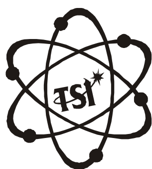
March 2009 Volume 3 Issue 1



Trade Science Inc.

BioCHEMISTRY

An Indian Journal

Review

BCAIJ, 3(1), 2009 [01-23]

Cancer is a disease of a disturbance of chromatin structure

Coral Wynter, Marcelo Alfonzo, Maria Fernanda Correa

Biomembrane Section and Molecular Biology of Infectious Agents Section, Institute of Experimental Medicine, Department of General Pathology and Physiopathology, Faculty of Medicine, Central University of Venezuela-50587, Sabana Grande, (CARACAS VENEZUELA)

E-mail : cvawynter@gmail.com, hmag5@hotmail.com, imataca@yahoo.com

Received: 4th February, 2009 ; Accepted: 9th February, 2009

ABSTRACT

Few of the 22,000 coding genes are transcribed in a normal cell as a silent gene is the default position. The role of non-coding RNAs (ncRNA) is essential in identifying the gene to be silenced or transcribed. It is proposed that transcription depends on nucleosome eviction around the promoter by a chromatin remodelling complex, attachment by a ncRNA, escorted by a protein, to the DNA sequence, followed by binding of transcription factors. The group of miRNAs is a special case of ncRNAs. There are two pathways for silencing the gene, one by methylation of lysine 9, H3K9me2, which is independent of polycomb group of proteins (PcG) and the other through lysine 27, H3K27me3, which is dependent on PcGs. Both silencing and transcription at a genetic locus must be reversible. A cancer is initiated when anyone of the ncRNAs, protein factors or coding genes is mutated or silenced epigenetically in the cancer stem cell. Thus any clinical intervention in this delicate process of genetic expression is fraught with difficulties and off-target consequences.

© 2009 Trade Science Inc. - INDIA

KEYWORDS

Cancer;
Epigenetics;
DNA methylation;
Chromatin modification;
Non-coding RNA.

INTRODUCTION

In order to find effective therapy for cancer, we need to understand the biological mechanism whereby a cancer develops. It has been difficult to link the different cellular and molecular manifestations of cancer, such as gross chromosomal rearrangements and eliminations, aneuploidy, dedifferentiation of cells, deletions of shorter DNA sequences labeled as Loss of Heterozygosity (LOH), chemical modification of cytosine bases, mostly involving methylation and demethylation of cytosine at promoter sites of specific genes but also across large sections of DNA as well as single-base mutations in key genes. For the past 40 years, despite an enormous amount of basic and clinical research, the explanation of all these phenomena has been elusive, until now. The basis of this new understanding began with the discovery of interfering RNA in plants, RNAi, which eventually led to microRNAs (miRNAs). The miRNAs, the small, 22 nucleotide (nt) endogenous noncoding RNAs (ncRNAs) are responsible for preventing translation of messenger RNAs (mRNAs) and organize their degradation in the cytoplasm. In the future as our knowledge of ncRNAs expands, miRNAs may actually be viewed as a special case of ncRNA.

Some 93% of the human genome maybe transcribed, from both strands of the DNA into these ncRNAs[1]. The intricate and complex system of ncRNA transcription allows for an integrated, tight and temporal control of translation of the 22,000 genes, coding

for human proteins. We think that all protein coding genes will eventually be found to have at least one or more ncRNAs, regulating its production. A mechanism for cancer must necessarily include an understanding of the role of these ncRNAs and their organization in the genome. More and more, cancer is being viewed as a disease of mutations both epigenetic and genetic, which result in the incorrect assembly of chromatin, thereby leading to the inappropriate transcription or silencing of a long or short section of DNA. This involves a complex machinery of ncRNA, protein intermediates, both enzymes and structural components, and the DNA code itself.

In this review, we have tried to examine most of the known factors that influence chromatin structure. Two aspects that have come into focus are the dynamic role of the nucleosome as it slips in and out from under the DNA chain and ncRNA. The review is divided into two halves, one that examines the newly discovered role of ncRNAs with their associated proteins and its impact on carcinogenesis. The second half looks at the multimeric proteins that reorganise the nucleosome, the PcG complexes, histone modifications of the side chains, and the methylation of CpG islands. Finally a mechanism of how a cancer may evolve, originating with cancer stem cells in colon cancer is described.

1. Role of noncoding RNA (ncRNA)

The idea of sequence specific ncRNA controlling transcription of coding genes was first proposed in the late 1960s[2] and then forgotten. Only 1.2% of the 2,851 billion bases of DNA are translated into 30,675 transcripts from 21,561 protein genes[3,4]. Introns make up 30% of the genome but were thought to be rapidly degraded, and therefore were ignored. The term ncRNAs cover all RNAs that are transcribed that do not code for a protein, including those transcripts that have a 7-methylguanosine cap at the 5' end and a polyA at the 3' end. NcRNA is made up of several classes of RNA, some of which have already been studied for decades, called the housekeeping ncRNAs. These are the ribosomal RNAs (rRNA), transfer RNAs (tRNA), both of which play an essential role in the translation of mRNA into proteins, telomerase RNA, small nuclear RNAs (snRNA), important in splicing out the introns and small nucleolar RNAs (snoRNA) which can methylate rRNA; all of these are transcribed by RNA polymerase I or III. The remaining ncRNAs as well as mRNA are transcribed by RNA polymerase II. A new comprehension of the enormous amount of mammalian DNA transcribed into short ncRNA surfaced with the third stage of the FANTOM project, which aimed to clone all full-length mouse cDNAs. This included all sequencing and mapping to the genome[5].

The researchers used a new technique based on the cap analysis of gene expression (CAGE) of the 7Me-guanosine of the ncRNAs with the aim of collecting every transcription start site[6]. In addition, every transcription termination site was also captured with gene identification signature (GIS) and gene signature cloning (GSC) ditag technology; all the methods allow for large throughput analysis and are described in detail[7]. cDNA libraries, extracted from a total of 35,000 mice to include every ncRNA from all tissues, delivered an incredible massive collection of 102,281 independent ncRNA sequences[8]. The ncRNAs had an extraordinary diversity ranging in size from 20 nt to 100 kb, lacking an open reading frame (ORF) and unspliced. Some 63% of the mouse genome was transcribed into RNA, although some 42% never left the nucleus and a huge proportion were non-polyadenylated. The total number of ncRNA transcripts is almost five times greater than the number of coding genes.

Two human libraries of hepatocarcinoma cell lines (HepG2) also yielded the same result using CAGE tags, indicating the same level of transcriptional diversity occurred in humans as in mice[9,10]. The results reveal a network of nested and overlapping transcripts on both DNA strands, where the intron on one strand can be the part of an exon on the antisense strand[3]. The same scenario was true for human chromosomes 21 and 22[11]. Previous gene tiling studies, which had yielded similar results, had been dismissed as artifacts or transcriptional noise from the microarray analysis, an assumption now known to be wrong. These results thoroughly debunk the notion of the existence of 'junk DNA'. It also unravels some of the mystery of the genome, as to why the nematode worm with 19,000 genes and the sea urchin with 23,000 genes have about the same number of genes as humans[12]. To account for the intelligence of human beings, it was once thought humans had at least 40-140,000 genes but the com-

plexity is due to the organization and precise programming of its protein coding genes by ncRNAs.

Sense/Antisense transcripts (S/AS)

A decade ago, it was suggested that antisense transcripts may regulate transcription of the coding sense transcript based on results found in bacteria[13]. The FANTOM project revealed there was a great amount of sense/antisense (S/AS) transcription of ncRNAs in mice. Some 70% of the total transcripts had a component that was antisense to another transcript. It was noted that there were 750 pairs of coding to coding genes, 1100 pairs of coding to non-coding genes and about 100 noncoding to noncoding genes overlapping at the 5′ end[14]. Similarly 900 pairs of coding to coding genes, 850 coding to non-coding genes and 150 noncoding to noncoding genes overlapping at the 3' end were found. Importantly there was no S/AS overlap of exon to exon. The expression of most S/AS pairs was positively correlated; that is, if the sense transcription was blocked so was the antisense transcription. This is an important observation in relation to a mechanism of cancer.

Many of the ncRNAs had extraordinary long transcripts without ORFs. Sixty-six regions were identified, each of which mapped outside known protein-coding loci and which had a mean length of 92 Kb[15]. It is known that long ncRNAs exist, such as the TSIX ncRNA, a 40 kb antisense to XIST inactivating the redundant X chromosome[16] and 108 kb AIR controlling imprinting of the IGF2R gene[17]. Alterations in the methylation status of the differentially methylated region (DMR) results in the loss of H19 and/or Igf2 biallelic expression and results in malignant cell growth[4]. In addition, a loss of H19 mRNA has been reported in many paediatric cancers, in particular Wilm's tumour. There are a total of 80 genes known to be imprinted in the human genome and perhaps many of these 66 regions are responsible for the imprinting of these genes with the help of a long ncRNA.

Another key finding is that the coding genes have more than one transcription start site (TSS) at the 5'UTR region. In mouse, 58% of CAGE tag mappings revealed that coding genes had two or more alternative promoters and the equivalent figure for humans was 52%[18,5]. This indicates that selection of the TSS for each coding gene is very complex and sophisticated and can change depending on the specific requirements. Further the FANTOM3 data also revealed a new role of the CpG islands which are often methylated on the cytosine and prevent transcription factors from binding to the promoter region near exon 1 of a coding protein. Methylated CpG islands in the promoter region usually indicate silencing of the gene. Here the CpG islands were often associated with bidirectional promoters, one operating in the 5' direction and the other in the 3' direction at the same point on the sense and antisense strands. This could be the mechanism of how the antisense strand initiates reactivation of the gene. Some 34% of CpG islands, located downstream of the TSS, function in a tissue-specific manner. The study of many CpG promoter regions in cancer may need to be reexamined as perhaps the more important aspect maybe the silencing of the antisense strand, not the sense strand.

The success of the FANTOM project in mice led onto the ENCODE project which analysed 30,000 Kb or just 1% of the human genome[19]. The 30Mb of DNA was not contiguous but distributed over 44 different genomic regions. Based on this limited data, it was concluded that virtually most of the genome was transcribed. Just as with the mouse genome, the vast majority of these transcripts was not polyadenylated and did not leave the nucleus. It is more useful to think of the genome encoding a network of RNA transcripts, rather than the production of a single protein. It revealed five types of *cis*-acting regulatory sequences, promoters, enhancers, silencers, insulators and locus control regions. Whole genome tiling arrays have identified other classes of short transcripts out of more than 450,000 ncRNAs. These include ncRNAs, which map on both genomic strands, both at the transcription start sites and transcription termination sites, of about half of the expressed known protein-coding genes.

In addition more than 10% of the full-length clones were pseudo-mRNAs, where some 600 transcripts were expressed from 20,000 pseudogenes in the human genome[20]. Nearly half of the pseudo-mRNAs were associated with transposons, that had become disabled due to a LINE1 retrotransposition. Some of the pseudo-mRNAs had a frameshift which allowed translation with different amino acids inserted. However, since much of its sequence is identical, the pseudo protein may share

many of the canonical protein's interaction and interfere with its function. One study implicated the pseudo neuronal nitric oxide synthase (pseudo-NOS) transcript as a natural antisense regulator of nNOS protein synthesis due to a stable duplex of RNA:RNA *in vivo* complex[21]. It raises the interesting question of whether the proteins from pseudogenes can also regulate oncogenes or tumour suppressor genes or can cause a cancer or if not, what mechanism is used to prevent their interference in the cell's biology.

Cancer potential of ncRNAs

It has long been known that antisense transcripts to oncogenes and tumor suppressor genes such as Wnt1[22], c-Myc[23], p53 and Bcl2 could control their expression but it was not known what to make of this information. Altered expression of 17 antisense transcripts from intronic regions as well as 13 miRNAs, including the *let-7* family, have been found in cancers of the prostate colon, breast, liver, lung, ovary and fetal tumours[4]. It may be necessary to re-examine antisense ncRNA transcripts of other oncogenes to see if patients had a mutation in the antisense ncRNA in the intron if no mutation was found in the coding gene.

Specific ncRNAs may yield much better information about the degree of malignancy of a cancer or predictive value of life expectancy than mRNA levels, or even miRNAs. It has been shown that a study of 23 antisense intronic ncRNA correlate with the degree of tumor differentiation in a prostate cancer[24]. In addition, another ncRNA, MALAT-1, with a size of 8,000 nt, was found to have three times the expression and able to predict metastasis and survival in non-small-cell lung cancer[25]. A ncRNA, named BC200, only 200 nt long, normally expressed in neurons but no other organ, was found at very high levels in breast tumours[26]. BC200 could be used as a molecular marker of invasive breast cancer as it was not expressed in benign breast tumours such as fibroadenomas. Due to the fact that cell differentiation is eliminated, it's possible that many other ncRNAs could play a prognostic role in other cancers.

