CHAPTER II

LITERATURE REVIEW

2.1 COLORECTAL CANCER

Colorectal cancer could grow in all sections of the large intestine. Even though its position to some extent influences the prognosis and obviously, the approach to surgical treatment, from a biological point of view, it could be considered less than one heading, i.e. colon cancer. Captivatingly, cancers of the small intestine are very rare and certainly signify a distinct disease. There is a good morphological and molecular genetical evidence that colon carcinoma develops through several precursor stages. The earliest recognizable preneoplastic changes result in hyperplastic or dysplastic crypts. It is generally agreed that adenomatous polyps are a precursor stage for many carcinomas. They are usually found as single benign tumors protruding into the lumen of the bowel and consist of a thickened, more or less disorganized epithelium. Multiple polyps are found in certain circumstances, e.g., in the familial cancer syndrome, 'familial adenomatous polyposis coli' (FAP). These polypous tumors are considered a type of adenoma and can be categorized into several substages according to the degree of growth and dysplasia. The incidence of colon cancer varies considerably across the world, with the highest incidences in Western industrialized countries. Colon cancer is usually a disease of older people. However, the incidence remains different between countries with an average of younger or older populations even

after adjustment for age. The causes for these differences in incidence are not really understood. The best evidence points to dietary factors as being responsible.

A large body of evidence supports the idea that accumulated genetic changes underlie the development of neoplasia. This multistep process is well illustrated by colorectal cancers, which typically develop over decades and appear to require at least seven genetic events for completion [13]. Even so, inheritance of a single altered gene can result in a marked predisposition to colorectal cancer in two distinct syndromes, Familial Adenomatous Polyposis (FAP) and Hereditary Nonpolyposis Colorectal Cancer (HNPCC). Recent evidence suggests that the genetic defect in FAP affects the rate of tumor initiation by targeting the gatekeeper function of the *APC* gene [14]. In contrast, the defect in HNPCC largely affects tumor progression by targeting the genome guardian of DNA mismatch repair [14]. Studies of these syndromes have provided unique insights into both inherited and sporadic forms of human tumors.

At least 50% of the Western population develop colorectal tumor by the age of 70, and in about 1 in 10 of these individuals, progression to malignancy ensues. As a result, colorectal cancer is the second leading cause of cancer death in the United States and first, when smoking related cancers are excluded [15]. Epidemiological studies have suggested that at least 15% of colorectal cancers occur in dominantly inherited patterns [16, 17]. The two best-defined familial forms are FAP and HNPCC. In the past few years, the genetic bases for both of these syndromes have been discovered, providing new insights into the nature of human neoplasia.

2.2 GENETIC OF FAMILIAL CRC

Between 2% to 5% of all colorectal cancers incidence, arise in the setting of inherited syndromes. These syndromes include Lynch syndrome, familial adenomatous polyposis, MUTYH-associated polyposis and certain hamartomatous polyposis conditions [18]. Each of the syndromes is associated with a high risk of CRC. In addition to the syndromes, up to 20% of CRC exhibit increased familial risk, likely related to inheritance. A number of less penetrant, but possibly more frequent susceptible genes might have involved in inheritance process of the disease.

Lynch syndrome, also known as a hereditary non-polyposis CRC (HNPCC), has an occurrence of approximately 1 in 300 people with CRC. In contrast, familial adenomatous polyposis is much less frequent and arises in approximately 1 in 7,000 people affected by CRC [19]. The genetic basis of Lynch syndrome is germline mutation in one of the genes in the DNA nucleotide mismatch-repair system, most commonly *MSH2* and *MLH1* (accounting for roughly two-thirds of the known mutations) and less commonly *MSH6*, *PMS1*, and *PMS2* [18, 20]. Almost all cases of Lynch-syndrome-associated colorectal tumours have microsatellite instability.MSI also occurs in about 15% of sporadic tumors, mainly by *MLH1* methylation silencing rather than by mutation, as in the familial cancers [21]. Microsatellites are repetitive DNA sequences comprising short reiterated motifs dispersed throughout the eukaryotic genome [22]. Microsatellite lengths are highly polymorphic in human populations, but appear stable during the life span of the individual. Two distinct modes of dinucleotide

microsatellite alterations in human colorectal cancer are: type A alterations, which are defined as length changes of ≤6 bp and type B changes, which are more drastic and involve modifications of ≥8 bp [23]. Both types of MSI were associated with *MSH2* or *MLH1* mismatch repair gene alterations.

