INTRODUCTION
INTRODUCTION

PART 1

Noncoding RNAs.

Central dogma of molecular biology recites that the transfer of genetic information is always from DNA to RNA to protein (Crick F. 1970); where DNA is the genetic material, proteins are functional moieties and the intermediary RNA is an essential accessory molecule facilitating the transfer of information from DNA to protein. However, the discovery of Reverse Transcriptase and Ribozymes (Latham JA, Cech TR. 1989; Goodman NC, Spiegelman S. 1971) have made the scientific community reorganize the concepts of transcription biology and have led to the strengthening of “RNA world” hypothesis. RNA world hypothesis proposes that early life was developed by making use of RNA molecules and not proteins for catalyzing the synthesis of important biological moieties (North G. 1987; Ganem B. 1987). Noncoding RNAs (ncRNAs) are non-protein coding RNAs usually transcribed by RNA pol II, lacking long ORFs and coding for functional RNA molecules (Eddy SR. 2002; Szymanski M et al. 2005; Costa FF. 2007; Michel U. 2000; Mattick JS, Makunin IV. 2006).

The completion of genome sequencing projects of human, mouse and other higher eukaryotes, have revealed that almost 98% of the transcriptional output of the genome constitutes ncRNAs. The human genome codes for only ~30,000 protein coding genes, which comprises less than 2% of the total 3.2 billion bases. Furthermore, the human genome sequencing has revealed that ~46% of the genome constitutes the repetitive sequences like LINEs and SINEs. The noncoding regions of protein-coding genes (intronic sequences, 5’ and 3’ UTRs) comprise ~25–27% of the genome, whose function is yet unclear. Recent studies have shown that the gene intronic sequences are the major sites of ncRNA generation (Lander ES et al. 2001; Venter CJ et al. 2001; Waterston RH et al. 2002). Similarly, ncRNAs from different organisms starting from prokaryotes to higher eukaryotes are being estimated using in silico methods, where the percentage of ncRNAs in total transcriptional output is in proportion with the complexity of the organism (Szymanski M et al. 2005).
Functions of noncoding RNA.

ncRNAs are implicated in a wide range of functions like gene silencing, gene transcription, DNA imprinting, DNA demethylation, chromatin structure dynamics and RNA interference etc. (Costa FF. 2005; Mattick JS, Makunin IV. 2006). The ncRNAs exert the above functions by complex mechanisms that involve RNA–DNA, RNA–RNA and RNA–protein interactions (Mattick JS. 2003). The complete set of interactions that occur inside a cell is known as “interactome”. The old concept of functional interactions (mostly) between proteins and specific regions of the DNA (e.g. transcription factor binding sites) has changed with the description of several ncRNAs that can interact with DNA, RNA and protein (Mattick JS. 2004; Mattick JS, Makunin IV. 2006).

ncRNAs have remained undiscovered and unexplored for about half a century, from the discovery of DNA structure (Watson JD, Crick FH.1953) as there are no specific methodologies to identify them from the genome or the transcriptome. Moreover, ncRNAs lack specific sequence features. Most of the ncRNAs have been identified serendipitously, while studying the differential gene expression. Now, Scientists have developed several methodologies to systematically identify ncRNAs like computational genome analysis, cDNA cloning and sequencing strategies, and high density oligonucleotide microarrays (Eddy SR. 2001). The characterization of newly identified ncRNAs can be performed by in vitro and in vivo knockdown experiments (RNAi and
knockout mice), and use of viral vectors such as lentivirus, retrovirus and adenovirus vectors (Huttenhofer A et al. 2005).

**Classification of noncoding RNAs.**

Regulatory RNAs comprise the small regulatory RNAs and untranslated RNAs. The untranslated large non-coding RNAs are mostly transcribed by RNA polymerase II and encode RNAs that lack open reading frames and function as their final products. They are involved in regulation of gene expression in eukaryotes, hence referred to as “riboregulators” (Rastinejad F, Blau HM. 1993; Michel U. 2002). The small regulatory RNAs can be classified into two classes i.e., microRNAs (miRNAs) and small interference RNAs (siRNAs).

MicroRNAs (miRNAs) are about 19- to 24-nucleotides long noncoding regulatory small RNAs that can silence target gene expression through base pairing to the complementary sequences in the 3’ untranslated region (3’UTR) of target genes. miRNA base pairing requires no perfect complementarity and they decrease the target protein translation without effecting the mRNA levels. miRNAs are evolutionally conserved and play an important regulatory role in embryogenesis, cell differentiation, proliferation and are involved in development and progression of cancers. Recent studies have proposed both oncogenic and tumor suppressor functions for miRNAs in higher eukaryotes. There are about 1000 human miRNAs predicted to this date, and it is estimated that they could target about 30% of all human transcripts (Lee RC, Ambros V. 2001; Lau NC et al. 2001; Lagos-Quintana M et al. 2001).

The siRNAs are 21–23 nucleotides in length and are generated by enzymatic cleavage of long double-stranded RNAs (dsRNAs) with Dicer (Elbashir SM et al. 2001b; Lee YS et al. 2004; Hammond SM et al. 2000; Elbashir SM et al. 2001a). siRNAs are known to base pair with perfect complementarily to its target mRNAs resulting in their degradation by endonucleases, thereby decreasing the target protein expression. This phenomenon is termed as RNA interference (RNAi) (Fire A et al. 1998) and is thought to have evolved as an antiviral defense mechanism by inactivating RNA viruses (Cullen BR. 2002; Kanzaki LI et al. 2008).

**Biogenesis of miRNAs and siRNAs.**

miRNAs and siRNAs can be most readily distinguished from each other by the unique aspects of their biogenesis (Figure 2). miRNAs are generated from a stem loop
structure formed by pairing of a single RNA strand. siRNAs are generated from the dsRNA that is formed by annealing of two complementary sense and anti-sense transcripts of a gene. These transcripts can be encoded from the same region (cis) or different region (trans) of the genome. The primary miRNA transcript is generally transcribed by RNA polymerase II, which is recognized by RNase III homologue- Drosha in the nucleus, which in association with DGCR8/Pasha cleaves the pri-miRNA to generate pre-miRNA (Denli AM et al. 2004; Gregory RI et al. 2004; Han J et al. 2004; Lee Y et al. 2003).

Figure 4. Proposed unified model of small interfering RNAs OR microRNAs mediated regulation of gene expression (Courtesy: Nelson P et al. 2003).

The pre-miRNAs ranges from ~70-100bp with stem loop structure of 19-24 bases and are transported into cytoplasm with the help of Exportin 5 (Bohsack et al. 2004; Lund E et al. 2003; Yi R et al. 2003). In cytoplasm, the stem loop structure of pre-
miRNAs or the long dsRNAs (siRNA precursors) is processed by the RNase III endonuclease called Dicer to generate small RNAs i.e., miRNAs or siRNAs respectively (Nelson P et al. 2003). Dicer acts in association with TRBP complex. These small RNAs are unwounded and presented for complementary regions of mRNAs by RNA inducing silencing complex (RISC) complex with the help of Argonaute proteins i.e, Ago2 in mammals (Hammond SM et al. 2000; Hammond SM et al. 2001; Hutvágner G, Zamore PD. 2002).