Tumor suppressor Gene P15 and Antisense ncRNA

Transcription of the p15 (INK4B) tumour suppressor gene, which encodes a cyclin-dependent kinase inhibitor, regulating cell cycle, is frequently deleted or hypermethylated in a wide variety of tumors, such as leukemia, melanoma, glioma, lung and bladder cancers. A long p15 transcript of 34.8 kb, antisense to p15 (p15 AS), was identified in a leukemia cell line[27]. Increased expression of p15 AS together with downregulated p15 expression was shown in 6/11 acute myeloid leukemia and 5/5 acute lymphoblastic leukemia, an example of negative correlation of the sense and antisense transcripts. By construction of plasmids into HeLa and HCT116 cells, it was shown that the antisense transcript had a strong cis effect, although a weaker trans effect on expression. However the mechanism did not involve the nuclease, DICER, necessary for miRNA processing. By chromatin immunoprecipitation, p15 *AS* induced histone H3 modification changes in the p15 promoter region and exon 1. The stable changes were a marked increase in H3K9me2 and reduced H3K4me2, histone markers for silencing. However no methylation of CpG DNA was found at the promoter sites.

Less direct evidence for the role of antisense ncRNA was found by the use of 21 nt double stranded RNA (dsRNA) to induce promoter regions of E-cadherin, p21 and VEGF into human prostate cancer cell lines[28]. The dsRNAs were designed to avoid rich CpG islands within gene promoters. E-cadherin was epigenetically silenced in HeLa cells due to methylation of the CpG island in the promoter region. But ds E-cadherin AS induced expression only after the methylation of CpG sites was stripped by demethylating agent, 5-azacytidine. The induction was sequence specific as they failed to activate gene expression with dsRNAs complementary to promoter sequences for p27, PTEN and APC tumor suppressor genes. The method of activation was due to an induced change of the methylation state on the lysine residue of the histone, specifically a loss of H3K9me3.

Recent evidence of repression of a gene by a ncRNA has surfaced with a study of human dihydrofolate reductase (DHFR)[29]. DHFR contains two promoters, with the major promoter being responsible for 99% transcription of the gene. The repression of the major promoter of the gene depended on the ncRNA initiated from the upstream minor promoter, which terminated within the second intron of the canonical DHFR. The specificity of the repression was due to the stable for-

mation of a complex between the ncRNA and the major promoter with the general transcription factor IIB. Of course, the effect was only observed when the regulatory ncRNA contained the sequence of the core major promoter. It formed a stable triplex structure of both strands of the DNA and RNA. However, how this induces an epigenetic change through chromatin remodelling of the nucleosomes, histone modification and/or CpG methylation is not known. How does this minor promoter become switched on? Another indication of its possible universality is that many genes have alternative promoters within the same 5'UTR. The APC gene, mutations of which cause Familial Adenomatous Polyposis, (FAP) also has two promoters, one major and one minor[30].

Ultraconserved regions (UCR) Encoding ncRNAs

Searches for ultraconserved regions of genomes across phyla and the animal kingdoms have been carried out for a long time as a method of discovering a meaningful role for 'junk DNA.' It wasn't until sequencing of whole genomes delivered a more defined method for comparison that ultraconserved regions of the genome could be identified. Now, comparative sequence analysis has delivered a number of ultraconserved genomic sequences (UCRs), virtually with 100% identity, across ten primate species. Half of these are located in noncoding areas while half have been designated exonic as they overlap mRNA of known protein coding genes[31]. Some 480 genomic regions with ncRNAs longer than 200 bp were analysed in both normal and cancer tissues[32]. They found that tumors of hepatocellular cancers, colorectal cancers and chronic lymphocytic leukemias (CLLs) could be differentiated by the pattern of transcribed UCRs, some up-regulated and some down-regulated. This indicated that transcribed UCRs are involved in the malignant process, whether as drivers of the process or innocent bystanders, is not known. Others have found that UCRs can act as enhancers of transcription.

It was also noted that specific miRNAs negatively regulated the transcription of UCRs[33]. Like miRNAs, it was found that UCRs are frequently located at fragile sites and genomic regions involved in cancers[34]. In particular miR-155, which is over-expressed in an aggressive form of CLL, was able to reduce the level of two nonexonic UCRs, uc.346A and uc.160 and upregulated another, uc.348 due to complementary sites with miR-155[32]. Another important UCR was uc.73A which was consistently up-regulated in colon cancers. If uc.73A was knocked out by the introduction of a synthetic miRNA, a siRNA, in COLO-320 colorectal cancer cell lines, then apoptosis of the cells was increased. If this is found to be true in cancers *in vivo*, it indicates that uc.73A can behave like an oncogene. It was concluded that two types of ncRNAs namely the miRNAs and the UCR act in a complex regulatory pathway in the progression of a cancer.

Pyknons

Other groups searched the human genome for a class of nucleotides with different criteria, namely those longer than 16 nt that occurred more than 40 times in the genome. Among the millions of discovered classifications, they found a subset of 128,000 patterns, which have additional non-overlapping instances in the untranslated and protein-coding regions of 30,675 transcripts from the 22,000 human genes[35]. The ncRNA patterns were called pyknons - a Greek word meaning dense, frequent and in series. They found that more than 90% of the coding proteins were associated with pyknons. The pyknons are arranged in different combinations in the untranslated regions, but mainly 3'UTRs. As an example, the 3'UTR of *birc4* (an apoptosis inhibitor) contains 100 instances of 95 distinct pyknons. Of these, 22 were also present in the 3'UTRs of another nine genes.

The typical pyknon length was similar to that of a miRNA, but was definitely different from UCRs, described above. In the 3'UTR, the inter-pyknon distance was on the average 20 nt, suggesting the possibility that the pyknons correspond to binding sites for small RNAs and act in a similar post-translational fashion to miRNAs. The pyknons covered intergenic and intronic genomes and appeared in different orders on both the sense and antisense strands. However 90% of the pyknons showed some overlap with repeat elements and may have some role in repressing transcription from these sites. However an interesting finding was that 85% of pyknons in the non-genic regions have not been found in rat or mouse genomes. These add another layer of control of gene regulation, but their role in cancer is

unclear[36].

It should be noted that another group of ncRNAs have been isolated in *Drosphila*, called the Piwi-associated small RNAs. These ncRNAs of about 25-30 nt bind to the Argonaute proteins, the Piwi clade. Small RNA partners of Piwi proteins have been identified in mammalian testes[37]. In mammals, Piwi-RNAs are only involved in the specialised germ cells and play a role in DNA methylation, specific to oogenesis[38]. No phenotypic abnormalities have yet been detected and they are not thought to play any role in tumorigenesis.

MicroRNAs and cancer

MiRNAs play an extensive role in carcinogenesis (for excellent reviews see[39-42]). MiRNAs are ncRNAs, can be polycistronic or monocistronic, capped with Me-7-guanosine at the 5' end, 3' polyadenylated and diced up with an endonuclease III and exported to the cytoplasm, finally ranging in size from 19 to 25 nt. They are transcribed as large precursors (pri-miRNA), spliced to a size of 60 to 110 nt from introns, exons or intergenic regions by RNA polymerase II. A subset of miRNAs has been found to be transcribed by RNA polymerase III, but only because they are transcribed from within an Alu locus[43]. Some 40% of miRNAs are encoded within introns of known genes and hence are coordinately expressed with the gene. Other miRNAs are transcribed on the antisense strand to the gene.

Briefly, the mechanism of RNA interference is a process whereby sequence specific, post-transcriptional gene silencing can be initiated by a short dsRNA, which can be either a miRNA or a synthetic siRNA, which mimics a miRNA. This can happen in two ways, either by post-transcriptional cleavage of mRNA through extensive complementarity in the coding region and/or the 3´untranslated region (UTR) or by translational repression of the mRNA, again in the 3'UTR. When the antisense strand of dsRNA or the guide strand is assembled onto a RNA-protein complex, called the RNA-induced silencing complex (RISC), it is the RISC that cleaves the targeted mRNA. Repressed mRNAs are degraded in special cytoplasmic bodies, called P-bodies[44], or the mRNAs can be re-released intact from the P-bodies to enhance translation[45]. Groups of mRNAs that form part of a metabolic pathway can be regulated at one time using a miRNA, which identifies it as a pleiotropic regulator of gene expression. Almost 700 miRNAs have been identified and sequenced in the human genome and another 500 miRNAs are awaiting confirmation from experimental data[miRNA registry at www.microrna.sanger.ac.uk/Software/Rfam/mirna/index][46].

It is thought that miRNAs regulate some 30% of the genome. In all cancers, an enormous variation in levels of expression of miRNAs for both mature miRNAs and precursors has been found, compared to levels in normal cells. Some miRNAs are elevated but most are down-regulated, which will change the concentration of known oncogenes and tumour-suppressors. In every cancer deregulation of miRNA concentrations contribute to further progression of the cancer. It is thought that the decreased expression of miRNAs contributes to dedifferentiation in tumours[47]. MiRNAs such as the cluster of miR-15a to 16-l, the cluster of miR-143 to 145 and the let-7 family function as tumour suppressor genes and the miR-155, miR-21 and miR17-92 clusters function as oncogenes. Those miRNA genes regulating hypoxia are often over-expressed in many different human cancers in order to supply as much oxygen to the growing cancer[48]. Likewise, the miRNA signature for angiogenesis should be found to be up-regulated in cancers, which require an extensive blood supply for continued growth. There is now an enormous amount of data on specific miRNAs and the changes in various types of cancers, listed in[42,40].

Further the CpG promoter site of some miRNAs can be hypermethylated in cancers and the hypermethyla tion is accompanied by histone modifications, similar to the epigenetic mechanisms used to shut down coding genes[49]. Thus genetic changes in miRNAs such as re-arrangements due to deletion, amplification or translocation or a single base mutation or epigenetic events can rapidly cause a transition from an indolent chronic myeloid leukaemia (CML) to an aggressive CML[50]. A similar scenario could operate in colon cancer where the progression from an adenoma to a cancer could be mediated by a change in transcription of a miRNA. Until now, few researchers have considered looking at single base mutation in the 3'UTR region of a coding gene. Re-examination of coding DNA in cancers where no mutation of the protein could be identified might be worth the search as inappropriate base pairing due to varia-

tions in the 3'UTR of the mRNA may be another mechanism for initiation of a cancer. Further polymorphisms in the 3'UTR, which made a weaker mRNA:miRNA pairing could be the basis of a predisposition for familial cancer[51].

However the most important role of miRNAs from a clinical point of view is that they are diagnostic and prognostic tools, able to classify cancers and predict outcome for cancer patients. By profiling each cancer in a tissue by a microarray analysis, miRNA expression profiles of human cancers can be classified according to their common derivation from embryonic endoderm[39]. The miRNA expression profiles in colon, liver pancreas and stomach all clustered together and also reflected their state of cellular differentiation[52]. Profiling miRNAs in B-cell lymphoma, breast cancer, papillary thyroid cancer and colorectal cancer can offer a reliable prognosis. Using a new technique called miRAGE, the largest analysis of miRNAs so far, a total of 274,000 RNA tags was found in colorectal cancer cell lines, with 200 of these being known mature miRNAs, 133 novel candidates and 112 uncharacterised miRNAs[53].

The usefulness of this approach is shown by the problem of diagnosis of metastatic cancers where the primary malignancy is unknown. These make up 5% of all cancers world-wide. Profiling of 200 miRNAs identified 11 out of 17 cancers as to their correct origin of primary cancer whereas the mRNA profile of 13,000 only identified 1 out of 17[54]. However it is not necessary to carry out such a large analysis of miRNAs. Microarrays of only a modest number may be sufficient to classify cancers. For example, the measurement of levels of only two miRNAs, miR-143 and miR-145, showed that a cancer was present in colorectal cells. In addition, miRNA profiles could also predict survival. A study of 144 patients with chronic lymphocytic leukaemia (CLL), the most common adult leukaemia, a unique signature of 13 miRNAs could differentiate those patients with a good or bad prognosis[52].

A germ-line mutation in the pri-miR-16-1/miR-15a precursor in a patient with familial CLL and breast cancer in first-degree members of the family was the first such case of a hereditary cancer due to a mutation in miRNA transcript[52]. Thyroid cancers, in which the upregulation of three miRNAs, miR-221, miR-222, and miR-146 was the strongest, showed loss of the oncogene, KIT, a tyrosine kinase receptor, active in cell differentiation. In half of the cases, the downregula tion was associated with germ-line, single-nucleotide polymorphisms in the two recognition sites in KIT for these three miRNAs[55]. Thyroid papillary carcinoma is the most common type of cancer of the thyroid and many familial cases have no known genetic basis. There are many examples of miRNA profiles underlying clinicopathological characteristics of cancer types that, before, were barely able to be distinguished by experienced pathologists.

