quad-SNP in Cancer

Among our SNPs, 98 SNPs had a favourable score of $\geq 0.5$ which translates into it having high probability of being genotyped. In order to query for more SNPs we relaxed the favourable score to $\geq 0.4$ score which included more SNPs in our study. Thus our final study was with 171 Quad-SNPs in 138 genes. Interestingly when these SNPs were also queried for their distribution in the HapMap population, we found that most of these SNPs had a minor allele frequency of $\leq 0.1$. Thus it can be inferred that they were rare alleles in the normal population.

3.4.2 Stringent data analysis was done through Beadstudio V3 version

Scanned images of genotyping from each sample was analysed using Beadstudio. A stringent score cut off of 0.8 was taken. Samples below this threshold were excluded from analysis. The OPA contained total of 1536 SNPs but we filtered out 171SNPs lying within PG4 motifs. 53 of these loci either did not achieve required Gentrain score or did not work among these samples and hence were removed from subsequent analysis. Four duplicate samples were included per plate to ascertain concordance of data. These were 98-99% concordant. Six samples were removed prior to analysis since they didn’t work in most samples (80%) and thus didn’t pass the missing-ness test. After removing duplicates, 121tumour and 121 controls remained for final analysis.

3.4.3 Cancer associated Quad-SNP allele are rare in normal population

All SNPs were subjected to basic allelic test using the PLINK software. The focus of PLINK is purely on analysis of genotype/phenotype data, so there is no support for steps prior to this (e.g. study design and planning, generating genotype or CNV calls from raw data). 11 SNPs were found to have minor allele absent from control population as listed in Table 3.1. Although not looked in detail but it can be speculated that these SNPs could be lethal in normal population. However, four of these were found to be monomorphic in our sample set and thus were removed from downstream analysis.
Table 3.2  **List of 11 SNPs genotyped using Illumina platform whose minor allele was absent in control samples.** Of these, four SNPs were found to have minor allele absent from both tumour and normal thus being monomorphic in the samples genotyped in our study.

Remaining seven SNPs were found to have minor allele absent in normal population but present in very low frequency in cancer tissues (Table 3.2). For example, in *CLCN2*, a chloride channel factor, the major allele “C” was present with an allele frequency of one in HapMap population; in other words, the minor allele was absent in the normal HapMap population. As expected we found all samples had the major genotype “CC” while one affected case had the “CA” genotype; the minor genotype “AA” was absent. Similar observations were made for the other six SNP PG4 detected in this study. Viewing the Genoplots from beadstudio shows one or two samples from the tumour cohort to have the minor allele (Figure 3.3). The red and the blue dots represent the samples with homozygote genotype, whereas the purple coloured dots denote the heterozygote genotypes. Minor allele was always seen to be present in the form of heterozygotes. The minor homozygote was absent from the population.
Figure 3.3  Genoplots from the Beadstudio genotyping module for six SNPs whose minor allele was absent from the control population. The purple dots represent the heterozygotic tumout samples which contain minor allele. Red and blue dots represent the two homozygotes.

On querying for the SNP frequency in normal population, it was observed that the major allele frequency is very high among the SNPs. The minor allele, if present, is a rare minor allele. Details of the allele frequency is summarised in Table 3.3.
Table 3.3  List of 7 Quad-SNP whose minor allele was found to be associated with cancer phenotype. Given are the details of their allelic distribution in HapMap population and our experimental cohort.

### 3.4.4 Validation of the genotype data

Since the genotyping was done on a large number of SNPs, we sought to validate our findings using a different platform. Cross-platform validation would increase the confidence in the genotype of the SNP in the samples. We selected SNAPSHOT using ABI platform for validating aforesaid observations.

SNAPSHOT works on the basic principle of single base extension of primer corresponding to the queried genotype. Primers are designed to (n-1) position of the template where “n” is the SNP position. Addition of fluorescent labelled ddNTPs ensures only single base extension from the primer. Based on these guidelines we analysed the peak intensities for the reaction.

For FZD1, two samples 2772T and 4283T had a G→T polymorphism according to the Illumina results. We prepared primers on the other strand and therefore expect a CA
genotype. Viewing SNaPshot results showed us a double peak, green and black, indicating CA genotype in the tumour sample. Simultaneously the normal samples i.e. 2772N and 4283N had only single green peak indicating homozygous state (Figure 3.4).