The advantages of RNA regulators over protein regulators are that they may require lower input of energy, limited resources and shorter time to synthesize a short ncRNA compared to proteins. Many of the characterized RNA regulators are expressed and function during specific developmental stages or under stress conditions. In addition, many ncRNAs act at a posttranscriptional level, which also would ensure a fast response to a developmental cue or environmental signal (Storz G et al. 2005).

MicroRNAs in cancer.

Several miRNA were reported to play a crucial role in the initiation and progression of human cancer, and those with a role in cancer are designated as oncogenic miRNAs (oncomiRs) (Esquela-Kerscher A, Slack FJ. 2006; Cho WC. 2007). For example, miR-15 and miR-16 induce apoptosis by targeting Bcl2 (Cimmino A et al. 2005). miRNAs from the miR-17-92 cluster modulate tumor formation and function as oncogenes by influencing the translation of E2F1 mRNA (Woods K et al 2007). The proto-oncogene c-Myc has been shown to activate expression of an miRNA cluster on human chromosome 13, and two miRNAs (miR-17-5p and miR-20a) from this cluster downregulate the expression of transcription factor E2F1 that activates cell cycle progression (O'Donnell KA et al. 2005). miR-21 modulates gemcitabine-induced apoptosis by phosphatase and tensin homolog deleted on chromosome 10-dependent activation of PI 3-kinase signaling (Meng F et al. 2007). miR-34a acts as a suppressor of neuroblastoma tumorigenesis by targeting the mRNA encoding E2F3 and reducing E2F3 protein levels (Welch C et al. 2007). The chromosomal translocations associated with human tumors disrupt the repression of High mobility group A2 (Hmga2) by let-7 miRNA (Yu F et al. 2007).
**Longer noncoding RNAs.**

The longer ncRNAs are the untranslated RNAs, usually transcribed by RNA polymerase II. They can be classified into 3’ polyadenylated and non-polyadenylated (Carninci et al. 2005; Cheng J et al. 2005; Kapranov P et al. 2005). Many of these ncRNAs are developmentally regulated, alternatively spliced and physiologically responsive in higher organisms. Currently, they are being discovered and characterized very rapidly from the genome and transcriptome of higher organism's including human. Recent studies showed a specific expression of ncRNAs in brain of human and mouse (Amadio JP, Walsh CA. 2006; Ravasi T et al. 2006). The first eukaryotic long ncRNA described was H19, which is identified as a maternally expressed gene in mice (Brannan CI et al. 1990; Bartolomei MS et al. 1991). The long ncRNAs are implicated in several cancers and diseases as well in functions like transcript regulation, epigenetic regulation (DNA demethylation), X-chromosome inactivation, imprinting and dosage compensation. In hitherto, several databases have been developed to list the known and unknown (putative ncRNAs) ncRNAs, like ncRNAdb http://biobases.ibch.poznan.pl/ncRNA/ (Szymanski M et al. 2007), RNAdb 2.0 http://research.imb.uq.edu.au/RNAdb (Pang KC et al. 2007) and fRNAdb http://www.ncrna.org/ (Kin T et al. 2006). These databases provide the ncRNA sequence information and function including programs to assess new sequences for the presence of ncRNAs.

**Noncoding RNAs in cancer.**

Previous studies have revealed that ncRNAs are involved, either directly or indirectly, in cancers. Some of them are over expressed in some cancers like BC1 and BC200 in breast and lung cancers, BIC1 and MALAT1 in NSCLC (Non Small Cell Lung Carcinoma), PCGEM1 and DD3 in prostate cancers.

The first ncRNA identified is H19 RNA in mice. H19 is an imprinted gene in the human, expressed from the maternal allele. It is extensively transcribed in fetal stages but is not translated into protein and functions as an RNA molecule. It is shown to be a candidate tumor suppressor gene. H19 over-expression reduces the tumorigenicity and growth of two embryonal tumor cell lines viz. RD and G401 (Hao Y et al. 1993). Recently, there are contradicting reports on the role of H19 RNA. H19 expression is elevated in some cancer types including lung, breast and bladder, which shows its oncogenic property (Ariel I et al. 2000; Lottin J et al. 2002). The MALAT-1 (Metastasis
Associated in Lung Adenocarcinoma Transcript 1) gene was identified from the stage I NSCLC (Non Small Cell Lung Carcinoma) tumors, can be used as a best prognostic marker for identification of early-stage NSCLC patients who are at high risk to develop metastases. The sequence analyses showed that MALAT-1 is conserved across species (Ji P et al. 2003). The prostate specific ncRNA gene, exclusively expressed in prostate tissue is PCGEM1 and is upregulated in prostate cancers. Among various prostate cancer cell lines analyzed, PCGEM1 expression was detected only in the androgen receptor-positive cell line LNCaP (Srikantan V et al. 2000). Over-expression of PCGEM1 correlates with increase in proliferation and colony formation which suggests its involvement in regulation of cell growth (Petrovics G et al. 2004). The overexpression of PCGEM1 in LNCaP cells results in the inhibition of apoptosis induced by doxorubicin (DOX), evident by the attenuated protein levels of cleaved caspase 7 and cleaved PARP in LNCaP cells stably overexpressing PCGEM1 compared to control LNCaP cells (Fu X et al. 2006). Another prostate specific ncRNA, DD3 was found overexpressed in prostate tumors compared to normal prostate (Bussemakers MJ et al. 1999). The steroid receptor activator RNA (SRA RNA) is involved in the regulation of gene expression mediated by steroid receptors. SRA RNA in complex with the SRC-1 protein (steroid receptor co-activator 1) acts as a strong co-activator of receptors for progestins, estrogens, androgens and glucocorticoids. It is highly expressed in breast, uterus and ovarian cancers (Lanz RB et al. 1999).

The cancer up-regulated drug resistant (CUDR) gene, the novel ncRNA gene involved in cancer cell drug resistance has been reported. It was found to be overexpressed in a doxorubicin-resistant subline of human squamous carcinoma A431 and A10A cells, which were also more resistant to drug-induced apoptosis. Stable transfection with the CUDR gene was found to induce resistance to doxorubicin and etoposide as well as drug-induced apoptosis in A431 cells. The overexpression of CUDR promoted anchorage-independent growth in A431 cells and showed down-regulation of caspase 3. Hence, suggesting that CUDR may likely regulate the drug sensitivity and promote cellular transformation at least through caspase 3-dependent apoptosis (Tsang WP et al. 2007). EGO is another ncRNA, whose transcript levels rapidly increases following interleukin-5 (IL-5) mediated stimulation of CD34(+) hematopoietic progenitors. RNA silencing of EGO results in downregulation of transcripts of major basic protein (MBP) and eosinophil derived neurotoxin (EDN), in developing CD34(+) hematopoietic progenitors (Wagner LA et al. 2007). A novel ncRNA gene, NAMA (noncoding RNA
associated with MAP kinase pathway and growth arrest) is weakly expressed in several human tissues, and the spliced forms are primarily detected in testis. NAMA is inducible by knockdown of BRAF, inhibition of the MAP kinase pathway, growth arrest and DNA damage in cancer cell lines (Yoon H et al. 2007).