Another important aspect is the prediction of those tumours which have metastasised. Application of miR-10b to cells could induce metastatic behaviour and measurement of miR-10b in metastatic breast cancer cell lines correlated with cell invasion[56]. Some 9 miRNAs were able to classify breast tumours into five subtypes; Luminal A, Luminal B, Basal-like HER2+ and Normal-like[57]. In addition one miR-155 was able to distinguish oestrogen-receptor (ER) positive tumours compared to non-ER tumours. This aspect is so important because the distinguishing molecular characteristics of each cancer will determine the drug treatment regime.

Processing of MiRNAs and cancer

Another important aspect of miRNA biology is that tumorigenesis can result from mutations in the enzymes processing miRNA[58]. Human DICER maps to chromosome 14, locus 14q32.13, an area, which is commonly deleted in lung adenocarcinomas, both smokers and non-smokers. LOH in a nearby region, 14q32.31 was found in 62% of atypical adenomatous hyperplasia, 65% of non mucinous bronchioloalveolar carcinoma, both early forms of invasive adenocarcinoma of the lung as well as 62% of stage II of adenocarcinoma but not the actual DICER locus[59]. This may indicate a lesion of the antisense ncRNA that regulates DICER. The levels of DICER were increased 10-fold in stage I of lung adenocarcinoma but stage II lung adenocarcinoma samples were similar to normal tissue and undetectable in a metastatic lung cancer cell line. This suggested only a transient upregulation of DICER in the very early stages of lung adenocarcinoma.

Reduced expression of the mRNA for DICER in

non-small cell lung carcinoma was found to be associated with a bad prognosis[60]. However empirical tests need to be carried out on each cancer as the biological characteristics differ in different tissues. In addition, a microarray analysis of 230 prostate specimens confirmed that DICER was increased in prostatic neoplasia and adenocarcinoma[61]. The increased DICER levels correlated with clinical stage and lymph node status. This was to be expected as there is a global increase in miRNA expression in prostate adenocarcinoma. Thus greater DICER activity is associated with aggressive cancer features and will be a very useful diagnostic marker for prostate cancer.

Quantification of miRNA precursors and miRNA final concentrations were measured in tissues from both hepatocellular and pancreatic cancers[62]. Results showed that in pancreatic cancer, miRNA increased levels were regulated at the rate of transcription whereas in hepatocellular cancer, the decreased miRNA expression resulted from a decrease in miRNA processing. No matter the reason for the low levels of miRNA, this condition promoted tumorigenesis. Further, the incomplete miRNA processing, caused a marked change in the transformed phenotype of the cancer cells.

MiRNAs and p53 gene

The guardian of the genome is an apt title for the transcription factor p53. It acts by repressing the transcription of several proteins that regulate different phrases of the cell cycle from G1 to cytokinesis. Aberrations in the p53 pathway, if p53 itself is not mutated, are found in most cancers. One of the most recent findings is that the protein p53 directly targets transcription of a miRNA, specifically upregulating miR-34a and miR-34b/c[63]. The p53 protein directly binds to the genomic region labelled the miR-34a promoter. This highly conserved perfect consensus p53 binding site in mouse, rat dog and cow was located just downstream of the transcription start site for miR-34[64].

The transcription factor encoded by the p53 gene is post-transcriptionally activated when DNA damage is detected. Double strand breaks may occur quite often as demethylation of the histone, H3K9me2, on the nucleosome at both enhancer and promoter sites of coding genes to allow the attachment of RNA polymerase II, releases hydrogen peroxide, a potent chemical[65]. The detection of DNA double stranded breaks activates ATM kinases which in turn phosphorylate p53. Phosphorylated p53 induces cell cycle arrest, senescence and also can promote apoptosis if the DNA damage is too severe.

MiR-34a is ubiquitously expressed in all tissues, encoded by its own transcript on chromosome 1p36. The other two, miR34b/c are expressed by a single transcript, mainly in the lung[66]. Phosphorylation of p53 protein in H1299 lung cancer cells induced a total of 32 miRNAs, but miR-34a was the most significant. To a lesser degree, miR-34b/c correlated with the concentration of p53 protein in mouse fibroblasts. Expression of miR-34a was a sufficient condition to immediately cause cell cycle arrest and apoptosis, triggered by p53 activation. In addition, the converse is also true-cell lines with normal levels of p53 protein also have low levels of miR-34. It is not known which proteins are the exact targets of the miR-34a/b/c upregulation. A massive reprogramming of gene expression occurred, with enrichment of genes involved in preventing cell cycle progression, preventing cellular proliferation, initiating apoptosis, DNA repair and angiogenesis. By acting through the miR-34s, p53 protein is enabled to regulate a large number of proteins simultaneously.

Therefore a mutation affecting the transcription of miR-34, or a mutation in the enhancer or promoter region of the p53 gene, which compromised binding to miR-34 would be a selective advantage for cancer cells. Deletion of miR34a in a neuroblastoma as well as in gliomas has been observed[67]. Region 1p36 is often deleted in pancreatic cancer and small non-small cell lung cancers. A similar role can be described for c-Myc. So there exists an intricate interaction of miRNAs and gatekeeper genes in preventing the progression of a cancer.

MiRNA let 7, K-Ras, RNA binding protein, let 28 and cancer

The complicated regulation of protein-coding oncogenes is observed on examination of the RAS proteins, with the realization that there are many ways to initiate a cancer by a mutation in this group alone. The RAS superfamily of small GTP binding proteins plays a prominent role in tumorigenesis. They function in signal transduction across the cell membranes, in particular

activating pathways that coordinate growth factors, transcription, cell survival, cell cycle progression, differentiation and senescence (for a review see[68]). More than 150 RAS-like genes have been identified in mammalian genomes. There is a very high frequency of mutations in the three major RAS proteins, H, K and N-RAS in cancers. For H-RAS, the frequency of mutations in salivary glands is 20%; for K-RAS the pancreas has 59% mutations and the large intestine 30%, and N-RAS mutations occur in the skin and nervous system at a rate of 17%.

It has been recently discovered that the miRNA let-7 negatively regulates RAS proteins. So if let-7 levels fall, RAS protein levels increase. From microarray analysis of tumour and normal samples, it was shown that let-7 expression was reduced in 12 of 12 lung cancers, four of six colon cancers as well as two of three breast cancers[69]. So let-7 is a tumour suppressor in lung cancers with its action mediated partially through upregulation of RAS. Various let-7 family members have been mapped to chromosomal regions, frequently deleted in lung cancers. So there are indications that let-7 may be useful as a diagnostic and prognostic marker for human lung cancers[68]. The exact role of let-7 and its interaction with RAS needs to be examined further, especially to see if there are other targets of let-7. One other known target of let-7 is the oncogene HMGA2, which codes for a small, non-histone chromatin associated protein that can alter the chromatin architecture[70].

Another clue in this complex scenario is that an RNA binding protein, Lin28B, was found to regulate let-7 biogenesis by inhibiting the processing of pri-let-7 in the nucleus[71]. In human hepatocellular carcinoma, Lin28B is overexpressed, decreasing let-7 concentrations and therefore allowing RAS protein levels to increase. It was found that Lin28B overexpression allowed cancer cell proliferation. It is complicated by the fact that two isoforms of Lin28B are expressed, differing in the 5'exon, the shorter form having no effect on let-7 processing. So this evidence points to one reason why miRNA levels may be generally downregulated in cancers. Typically, it is usual to find no more than 20% of colon cancers will have a K-RAS mutation, considered a significant result[72]. However, these three results taken together indicate that the RAS/BRAF/MEK/ERK pathways may be disrupted in all colon, lung and breast cancers. Examination of the most common mutation in colon cancer at K-RAS, a G12V mutation, may not give a true picture of the extent of the disruption of this pathway as the miRNA, let-7, could be down-regulated or the processing enzyme, Lin28B, maybe over-expressed, producing the same effect. So we must begin to look at the entire pathway, including all known enzyme inhibitors, rather than simply the presence or absence of one mutation in an oncogene or a tumour suppressor gene.

2. Multimeric Proteins Effecting Chromosome Structure

Assembly of nucleosomes

How does the cell faithfully reproduce the 3 billion bases in the nucleus, repairing any nucleotide errors, removing all the nucleosomes and then replacing them as the replicating fork travels along the DNA chain? How does the genome, at the same time maintain all the epigenetic marks, both temporary and permanent ones? After cell division, how does the cell keep alive a very small number of transcription sites for the few genes which maintain the specificity of the tissue and shut down the rest? Any error in this complicated regulation of housekeeping genes and genes defining an organ, could eventually lead to a cancer and other diseases. We now realise this is a dynamic process, which in the brain can lead to an increase of neuronal connections but a marked decrease in active transcription sites, with ageing. The standard picture of the nucleosome is that of a double- stranded DNA, with 147 nucleotides wrapped twice over a dimer, made up of two identical complexes, each consisting of four histone proteins H2A, H2B, H3 and H4 and about 50 nt separate the two nucleosomes. So what factors initiate the unwinding of DNA from the histones in the nucleosome, what regulates the binding of RNA polymerase II to this complex and the adding of ribonucleosides to form unspliced heterogenous nuclear RNA?

Part of the answer lies with the discovery that maintenance of gene expression is probably carried out by a continuous destabilisation and reformation of the nucleosomes at the site of active transcription. This may also involve regulatory sequences in the DNA, incorporation of histone variants, the modification of histone tails and the binding of chromatin-associated proteins[for

a review see[73]. Part of this new thinking involves a modified concept of the histone 'code.' The chemical changes on the histones do not stipulate a code, as such, which is read and edited by active enzymes, but rather the histone modifications play a more subtle role in the ease of eviction and reformation of the nucleosomes.

An important discovery in understanding chromatin activation was the observation that transcription factors bind to genomic sites where the nucleosomes have been removed, even in humans[74]. The transcription machinery is rather bulky and is incompatible with attachment to the DNA chain if it is wrapped around the nucleosome. So the nucleosome is evicted at the transcription start site. The binding of RNA polymerase II prevents the nucleosome from occupying the -1 position[75]. It is possible the nucleosome is evicted or even slides forward or backward along the chain. Eukaryotic cells contain at least five classes of chromatin remodelling complexes: SWI/SNF, ISWI, CHD(Mi-2), INO80 and SWR1. These 2 MDa multimeric complexes have a catalytic ATPase subunit that provides the energy in shifting the nucleosome out of the way in order to liberate the DNA chain for attachment to transcription factors[76]. The SWI/SNF remodelling factors are also involved in stem cell differentiation and/or senescence[77].

The Isw2 chromatin remodelling enzyme in yeast, *Saccharomyces cerevisiae,* is involved in shifting nucleosomes to a position with an energetically unfavourable DNA sequence[78]. Thus Isw2 binds to both 5' and 3'ends of genes, just downstream from the transcription start site but appears to act in the direction of 3' to 5'. From examination of mutants, it was found that Isw2 serves to shift nucleosomes onto adjacent intergenic regions. The general mode of action of Isw2 is to repress transcription from promoters[79]. Deletion of Isw2 results in the production of a noncoding, antisense transcript from the gene on the opposite strand. One can begin to think of how this could function in the human genome. If the role of the equivalent protein, ISWI in humans, is found to operate in the same manner, then ISWI would prevent the transcription of the antisense ncRNA, perhaps forcing transcription from an alternative promoter. Therefore a mutation in the antisense ncRNA could play a role in the initiation of a cancer.

However, histone variants may also be involved in the process of nucleosome eviction. There are different variants of H1, H2A and H3, each of these is thought to have specific properties and functions[80]. Deposition of the histone variant H2A.Z by the ATP-dependent INO80 into nucleosomes[81] may also facilitate nucleosome eviction or repositioning by destabilizing the nucleosome structure. So the chromatin is maintained in an active state by rapid histone turnover with continuous histone replacement from H3.3 to H3 and back again[73]. The default position is the silent state. Other evidence points to acetylation of the histone by histone acetyl transferases (HAT) enzymes as contributing to the destabilizing event and releasing the nucleosome from the DNA chain. It is also possible that the DNA sequence itself is an important factor in the ease of nucleosome assembly and stability.

Polycomb group of proteins, PRC1 and PRC2 repressive complexes

The paradigms used to understand chromatin formation have to a large extent been developed from our discovery of the biological processes used in plants, the position effect variegation (PEV) and the ciliated protoza, *Tetrahymena,* the lowly worm *Caenorhabiditis* and the fruit fly *Drosophila*; hence, the names of the protein subunits. There are protein complexes that maintain stable states of gene expression, the Polycomb group (PcG) that maintain the 'off' state. The PcG components are involved in all aspects of epigenetic regulation[82-84]. The PcG complexes are responsible for locking up large areas of the chromosome, through the formation of bunches of nucleosomes, locked together, especially after their initial expression in the embryo has been completed. In the stem cells, they play a role in keeping genes involved in differentiation of the tissue silent and also maintaining its self-renewal capacity. PcG proteins form large multimeric complexes with a variety of proteins, of two major types, both repressive, PRC1 and PRC2. It is impossible to separate PcG complexes from histone modifications on the nucleosomes as they are intimately involved and carry out many but not all of the enzymic activities.