Familial adenomatous polyposis (FAP) is the best known familial CRC syndrome. Affected individuals will inevitably develop CRC by the age of 20-30. FAP is a classical, rare, mainly inherited disease that affects one in 13,000 births [24]. FAP was one of the first relatively common Mendelian diseases whose gene responsible for the disease, Adenomatous polyposis coli [25], was identified by positional cloning and linkage analysis. *APC* is located on chromosome 5q21 which acts as a tumor suppressor [21]. It inhibits the Wnt signalling pathway in which the key component of this pathway is β-catenin [26]. *APC* acts as a negative regulator of Wnt signalling by targeting β-catenin for degradation. Mutation in *APC* impaired the interaction between *APC* and β-catenin, therefore excess β-catenin is translocated to the nucleus. β-catenin then activated the transcription of growth factor regulatory genes through its interaction with the transcription factor T-cell factor 4 (TCF4). About 80% of all CRCs have their first mutation in this gene, which is therefore a key initiator of carcinogenesis [14]. Many *APC* mutations have now been identified, both in the germ line in FAP patients, and somatically in sporadic CRCs. There are two relatively common germ line mutations at two different 5 base pair repeat positions where the mutation is up to 1,000 times that at ordinary single base pair changes. CpG positions, often with the C methylated, have about 40 times that rate. In FAP

germline mutations are located throughout the entire *APC* gene, and more than 90% of mutations introduce a premature stop codon that results in a truncated protein product [27]. However, APC mutations are not detected in 10-50% of FAP patients. Recently Thean et al [28] has searched for a new cancer gene by performing genome-wide genotyping on members of an APC mutation-negative FAP variant family and ethnicity-matched healthy controls. Their results showed no common copy number change was found in all affected members using the unaffected members and healthy controls as baseline. Though a 111 kb copy number variable (CNV) region at 3q26.1 was shown to have copy number loss in all eight polyps compared to match lymphocytes of two affected members [28]. A common region of loss in all polyps, which are precursors to CRC, is likely to harbor disease-causing gene in accordance to Knudsen's "two-hit" hypothesis. Through these result CNVs showed to have an important role in familial CRC.

The goal of cancer genetics is to discover all variant parallels that predispose to neoplasms. To this end, single nucleotide polymorphisms (SNPs) have been the most widely studied form of genetic variation and, by using massive whole genome studies (genome-wide association (GWA) studies), many common SNPs have been shown to be associated with cancer and other complex traits. However, the results of these efforts have not explained much of the heritability of disease [29]. CNVs, especially the smaller variants, have been essentially hidden from view until recently; thus, only a handful of studies have found an association of CNVs with cancer.

2.3 CNVs IN FAMILIAL CRC

Human cancer is caused in part by irreversible structural mutations. These can produce changes in DNA copy number at distinct locations in the genome [30]. Deletions, insertions, duplications, and complex multi-site variants of DNA segments, collectively termed copy number variants (CNVs) or copy number polymorphisms (CNPs), are found in all humans [31].Recent genome-wide studies have shed light on CNVs, an unexpectedly frequent, dynamic and complex form of genetic diversity, and have quickly overturned the idea of a single diploid human 'reference genome'. Although the characterization of the extent and location of these regions in healthy genomes is far from complete, many groups are actively trying to determine the clinical impact of CNVs in patient populations.