**Noncoding RNAs in Diseases.**

In both schizophrenia and bipolar illness, susceptibility loci are present within the disabled in schizophrenia 1 (DISC1) gene and in the large antisense DISC2 RNA that modulates its expression (Millar JK et al. 2000; Millar JK et al. 2004). DISC1 encodes a novel multifunctional scaffold protein, whereas DISC2 is a putative noncoding RNA gene antisense to DISC1 (Kamiya A et al. 2005). DISC1 was found to be a key regulator that orchestrates the tempo of functional neuronal integration in the adult brain and demonstrates essential roles of a susceptibility gene for major mental illness in neuronal development, including adult neurogenesis (Chubb JE et al. 2008). Another ncRNA, PRINS (Psoriasis susceptibility-related RNA Gene Induced by Stress) was expressed higher in the uninvolved epidermis of psoriatic patients compared to both psoriatic lesions and healthy epidermis, suggesting a role for PRINS in psoriasis susceptibility. PRINS is regulated by the proliferation and differentiation state of keratinocytes. The precipitation of psoriatic symptoms by treatment with T-lymphokines resulted in downregulation of PRINS expression in the uninvolved psoriatic but not in healthy epidermis. Hence, the elevated PRINS expression in the epidermis may contribute to psoriasis susceptibility in patients (Sonkoly E et al. 2005).

**Noncoding RNAs in Brain.**

The ‘brain cytoplasmic’ RNAs BC1 and BC200 are found highly expressed in the rodent and human brains, respectively. BC1 gene encodes a 154 nt BC1 RNA, transcribed by RNA pol III. Evolutionarily BC1 is derived to be a product of retroposition, which led to the creation of rodent ID sequences SINEs, similar to Alu sequences in primates (Martignetti JA, Brosius J. 1993). The 200 nt long BC200 RNA is of different evolutionary origin from BC1, but the expression pattern is very similar to BC1 and is believed to have an analogous function in humans. BC1 RNA is selectively expressed in tumor cells, but not in the corresponding normal tissues (Chen W et al. 1997a). Similarly, BC200 RNA is expressed in human carcinomas of the breast, cervix, esophagus, lung, ovary, parotid, and tongue, but not in the corresponding normal tissues (Chen W et al.
BC1 and BC200 RNAs are expressed almost exclusively in neural tissues, like hippocampus and transported to dendrites. BC1 knockout mice appeared normal, however showed several behavioral changes and lower survival rates when compared to normal controls. This provides evidence that BC1 ncRNA is very important for brain function and longevity (Lewejohann et al. 2004).

**X-chromosome inactivation.**

Dosage compensation of X-linked gene products is important for proper development and is achieved by epigenetic transcriptional silencing of a single X-chromosome in mammals. The XIST, a ncRNA of 17 kb in size is required for X inactivation and chromosomal silencing (Wutz A et al. 2002), which randomly coats one of the two X chromosomes in cis fashion. The Xist/XIST RNA regulates X-inactivation in mice and humans, respectively (Herzing LB et al. 1997; Plath K et al. 2002). Later in mice, Tsix is discovered from the same genomic locus but in the opposite orientation (Lee et al. 1999), which is controlled by imprinting. Tsix functions as a “guardian” that protects the X-chromosome from silencing by blocking Xist RNA (Stavropoulos et al. 2001). Similarly, the roX1 and roX2 ncRNAs in flies have been shown to associate with the chromatin of the X-chromosome in males and could be an ancestor of XIST (Meller VH et al. 2000).

**Noncoding RNAs in Stem cells.**

Stem cells are basic cells of all multicellular organisms having the potency to differentiate into a wide range of specialized cell types. Noncoding RNAs specifically small RNAs are shown to play important roles in stem cells. Genetic studies of various components of RNA silencing in several model organisms have demonstrated the crucial roles of these silencing pathways in both early embryonic development and ES cells. In the most obvious instances, genetic ablation of an RNA-silencing factor induces embryonic lethality (Stadler BM, Baker HR. 2008). In fruit flies, homozygous mutation of the miRNA-processing enzyme Dicer-1 is lethal and in mice, disruption of the single Dicer protein leads to embryonic lethality (Kloosterman WP, Plasterk RH. 2006). The defective embryos of the Dicer mutant mice have greatly reduced Oct4 expression implying a lack of stem cells. The large scale cloning and sequencing to delineate the miRNA profile in several human and murine ES cell lines has demonstrated that the ES cells express a unique repertoire of enriched miRNAs. In zebrafish, miR-430 gene family
expression is critical for its early development (Giraldez AJ et al. 2006). The targeted deletion of Dicer in the limb mesoderm of the mouse embryo resulted in developmental defects and the formation of a smaller limb (Kloosterman WP, Plasterk RH. 2006). Additionally, experiments using isolated populations of hematopoietic stem cells have demonstrated roles for specific miRNAs in lineage differentiation, and evidences suggest that miRNAs are important for differentiation of somatic stem cells in several other tissues (Lakshmipathy U, Hart RP. 2007). A specific miRNA profile signature (expression pattern) is identified in pluripotent cells, comprising a small subset of differentially expressed miRNAs in hES cells. The specific expression pattern of miRNAs in the progression from hES cells to differentiated cells suggests a role for selected miRNAs and their clusters in maintenance of the undifferentiated and pluripotent state (Lakshmipathy U et al. 2007).

Natural antisense transcripts.

Natural antisense transcripts (NATs) are endogenous RNA molecules that exhibit partial or complete complementarity to other transcripts, through which they may contribute to the regulation of molecular expression at various levels. There are two types of NATs in eukaryotic genomes: cis-encoded antisense NATs, which are transcribed from the opposite strand of the same genomic locus as the sense RNA and have a long (or perfect) overlap with the sense transcripts; and trans-encoded antisense NATs, which are transcribed from a different genomic locus of the sense RNA and have a short (or imperfect) overlap with the sense transcripts. Cis-NATs are usually related in a one to-one fashion to the sense transcript, whereas a single trans-NAT may target several sense transcripts (Werner A. 2005; Lavorgna G et al. 2004; Werner A, Berdal A. 2005).

RNA transcribed from the sense gene may interact with the antisense RNA, possibly leading to various cellular responses. NATs are involved in functions like gene regulation (mainly in viruses and prokaryotes), genomic imprinting, RNA interference, translational regulation, alternative splicing, X-inactivation and RNA editing (Yin Y et al. 2007). In addition, accumulating evidences suggest that NATs may play a key role in a range of human diseases (Lavorgna G et al. 2004).

Most of the NATs are reported to code from the intronic regions of the protein coding transcripts. However, several NATs are identified in the non-protein coding transcripts (ncRNAs). NATs can be classified according to the origin of interacting
transcripts into coding-coding, coding-noncoding and noncoding-noncoding transcripts. Mostly the antisense overlap between two protein-coding genes is restricted to their 3’ or 5’ UTRs (untranslated regions). However, several noncoding RNA transcripts are found to interact with the whole sense and antisense transcript. This pairing may lead to the generation of siRNAs by enzymatic digestion of cellular Dicer. The siRNAs inturn elicits RNA interference by integrating into RISC complex (Tijsterman M, Plasterk RH. 2004; Collins RE, Cheng X. 2005).

