Cancer is a complex disease in which a group of cells display uncontrolled growth, invasion that intrudes upon and destroys adjacent tissues and spreads to other locations in the body via lymph or blood. It is a gene disorder that occurs in somatic tissues. Hereditary or acquired anomalies in the regulation of the genes responsible for controlling cell reproduction can lead to cancer. Multiple genes in diverse pathways are involved in its initiation, progression, invasion and metastasis. The first section of this chapter provides a general overview of the biology behind cancer, particularly colorectal cancer which is a commonly diagnosed cancer in western countries. The second section presents a brief description of genetic regulation and the last section describes the role microRNAs in gene regulation and their influence in cancer formation.

2.1 Colorectal Cancer

Colorectal cancer, commonly known as bowel cancer, originates from the inner lining of the colon or the rectum called the mucosa. In most cases, colorectal cancer progresses slowly over a period of 10 to 15 years. It may be present without symptoms for several years. The tumor typically begins as a noncancerous polyp on the inner lining of the colon or rectum (see Figure 2.1). This tumor can be benign or malignant. Benign polyps are
Chapter 2

not cancer and are not life threatening. Malignant tumors are cancer. It
invades nearby tissues and spreads to other part of the body. The polyps are
an early warning sign that colorectal cancer may develop. A polyp may or
may not change into cancer. The chance of the polyp turning cancer depends
upon the kind of polyp. For example, a type of polyp known as an adenoma
can become cancer. If polyps are not removed surgically, they can become
malignant over time. Thus screening and removing polyps from large
intestine reduces the risk of developing of colorectal cancer.

Once cancer forms in the large intestine, it can grow through the
lining and into the wall of the colon or rectum [2]. The cancer cells may
invade and destroy adjacent tissues and may break away from the tumor and
spread via blood or lymph vessels to form new tumor in different locations
of the body. The process through which cancer cells break away from
primary (original) tumor and travel to distant parts of the body through
blood or lymph is called metastasis.

![Colon Cancer and Polyp](http://www.medicinenet.com/colon_cancer)
2.1.1 Stages of Colorectal Cancer

The extent to which a colorectal cancer has spread in the body is described as its stage. Staging is one of the most important factors in determining the choice of treatment and in assessing prognosis. It is also useful in predicting the probability of the cancer recurring after surgical removal. Colorectal cancer develops through five definable stages (0-4):

- **Stage 0 (in situ)** – Abnormal cells are found in the innermost lining of the colon and hasn't moved from where it started.
- **Stage I (Local)** – In this stage, cancer has extended beyond the innermost layer of the colon into the middle layers of the colon.
- **Stage II (Local)** – Cancer has grown beyond the middle layer of the colon, but has not extended through the wall to invade nearby tissue.
- **Stage III (Regional)** – Cancer has spread through outer most layer of the colon wall and has invaded nearby tissue, or has spread to nearby lymph nodes.
- **Stage IV (Distant)** – Cancers has spread through blood or lymph nodes to distant organs, such as the liver or lung.

Staging helps in determining whether the treatment may be helpful in preventing or decreasing the likelihood of a cancer recurrence. Stage I colon cancers have survival rates ranging from 80 to 95 percent. Stage II cancers have a survival rate of 55 - 80 percent. A stage III tumor has about a 40 percent chance of cure and a patient with a stage IV colon cancer has only a 10 percent chance of a cure.
2.1.2 Risk factors for Colorectal cancer

The exact cause of colorectal cancer is still unknown. However, researchers have found that certain risk factors that may increase a person's chance of developing colorectal cancer. The factors that increase the risk factor of colorectal cancer includes

- **Age**- The risk of developing colorectal cancer increases with age. Although this disease can affect people of all ages, most people who develop colorectal cancer are over age 50.

- **Personal History**- A person who has treated for colorectal cancer has an increased risk for developing colorectal cancer in future.

- **Family History**- A person, whose one or more close relatives (parents, siblings, or children) has had colorectal cancer, is at a risk for developing colorectal cancer.