As with SNPs, CNVs that are found frequently in the healthy population (common CNVs) are very likely to have a role in cancer etiology. CNVs are generally thought to be depleted in gene regions. However Shlien et al. (2009) reported that 49 cancer genes are overlapped by a CNV in more then one person in a large reference population. In their analysis, using the Affymetrix 500K array, only CNVs directly overlapping a cancer gene were selected (either both breakpoints were inside the genomic interval containing the gene, both were outside the interval or one breakpoint was inside while the other was outside). Nevertheless these are probably an underestimation of the genuine number of common cancer CNVs as many smaller variants are missed at the resolution of this array (mean size of CNVs found using the array is 206 kb) and there are more

additional and more distal CNVs that have long range effect on cancer gene transcription [8].

An ongoing studies performed by a colleague, Venkatachalam et al. (2010) [12] has shown that in cases of unknown causes of familial CRC (MSS) CNVs occurs in regions that affecting microRNA (miR/miRNA) genes. These discoveries lead to the possibility that not only CNVs of large size occur but also smaller CNVs, that might has been missed by many techniques of screening, contribute to the predisposition of familial CRC.

Recent discovery of thousands of members of the class of noncoding RNAs (ncRNAs, which are genes without a significant open reading frame [ORF]), has added the genomic complexity of the cancer cells than previously anticipated. At present, cancer is considered a complex genetic disease involving structural and expressive abnormalities of both coding and noncoding oncogenes (OG) and tumor suppressor genes (TSG). Data accumulated in the last couple of years show that alterations in microRNA (miRNA) genes play a critical role in cancer initiation and progression. The genetic identification of hot spots for chromosomal abnormalities showed that miRNAs, a small ncRNA class of genes, frequently reside in such genomic regions.

2.4 MicroRNA

MicroRNAs (miRNAs) are functional RNA molecules that are transcribed from the DNA sequence of RNA genes, but not translated into protein. Several hundred genes within our genome have been shown to encode miRNAs. Lewis et al., using computational target predictions, have shown that 5,300 human genes

(~30% of the human gene set) are implicated as miRNA targets, therefore putting miRNA one of the most abundant classes of regulatory genes in humans [32].

2.4.1 MicroRNA DISCOVERY

The first miRNA was first discovered by Ambros et al. in 1993. On the genetic screening of the round worm Caenorhabditis elegans (C. elegans), one gene, *lin-4*, did not encode a protein but, instead, a 22-nucleotide small RNA [33, 34]. In 2000, another 22-nucleotide miRNA, *let-7*, was discovered in C. elegans. This miRNA was found to be involved in coordinating developmental timing and conserved across species, including humans, thus suggesting that it has an important biological function. Subsequently, many small regulatory RNAs similar to *lin-4* and *let-7* were identified in almost all multicellular organisms and were named miRNAs.

2.4.2 BIOGENESIS OF microRNA

MicroRNA maturation involves three major steps: transcription of pri-miRNA, cleavage in the nucleus to form pre-miRNA, and a final cleavage in the cytoplasm to form mature microRNA [35]. Pri-miRNA, found as independent transcripts or within the introns of another gene, is synthesized from DNA by RNA polymerase II and may be up to 1 kb in length, forming hairpin loops [36]. As an example, *lin-4* that lies within a host gene, is lined with conventional splice donor and acceptor sites, and is transcribed as an intron. While the genomic organization of *let-7*miRNAs are described as gene clusters found on chromosome 9 and 17 by Lagos-Quintana et al. [37].Subsequent of the transcription, pri-

miRNA is cleaved by the RNaseDrosha on the non-loop end and forming 60–70 bp length precursor microRNA (pre-miRNA). Pre-miRNA then exported from the nucleus into the cytoplasm of a cell via a transporter on the nuclear membrane. The nature of this transporter was previously suspected but remained unknown until 2003, when two independent teams published their finding that the transporter is RanGTP dependent Exportin 5 [38-40]. In the cytoplasm, pre-miRNA is cleaved by Dicer forming a miRNA-microRNA duplex that is unwound by a helicase. It releases two mature microRNAs of which one or both may be active [41].