There are actually two different mechanisms for silencing genes, one based on the polycomb group of

proteins which recognise the histone mark of H3K27me3. The other utilizes the histone mark of H3K9, without the involvement of PcG proteins. The former is involved in X-chromosome inactivation in females as well as HOX expression during development. In humans, PRC2 is responsible for binding the other PcG complex, PRC1, to the targeted gene promoter site. These are labelled the wrong way round as it is PRC2 that first binds to chromatin. PRC2 contains four core components, a histone methyl transferase, called EZH2, which trimethylates H3K27, a strong epigenetic mark for transcription repression[85]. The other three components are protein Embryonic Ectoderm Development EED, SUZ12 and EZH1 all of which aid EZH2 in its histone methyl transferase activity. In addition, EED, SUZ12 and EZH2, which also contains a zinc finger domain, all have RNA-binding activity[86].

The other repressive PcG complex core, PRC1, is composed of four major protein groups; Pc (a chromodomain which reads the lysine methyl mark on histone), dRING, (a E3 ubiquitin ligase at histone H2AK119), the Psc (essential for H2A ubquitylation enzyme) and PH, a zinc finger domain that might be responsible for the spreading of PcG complexes along the genome. These multiprotein complexes of PcG, control transcription through a series of steps that involve nucleosome modification by chromatin remodelling and histone modifications. In *Drosophila,* there is only one Pc but in mouse and humans this has expanded to five genes, named after the N-terminal chromodomains which bind to methylated histones, Cbx2, Chx4, Cbx6, Cbx7 and Cbx8[87]. A recent and important discovery is that a ncRNA was required for recruiting the PcG complex to chromatin but it was the ncRNA that determined specificity, as the Cbx proteins could bind all types of RNA[88]. It is not clear whether this only required PRC2 or also includes PRC1, as both complexes have zinc finger domains, which can bind RNA.

Much less is known about the large protein complex making up the active transcription unit in humans. This could be because the situation in mammals is different to *Drosophila*, where active transcription is much more dependent on nucleosome displacement as described above. Key to this complex is the SNF2-like DNA helicase or ATPase which can change chromatin structure by altering the histone composition within the nucleosome. Several of these complexes have been described[89]. The active transcription unit in *Drosophila* is a heterogeneous group of TrxG proteins. In humans, they are made up of SWI/SNF complex, a NURF complex and MLL1-3. The SWI/SNF and NURF complexes are enzymes that mobilize the nucleosome to open up or to assume a different shape to allow access of the transcriptional machinery by expending the energy of ATP[83]. It must also include a HAT, to acetylate the histones for ease of nucleosome eviction. TrxG also includes nucleosome remodellers such as Brahma and Kismet. Other proteins that belong to TrxG include DNA binding proteins, a GAGA factor (GAF) and Zeste, another histone lysine methyltransferase (HMT) for lysine 4, H3K4me and another enzyme a demethylase which can remove methyl groups from lysine, H3K4 demethylase (Lid)[90]. There may very well be other proteins in vertebrates which are involved in maintaining an active chromatin. Further studies may show they are also dependent on specific sequences of the genome, which are related to the Ultraconserved Regions (UCRs). It is possible that a certain combination of these UCGs mark a site which should be generally silenced or constitutively expressed or alternatively a site that can oscillate between both conditions.

CpG methylation

One of the first advances in understanding tumour genetics was the realisation that a gene could be silenced by methylation of the DNA, specifically on the 5'-cytosine of a CpG dinucleotide, upstream of the promoter site. If the silenced gene was a tumour suppressor, the resulting loss of transcription could result in a cancer (for reviews see[91-94]). A CpG methylation is catalysed by DNA methyltransferases (DNMT) by transferring a methyl group from S-adenosylmethionine to the cytosine ring. In vertebrates about 30% of all genes contain a CpG island, about 5,000-7,000 genes; the rest are silenced by a different mechanism. Approximately 70% of all CpG sites are methylated in normal human cells, as it is the main mechanism for shutting down unwanted retrotransposons, repetitive DNA, satellite DNA and any foreign DNA. There are two major events in cancer, global hypomethylation, which most often occurs in the later stages of tumour growth, and metastases. The other event is hypermethylation of spe-

cific genes, although this can include large areas of the genome as well, at times a total length of 4 Mb[95]. These two processes, hyper- and hypomethylation occur simultaneously, although we have much more data on hypermethylated genes[94].

Hypomethylation can cause 70% of normal methylated sites to be demethylated but specific hypomethylation of some genes can be observed[96]. In a study of monozygotic twins, it was found that aging is due to the massive loss of DNA methylation, as well as hypermethylation of a few specific genes, very similar to tumour development, called epigenetic drift[97]. Moreover, the link between ionizing radiation and chromosome instability has been reported. Chronic UV exposure resulted in global DNA hypomethylation in mice[98]. The release from transcript repression caused by hypomethylation allows a large number of transposons, repeat elements such as LINE1s and Alus to be expressed, causing deletions and translocation as well as a greater number of mitotic combinations[99].

However this is not the only reason for chromosomal aneuploidy and chaos. New results have shown that deletion of the H3K9Me3 mark made by mammalian histone methyl transferase G9a led to widespread genomic instability due to centrosome disruption and not genetic expression[100]. Different histone methyl transferases act at H3K9. Essentially G9a carries out the monomethylation and dimethylation reaction, whereas the trimethylation mark is carried out by Suv39h1 and Suv39h2. The latter two enzymes have redundant roles in the structural organisation of the genome and do not cause genomic instability. The basis of aneuploidy may have a separate origin. The nuclear relocation of the lipid phosphatase, PTEN, which regulates PI3K signalling, was disrupted by interference in its ubiquitylation state. This resulted in an increased number of chromosomal fragments with breakage at the centromeres and chromosomal translocations[101]. Therefore it is thought PTEN might function as guardian of the chromosomal stability. However, more research focused on hypomethylation is needed to tease out the details of this mechanism to understand exactly why some cancers remain diploid.

There are over 100 genes that are hypermethylated and hypomethylated in cancer and it could be as high as 400 genes[http://www.pubmeth.org][102]. Most hypermethylated promoters are occupied as well by methyl CpG binding proteins (MBD), which are made up of MeCp2, MBD1 and MBD2. MBD3 is part of the NuRD complex and is recruited to CpG methylated DNA through association with MBD2. The MBD pattern of tumour cells is highly characteristic of the cancer[103]. Half of the genes that cause familial cancers due to mutations are hypermethylated in sporadic cancers. Measurement of certain CpGs can also play a role in prognosis. CpG methylation of a homeodomain transcription factor was able to predict the possibility that distant metastases in breast cancer would reappear[104]. Moreover, the hypermethylation of a DNA repair protein, MGMT, in untreated patients with astrocytoma, was a marker of a poor prognosis, related to a great number of mutations occurring due to lack of repair[105]. The CpG methylation pattern of a few genes in prostate cancer could distinguish between late primary stage tumour, androgen-independent tumour and metastases[106].

Insulator CTCF binding factor

Within a cluster of genes which are silenced, some can be deselected by the binding of CTCF, a conserved protein named from CCCTC-binding factor, which recognises long and diverse nucleotide sequences[107]. However in humans, this insulator is a conserved protein CTCF, with an 11 Zn-finger domain which can bind to 13,800 sites of a 20-mer consensus sequence of DNA. This plays an essential role in protecting key genes from inadvertent CpG methylation and silencing of the promoter site. Half of the CTCF binding sites in the human genome were found at sites remote from the transcription site, while the other half were found in intergenic regions, where it was considered the insulator was segregating alternative promoters, as 52% of human genes possess several promoters.

CTCF is a candidate tumour suppressor gene as point mutations and Loss of Heterozygosity (LOH) at 16q22 produced a variety of cancers in breast, prostate and Wilm's tumours[108]. The 11-Zn-finger protein can use different combinations of the Zn finger domain to bind different DNA target sequences. CTCF binding sites were very far from promoters with an average distance of 48,000 nt. Only about 20% of CTCF sites were near transcription start sites. A CTCF was bound

to the promoter site of retinoblastoma gene (Rb) to prevent its aberrant methylation. In addition, mutations of the CTCF-binding site on the Rb promoter can induce CpG hypermethylation that can spread to the other CpG islands of the promoter[109]. This could be a part explanation of the classification of some colon cancers into a CpG Island Methylator Phenotype (CIMP) where a large number of genomic regions are shut down due to hypermethylation[110]. A mutation in the CTCF protein could prevent its efficient binding to promoter regions of hundreds of genes.

Maintenance of active chromatin

The following description is an hypothesis of how an active transcription site might be maintained through histone variants and enzymatic additions to the carboxyl or amino terminal of the tails of histones. At this stage, it is not known the exact ordering of events nor whether all enzymatic changes occur on all the histones, nor the specificity of the different methylases and acetylases. At the regulatory region or promoter site of the active gene locus, the variant histone, H3.3 is present (Figure 1). Some regions of the genome are 'hot' as they have a high rate of turnover of nucleosome assembly and disassembly, in particular promoter regions[73]. The signature of the methylated lysine reside on H3.3K4me recruits chromatin remodelling factors CHD1 and hence destabilises the nucleosome[111]. H3K4me2 definitely correlates with active chromatin in all studies[112]. Part of this process also involves the acetylation and methylation of other lysine groups on the same histone, H3K9Ac and H3K79me. Because these modifications make the nucleosomes unstable, they are continually evicted and reformed with the interchange of histone variants, H3.3 and H3.1 which are escorted to the site with histone chaperone, HIRA. The particular modification, H3.3K4me, is very important in this process because the substitution of the lysine with glutamic acid fails to maintain the active transcriptional memory in nuclear transplant embryos[113]. Another nucleosome variant, H2A.Z, together with H3.3, also appears to be important in nucleosomal destabilization and keeping transcription open[114].

The addition of an acetyl group on the lysine of histone H3K9Ac by HAT enzymes is associated with active transcription as RNA polymerase II was always

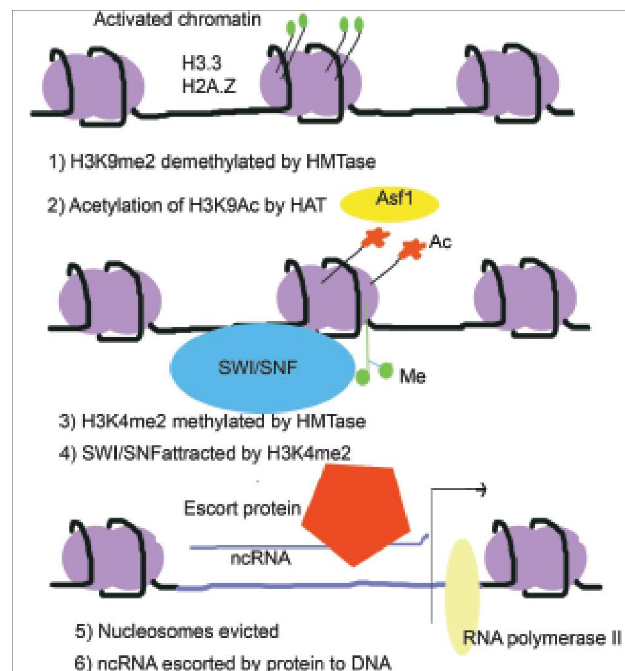


Figure 1 Maintenance of active chromatin. It is postulated that Histone 3.3 or H2A.Z is dimethylated at H3K9me2 by a histone methyl transferase (HMTase). The histone is acetylated at H3K9Ac by histone acetyl transferase (HAT), perhaps escorted by Asf1 protein. The histone is also methylated at H3K4me2 by another HMTase. A chromatin remodelling protein, SWI/SNF is recruited by the H3K4me2 mark. The H3.3 containing nucleosome is evicted from the promoter site by SWI/SNF, an ATPase. The noncoding RNA, with sequence specific to the DNA is escorted to the site by an unknown protein, allowing RNA polymerase II to bind at the transcription start site.

found at these sites. It is thought the acetyl group opens up the histone due to the neutralization of the positive charge on the lysine group on the histone. Another novel modification is acetylation of H3K36Ac in yeast, *Tetrahymena* and mammalian cells[115]. There is also another destabilizing histone modification associated with H3.3 which acetylates K56[73]. Modulation of nucleosome stability is a key mechanism to epigenetic regulation. It is thought that the same lysine site may compete for acetylation or methylation and that this could be the switch between an active or silent gene locus. The same applies to H3K9, which can be acetylated or methylated but never both.