The antisense overlap may involve alternative polyadenylation creating several variants of the transcript that differ in their 3’ termini length. For example, TP53BP1 sense–antisense locus is abundant in 6.3 kb transcript, which has no potential overlap with its small natural antisense transcripts (76P transcripts). The NATs have overlapping regions only to the less abundant and comparatively long transcripts of 10.5 kb or 6.8 kb in size, which are alternatively polyadenylated TP53BP1 transcripts (Yelin R et al. 2003).

The completion of genome sequencing projects of higher organisms including human and mouse and advancements in the transcriptome i.e., full length cDNA sequencing and EST sequencing have revealed that antisense transcription is widespread in mammals. In humans, 20-25% of all genes were found to have an antisense counterpart (Lehner B et al. 2002; Shendure J, Church GM. 2002; Yelin R et al. 2003). Similar results were reported for the mouse genome, where ~2400 sense–antisense gene pairs have been identified (Kiyosawa H et al. 2003). Several databases like NATsDB, http://natsdb.cbi.pku.edu.cn/ (Zhang Y et al. 2007) and antiCODE, http://bioinfo.ibp.ac.cn/anticode or http://www.anticode.org (Yin Y et al. 2007) have developed to enable efficient browsing, searching and downloading of this currently most comprehensive collection of SA genes. Several ncRNAs are reported to exist as SAST pairs. Tsix, the antisense transcript of Xist is known to regulate Xist expression thereby regulating dosage compensation in mouse (Lee JT et al. 1999). DISC2, a ncRNA that is antisense to DISC1, functions in an antagonistic manner in regulating DISC1 protein expression. They are reported to be highly expressed in schizophrenia patients (Millar JK et al. 2000).

The renowned gene expression methodologies like Northern blots or RNase protection assays are used to measure the sense-antisense (SA) RNA transcripts. However, the above methods lack sensitivity and require large amounts of RNA sample. Therefore, they are not suitable in most cases especially when analyzing low expression levels of RNA. Recently, the reverse transcription-polymerase chain reaction (RT-PCR)
has become the method of choice for the detection and measurement of SA RNA expression patterns in both cells and small quantities of tissue, which is known as strand specific reverse transcriptase assay (Reis EM et al. 2004). Here, the gene specific oligos (sense and reverse) are used to synthesize the strand specific cDNA using RT at temperatures at 50°C followed by expression analysis using semi-quantitative or quantitative/real time RT-PCR with the internal GSPs (Haddad F et al. 2007).

**Repeat elements.**

In humans, the repeat sequences account for at least 50% of the total genome, of which 90% constitutes the transposable elements i.e. 45% of the total genome (Lander ES et al. 2001). In mammals, almost all transposable elements fall into one of four types, of which three transpose through RNA intermediates and one transposes directly as DNA. These are long interspersed elements (LINEs), short interspersed elements (SINEs), LTR retrotransposons and DNA transposons. LINEs are about 6 kb long, harbor an internal polymerase II promoter and encode two open reading frames (ORFs). Upon translation, a LINE RNA assembles with its own encoded proteins and moves to the nucleus, where an endonuclease activity makes a single-stranded nick and the reverse transcriptase uses the nicked DNA to prime reverse transcription from the 3' end of the LINE RNA (Lander ES et al. 2001; Weiner AM. 2002). Reverse transcription frequently fails to proceed to the 5' end, resulting in many truncated, nonfunctional insertions. The most common LINE elements in human are Line1 or L1 elements that induce retrotransposition. SINEs are wildly successful freeloaders on the backs of LINE elements. They are short (about 100±400 bp), harbour an internal polymerase III promoter and encode no proteins. These non-autonomous transposons are thought to use the LINE machinery for transposition. Indeed, most SINEs 'live' by sharing the 3' end with a resident LINE element (Okada N, Hamada M. 1997). Alu is the only active SINE in the human genome and is derived from 7SL RNA of signal recognition particle (Häsler J, Strub K. 2007).
Hallmarks of Cancer.

Cancer is a state of uncontrollable growth of cells with malfunctions in the cell cycle regulation. It is a genetic disease, where the injured or mutated cells that are normally destroyed are allowed to proliferate by accumulating mutations in the genome (Story M, Kodym R. 1998). The activation of oncogenes or inactivation of tumor suppressor genes play key role in progression of cancer phenotype. Proto-oncogenes normally act at different levels of cell proliferation, but can promote tumor growth when mutated (Vermeulen K et al. 2003). The cancer cell genotype is a manifestation of six essential alterations in cell physiology that collectively dictate malignant growth like self-sufficiency in growth signals, insensitivity to growth-inhibitory (antigrowth) signals, sustained angiogenesis and tissue invasion & metastasis, as depicted in Fig.5 (Hanahan D, Weinberg RA. 2000).

![Figure 5. The hallmarks of cancer (Courtesy: Hanahan D, Weinberg RA. 2000).](image)

Neoplastic cell transformation is an analogue of cancer observed in the cell culture *in vitro*, which involves perturbation of multiple genetic growth controls. Transformation can result from an infection with an oncogenic RNA retrovirus or DNA virus, which induces the activation of oncogenes or inactivation of tumor suppressor genes. Chemical carcinogens can also contribute to malignant transformation through
their ability to mutate cancer-related genes. Transformation of cells in culture is a multi-step process involving one or many of the above factors. Extensive reports suggest that rodent cells require at least two introduced genetic changes to acquire tumorigenic competence, while their human counterparts are more difficult to transform (Hahn WC et al. 1999; Hartwell L. 1992).

The adherent cells that are transformed in vitro exhibit following growth characteristics like, high growth proliferation, loss of contact inhibition (i.e., cells grow in piling up nature), formation of foci in culture, exhibit “anchorage independent” growth, potential to divide indefinitely, require less growth factors for normal growth and they generate solid tumors in vivo in congenitally immuno-compromised animals like athymic nude mice. The above properties of transformed cells can be examined experimentally to assess the degree of transformation (Todaro GJ, Green H. 1963).

Cancer and Epigenetics.

‘Epigenetics’ refers to heritable changes in gene expression that do not result from alterations in the gene nucleotide sequence. Epigenetic changes include DNA methylation and histone deacetylation that play a role in maintaining the chromatin structure and thus play a role in normal mammalian development (Holliday R. 1994; Esteller M. 2008). DNA methylation in the promoter region of genes, where the gene transcription is initiated, leads to silencing or inactivation of genes. It has been observed that this phenomenon generally occurs in tumor suppressor genes (Baylin SB et al. 1998; Magewu AN, Jones PA. 1994; Issa JP et al. 1994). This process is often deregulated in tumor cells. There are a number of tumor types, particularly hematopoietic malignancies such as myelodysplastic syndromes (MDSs), in which hypermethylation occurs and results in leukemic transformation (Zion M et al. 1994; Sun B et al. 2001; Rachmilewitz EA. 2000).

DNA Methylation and Cancer.