Mature microRNAs inhibit protein expression in two different ways. Firstly, mature microRNA act through the RNA-induced silencing complex (RISC) to target and cleave mRNA [42]. RISC, well known in its association with small interference RNA (siRNA), joins with Argonaute 2 proteins Gemin 2 and Gemin 3 when it is charged by microRNA. Both microRNA and siRNA are small RNAs associated with RISC, but they differ in that siRNA matches exactly to its target mRNA and leads to cleavage of the mRNA [36]. Without perfect complimentary of miRs towards their target mRNA they can cleave their target, though the introduction of a synthesized miR that has a perfect complimentary has an action identical to that of siRNA [36, 43].

Secondly, microRNAs act through translational inhibition. When the microRNA does not perfectly align with its mRNA, target translational inhibition occurs [44]. In 1999, Olsen and Ambros showed that the inhibition of protein translation by miRNA through the isolation of *lin-4* microRNA from cytoplasmic

ribosomal complexes, along with *lin-14*miRNA [45]. Their study also demonstrated that the *lin-4* mechanism of action was not mediated by cleavage or destabilisation of the mRNA, because the level of *lin-14* remained steady and the poly-A tails were not shortened [45].

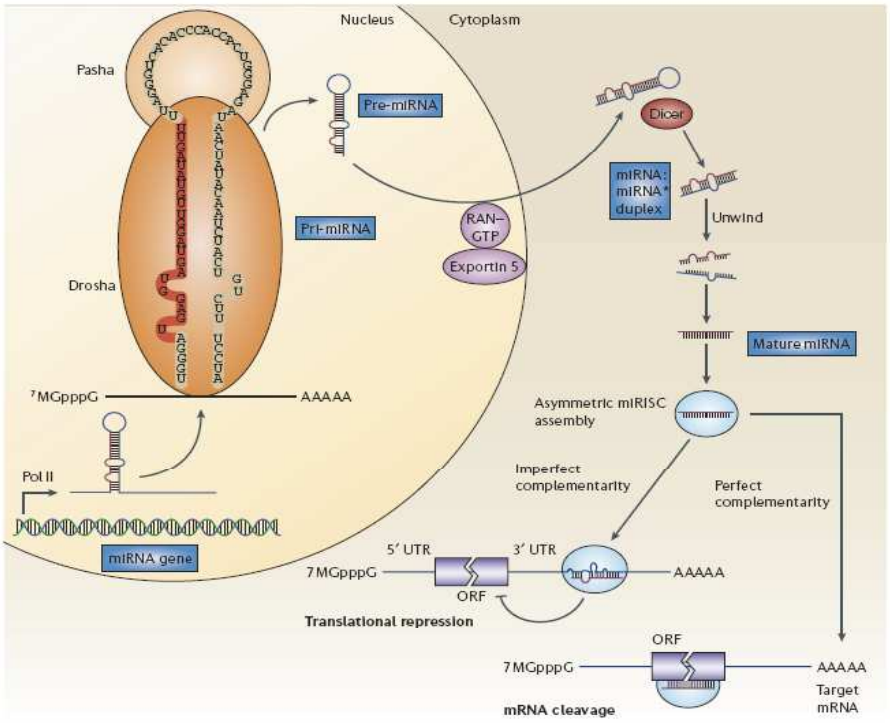

Figure 1.The biogenesis of microRNAs. MicroRNA (miRNA) genes are generally transcribed by RNA Polymerase II (Pol II) in the nucleus to form large pri-miRNA transcripts, which are capped (7MGpppG) and polyadenylated (AAAAA). These pri-miRNA transcripts are processed by the RNase III enzyme Drosha and its co-factor, Pasha, to release the ~70-nucleotide pre-miRNA precursor product. (Note that the human let-7a-1 miRNA is shown here as an example of a pre-miRNA hairpin sequence. The mature miRNA sequence is shown in red.) RAN–GTP and exportin 5 transport the pre-miRNA into the cytoplasm. Subsequently, another RNase III enzyme, Dicer, processes the pre-miRNA to generate a transient ~22- nucleotide miRNA:miRNA* duplex. This duplex is then loaded into the miRNA associated multiprotein RNA-induced silencing complex (miRISC) (light blue), which includes the Argonaute proteins, and the mature single-stranded miRNA (red) is preferentially retained in this complex. The mature miRNA then binds to complementary sites in the mRNA target to negatively regulate gene expression in one of two ways that depend on the degree of complementation between the miRNA and its target. miRNAs that bind to mRNA targets with imperfect complementary block target gene expression at the level of protein translation (lower left). However, recent evidence indicates that miRNAs might also affect mRNA stability (not shown). Complementary sites for miRNAs using this mechanism are generally found in the 3′ untranslated regions (3’ UTRs) of the target mRNA genes. miRNAs that bind to their mRNA targets with perfect (or nearly perfect) complementation induce target-mRNA cleavage (lower right). miRNAs using this mechanism bind to miRNA complementary sites that are generally found in the coding sequence or open reading frame (ORF) of the mRNA target. (Taken from [46])