So it is thought that one of the members of the SWI/SNF group, an ATPase remodelling complex, such as Brahma (BRM) or Brahma-related gene 1 (BRG1), binds to acetylated H4 tails and destabilise the nucleosome. The different ATPase remodelling complexes bind different hormone and transcription factors as only

BRG1 and not the ISWI complex was not found at the steroid hormone active site[116]. Also another histone escort protein, called anti-silencing function protein (Asf1), escorts H3 and H4 to sites of assembly[117]. It can also act as a nucleosome eviction agent, knocking off the H3-H4 dimer from DNA. Asf1 facilitates the acetylation of histone H3K56 by the unusual histone acetyl transferase, Rtt109[118]. So histone modification of the tails can anchor the nucleosomes together or tear them apart.

It is postulated that a short ncRNA, synthesized in cis or trans, bound to an escort protein can attach to the exposed DNA strand, forming a triplex, and begin the initiation of transcription. Little experimental evidence is available for this stage of the process. It is assumed the sequence of ncRNA determines if the ncRNA is a signal for transcription or silencing, perhaps specified by binding to different promoter sites on the one coding gene. The transcription factors and RNA polymerase II attaches to the DNA and the coding gene is transcribed into pre-RNA, processed by the removal of the introns and finally, after capping and polyadenylation, the mRNA is transported to the cytoplasm for translation. Perhaps there is also a counting system, as associated with the X chromosome, so that depending on how frequently the gene is transcribed, the modification on the histone can vary to make it more easily accessible. There are hundreds of modifications of the histone amino acids, phosphorylation, sumoylation, ubiquitination, proline isomerisation, besides methylation and acetylation. This would mean the gene can adapt in many different ways to the environment as conditions change.

Active chromatin coverted to silent chromatin

More is known about the permanent silencing of chromatin than the conditions for maintaining transcription, due to an intensive study of tumor suppressor genes, hypermethylated and shut down in cancer. Just as some histone variants are involved with active transcription, other histone variants are associated with silencing. These silencing nucleosomes are H3.1 and H3.2 in flies and mammals, which make up a large proportion of their genome. We have based the following on events established for shutting down the second X chromosome in females[119]. The first step may be the binding of the ncRNA to the exposed DNA double strand, free of nucleosomes (Figure 2). Recent evidence found that ncRNA could co-localise to the DNA[120]. This is a signal for the reinsertion of the nucleosomes, which are realigned in dimers of H3-H4 and H2A-H2B, and chromatin remodelling is carried out by one of the ATPase complexes, SWI/SNI. Most of the epigenetic histone modifications are carried out on the H3-H4 dimer. This formation would partly exclude the binding of RNA polymerase II and transcription factors at the promoter site.

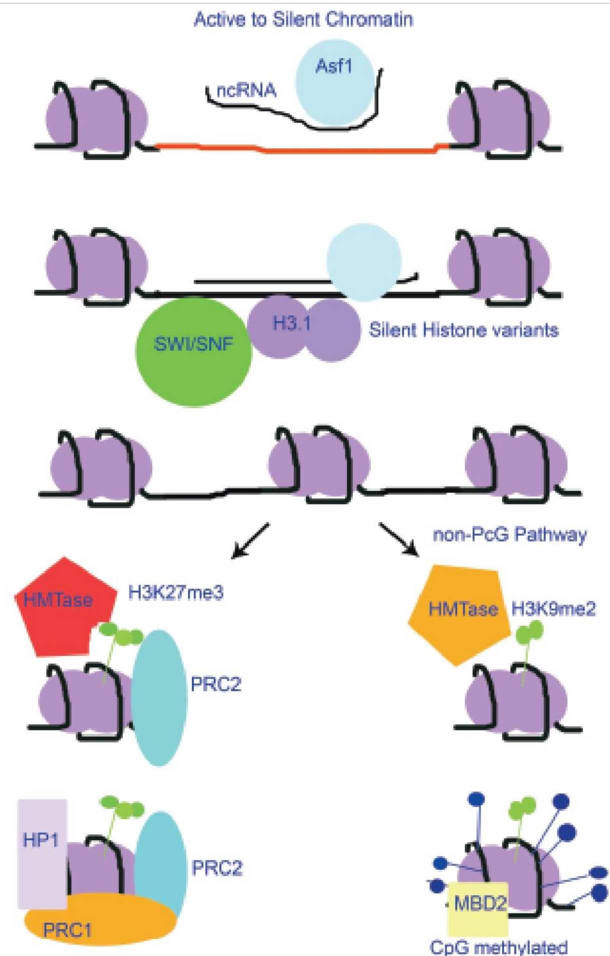


Figure 2 Active chromatin converted to silent chromatin. It is postulated that non-coding RNA with sequence specific to the DNA is escorted to the promoter site. This recruits nucleosomes H3.1 and chromatin remodelling protein, NuRF. The DNA is wrapped around the nucleosome. There are two pathways to the silent state, one is through the PcG proteins, the repressive complexes, PRC2 and PRC1 via methylation of H3K27me3, which may act on promoter sites lacking in CpG islands. This attracts HP1α which links the nucleosomes together like a bunch of grapes. The other pathway utilizes methylation of H3K9me2. This mark then attracts methyl binding proteins, (MBD) and DNA methyl transferase (DNMT) which methylates the CpG islands in the DNA, to silence the gene.

A nucleosome remodelling and deacetylase complex (HDAC) must act on the DNA sequence for silencing. The silenced transcription might require a dedicated remodelling complex, or the same ATPase complex, that can react to both active and silencing histone signals. In humans, the combined silencing complex, NuRD, is known to contain at least seven polypeptides, including histone deacetylase (HDAC1 and 2), H4 interacting proteins (RbAp46/48), methyl-binding protein 3 (MBD3), MTA-family members (MTA1-3) and a SNF2-like chromatin-remodelling ATPase (Mi-2/CHD4)[121]. Further, another methyl binding protein, MBD2, can also recruit NuRD through other DNA specific binding proteins.

There are two mechanisms for silencing genes, one based on the PcG group with the histone mark H3K27me3 and the other based on the histone mark of H3K9. The H3K27me3 mark appears strongly over the promoter sites of genes which lack a CpG island. It should be noted, however, that there are situations in which both H3K9 and H3K27 methylation can silence gene expression without any involvement of CpG DNA methylation[122]. The binding of ncRNA to the DNA recruits a histone demethylase(HMT), one for H3K4me2 and another demethylase for H3K36 or other activating methylated marks. The polycomb repressive complex, PRC2, is also transiently recruited by the presence of ncRNA where PRC2 carries out a trimethylation H3K27me3 and a deacetylation, H4K16. One of the subunits of PRC2 is a histone deacetylases, RbAp48 and one of the other subunits is a histone methyltransferase, EZH2, which also depends for its activity on the presence of EED. H3K27me3 is very important as it is a recognition site for the second PcG complex, PRC1. While the PRC2 complex is sitting on the DNA, PRC1 is also transiently recruited by the ncRNA. PRC1 contains a Ring1B enzyme, a ligase which ubiquitinates H2AK119ub, allowing compaction of the nucelosomes. This is unusual as ubiquitin is a large compact globular molecule of 76 amino acids. Ubiquitination of the other histone, H2B, may have the opposite effect, that of gene activation[123].

In addition, at least in the inactivation of the X chromosome, the PRC1 complex can be recruited to the locus, independently of the H3K27me3 signal, carried out by PRC2. That is, Ring1B can act independently of PRC2 in silencing the DNA[124]. To shut down the system, the antisense, ncRNA binds to histone escort protein Asf1, which transports it to the open chain of DNA, free of nucleosomes, again to form a triplex of two DNA molecules and the RNA. As observed in *Drosophila*, the Asf1 protein is able to form a complex with H3 and H4. In addition there are species differences, with H4 often much less modified than H3[115].

Methylated histone, H1K26me is able to attract chromo protein HP1α, (the alternative name is Cbx5), and bind other nucleosomes through another HP1α link. Finally the PRC2 complex is able to bind DNA methyl transferase, DNMT3a and methylate the cytosine CpG to lock in the repression. NuRD plays a role in this activity, as it is postulated that the remodelling activity of NuRD may facilitate access of DNMTs to chromatin template for deposition of methyl groups at CpG sites[121]. In addition there is a gain of methylation of another lysine, H3K9me, but both mono and di-methylated forms are repressive. The intimate link between NuRD and DNA methylation was postulated since the identification of MBD2 and MBD3 within the MeCP1/NuRD complex. It was previously shown that the methyl binding proteins are not a subunit of NuRD but recruit NuRD to methylated DNA[125].

There is some indication of how the silencing system may operate in cancer due to the research on imprinting of the IGFR, IGF and AIR genes and more recently the HOX genes. The use of a microarray at a 5bp resolution of the HOX genes revealed how antisense ncRNA controls remodeling of the chromatin through methylation[126]. In mammals there are 39 HOX genes but 407 discrete transcribed regions were found in just four HOX loci A, B, C, D. Some 101 of these were exons and 75 were introns, leaving 231 ncRNAs mainly from the intergenic regions. Some 74% of these ncRNA were antisense transcription from HOX genes. Surprisingly, the ncRNAs were enriched for specific DNA sequence motifs, which were considered to be regulatory sites, either DNA or RNA, operating in the same general area, in cis.

The region could be divided into two separate domains, that either allowed transcription or were silenced. Both HOX and ncRNA transcription were in a broad domain occupied by RNA polymerase II and H3K4me2, a mark of expression. However the silent

regions were occupied by PRC2 and the histone mark of silencing H3K27me3. A specific ncRNA at the boundary between active and inactive domains in the HOXC locus was examined more closely. It was found that a previously unknown ncRNA, labeled HOTAIR ncRNA, transcribed from the HOXC, silenced the HOXD locus by targeting PRC2 with its components of SUZ12 and EZH2 to methylate the histone tails. It has been postulated that ncRNAs may guide specific histone modification activities to discrete chromatin loci. The HOTAIR ncRNA was required for methylation on H3K27me3 by its HMTase activity, associated with PRC2 and thus for transcriptional silencing of the HOXD locus[126].

In another example, which applies to *Drosophila* and may be less relevant for humans, transcription of three elements of a ncRNA, all exactly the same sequence at an active site called TRE in the promoter region of the Ubx gene, were able to each bind a HMTase, Ash1[127]. The ncRNA transcript of the TRE bonded to the DNA copy and with Ash1 was able to cause the transcription of the gene Ubx. This occurred as the Ash1 carried out trimethylation of histone lysines, H3K4, H3K9, and H4K20, signals for transcriptional activation, by sitting on the RNA:DNA scaffold and enzymatically converting the histones. Thus a previously silent transcription site was converted into an active transcription site[4]. It is similar to the mechanism, first discovered in a mouse, where the long ncRNA XIST that can act in trans to regulate the chromatin domain of the extra X chromosome.

There is clear evidence that small ncRNAs can regulate gene expression in cancer cells. Small 21-nt dsRNAs targeting selected promoter regions of human genes E-cadherin, p21WAF1/CIP1 (p21), and VEGF in human prostate cancer cell lines were able to cause long-lasting and sequence-specific induction of the targeted genes. E-Cadherin is epigenetically silenced in cancer HeLa cells because of aberrant methylation of its CpG promoter. However three other 21 nt dsRNAs from other sites in the promoter region did not initiate expression of E-Cadherin, indicating this effect was sequence specific. In addition other dsRNAs targeting promoter sites of coding genes p27, PTEN and APC did not cause increased gene expression, as the researchers were just unlucky in their choice of sequences. The Ago family of proteins are key regulators of RNA silencing. Humans have a total of four closely related Ago proteins (Ago1-4) that interact with the trigger of dsRNA and function in target recognition. Further Ago2 was necessary for this activation, pointing to the possibility that stand separation and removal of one passenger stand of the dsRNA was important for gene activation to occur[28]. The Ago proteins may function to deal with exogenous DNA and may not play a role *in vivo* in carcinogenesis.

Cancer, histones, PcG and DNA methylation

Some mutations have been found in the nucleosome remodelling proteins. Not all proteins in the epigenetic process will be found to undergo a mutation, as many which do occur would be lethal to the cell, provoking apoptosis. The cell can only survive if mutations take place in redundant proteins acting in a pathway, which may also have an alternative and functional back up system, even though it may be inefficient. Only some of the more recent oncogenes and tumour suppressor genes will be mentioned, as the list is getting longer every day with the discovery of miRNAs and ncRNAs. Among the many members of the SWI/SNF family, four are involved with cancer development: PASG/LSH, BRG1, HLTF and SNF5. The smallest of the complexes, SNF5, shows inactivating somatic and germline mutations and plays a role in p53 dysfunction as well as p21CIP/WAF1 and p16INK4a in malignant rhabdoid tumours[128].

Another ATPase complex containing BRG1 (sometimes labelled SMARCA4) has been shown to be required for cell cycle control, apoptosis and cell differentiation in several biological systems. BRG1 plays a role in negative regulation of cell cycle progression by binding to the Rb[129] and the tumour suppressor protein p53. Other SWI/SNF members bind to c-MYC and BRCA1. Both Rb and p53 are master genes in the control of cell cycle exit that can lead either to cell differentiation or to cell death. Forced expression of BRG1 caused cell death through activation of p53 and should be considered as a tumour suppressor gene[130]. Mutations of the coding sequence of BRG1 were found in 24% of lung cancer cell lines[131].