(a)

(b) GGGGAGAACGCCGGTGCT(A/C)GGGCGCGCCGCGGG (rs2232157)

Figure 3.4 3.4(a) SNaPshot results for rs 2232157 showing heterozygous genotype in the tumour tissues and homozygous major genotype in normal tissue. T denotes tumour and N denotes normal. 3.4(b) Sequence of the PG4 motif in which the SNP, rs2232157, resides. The polymorphic base has been highlighted by red font.

For the remaining six SNPs Illumina genotyped alleles could not be validated by SNaPshot and thus were considered erroneous.

3.4.5 FZD1 contains a G4 motif in its promoter

After SNaPshot validation of FZD1 we decided to validate the structural integrity of the PG4 motif in the FZD1 promoter and to elucidate the effect of the SNP on the structure. We synthetically synthesized the oligos with presence of SNP different alleles within PG4. Circular dichroism is a method to precisely detect the presence of G quadruplexes (37;38) and its conformational polymorphisms (parallel/ anti parallel). We thus tested in-vitro for the formation a Quadruplex structure by the G-rich oligos in presence of K+ cation (for stability). The sequence adopted a peak at 260nm indicating the presence of a parallel G-quadruplex [Figure 3.5(a)]. SNP in FZD1 was present within the loop region of the PG4 sequence.
[Figure 3.4(b)], thus we do not expect a drastic difference in CD and melting profiles. In line with our hypothesis we found that both the oligos with variant allelic forms, adopted the quadruplex structure with subtle change in stability [Figure 3.5(a)]. Similarly the melting temperature of the quadruplex showed almost negligible change with “A” allele having melting point at 64.2°C and “C” allele having 63.2 °C.

3.4.6 Functional validation of the FZD1 variant

FZD1 is a key signalling molecule of the WNT/β-catenin pathway, which plays an important role in cancer etiology (39). Under normal circumstances the β-catenin is inhibited by glycogen synthase kinase (GSK), adenomatous polyposis coli (APC) and associated proteins like axin/conductins. When signalling cascades break, wnt binds to frizzled family proteins, thereby activating Dishevelled which suppresses the β-catenin downregulation (40). The free β-catenin binds to other transcription factors thereby increasing expression of genes like c-myc, cyclin D1 etc resulting in cancer.

In case of FZD1, the major allele “C” was present with an allele frequency of one in normal individuals; all 116 unaffected samples had the major genotype as “CC”. In affected cancer tissues, two out of 116 had C to A substitution. The functional promoter of FZD1 (41) was cloned upstream of luciferase either with or without the Quad-SNP (C to A substitution). In order to test expression of the promoter we selected cancer cell line derived from the human larynx-HEP2. The reporter construct harbouring the detected substitution gave >5-fold decrease in expression relative to the promoter construct without the substitution. In order to further check whether the adverse effect on gene expression was found on other types of cancers we checked expression of the reporter construct in fibroblasts cells and breast carcinoma cells. Here also we found that the substitution resulted in very low promoter activity of FZD1: about 12.4 fold and 3.24 fold less in HT1080 cells and MDAMB cell lines respectively (Figure 3.5).
Figure 3.5  

3.5(a) Left panel shows a typical CD plot for a G quadruplex structure. A peak is observed near 260nm. 3.5(b) Green coloured bars indicate the “A” allele and shows a repression in promoter activity as compared to the “C” (black) allele consistently in all the 3 cell lines.

3.5 Discussion

SNPs studied with respect to cancer are typically of two types. First are those SNPs which are silent and do not affect the phenotype, but are found associated to cancer. This is particularly because of the reason that SNPs are usually not randomly inherited, but as blocks or haplotypes from one generation to another. Second type are those that have a very low minor
allele frequency but with high penetrance of the allele. These are the disease; cancer in our case, causing SNPs. Mechanism by which SNPs cause the disease can be many viz. change in coding sequence, alteration of transcription factor binding site, disruption of regulatory sequences etc. Non-B DNA structures are known to act as gene regulatory regions and many times cause instability (2;42) leading to evolution or diseases (2;43). Genomic instability is the hallmark of cancer (14) and is frequently caused by these DNA secondary structures (9). SNPs have the likelihood of altering these structural motifs, thereby altering downstream gene regulation (33). These dis-regulated genes may lead to loss of cellular homeostasis and hence cancer.