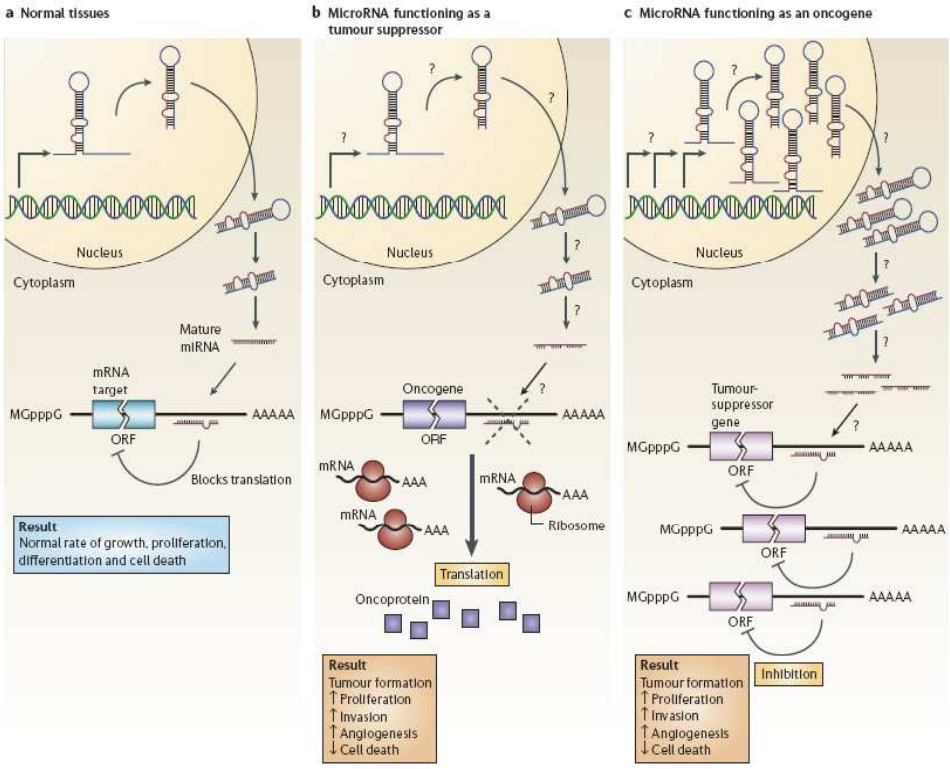

Figure 2.MicroRNAs can function as tumor suppressors and oncogenes. A. In normal tissues, proper microRNAs (miRNA) transcription, processing and binding to complementary sequences on the target mRNA results in the repression of target-gene expression through a block in protein translation or altered mRNA stability (not shown). The overall result is normal rates of cellular growth, proliferation, differentiation and cell death. **B.** The reduction or deletion of a miRNA that functions as a tumor suppressor leads to tumor formation. A reduction in or elimination of mature miRNA levels can occur because of defects at any stage of miRNA biogenesis (indicated by question marks) and ultimately leads to the inappropriate expression of the miRNA-target oncoprotein (purple squares). The overall outcome might involve increased proliferation, invasiveness or angiogenesis, decreased levels of apoptosis, or undifferentiated or de-differentiated tissue, ultimately leading to tumor formation. **C.** The amplification or over expression of a miRNA that has an oncogenic role would also result in tumor formation. In this situation, increased amounts of a miRNA, which might be produced at inappropriate times or in the wrong tissues, would eliminate the expression of a miRNA-target tumor-suppressor gene (pink) and lead to cancer progression. Increased levels of mature miRNA might occur because of amplification of the miRNA gene, a constitutively active promoter, increased efficiency in miRNA processing or increased stability of the miRNA (indicated by question marks). ORF, open reading frame. (Taken from [46])

2.5 MicroRNA IN CANCER

Shortly after the discovery that microRNA deletions were associated with CLL, microRNA abnormalities were also reported in pediatric Burkitt lymphoma, lung carcinoma, and large-cell lymphoma [47]. For example, miR-155, located in

the BIC gene, is highly expressed in many types of B cell lymphomas, including diffused large B-cell lymphoma and pediatric Burkitt lymphoma [48]. Takamizawa et al. [49] reported decreased expression of *let-7*miRNA in non-small–cell lung cancers, ranging from adenocarcinomas, to squamous carcinomas, to adenosquamous carcinomas, to large cell carcinomas of the lung. These authors also found that depressed *let-7* expression is correlated with poorer patient prognosis.