Mutations in mouse chromatin-remodelling protein SNF2, the Lsh genes and human ATRX, have signifi-

cant effects on global DNA methylation patterns[121]. Another of the chromatin remodelling enzymes is a chromodomain helicase DNA binding protein called CHD. An examination of the CpG islands for the nine members of the family of CHD revealed that only one CHD5 promoter was methylated[132]. This occurred particularly in gliomas, colon and breast carcinomas. So chromatin modifiers represent an additional level of transcription regulation that can be disturbed and provoked into tumorigenesis[90].

Since the proliferating cells are shutting down genes regulating differentiation and tumour suppressor genes, it would be expected that there would be elevated levels of PcG factors. This is exactly what happens as most cancer cells have elevated levels of PcGs, all the histone deacetylases, such as SIRT1 and an essential component of PRC2, the EED protein, while these proteins are undetectable in normal tissue. In humans, the PRC1 complex contains the oncogene, BMI-1, which has a RING-finger domain. BMI-1 is important in the maintenance of stem cells, in particular the self-renewal property of stem cells, but plays no role in differentiation. Decreases in BMI-1 levels downregulate the expression of tumour-suppressor genes, such as p16INK4a and p14ARF, both of which are often inactivated by epigenetic means in colon cancer[72]. The human BMI-1 gene is located on chromosome band 10p13 and chromosomal translocation involving this region has been reported to occur in infant leukaemia and T-cell malignant lymphoma[133]. Microarray analysis of 71 prostate carcinomas indicated that patients with tumours having increased levels of both BMI-1 and the histone methyl transferase EZH2 manifest clinically aggressive disease phenotypes. It was an indication that patients were significantly more likely to relapse and develop disease recurrence after radical prostate surgery[134]. In addition, microarrays of PcG expression, in a similar fashion to microarrays of miRNAs, might be valuable as prognostic markers of patient survival from a cancer[135].

Many mutations in the enzymes chemically modifying histones can cause a cancer. Acetylation is present at low levels in the normal genome but in cancer global hypoacetylation or hyperacetylation can occur. Altered histone acetyl transferases (HATS) occur in leukaemias due to fusion proteins[136]. Only one mutation has been found in the genes coding for HDAC2 in cancer but they are often over-expressed[137]. A loss of monoacetyla tion at H4K16Ac and a loss of trimethylation at H4K20 me3 have been found in cancer cell lines of lymphocytes, breast, lung and colon as well as in lymphomas and colon carcinomas[136]. This was related to hypomethylation of repetitive DNA areas, which indicates a loss of differentiation and active transcription of inappropriate genes. Other results suggested that histone H3K9 deacetylation appears to play a crucial role in transcriptional repression of E-cadherin in colorectal cancers[138].

Another histone methyl transferase (MLL) is translocated in haematological malignancies with over 100 different translocations around chromosome 11, where the MLL gene is located[139]. It has been known for a long time that the histone methyl transferase EZH2 is upregulated in several tumours, lymphoma, prostate, melanoma and breast[84]. Overexpression of histone methyl transferase, EZH2, occurs in late stages of prostate cancer and is indicative of a poor prognosis[140]. In cancer cell lines of colon or prostate, there is evidence that EZH2 premarks the gene that is to be shut down in the next cell division; it initiates a de novo methylation. The activated EZH2, bound by PcG to the DNA, trimethylates the histone within the nucleosome, specifically H3K27me3[141]. In the next cell division, DNMT methylates the CpG site on the DNA due to the H3K27me3 signal. Thus levels of activity of the DNMTs are overexpressed during neoplastic transformation, especially in solid tumours, but they do not play any role in the selection of the genetic sites to be closed down. H3K27me3 epigenetic mark correlates with the distribution of PcG, sometimes over domains that are hundreds of kilobases in size. Thus altered expression of EZH2, SUZ12 as well as BMI-1 has been shown to occur in very different cancers in humans[82].

One of the mysteries of cancer that needs resolving is why epigenetic inactivation occurs on a functioning allele when the other allele is mutated or even deleted. This action inevitably drives the cell towards carcinogenesis. This has been documented with the mismatch repair gene, MLH1[142] and CDKN2A (the gene that encodes p16INK4a) in HCT116 cells and many others genes[143]. This procedure appears to take place in a time-dependent manner so that the cell could be adapting to a genetic switch as a result of the mutation[144].

The other mystery is why this occurs in only one specific tissue and not every tissue in the body, if it is an inherited mutation. Acute Myeloid Leukemia, an inherited condition, in 50% of cases is caused by a translocation around chromosome 11. Presumably every cell in every other organ also suffers from such deleterious translocations but effectively eliminates these stem cells or their function is neutralised. The same applies to an inherited cancer in the colon cancer disease FAP, where the germline mutated APC gene in the Wnt pathway, is often inactivated. In FAP, every cell in the body should be clogged with large amounts of aberrant mRNA and dysfunctional APC protein. Possibly silencing of the mutation also silences a miRNA specific for differentiation of that particular tissue. Some evidence points to the concept that different genetic events in the stem cell will determine the phenotypic profile of the tumour. Studies on papillomas and squamous carcinomas provide support for this interpretation[145]. Obviously some mutations will automatically lead to apoptosis and stem cell elimination as the result is catastrophic for survival of the cell. Other mutations, far more dangerous, allow the cell to survive because of redundant mechanisms that can compensate.

A related aspect which urgently requires an explanation is the question of pathway addiction[143]. Again the example is FAP, where the APC gene disturbs the Wnt pathway. However this wouldn't matter if members of the secreted frizzled-related gene family SFRP1, 2, 4 and 5 that encode Wnt antagonists remained active. Instead, cells carrying abnormal activation of Wnt signaling, such as activating mutations of β-catenin, also have these SFRP genes epigenetically silenced. A similar situation develops with the HIC-1 gene in mice[143].

There must be crosstalk taking place between CpG methylation and PcG-directed activity, modifying histones in the stem cells. The high rate of CpG DNA methylation in some colon cancers (CIMP+) is thought to be due to abnormal de novo methylation which occurs in the very early stages of the cancer[144]. About 370 genes in colon cancer were methylated at least once out of a total of 4500 genes examined. Many of these would have been methylated due to the induction of high levels of methyl transferase, EZH2 and not necessarily related to growth requirements. Genes in cancer cell lines, silenced in association with the H3K9 histone mark, undergo DNA methylation, so the two are linked in some manner[146]. However since genes can also be silenced by the H3K27me3 mark without DNA methylation, measurement of CpG methylation alone will give a considerable underestimation of the frequency of gene silencing in CIMP+ cancers. The majority of genes which lack a CpG island in their promoters were silenced by the H3K27me3 mark. In prostate cancer cells, 5% of promoters were silenced by H3K27me3, which included 16% of CpG islands and 84% of non-CpG island[147]. The situation was complicated as some genes contained both marks, H3K27me3 and CpG DNA methylation, although CpG methylation operated through a pathway independent of PcG complexes. Thus both CpG methylation and Pc-mediated H3K27me3 can target the same genes. SW48 a colon cancer cell line which is CIMP+ had very few H3K 27me3 marks. It was concluded that different epigenetic mechanisms may be specific for tissue and cancerous cell lines[147].

Cancer stem cells

Recent studies have provided evidence of the existence of a population of stem cells located within a niche in each tissue. Stem cells are characterised by an asymmetric division resulting after division into one stem cell and one proliferative daughter cell. The stem cells continuously repopulate the tissue and replace the mature cells in each organ[148]. In the human colon, the stem cells are located at the base of each crypt and there are millions of crypts[149]. Proliferative daughter cells of the migrating compartment, after fully differentiating, move on their way up the crypt to the surface, where they are sloughed off into the lumen after 5-7 days. It is thought that the first genetic dysregulation that leads to a cancer occurs in one of the stem cell, leading to what has been labelled the cancer stem cell. Clonal origin of colon cells was proven a while ago, when an XO/XY individual who had an FAP had colon crypts which were either XO or XY. The adenomas were from either from an XO clone or an XY clone[150]. The origin of a cancer in a stem cell could explain why many anti-cancer drug treatments reduce the tumour to a small size but the cancer remerges because the drugs have not been able to eliminate the cancer stem cell population.

One mutated stem cell begins to take over and replace the non-mutant cells in the niche, called niche succession[151]. But every now and again to maintain the population, there is a stem cell division in which no proliferative stem cells are produced and both stem cells remain as stem cells. It is calculated that a stem cell becomes extinct every 8 years, about 9-10 times in a lifetime. So even if a mutation does not give a growth advantage, the mutation may hitch a ride through niche succession. If for example, an APC mutation occurs in one stem cell, only one lineage survives the niche. Thus the APC alteration will become fixed or extinct. The monocryptal adenomas have been known for a long time. The resulting APC+/- gives the cell a growth advantage, as it is able to expand at a faster rate[148].

Then a second mutation or epigenetic loss may occur, with perhaps the proliferative cell with both mutations able to move up the crypt. Niche succession occurs again, followed by clonal conversion to a crypt filled with APC-/- cells[151]. Over time the adenoma takes over the crypt leading to fission of crypts simply due to expansion and another mutation perhaps in a miRNA leads to dedifferentiation and an adenocarcinoma[152]. These mutations could involve any number of different coding genes and intergenic regions and DNA coding for ncRNA as described above. However the idea of niche succession, with an 8 year period where a cancer stem cell becomes fixed, would explain why cancers develop over a very long period of time, even with patients who have one inherited allele which is mutated. There is tremendous new interest in the relationship between stem cells, progenitor cells and cancer cells and exactly what epigenetic changes are occurring in each population and how does this influence the stem cell. Somewhere in this mix, the wires are crossed and perhaps the semi-active chromatin state gets translated into a permanent silencing of hundreds of genes[153].

APC gene and colon cancer

A close examination of the APC gene, might throw more light on cancer aetiology. More than 300 known mutations in APC gene are thought to affect the Wnt signalling pathway, of which the important regulator is β-catenin, which after translocation to the nucleus can switch on several Wnt-specific target genes[154]. Mostly APC protein is found in the cytoplasm but it shuttles between the nucleus and cytoplasm. A recent study showed that miR-135a and miR-135b both bind to the 3' UTR of mRNA for APC and suppress its expression[155]. This would also induce increased β-catenin and Wnt-signalling. Increased miR-135a and b expression was also found in colon cancer cells, despite the state of regulation of the APC gene, that is whether it was mutated, hypermethylated or wild type. This didn't occur in normal colon cells. The gene, LEMD1, where miR-135b is located in the first intron was also upregulated. MiR-135a is encoded by two copies in the human genome, located in the first intron of STAB1 on 3p21 and in intron 5 of RMST gene on 12q23. It was not known what was responsible for the upregulation of the miR-135. The biallelic APC mutation did not seem to be the important factor- the pathogenesis was driven by upregulation of the miR-135a and b. Thus the APC mutations and the increased expression of miR-135 worked in synergy to upregulate the Wnt pathway to an even greater extent. Perhaps the mystery of an inherited gene mutation only causing a cancer in one organ could be related to the downregula tion of certain miRNAs needed for differentiation in that organ.

However recent research shows that the protein APC can bind directly to DNA and prevent DNA replication[156]. The inhibition was not due to a ncRNA as it was independent of transcription. The inhibition mapped to the carboxy terminal of the 2843 aa long protein, a different site to that which binds to β-catenin. Phosphorylation of the serine or threonine in the binding motif also reduced inhibition of DNA replication. However the inhibition of DNA replication would be expected to prevent proliferation of cells and prevent cancers from developing.

There are at least three distinct pathways to an inherited cancer in the colon. One induced by a mutation in APC, producing FAP, as discussed. Another is hereditary nonpolyposis colorectal cancer (HNPCC), where there is a familial mutation in one of the DNA repair enzymes, MLH1, MSH2, MSH6 or PMS2. This results in random mutations of repeat runs of bases as they fail to be repaired, affecting genes such as TGFβIIR, E2F4 and Bax. The failure to repair DNA would result in an early or late appearance of cancer, even in inherited form as it would depend on exactly

which genes became dysfunctional. A third pathway is through a hyperplastic polyposis (HPP) leading to a serrated sessile adenoma[157]. This serrated pathway typically has an increase in microsatellites (MSI-H), BRAF or K-RAS mutation and hypermethylation of hundreds of genes. The HNPCC and HPP do not have any APC mutation, both sporadic and familial cases accounting for 20-30% of cancers[158]. The germline mutation causing HPP colon cancer, is not known but must involve an epigenetic mutation, perhaps a chromatin remodeller, or a histone modifying enzyme but not a PcG related protein.