DNA methylation occurs by covalent addition of a methyl group at the 5’ carbon of the cytosine ring, resulting in 5-methylcytosine (Bird A. 2002). These methyl groups project into the major groove of DNA and effectively inhibit transcription. In mammalian DNA, 5-methylcytosine is found in approximately 4% of genomic DNA, primarily at cytosine–guanosine dinucleotides (CpGs), which are referred to as CpG islands (Gardiner-Garden M, Frommer M. 1987; Strathdee G, Brown R. 2002). The genome
consists of ~30,000 CpG islands and 50-60% of these are associated with genes usually within the promoter region (Costello JF, Plass C. 2001). DNA methylation helps to maintain transcriptional silence in nonexpressed or noncoding regions of the genome. For example, the condensed pericentromeric heterochromatin is transcriptionally inactive, as it is heavily methylated (Jones PA, Baylin SB. 2002). DNA methyl transferases (DNMTs) are enzymes that catalyze the addition of methyl groups to cytosine residues in DNA (Bestor TH. 2000). The DNA methylation inhibitors like, azacitidine and decitabine can induce functional re-expression of aberrantly silenced genes in cancer causing growth arrest and apoptosis in tumor cells, especially in leukemias, indicating the reversal of DNA methylation (Pinto A, Zagonel V. 1993; Momparler RL et al. 1986).

The methylation status of a DNA sequence can be determined using sodium bisulfite treatment. Incubation of the target DNA with sodium bisulfite results in conversion of unmethylated cytosine residues into uracil, leaving the methylated cytosines unchanged. Therefore, bisulfite treatment gives rise to different DNA sequences for methylated and unmethylated DNA. This can be detected by performing methylation specific PCR or sequencing using methylated or unmethylated primers (Herman JG et al. 1996).

**Genomic instability and Cancer.**

Genomic instability (GI) is a hallmark of cancer cells and is hypothesized to be a driving force behind multi step carcinogenesis (Nowell PC. 1976). A number of genetic changes are required for a normal cell to become tumorigenic. Genomic instability causes increase in the accumulation of genetic changes (mutations), which in accordance with subsequent selection for cell proliferation (growth) may lead to carcinogenesis (Foulds L. 1954; Fearon ER, Vogelstein B. 1990).

Cancer cell exhibit two distinct types of GI namely microsatellite instability and chromosomal instability (Loeb KR, Loeb LA. 2000; Lengauer C et al. 1997). Microsatellite instability particularly affects DNA repeat sequences, most often of the dinucleotide or mononucleotide repeat variety, which arise from defects in the DNA mismatch repair system (Kolodner RD. 1995; Modrich P. 1997). Chromosomal instability refers to aneuploidy i.e. gains or losses of entire chromosomes; or include translocations, amplifications and deletions that are particularly visualized by cytogenetics or comparative genomic hybridization (Lengauer C et al. 1997).
Genomic instability is assessed by techniques like flow cytometry, fluorescent *in situ* hybridization, CGH, and allelotyping. These techniques even though informative are cumbersome to perform and impractical in assessment of clinical cases. Therefore GR Anderson and his co-workers have developed a PCR based method called Inter Simple Sequence Repeat PCR (ISSR-PCR) to quantitate genomic instability in cancer biopsies, where the sample is limiting (Basik M et al. 1997). It is a rapid and more sensitive technique for assaying GI. The genomic DNA from samples is used as a template for PCR amplification with a single primer based on repetitive DNA sequences such as CA or CG repeats. The primers are anchored at the 3’ end with unique sequences to prevent slippage, such as (CA)$_8$RY, (CA)$_8$RG or (CG)$_4$RY. This technique produces amplification of DNA sequences typically less than 2 kb in size, which are present between relatively close and inverted primer-binding repeat sequences. The GI can be quantitated by GI index, which is calculated by dividing the number of altered bands in the test (tumor tissue) by the total number of bands in the corresponding control (normal tissue) (Anderson GR et al. 2001).
Part III

Cell Cycle regulation and homeostasis.

Cellular homeostasis is regulated by cell proliferation, growth arrest and apoptosis. Under normal conditions, growth-regulating mechanisms endeavor to maintain homeostasis. Disturbances in such homeostasis may result in hyperplasia or neoplasia (Bergman PJ, Harris D. 1997). Hyperplasia is dependent on growth stimulus and reverses once stimulus is removed. Cancer cells (neoplasia) are characteristically independent of growth stimulus due to mutations of intracellular signal pathways. Such independence facilitates re-entry into the Cell Cycle, irrespective of positive or negative external stimulus (Molinari M. 2000).

Cell Cycle is an intricate sequence of events through which a cell duplicates its genome, grows and divides into two daughter cells. The growth stimulus begins with the release of growth factors from a cell, which bind to receptors on the cell membrane and signals are conveyed through the membrane to cytoplastic proteins. These in turn signals the release of transcription factors within the nucleus. This sequence of events pushes the cell through the cell cycle (Vermeulen K et al. 2003). The eukaryotic cell cycle can be divided into following phases: G1 phase (Gap1 phase), S phase (synthesis phase), G2 phase (Gap2 phase) and M phase (Metaphase). Progression through the cell-division cycle is driven by activation and inactivation of cyclin-dependent kinases (CDKs), which are regulated by association with varying expression of Cyclin proteins (Hartwell LH, Kastan MB. 1994; Liebermann DA et al. 1995). Cdk activity is checked by cell cycle inhibitory proteins called cdk inhibitors (cki). Two distinct Cki families are namely, ink4 family includes p15, p16, p18 and p19, which inhibits cdk4 /6; and Cip/Kip family includes p21, p27 and p57, which inhibits cdk-cyclin complexes (Vermeulen K et al. 2003).

Restriction points.

Following mitosis, daughter cells may remain temporarily quiescent (G0) until the growth factor stimulation is sent to enter into G1 phase to go through the full cycle again. Having passed the first restriction point i.e., Retinoblastoma gene (Rb) in late G1, the cell will enter the S phase where DNA synthesis occurs. This phase should be completed as quickly as possible since DNA is very sensitive to damage, when bases are exposed and DNA strands are separated. The cells cope with the required
monitoring and maintenance of genomic integrity by means of a complex network of DNA repair pathways and the cell-cycle checkpoints (Hoeijmakers JH. 2001; Wood RD et al. 2001). The repair pathways include homologous recombination repair (HR), which is virtually error free and requires a template; and the more error prone non-homologous end joining repair (NHEJ), which do not require any template (van Gent DC et al. 2001). The cell-cycle checkpoints are biochemical signaling pathways that sense various types of structural defects in DNA or in chromosome function and induce a multifaceted cellular response that activates DNA repair and delays cell-cycle progression (Bartek J, Lukas J. 2001; Hartwell LH, Kastan MB. 1994; Zhou BB, Elledge SJ. 2000; Khanna KK, Jackson SP. 2001). Depending on the position and functions within the checkpoint cascade, the components of the cell cycle checkpoints have been sub-classified into DNA damage sensors, signal transducers and effectors (Zhou BB, Elledge SJ. 2000).