Michael et al. in 2003 published the first study of miRNA in colon cancer, identified miR-143 and miR 145 as potential factors in colon tumorigenesis [50]. Total RNA from matched normal and colon adenocarcinoma were collected by these researchers and they cloned the 18-22 nucleotide RNA fragments, which sequences had been compared to microRNA databases. The result was that miR-143 and miR-145, being located close together on chromosome 5, is expressed at reduced levels in colon cancer epithelial cells [50]. The targets of miR-143 and miR-145 remain speculative but there are several candidates such as *ERK5* and *IRS1*.

The *let-7* family consists of 14 isomers, which majorities are dysregulated in colorectal cancer. Akao et al. has reported down regulation of *let-7a-1* in colon cancer tumors and colon cancer cell lines. Reduced expression of *let-7a-1* in colon cancer cell line shows increased *RAS* and *c-myc* protein expression level lead to increase cell growth [51].

Not only are microRNAs in CRC down regulated, some of them also are up regulated as shown by Bandres et al [52]. Using RT-PCR and microarray

analysis, they demonstrate the elevated levels of *miR-31*, *miR-96*, *miR-31*, *miR-135b*, and *miR-183* in colorectal tumor cells. The authors also find the association of cells showing up regulation of miRNA which also harbored oncogenic mutation of either *KRAS* or *BRAF* [52].

Thousands of scientific groups has been studying cancer predisposition, however many remain unaware about the genetic basis in the majority of familial cases. In the light of new findings, miRNAs represent perfect candidates for cancer predisposition genes. MiRNAs work in a specific way that a small variation in the expression of a specific miRNA has effects on hundreds of target mRNAs and it leads to possible functional consequences. For example, variations in miRNA expressions that were inherited could represent the basis of predisposition in various types of familial cancer with unknown pathogenesis such as familial CLL or familial prostate cancer [53]. This hypothesis was supported through the finding of a mutation in the primary transcript of *miR-16* in a patient with familial CLL and familial breast cancer that induced a lower expression of the gene [54]. With the development of a CLL-like disease in a spontaneous mouse model with mutations of the *miR-15a/miR-16-1*, suggesting that the altered expression of *miR-15a/miR-16-1*is the likely to be the main molecular lesion in CLL [55]. These studies give a new perspective towards studying cancer predisposition that miRNAs are involve as cancer predisposing genes.