Another pathway that occurs in spontaneous cancers also shows hypermethylation, with CIMP+, similar to HPP. The stem cell is thought to maintain a set of embryonic genes, which are kept in a poised low state of transcription to prevent differentiation. These genes have CpG promoters which are kept in a state of bivalent promoter domains. They contain both repressive mark on the histones, of H3K27me3 and the active mark of H3K4me2[159]. It is possible that something interferes in the transmission from temporary to permanent silencing, so that the bivalent promoter state is wrongly converted into a permanent shut down in these genes and this state is transferred to the proliferating cells. As discussed, there is evidence there are two states of transcriptional silencing, one mediated by CpG methylation and the mark of H3K9me3, and another where the genes lack a CpG island but contain an enrichment of H3K27me3[160]. So, CpG silencing is a higher order, where the gene locus is double- locked to ensure no transcription takes place. In a colon cancer cell line HCT119 which had a double knock out of DNMT1 and 3b, the CpG controlled genes could be left in a bivalent promoter state if treated first with histone demethylases, followed by treatment with histone deacetylases. These particular CpG promoters had two marks, that of the repressive H3K27me3 as well as the active mark H3K4me2, ready for low level expression. This is important for drug administration in patients so that CpG methylation must be removed first by 5'azacytidine in order to then allow inhibitors of histone deacetylases (TSA) to act to allow a low level of re-expression of the genes.

CONCLUSION

There are some general conclusions we can make about the initiation of a cancer from clinical studies. For an oncogene or a tumour suppressor gene, the mutation must affect function. As an example a mutant of the human DNA repair gene, O(6)-methylguanine-DNA methyl-transferase (MGMT), with eight mutations within the active site, was as active as the wild-type protein and was tolerated with no ill effects, without the production of a cancer, at least in a cell culture[161]. Function must include the ability of the antisense ncRNA which regulates the coding gene to fulfil its regulating function. We have not mentioned repair of mismatched nucleotides, nor that of double strand breaks, which are responsible for the initiation of many cancers.

From the above discussion, it is clear that the initiation and development pathway of a cancer is extremely complex. New discoveries have highlighted the role of ncRNAs, both sense and anti-sense. These are utilised not in the form of siRNAs and RITS complexes as in *Drosophila, Tetrahymena, and Sacharromyces,* but binding to DNA to form DNA:RNA triplexes. As described above, there are a thousand and one ways that a cancer could begin. Relapses after many cancer-free years from leukemias and solid tumours such as breast cancer usually have tragic results. One of the main challenges for the future will be the development of methods for killing the rogue cancer stem cells that re-emerge after a previous successful drug treatment.

Recent reviews have highlighted the possibility of cancer therapy by use of antisense RNA or siRNA to knock out particular coding genes. This would be a mistake, as we know so little about the genome and its system of regulation of transcription and silencing, and in particular the reason for a cancer arising in one specific organ. As we have seen, it is complicated, a delicate balance, with many layers of control, with every action having a reaction, both negative and positive[162]. The human genome has evolved over millions of years to protect itself from any injection of foreign DNA from whatever source. Such crude interference as introduction of a siRNA could initiate a cancer in another organ.

At the moment the best information that can be obtained from the present knowledge is a systematic method of assessing the extent of the growth of the

cancer, and its likely prognosis through the use of arrays of miRNAs, PcGs or even perhaps methylated CpG islands of specific genes. A better approach for a cure would be the development of individualised antibodies against cancer cells and aberrant stem cells. We should be well aware of clinical cases and trials that have already failed. The application of the siRNA therapy can go badly wrong, as indicated by an experiment on eight young patients suffering from X-linked severe combined immunodeficiency (SCID)[163]. The patients, after injection with a retroviral-based therapy, the lentivirus vector containing the corrected gene, initially improved. However, 30-34 months later, two patients developed leukaemia, indicating a serious problem with the delivery system as the proviral integration site disrupted a key gene, the LMO-2 gene. There may be other serious problems with gene therapy, such as the generation of chromosome instability and miRNA dysfunction. Thorough evaluation is even more urgent now that human clinical trials based on siRNAs are already underway and expensive legal action may result. With sufficient dedication and resources, we can find a cure for cancer, but the route may not be through the genome- as we have tried to show there are just too many unknowns.

REFERENCES

[1] P.Carninci, J.Yasuda, Y.Hayashizaki; Curr.Opin. Cell Biol., **20**, 274-280 **(2008)**.

[2] R.J.Britten, E.H.Davidson; Science, **165**, 349-357 **(1969)**.

[3] M.C.Frith, M.Pheasant, J.S.Mattick; Eur.J.Hum. Genet., **13**, 894-897 **(2005)**.

[4] K.V.Prasanth, D.L.Spector; Genes.Dev., **21**, 11-42 **(2007)**.

[5] P.Carninci, A.Sandelin, B.Lenhard, et al.; Nat. Genet., **38**, 626-635 **(2006)**.

[6] Y.Okazaki, M.Furuno, T.Kasukawa et al.; Nature, **420**, 563-573 **(2002)**.

[7] R.Kodzius, M.Kojima, H.Nishiyori et al.; Nat. Methods, **3**, 211-222 **(2006)**.

[8] J.Yasuda, Y.Hayashizaki; Adv.Cancer Res., **99**, 77-112 **(2008)**.

[9] P.Carninci, T.Kasukawa, S.Katayama et al.; Science, **309**, 1559-1563 **(2005)**.

[10] J.Cheng, P.Kapranov, J.Drenkow et al.; Science, **308**, 1149-1154 **(2005)**.

[11] S.Cawley, S.Bekiranov, H.H.Ng et al.; Cell, **116**, 499-509 **(2004)**.

[12] M.Pheasant, J.S.Mattick; Genome Res., **17**, 1245-1253 **(2007)**.

[13] C.Vanhee-Brossollet, C.Vaquero; Gene., **211**, 1-9 **(1998)**.

[14] S.Katayama, Y.Tomaru, T.Kasukawa et al.; Science, **309**, 1564-1566 **(2005)**.

[15] M.Furuno, K.C.Pang, N.Ninomiya et al.; PLoS Genet, **2**, e37 **(2006)**.

[16] J.A.Erwin, J.T.Lee; Curr.Opin Cell Biol., **20**, 349-355 **(2008)**.

[17] C.I.Seidl, S.H.Stricker, D.P.Barlow; Embo.J., **25**, 3565-3575 **(2006)**.

[18] K.Kimura, A.Wakamatsu, Y.Suzuki et al.; Genome Res., **16**, 55-65 **(2006)**.

[19] E.Birney, J.A.Stamatoyannopoulos, A.Dutta et al.; Nature, **447**, 799-816 **(2007)**.

[20] M.C.Frith, L.G.Wilming, A.Forrest et al.; PLoS Genet., **2**, e23 **(2006)**.

[21] S.A.Korneev, V.Straub, I.Kemenes et al.; J. Neurosci., **25**, 1188-1192 **(2005)**.

[22] M.R.Eccles, G.Grubb, O.Ogawa, J.Szeto, A.E. Reeve; Oncogene, **9**, 2059-2063 **(1994)**.

[23] P.Celano, C.M.Berchtold, D.L.Kizer et al.; J.Biol. Chem., **267**, 15092-15106 **(1992)**.

[24] E.M.Reis, R.Louro, H.I.Nakaya, S.Verjovski-Almeida; Omics, **9**, 2-12 **(2005)**.

[25] P.Ji, S.Diederichs, W.Wang et al.; Oncogene, **22**, 8031-8041 **(2003)**.

[26] A.Iacoangeli, Y.Lin, E.J.Morley et al.; Carcinogenesis, **25**, 2125-2133 **(2004)**.

[27] W.Yu, D.Gius, P.Onyango et al.; Nature, **451**, 202-206 **(2008)**.

[28] L.C.Li, T.Steven, Okino, Hong Zhao et al.; PNAS, **103**, 17337-17342 **(2006)**.

[29] I.Martianov, A.Ramadass, A.Serra Barros, N. Chow, A.Akoulitchev; Nature, **445**, 666-670 **(2007)**.

[30] T.Tsuchiya, G.Tamura, K.Sato et al.; Oncogene, **19**, 3642-3646 **(2008)**.

[31] G.Bejerano, M.Pheasant, I.Makunin et al.; Science, **304**, 1321-1325 **(2004)**.

[32] G.A.Calin, C.G.Liu, M.Ferracin et al.; Cancer Cell, **12**, 215-229 **(2007)**.

[33] S.Rossi, C.Sevignani, S.C.Nnadi, L.D.Siracusa, G.A.Calin; Mamm Genome, **(2008)**.

[34] G.A.Calin, C.Sevignani, C.D.Dumitru et al.; Proceedings of the National Academy of Sciences, **101**, 2999-3004 **(2004)**.

[35] I.Rigoutsos, T.Huynh, K.Miranda, A.Tsirigos, A. McHardy, D.Platt; Proc.Nat.Acad.Sci., U.S.A., **103**, 6605-6610 **(2006)**.

[36] A.Meynert, E.Birney; Cell, **125**, 836-838 **(2006)**.

[37] A.A.Aravin, G.J.Hannon, J.Brennecke; Science, **318**, 761-764 **(2007)**.

[38] H.Kawaji, M.Nakamura, Y.Takahashi et al.; BMC Genomics, **9**, 157 **(2008)**.

[39] G.Calin, C.Croce; Nature Reviews Cancer, **6**, 857-866 **(2006)**.

[40] J.B.Cowland, C.Hother, K.Gronbaek; Apmis, **115**, 1090-1106 **(2007)**.

[41] B.Zhang, X.Pan, G.P.Cobb, T.A.Anderson; Developmental Biology, **302**, 1-12 **(2007)**.