Negative growth rate control on the other hand is regulated by a restriction point in the G1 phase that determines the progression of cell cycle. The Rb protein acts as a switch, controlling the cell entering into S phase from G1. At this point, the cell is either committed to complete the cell cycle irrevocably or arrest. Progression through cell cycle is determined by phosphorylation of the Rb protein, which is being initiated by cdk. In its hypophosphorylated state, Rb binds with and inhibits the transcription factor E2F (Sellers WR, Kaelin WG. 1996). Cdk combines with cyclin to form the cyclin-cdk complex, which phosphorylates the Rb protein (Lundberg AS, Weinberg RA. 1998; Morris EJ, Dyson NJ. 2001). The phosphorylated Rb protein releases bound E2F, which activates the transcription of genes required for entering S phase (Harbour JW, Dean DC. 2000).

**DNA damage and cell cycle checkpoints.**

The ordered mechanisms described above, becomes disrupted when cells are damaged i.e., DNA damage caused by DSBs with UV or IR or chemical agents. When the DNA damage is irreparable, checkpoints eliminate such potentially hazardous cells by permanent cell-cycle arrest or cell death. The malfunction of checkpoints leads to progressive accumulation of mutations and chromosomal aberrations. These, when accompanied by proliferation signals may result in genetic syndromes and diseases including cancer (Bartek J, Lukas J. 2001; Khanna KK, Jackson SP. 2001).

Three of the best defined checkpoints in the cell cycle are the G1-S phase transition, the intra-S phase checkpoint and the G2-M phase transition. All these cell
cycle transitions are negatively regulated in response to DNA damage. Conceptually, checkpoint control pathways consist of three elements: stimuli (i.e., different types of DNA damage), signal transduction machinery, and its targets (i.e., different basal cell cycle regulators). The checkpoints may possess three types of activities include (1) surveillance mechanisms that can detect DNA damage, (2) signal transduction pathways that transmit and amplify the signal to the replication or segregation machinery, and (3) the repair activities (Hartwell L. 1992). The loss of cell cycle control will lead to several physiological diseases, like Ataxia telangiectasia, Nijmegen breakage syndrome, Li-Fraumeni syndrome, Fanconi anemia and Bloom’s syndrome. These are all associated with defects in cell cycle checkpoints and cancer susceptibility (Bartek J, Lukas J. 2001).

The G1-S Checkpoint.

DNA damage triggers a rapid cascade of phosphorylation events involving the ATM (ataxia telangiectasia mutated) and ATR (ATM and Rad3-related) kinases. ATM is mostly activated by DSBs induced by IR and ATR upon UV radiation (Durocher D, Jackson SP. 2001). ATM and ATR activate Chk1 and Chk2 (checkpoint kinases) respectively, which phosphorylates the downstream molecules like p53 and MRN complex (Kastan MB, Lim DS. 2000; Shiloh Y. 2001). The activated ATM induces the phosphorylation of several downstream molecules like, H2AX, p53, 53BP1, MRN complex (Mre11/Rad50/Nbs1) etc. In higher mammals the G1 checkpoint response against DNA damage exists in two waves of action (Mailand N et al. 2000; Bartek J, Lukas J. 2001).

The first G1/S delay occurs within minutes after damage and employs post-translational modification of proteins. It is an immediate delay response which lasts for only several hours. The key step in this pathway is the Chk2/Chk1- triggered phosphorylation of the Cdc25A phosphatase, which primes Cdc25A for ubiquitination (Ub) and rapid destruction by the proteasome. The absence of Cdc25A phosphatase activity ‘locks’ the CDK2 kinase in its inactive form by phosphorylation on inhibitory threonine 14 (T14) and tyrosine 15 (Y15). This delay response is independent of p53 (Mailand N et al. 2000; Nyberg KA et al. 2002).

The second G1/S delay mechanism involves ATM/ATR and is p53 dependent. This delay is slow and some times irreversible and may function to remove cells from the cell cycle. Here, ATM/ATR phosphorylates p53 directly or through chk1/chk2. Thus, phosphorylation stabilizes the p53 protein by preventing its interaction with Mdm2, which
is a specific inhibitor of p53 transactivation domain and is a p53 ubiquitin ligase. Phosphorylation of both p53 and Mdm2 also inactivates nuclear export of p53, leading to accumulation of a stable and transcriptionally active p53 protein in the cell nucleus. This in turn results in phosphorylation of p21 causing transcriptional induction of p21, which then inhibits the Cdk2-cyclin E complex (Ekholm SV, Reed SI.  2000; Sherr CJ, Roberts JM. 1999).

When accumulated to a threshold level, the p21 can stoichiometrically bind and inhibit all cellular S-phase promoting cyclinE–CDK2 complexes, which causes failure to load Cdc45 (an attractant for DNA polymerases) on chromatin. This leads to rapid blockade of the DNA replication origins, thereby securing the maintenance of the G1-S arrest (Fig. 6). Another important consequence of increase in nuclear p21 is inhibition of both cyclinD/CDK4/6 complexes, which leads to hypophosphorylation of Rb and inhibition of the E2F-dependent transcription of S-phase genes (Bartek J et al. 1996; Harbour JW, Dean DC. 2000). All the major G1/S checkpoint transducers and effectors qualify as either tumor suppressors or proto-oncogenes, and their loss-of-function by mutations or overexpression have been identified in many types of human malignancies (Bartek J, Lukas J. 2001).

![Figure 6. Schematic representation the G1-S checkpoint pathway. (Courtesy: Nyberg. KA et al. 2002).](image-url)
Molecules involved in G1-S checkpoint control.

The *ATM* gene was identified as the gene mutated in the recessive human genomic instability syndrome *ataxia telangiectasia* (A-T). A-T syndrome is characterized by hypersensitivity to ionizing radiation, progressive neuro-degeneration, immunodeficiency, chromosomal instability and cancer. ATM belongs to conserved family of serine/threonine kinases and shows similar structure and has sequence homologies to PI3K (phospho inositol kinase). In normal conditions, ATM is present as a homodimer in which the kinase domain is physically blocked by its tight binding to an internal domain of the protein surrounding serine 1981. The introduction of a DNA DSB leads to a conformational change in the ATM protein. This stimulates the kinase to phosphorylate serine 1981, causing the dissociation of the homodimer (Durocher D, Jackson SP. 2001). The activated ATM monomer can now phosphorylate its numerous substrates, like p53, CHK2, H2AX (Burma S et al. 2001), NBS1 (Nijmegen breakage syndrome 1), BRCA1 (breast cancer 1) and SMC1 (structural maintenance of chromosomes 1) at the sites of DNA breaks (Lee JH, Paull TT. 2004; Shiloh Y. 2003).

ATR, like ATM belongs to conserved family of serine/threonine kinases and show homology with PI3K. There is not much known about ATR as its knockdown has lead to embryonic lethality, which suggests its importance in cellular pathways. Seckel syndrome is a recessive disorder caused by mutations in the *ATR* gene or in ATR signaling (Shiloh Y. 2001).