2.6 DNA COPY NUMBERABNORMALITIES AFFECTING microRNA GENES

Alteration in DNA copy number is one of the mechanisms that modify gene expression and function. DNA dosage alterations occurring in somatic cells are frequent contributors to cancer. The first example of the miRNA gene with DNA copy number alteration in cancer was reported in CLL patients. It was found that *mir-16-1* and *mir-15a* at 13q14 were deleted in more than 50% of the CLL patients, with concurrent reduced expression in ~65% patients [56]. Additional studies demonstrated that these two miRNAs suppressed *BCL2* expression and may serve as tumor suppressor genes in this disease [57]. Deletion of *mir-16-1* and *mir-15a* deletion were later identified in epithelial tumors, such as pituitary adenomas [58], and in ovarian and breast cancers [59]. In 2004, amplification of *C13orf25* at 13q31-32 was first reported in lymphoma patients [60]. Most interestingly, this amplified region contains seven miRNAs as a polycistronic cluster, and the expression of primary and mature miRNAs derived from this locus was increased in this type of lymphoma [61]. We now know that this miRNA cluster actually serves as an oncogene in human cancer through disturbed balance between cell death and proliferation via the proto-oncogene *c-Myc* mediated pathway [62, 63]. Using public database-retrieval and bioinformatics based approaches, Calin et al. compared 186 miRNA loci to the sequences of previously reported nonrandom genetic alterations and discovered that miRNA genes frequently reside in fragile sites, as well as in minimal regions of loss of heterozygosity, minimal regions of amplification, or common breakpoint regions

[11]. Recently, this finding was experimentally confirmed by an array-based comparative genomic hybridization (aCGH) study in 227 human tumors [59]. Common fragile sites are specific chromosomal areas that are susceptible to form gaps and breaks when cells are exposed to stresses that affect DNA synthesis [64]. The latest compendium comprises 103 Common fragile sites [65] in colon cancer includes: chromosome 16q23 [66], 8q24 (128.14-128.62 Mb) [67], 8p21 [68], 15q15 [69], 2q31–33 [65], 3p14.2 [70], 5q21 [65], Xp23.3 [64], 17q21 [71] and 6q26 [72]. The rearrangements are usually one or more large deletions of tens to hundreds of kilobase pairs within the fragile sites. To date, all fragile sites that have been cloned are practically AT-rich and they do not have expanded di- or trinucleotide repeats [64]. A more recent study has suggested that genomic copy number loss may account for the downregulation of approximately 15% of miRNAs in advanced ovarian tumors. These findings support the notion that DNA copy number alterations of miRNAs are highly prevalent in cancer and may account in part for the frequency of miRNA gene dysregulation.

2.7 OLIGONUCLEOTIDE ARRAY COMPARATIVE GENOMIC HYBRIDIZATION (Oa-CGH)

Microarray-based comparative genomic hybridization (arrayCGH) methods have been widely used to investigate chromosomal abnormalities, such as segmental copy number alterations, associated with cancer, developmental disorders, and population studies of normal copy number variation [73]. In aCGH, DNA from a test (e.g. tumor) and a reference genome (genomic DNA from a normal individual) are differentially labeled and hybridized to a representation of

the genome, which was originally a metaphase chromosome spread. The fluorescence ratio of the test and reference hybridization signals is determined at different positions along the genome and provides information on the relative copy number of sequences in the test genome compared with a normal diploid genome. Aberrations leading to gains or losses of part of the genome can be detected by CGH and include interstitial deletions and duplications, non-reciprocal translocations and gene amplifications. NimbleGen provide arrays containing 385 K oligonucleotides that photolithograhically synthesized on the array. The array production is extremely flexible such that each array produced can have a different set of oligonucleotides on it. The oaCGH oligonucleotides are designed to be isothermal and vary in length (http://www.nimblegen.com/).