[42] E.Barbarotto, T.D.Schmittgen, G.A.Calin; Int.J. Cancer, **122**, 969-77 (**2008**).
[43] G.M.Borchert, W.Lanier, B.L.Davidson; Nat. Struct.Mol.Biol., **13**, 1097-1101 (**2006**).
[44] I.Bruno, M.F.Wilkinson; Cell, **125**, 1036-1038 (**2006**).
[45] P.R.Boag, A.Atalay, S.Robida, V.Reinke, T.K. Blackwell; J.Cell.Biol., **182**, 543-557 (**2008**).
[46] S.Griffiths-Jones, H.K.Saini, S.Van Dongen, A.J. Enright; Nucleic Acids Res., **36**, D154-158 (**2008**).
[47] C.V.A.Wynter; Medical Hypotheses, **66**, 612-635 (**2006**).
[48] R.Kulshreshtha, M.Ferracin, S.E.Wojcik et al.; Mol.Cell Biol., **27**, 1859-1867 (**2007**).
[49] A.Lujambio, M.Esteller; Cell Cycle, **6**, 1455-1459 (**2007**).
[50] J.V.Melo, D.J.Barnes; Nat.Rev.Cancer, **7**, 441-453 (**2007**).
[51] G.A.Calin, C.M.Croce; Cancer Res., **66**, 7390-7394.
[52] G.A.Calin, M.Ferracin, A.Cimmino et al.; The New England Journal of Medicine, **353**, 1793-1801 (**2005**).
[53] J.M.Cummins, Y.He, R.J.Leary et al.; Proceedings of the National Academy of Sciences, **103**, 3687-3692 (**2006**).
[54] J.Lu, G.Getz, E.A.Miska et al.; Nature, **435**, 834-838 (**2005**).
[55] H.He, K.Jazdzewski, W.Li et al.; Proceedings of the National Academy of Sciences, **102**, 19075-19080 (**2005**).
[56] L.Ma, J.Teruya-Feldstein, R.A.Weinberg; Nature, **449**, 682-688 (**2007**).
[57] C.Blenkiron, L.D.Goldstein, N.P.Thorne et al.; Genome Biol., **8**, R214 (**2007**).
[58] M.S.Kumar, J.Lu, K.L.Mercer, T.R.Golub, T.Jacks; Nat.Genet., **39**, 673-677 (**2007**).
[59] S.Chiosea, E.Jelezcova, U.Chandran et al.; Cancer Research, **67**, 2345-2350 (**2007**).
[60] Y.Karube, H.Tanaka, H.Osada et al.; Cancer Sci., **96**, 111-115 (**2005**).
[61] S.Chiosea, E.Jelezcova, U.Chandran et al.; Am.J. Pathol., **169**, 1812-1820 (**2006**).
[62] E.J.Lee, Y.Gusev, J.Jiang et al.; Int.J.Cancer, **120**, 1046-1054 (**2007**).
[63] H.Hermeking; Cancer Cell, **12**, 414-418 (**2007**).
[64] T.C.Chang, E.A.Wentzel, O.A.Kent et al.; Mol.Cell, **26**, 745-752 (**2007**).
[65] B.Perillo, M.N.Ombra, A.Bertoni et al.; Science, **319**, 202-206 (**2008**).
[66] G.T.Bommer, I.Gerin, Y.Feng et al.; Curr.Biol., **17**, 1298-1307 (**2007**).
[67] C.Welch, Y.Chen, R.L.Stallings; Oncogene, **26**, 5017-5022 (**2007**).
[68] K.Rajalingam, R.Schreck, U.R.Rapp, S.Albert; Biochim.Biophys.Acta, **1773**, 1177-1195 (**2007**).
[69] S.M.Johnson, H.Grosshans, J.Shingara et al.; Cell, **120**, 635-647 (**2005**).
[70] C.Mayr, M.T.Hemann, D.P.Bartel; Science, **315**, 1576-1579 (**2007**).
[71] S.R.Viswanathan, G.Q.Daley, R.I.Gregory; Science, **320**, 97-100 (**2008**).
[72] C.V.A.Wynter, T.Kambara, M.D.Walsh, B.A. Leggett, J.Young, J.R.Jass; International Journal of Cancer, **118**, 907-915 (**2006**).
[73] S.Henikoff; Nat.Rev.Genet, **9**, 15-26 (**2008**).
[74] A.Barski, S.Cuddapah, K.Cui et al.; Cell, **129**, 823-837 (**2007**).
[75] D.E.Schones, K.Cui, S.Cuddapah et al.; Cell, **132**, 887-898 (**2008**).
[76] A.Saha, J.Wittmeyer, B.R.Cairns; Nat.Rev.Mol. Cell Biol.M, **7**, 437-447 (**2006**).
[77] M.Buszczak, A.C.Spradling; Cell, **125**, 233-236 (**2006**).
[78] I.Whitehouse, T.Tsukiyama; Nat.Struct.Mol.Biol., **13**, 633-640 (**2006**).
[79] I.Whitehouse, O.J.Rando, J.Delrow, T.Tsukiyama; Nature, 450, 1031-1035 (**2007**).
[80] R.K.Ng, J.B.Gurdon; Nat.Cell Biol., **10**, 102-109 (**2008**).
[81] G.Mizuguchi, X.Shen, J.Landry, W.H.Wu, S.Sen, C.Wu; Science, **303**, 343-348 (**2004**).
[82] R.Kanno, H.Janakiraman, M.Kanno; Cancer Sci., **99**, 1077-1084 (**2008**).
[83] B.Schuettengruber, D.Chourrout, M.Vervoort, B. Leblanc, G.Cavalli; Cell, **128**, 735-745 (**2007**).
[84] A.Sparmann, M.Van Lohuizen; Nat.Rev.Cancer, **6**, 846-856 (**2006**).
[85] F.M.Raaphorst; Human Molecular Genetics, **14**, R93-R100 (**2005**).
[86] T.M.Hall; Curr.Opin Struct.Biol., **15**, 367-373 (**2005**).
[87] K.Tajul-Arifin, R.Teasdale, T.Ravasi, D.A.Hume, J.S.Mattick; Genome Res., **13**, 1416-1429 (**2003**).
[88] E.Bernstein, E.M.Duncan, O.Masui, J.Gil, E.Heard, C.D.Allis; Mol.Cell Biol., **26**, 2560-2569 (**2006**).
[89] D.Bourchis, T.H.Bestor; Bioessays, **24**, 297-299 (**2002**).
[90] H.Özdag, A.E.Teschendorff, A.A.Ahmed et al.; BMC Genomics, **7**, 90-105 (**2006**).
[91] C.K.Fog, K.T.Jensen, A.H.Lund; Apmis, **115**, 1060-1089 (**2007**).
[92] P.A.Jones, S.B.Baylin; Cell, **128**, 683-692 (**2007**).
[93] A.H.Ting, K.M.McGarvey, S.B.Baylin; Genes and Development, **20**, 3215-3231 (**2006**).
[94] M.Esteller; Nat.Rev.Genet., **8**, 286-298 (**2007**).
[95] J.Frigola, J.Song, C.Stirzaker, R.A.Hinshelwood, M.A.Peinado, S.J.Clark; Nature Genetics, **38**, 540-549 (**2006**).
[96] M.J.Hoffmann, W.A.Schulz; Biochem.Cell Biol., **83**, 296-321 (**2005**).
[97] M.F.Fraga, E.Ballestar, M.F.Paz et al.; Proc. Natl.Acad.Sci., U.S.A., **102**, 10604-10609 (**2005**).
[98] A.Mittal, C.Piyathilake, Y.Hara, S.K.Katiyar; Neoplasia, **5**, 555-565 (**2003**).
[99] A.R.Karpf, S.Matsui; Cancer Res., **65**, 8635-8639 (**2005**).
[100] Y.Kondo, L.Shen, S.Ahmed et al.; PLoS one, **3**, e2037 (**2008**).
[101] W.H.Shen, A.S.Balajee, J.Wang; et al.; Cell, **128**, 157-170 (**2007**).
[102] R.A.Hinshelwood, S.J.Clark; J.Mol.Med., (**2008**).
[103] L.Lopez-Serra, E.Ballestar, M.F.Fraga, M. Alaminos, F.Setien, M.Esteller; Cancer Research, **66**, 8342 (**2006**).
[104] S.Maier, I.Nimmrich, T.Koenig et al.; Eur.J. Cancer, **43**, 1679-1686 (**2007**).
[105] P.Hau, R.Stupp, M.E.Hegi; Dis.Markers, **23**, 97-104 (**2007**).

[106] M.Camoriano, S.R.Kinney, M.T.Moser et al.; Cancer Res., **68**, 4173-4182 (**2008**).
[107] T.H.Kim, Z.K.Abdullaev, A.D.Smith et al.; Cell, **128**, 1231-1245 (**2007**).
[108] G.N.Filippova, C.F.Qi, J.E.Ulmer et al.; Cancer Research, **62**, 48-52 (**2002**).
[109] I.A.De La Rosa-Velazquez, H.Rincon-Arano, L. Benitez-Bribiesca; Cancer Res., **67**, 2577-2585 (**2007**).
[110] J.P.Issa; Nat.Rev.Cancer, **4**, 988-993 (**2004**).
[111] R.K.Ng, J.B.Gurdon; Cell Cycle, **7**, 1173-1177 (**2008**).
[112] R.J.Sims, C.F.Chen, H.Santos-Rosa, T.Kouzarides, S.S.Patel, D.Reinberg; J.Biol.Chem., **280**, 41789-41792 (**2005**).
[113] S.B.Hake, B.A.Garcia, E.M.Duncan et al.; J.Biol.Chem., **281**, 559-568 (**2006**).
[114] C.Jin, G.Felsenfeld; Genes Dev., **21**, 1519-1529 (**2007**).
[115] B.A.Garcia, S.B.Hake, R.L.Diaz et al.; J.Biol. Chem., **282**, 7641-55 (**2007**).
[116] T.A.Johnson, C.Elbi, B.S.Parekh, G.L.Hager, S. John; Mol.Biol.Cell, **19**, 3308-3322 (**2008**).
[117] R.Natsume, M.Eitoku, Y.Akai, N.Sano, M. Horikoshi, T.Senda; Nature, **446**, 338-341 (**2007**).
[118] J.Schneider, P.Bajwa, F.C.Johnson, S.R.Bhaumik, A.Shilatifard; J.Biol.Chem., **281**, 37270-37274 (**2006**).
[119] D.Umlauf, P.Fraser, T.Nagano; Biol.Chem., **389**, 323-331 (**2008**).
[120] K.Murakami, M.Oshimura, H.Kugoh; J.Hum. Genet., **52**, 926-933 (**2007**).
[121] L.Morey, C.Brenner, F.Fazi et al.; Mol.Cell Biol., (**2008**).
[122] E.Vire, C.Brenner, R.Deplus et al.; Nature, **439**, 871-874 (**2006**).
[123] T.Akasaka, M.van Lohuizen, N.Van der Lugt et al.; Development, **128**, 1587-1597 (**2001**).
[124] S.Schoeftner, A.K.Sengupta, S.Kubicek et al.; Embo.J., **25**, 3110-3122 (**2006**).
[125] Q.Feng, Y.Zhang; Genes.Dev., **15**, 827-832 (**2001**).
[126] J.L.Rinn, M.Kertesz, J.K.Wang et al.; Cell, **129**, 1311-1323 (**2007**).
[127] T.Sanchez-Elsner, D.Gou, E.Kremmer, F.Sauer; Science, **311**, 1118-1123 (**2006**).
[128] J.Chai, A.L.Charboneau, B.L.Betz, B.E. Weissman; Cancer Research , **65**, 10192-10198 (**2005**).
[129] D.L.Burkhart, J.Sage; Nat.Rev.Cancer, **8**, 671-682 (**2008**).
[130] M.A.Napolitano, M.Cipollaro, A.Cascino, M.A. Melone, A.Giordano, U.Galderisi; J.Cell Sci., **120**, 2904-2911 (**2007**).
[131] P.P.Medina, O.A.Romero, T.Kohno et al.; Hum Mutat , **29**, 617-622 (**2008**).
[132] S.Mulero-Navarro, M.Esteller; Epigenetics, **3**, (**2008**).
[133] S.Bea, F.Tort, M.Pinyol et al.; Cancer Res., **61**, 2409-2412 (**2001**).
[134] G.V.Glinsky; J.Clin.Oncol., **26**, 2846-53 (**2008**).
[135] G.V.Glinsky, O.Berezovska, A.B.Glinskii; J.Clin.Invest., **115**, 1503-1521 (**2005**).
[136] M.F.Fraga, E.Ballestar, A.Villar-Garea et al.; Nature Genetics, **37**, 391-400 (**2005**).
[137] S.Ropero, M.F.Fraga, E.Ballestar et al.; Nat. Genet., **38**, 566-569 (**2006**).
[138] Y.Liu, Y.Hong, Y.Zhao, T.M.Ismail, Y.Wong, K.W. Eu; Cancer Invest., **26**, 575-582.
[139] E.McCormack, O.Bruserud, B.T.Gjertsen; Oncogene, **27**, 3765-3779 (**2008**).
[140] S.Varambally, S.M.Dhanasekaran, M.Zhou et al.; Nature, **419**, 624-629 (**2002**).
[141] Y.Schlesinger,R. Straussman, I.Keshet et al.; Nat. Genet., **39**, 232-236 (**2007**).
[142] A.McGivern, C.V.Wynter, V.L.Whitehall et al.; Fam.Cancer., **3**, 101-107 (**2004**).
[143] S.B.Baylin, J.E.Ohm; Nat.Rev.Cancer, **6**, 107-116 (**2006**).
[144] I.Keshet, Y.Schlesinger, S.Farkash et al.; Nat. Genet., **38**, 149-153 (**2006**).
[145] J.Perez-Losada, A.Balmain; Nat.Rev.Cancer, **3**, 434-43 (**2003**).
[146] Y.Kondo, L.Shen, P.S.Yan, T.H.Huang, J.P.Issa; Proc.Natl.Acad.Sci., U.S.A., **101**, 7398-7403 (**2004**).
[147] Y.Kondo, L.Shen, A.S.Cheng et al.; Nat.Genet., **40**, 741-50 (**2008**).
[148] E.J.Oakley, G.Van Zant; Leukemia, **21**, 612-621 (**2007**).
[149] E.D.Williams, A.P.Lowes, D.Williams, G.T.Williams; Am.J.Pathol., **141**, 773-736 (**1992**).
[150] M.R.Novelli, J.A.Williamson, I.P.Tomlinson et al.; Science, **272**, 1187-1190 (**1996**).
[151] A.Humphries, N.A.Wright; Nat.Rev.Cancer, **8**, 415-424 (**2008**).
[152] J.L.Tsao, J.Zhang, R.Salovaara et al.; Am.J. Pathol., **153**, 1189-1200 (**1998**).
[153] B.E.Bernstein, A.Meissner, E.S.Lander; Cell , **128**, 669-681 (**2007**).
[154] A.Klaus, W.Birchmeier; Nat.Rev.Cancer, **8**, 387-398 (**2008**).
[155] R.Nagel, C.Le Sage, B.Diosdado et al.; Cancer Res., **68**, 5795-5802 (**2008**).
[156] J.Qian, A.A.Sarnaik, T.M.Bonney et al.; Gastroenterology , **135**, 152-162 (**2008**).
[157] C.V.A.Wynter, M.D.Walsh, T.Higuchi, B.A. Leggett, J.Young, J.R.Jass; Gut., **53**, 573-580 (**2004**).
[158] J.E.East, B.P.Saunders, J.R.Jass; Gastroenterol. Clin.North Am., **37**, 25-46 (**2008**).
[159] B.E.Bernstein, T.S.Mikkelsen, X.Xie et al.; Cell, **125**, 315-326 (**2006**).
[160] K.M.McGarvey, L.Van Neste, L.Cope et al.; Cancer Res., **68**, 5753-5759 (**2008**).
[161] L.A.Loeb, K.R.Loeb, J.P.Anderson; PNAS, **100**, 766-781 (**2003**).
[162] A.L.Gartel, E.S.Kandel; Biomol.Eng., **23**, 17-34 (**2006**).
[163] S.Hacein-Bey-Abina, C.Von Kalle, M.Schmidt et al.; N.Engl.J.Med., **348**, 255-256 (**2003**).