One of the earliest modifications of chromatin in the damage response is phosphorylation of histone H2AX at Ser139 (referred to as γH2AX) by members of the PI3 family of kinases like phospho ATM (Stiff T et al. 2004; Burma S et al. 2001; Ward IM, Chen J. 2001). Phosphorylation of H2AX reaches half its maximum by 1-3 minutes soon after induction of DNA damage. Several thousand molecules of γH2AX bind to the site surrounding DNA double strand break (DSB), which extends to several kilobases on the chromatin. This permits microscopic visualization as discrete nuclear foci (Shroff R et al. 2004). Therefore, phosphorylation of histone H2AX has been used as a marker for DNA double strand breaks for evaluating the genotoxic stress (Dmitrieva N et al. 2000). Additionally, γH2AX recruits many other repair or checkpoint proteins to the damaged sites, such as the Mre11/Rad50/Nbs1 (MRN) complex, BRCA1, 53BP1 etc. (Ward IM et al. 2003; Fernandez-Capetillo O et al 2002; Fernandez-Capetillo O et al. 2003).

The p53 protein is popularly known as the ‘guardian of the genome’, because it mediates the cell’s response to certain types of injury, relaying signals from both the
DNA-damage response pathway and the oncogenic (tumour promoting) stress pathway. Defects in these pathways are associated with a predisposition to cancer (Liu Y, Kulesz-Martin M. 2001). The critical role of the p53 gene in maintaining the integrity of the genome is evident in that p53 is the most commonly altered gene in human cancer, with a mutation frequency exceeding 50% (Harris CC, Hollstein M. 1993). Most mutations are missense mutations within the evolutionarily conserved DNA binding domain (Hollstein M et al. 1994). Mutation of the p53 gene results not only in loss of p53 function but also in gain of oncogenic functions (Dittmer D et al. 1993) and in adoption of a dominant-negative conformation able to inactivate the protein product of the normal allele through heterotetramerization (Milner J, Medcalf EA. 1991). Germ line mutations of p53 have been found in Li-Fraumeni syndrome, an inherited disorder with a high risk of developing a variety of cancers at an early age (Malkin D et al. 1990).

p21Cip1/WAF1 inhibits cell cycle progression by binding to G1 cyclin–CDK complexes through its N terminal domain. The gene product also binds proliferating cell nuclear antigen (PCNA) through its C-terminal domain and blocks the ability of PCNA to activate DNA polymerase δ, the principal replicative DNA polymerase (Waga S et al. 1994; Chen J et al. 1995; Luo Y et al. 1995). In normal fibroblasts, these cell cycle regulators form quaternary complexes consisting of p21Cip1/WAF1, PCNA, cyclin and CDK (Li R et al. 1994; Zhang H et al. 1994). In addition, recent studies have shown that p21Cip1/WAF1 promotes the association of Cdk4 with D-type cyclins, and targets Cdk4 and cyclin D1 to the nucleus by its bipartite nuclear translocation signal (LaBaer J et al. 1997). Thus, the cell cycle inhibitory activity of p21Cip1/WAF1 is intimately correlated with its nuclear localization. The cytoplasmic p21Cip1/WAF1 forms a complex with the apoptosis signal regulating kinase 1 (ASK1) and inhibits stress-activated MAP kinase cascade. Expression of a deletion mutant of p21Cip1/WAF1 lacking the nuclear localization signal (ΔNLS-p21) did not induce cell cycle arrest or monocytic differentiation, but led to an apoptosis-resistant phenotype mediated by binding to and inhibition of the stress-activated ASK1 activity. Thus, cytoplasmic p21Cip1/WAF1 itself acted as an inhibitor of apoptosis (Asada M et al. 1999).

**Intra S checkpoint.**

In contrast to G1/S and G2/M transitions, S phase can only be delayed and never permanently blocked in the presence of DNA DSBs. Intra-S checkpoint is independent of p53 or p21. The intra-S-phase checkpoint mechanisms induced by DSBs involves the
Cdc25A-degradation pathway (similar to G1/S transition) i.e., ATM–Chk2–Cdc25A–CDK2–Cdc45 cascade, in slowing down ongoing S phase (Falck J et al. 2001; Kastan MB. 2001). This checkpoint predictably affects both the early- and late-firing origins of DNA replication (Larner JM et al. 1999). Replication forks slow or stall when they encounter DNA adducts, creating a physical impediment, or when dNTP pools are limiting. Replication disruption elicits four cellular responses namely, a block to initiation of replication (origin firing), slowing of elongation, and maintenance of slowed or stalled replication forks. Inhibition of CDK2 activity through Cdc25A degradation leads to a several-hour delay of S-phase progression, a timing that correlates well with the transient intra-S-phase checkpoint response (Fig. 7) (Falck J et al. 2001; Bartek J, Lukas J. 2001).

**Figure 7.** The schematic representation of intra S checkpoint. (Courtesy: Bartek J, Lukas J. 2001)

**G2-M checkpoint:**

Checkpoints in G2 phase require the cyclin-cdk-cki system to eliminate damaged cells, which have passed through from G1 phase. G2/M checkpoint prevents chromosome segregation if the chromosomes are not intact (Bartek J, Lukas J. 2001). DNA damage in mammalian cells causes the activation of ATR, which in turn
phosphorylate Chk1 (Guo Z et al. 2000). Chk1 then negatively regulates Cdc25C by phosphorylating it at Ser 216. This phosphorylation of Cdc25C creates a binding site for 14-3-3 proteins. In the bound state Cdc25C is catalytically inactive or translocated to cytoplasm and lacks the phosphatase activity. Therefore, the downstream Cdc2 is inactive and blocks the binding of Cdc2-cyclin B complex, thereby arresting the cells in G2/M phase (Fig. 8) (Peng CY et al. 1997; Sanchez Y et al. 1997; Nyberg KA 2003).

Figure 8. Schematic representation the G2-M checkpoint pathway. (Courtesy: Nyberg. KA et al. 2002).
Common Fragile sites.

Common fragile sites (CFSs) are large chromosomal regions that preferentially exhibit gaps or breaks after DNA synthesis is partially perturbed (Fig. 9) (Iliopoulos D et al. 2005). In cultured cells, CFSs are induced upon induction of replicative stress, such as folate deficiency or treatment with aphidicolin, an inhibitor of DNA polymerase α and other polymerases (Glover TW et al. 1984). Aphidicolin induces chromosomal aberrations. However, its action is dose dependent. Aphidicolin, at concentrations that did not greatly affect mitotic index, induces a striking number of chromosome gaps and breaks at hot spots mainly in common fragile sites. Aphidicolin represents a novel tool for detection of hot spots on human chromosomes through the mechanism of DNA polymerase inhibition (Glover TW et al. 1984; Pedrali-Noy G et al. 2000). The four most frequently expressed CFSs are aphidicolin-sensitive sites– FRA3B (3p14.2), FRAXB (Xp21.1), FRA16D (16q23.2) and FRA6E (6q26) (Sutherland GR, Baker E. 2000).