G quadruplexes is one of the very well characterised non B DNA structure (44) with many faceted functional activities (45;46). PG4 motifs have been identified in the promoter of many genes related to cancer(27;47) and have been observed to be prevalent in promoter of oncogenes (48). Our study thus focussed on oncogenic promoter PG4 motifs. The Quad-SNP associated with cancer phenotype was shown to alter the gene expression and the structural stability of the motif. We could perform the Illumina Goldengate experiment with only a fraction of the oncogenes due to constraint of assay design. Inclusion of more SNPs for study would enhance the potential of capturing more structure altering cancer causing Quad-SNPs. Some of our Goldengate results of positive association could not be validated. This could be partly because of two reasons; erranous genotype calling by Illumina Beadstudio or erroneous primer extension in SNaPshot. G-quadruplexes are known to hinder the extension of primer during PCR amplifications (49) . This can lead to allele specific extension for one of the allelic forms whereas amplification of the other is compromised, leading to detection of only one allele.

Population genomics of disease etiology can act as a major indication for selection of the allele and evolution in the population. HapMap consortium contains data for a spectrum of SNPs in representative global population (50). The disease causing allele of the causative SNP is seen to be absent in the normal population. It is only present in 2 samples of our tumour populations. It is interesting to note that when present, the minor allele, exists in a heterozygote. The minor homozygote is totally absent from the population, both cancer and control. This indicates that the minor allele is less frequent but has high penetration level. Thus cancer onset is there when the allele is present as a heterozygote. The minor homozygote might have been de-selected in evolutionary passage or the genotype might be lethal and hence doesn’t exist. Chance retention may lead to life-threatening disease states.
like Cancer, which although in small amount in our study show its occurrence only in cancer affected individuals.

At any polymorphic locus, the allele frequency in our subject population is an important consideration to address in order to identify all the allelic forms at that locus. In case of alleles with high frequency it is not difficult to detect in a sizeable number of individuals. But in case the allele is a rare allele we may miss it in a small population. The study done here can be considered a preliminary investigation of role of Quad-SNPs in regulation of oncogenic promoter through quadruplex alteration. Further validation of frequency and distribution of the allele in cancer population has to be done undertaking a large population. Probably, a study with large number of case and control may be better suited to highlight its importance in cancer.

Several pathways have been implicated in cancer etiology. One of the important pathways is the Wnt signalling pathway (39). Under normal conditions a complex of GSK, APC and associated proteins breakdown β-catenin and hence reduce its level. Wnt binds to receptor FZD1, LRP5/6 and other proteins to up-regulate Dishevelled which suppresses the breakdown of β-catenin (51) and thereby leads to oncogene activation. Members of low density lipoprotein family LRP1, LRP2, LRP5/6 play a role in regulating this canonical pathway. LRP5/6 interacts with Frizzled family and act as a co-receptor for Wnt mediated signal transduction. But despite having similar core domains LRP1 cannot substitute LRP5/6 in this signal transduction (52). mLRP4T100 (53) is a mini co-receptor of LRP1. It was shown to interact to FZD1 in a specific manner which disrupted the formation of FZD1-LRP5/6-Wnt complex (54). This led to cessation of the Dishevelled mediated β-catenin stabilisation and hence suppression of the oncogenes. Thus different binding of proteins to FZD1 supresses cancer. It has been shown that allelic change of rs2232157 leads to formation and disruption of transcription factor (41) binding site. In a similar manner the C allele may be leading to increase in expression of FZD1 which binds to the mLRP4T100 and thus resists cancer, whereas the ‘A’ allele expresses a form of FZD1 where LRP5/6 can bind and induce cancer. FZD1 levels have been known to have a low expression in thyroid follicular tumours (55). Similar mechanism may be in place for Head and neck cancer etiology too.
3.6 Reference List


technologies to study cancer. Oncogene., 25, 1594-1601.

35. Fan, J.B., Gunderson, K.L., Bibikova, M., Yeakley, J.M., Chen, J., Wickham, G.E.,

36. Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N. and


40. Gazit, A., Yaniv, A., Bafico, A., Pramila, T., Igarashi, M., Kitajewski, J. and Aaronson, S.A.
(1999) Human frizzled 1 interacts with transforming Wnts to transduce a TCF
dependent transcriptional response. Oncogene., 18, 5959-5966.

association analysis of frizzled 1 (FZD1) promoter haplotypes with femoral neck
geometry. Bone., 46, 1131-1137.


44. GELLERT, M., LIPSETT, M.N. and DAVIES, D.R. (1962) Helix formation by