![Figure 9. Human G-banded metaphase chromosomes with breaks at fragile sites FRA3B and FRA16D (arrows). (Courtesy: Iliopoulos D et al. 2006).](image)

The CFSs are the regions of profound genomic instability that are distributed throughout the genome, sometimes extending for over 10 Mbs (Smith DI et al. 2007). They are the “hot spots” for increased sister chromatid exchanges (SCE), translocations and deletions (Glover TW, Stein CK. 1987; Glover TW, Stein CK. 1988; Wang L et al. 1997). They are preferred sites of plasmid integration (Rassool FV et al. 1991) and may be favored targets for DNA virus integration in vivo (Thorland EC et al., 2000; Wilke CM et al., 1996). The possible involvement of miRNA in cancer was assessed by Calin et al. has revealed that many miR genes are frequently located at the fragile sites, in minimal
regions of amplifications or common breakpoint regions. 98 out of 186 miRNAs are found to be located in cancer–associated genomic regions or in fragile sites (Calin GA et al., 2004). There are 90 described CFSs throughout the human genome and these sites vary in their frequency of expression. Aphidicolin induces expression of 83 of the 90 described CFSs (Smith DI et al. 2007). All fragile sites are relatively AT-rich and have no expanded di- or tri-nucleotide repeats. CFSs are unstable in tumors i.e., the frequent chromosome breaks and rearrangements in cancers correlate to the location of fragile sites. The replication checkpoint kinase ATR is critical for maintenance of fragile site stability. ATR deficiency results in fragile site expression with and without addition of replication inhibitors, such as aphidicolin (Casper AM et al. 2002). Another group has shown that BRCA1 deficient mouse and human cells also have elevated fragile-site expression. They have also demonstrated that BRCA1 functions in the induction of the G2/M checkpoint; after aphidicolin induced replication stalling that involves altered fragile-site stability (Arlt MF et al. 2004).

Fragile histidine triad gene (FHIT).

FRA3B was the first fragile site to be mapped and cloned and it was found to lie within the FHIT (fragile histidine triad) gene (Glover TW et al. 2005). FHIT gene encompasses the most common fragile site at 3p14.2 that spanned over 1.5 Mbs, where the final processed transcript was only 1.1 kb coding for 16 kDa protein (Zimonjic DB et al. 1997). Hence, 99% of this gene corresponded to intronic sequences probably implicating the importance of these introns in FHIT function. The FHIT protein encodes for AP3A hydrolase (EC 3.6.1.29), preferably acting on substrates- AP3A & AP4A and cleaving them asymmetrically into ATP and AMP (Barnes LD et al. 1996; Blackburn GM et al. 1998). The FHIT protein shows 69% similarity to a *Schizosaccharomyces pombe* enzyme Ap4A asymmetric hydrolase (Robinson AK et al. 1993).

FHIT was found to be a bona-fide candidate tumor suppressor. The expression of FHIT was frequently abrogated in multiple tumor types like gastric carcinomas (Baffa R et al. 1998) and lung cancers (Sozzi G et al. 1998). The deletions at FRA3B genomic locus inactivate the FHIT gene, and partly based on these deletions it was suggested to be a tumor suppressor gene. FHIT is inactivated by a number of known mechanisms, including gross deletions and chromosomal rearrangements by methylation (Wu Y et al. 2006; Lee EJ et al. 2006). The analysis of FHIT in a number of cancer-derived cell lines revealed frequent homozygous deletions in the large FHIT introns; however the FHIT
exons are found intact (Wang L et al. 1998). The recent studies showed that FHIT-deficient mice have increased NMBA induced gastric tumors and this susceptibility can be rescued by the introduction of a functional FHIT allele (Dumon KR et al. 2001). In addition, overexpression of FHIT has been shown to suppress the growth of different cancer cell lines both in vitro and in vivo (Siprashvili Z et al. 1997). The exogenous FHIT protein expression induces apoptosis, directly or indirectly, in cancer cells (Roz L et al. 2002; Sevignani C et al. 2003). Thus, loss of FHIT appears to play a role in mutagen sensitivity and cancer susceptibility. FHIT is frequently deleted in numerous cancers and cancer cell lines, such as lung, digestive tract, kidney, breast, liver, and pancreatic cancers; and FHIT protein is absent or reduced in most cancers (Croce CM et al. 1999).

The ortholog of human FRA3B in mouse is Fra14a2, which encompasses the murine Fhit (Glover TW et al. 1998; Shiraishi T et al. 2001). Mouse Fhit is similar to that of human FHIT in following characteristics, (1) high AT rich sequences (2) presence of large intronic regions (3) functions as a tumor suppressor gene. The entire length 1.6 Mbp, of FHIT and Fhit gene sequences were compared by using the ADVANCED PIPMAKER program. It showed that 81.8% of FHIT sequence, excluding repetitive elements, was homologous to Fhit sequence in conserved positions. Exons 3 and 10 in FHIT, both noncoding exons, did not have homologous regions in Fhit sequence, whereas other exons in human sequences had homologous exons in mouse sequences (Matsuyama A et al. 2003). Fhit knockout mice are more susceptible to tumor formation (Fong LY et al. 2000) and delivery of Fhit in viral vectors prevents and reverses the cancer development (Ishii H et al. 2001; Dumon KR et al. 2001). In tumors associated with environmental carcinogens, alterations in the Fhit gene occur quite early in the development of cancer. However, in other cancers, Fhit inactivation seems to be a later event, possibly associated with progression to more aggressive neoplasias. Thus, detection of FHIT/Fhit expression by immunohistochemistry in premalignant and malignant tissues may provide important diagnostic and prognostic information (Croce CM et al. 1999).
Genesis of work.

According to Loeb and his colleagues, the cancer cells exhibit mutator phenotype. It says that normal mutation rates are insufficient to account for the multiple mutations observed in cancer cells, and therefore, mutations that increase mutation rates are essential to account for the large numbers of mutations observed in human tumors (Loeb LA et al 1998). The recent evidences suggest that as the tumor progresses it accumulates more and more mutations to become highly malignant. If the cell cannot repair its damaged DNA, usually it is deleted by apoptosis or killed. However, if the mutations are very large in numbers and affects multiple pathways the cell survives primarily with its growth proliferation signals and the transformation phenotype (Bartek J, Lukas J. 2001). Earlier studies revealed that most cancers arise from a single cell that is transformed into a cancer initiating cell (cancer stem cell), which has a capacity to proliferate and form tumors (Terzis et al 2005).

Our lab is working on identification and study of oncogenes implicated in melanoma development and progression. For this purpose, a cDNA library was generated from CloneM3 (a subclone from the Cloudman mouse melanoma cell line) cells using Automatic Directional Cloning (ADC) approach in γpCEV27 vector and the cDNA library was assayed for competent transforming genes by transfecting it into NIH 3T3 cells. In the process, we hit upon this novel noncoding RNA named M3TR (CloneM3 transforming RNA) that was transforming cDNA. M3TR was found to possess high transforming potential in vitro and in vivo, when ectopically expressed.

AIM of the Thesis:

The work in this thesis aims to address the role of M3TR derived from human neuro-epithelioma cell line, and its effects on cell growth and regulation with following approaches:

Objectives:

1. Cloning and molecular characterization of hM3TR.
2. Analyzing the biological functions of hM3TR by ectopic expression studies.
3. Deriving molecular clues to understand M3TR function.
4. Deciphering the signaling pathways elicited by M3TR, and
5. Identifying the transcriptional targets of M3TR.
The thesis work is divided into three chapters. The first chapter deals with the first two objectives- Cloning, functional and molecular characterization of M3TR from human. The second chapter deals with understanding the molecular clues or mechanisms for M3TR functionality. The third uncovers signaling mechanisms elicited by M3TR and identifying the downstream transcriptional targets of M3TR.