- **Diet**- A diet that is high in red and processed meat and low in fresh vegetables and fruits increases the risk of colorectal cancer.

- **Physical Inactivity**- The people who follows a sedentary lifestyle may have an increased risk of developing colorectal cancer. Regular exercise will reduce the risk of developing colorectal cancer.

- **Smoking**- Tobacco smoking, particularly long-term smoking increases the risk of colorectal cancer.

- **Alcohol**- Alcoholic drink, especially drinking heavily, may be a risk factor.
2.1.3 Cancer Genes

Genetic instability is a hallmark of almost all cancers. It refers to a set of events capable of unscheduled alterations, either in temporary or permanent nature, within the genome. Alterations in three types of genes such as oncogenes, tumor suppressor genes and DNA mismatch-repair genes are responsible for the development of cancer [18].

2.1.3.1 Oncogenes

Oncogenes function as a positive growth regulators and has the potential to cause cancer. Oncogenes are altered forms of normal cellular genes called proto-oncogenes which produce proteins that regulate cell growth and division. When mutated, oncogenes typically produce more proteins, results in the alteration of the pathway of cell growth and proliferation. This may lead to abnormal growth of cell. For example, K-ras gene is an oncogene that is mutated in colon cancer cells.

2.1.3.2 Tumor Suppressor Genes

Tumor suppressor genes or anti-oncogenes function as a negative growth regulator and suppress tumor formation. They regulate cell growth, differentiation and promote cell suicide (apoptosis). When mutated; tumor suppressor genes produce less of their protein. Thus, apoptosis does not occur and abnormal cell growth results. Tumor suppressor genes such as DCC(Deleted in Colon cancer) and p53 are mutated in colorectal cancer.

2.1.3.3 DNA Mismatch-repair Genes

Mismatch-repair genes (MMR) play a central role in maintaining genomic stability by repairing damaged DNA. When these genes are
mutated, repair does not occur and the cell is more prone to become cancerous. Germline mutations in DNA mismatch-repair genes are associated with the inherited cancer syndrome, hereditary non-polyposis colorectal cancer (HNPCC)

2.1.4 Colon Carcinogenesis

Carcinogenesis, also called tumorigenesis, is the molecular process by which cancer develops. There are four distinct sequential mutations described in the development of colon cancer. This includes mutations of the APC (adenomatous polyposis coli), K-ras, DCC (deleted in colon cancer), and p53 genes. Each mutation causes progressive changes in the colonic epithelium. During initiation phase mutation of APC typically occurs and is sometimes inherited. Mutations in APC lead to benign polyp formation. These polyps can remain inactive for several years. When one cell in this polyp develops a second mutation, in the K-ras oncogene, it grows at a faster rate resulting in a larger tumor or intermediate adenoma. Mutation of tumor suppressor gene DCC represents the third step in genetic pathway. Loss of DCC plays a role in tumor progression, invasion and metastasis. Mutations of p53 lead to late adenoma and finally carcinoma.

2.1.5 Treatment

Treatment options of colorectal cancer depend on the stage of the tumor as well as the general state of the patient like age, medical history, overall health etc. In general, treatments include:

1. Surgery – The tumor and the nearby tissues in the diseased area are removed. In addition to removal of the primary tumor, surgery is
often necessary for estimating the penetration of disease and whether it has metastasized

2. Chemotherapy - Chemotherapy is the treatment of cancer with drugs that can kill cancer cells and thus decrease the chance of the tumor reoccurring elsewhere in the body. It targets all rapidly dividing cells and is not specific to cancer cells. Therefore chemotherapy may harm healthy tissues, especially those that have a high replacement rate.

3. Radiation therapy - High-energy radiation is used to destroy cancer tissue. Radiation destroys any remaining cancer cells after surgery and reduces the chance of cancer spread or recurrence. Although radiation is occasionally used for the treatment for colorectal cancer, in some cases radiation is used in conjunction with chemotherapy treatments to gain better results.

Cancer treatment aims at the complete removal of the cancer without damage to the rest of the body. To some extent, this can be accomplished by surgery, but invasion and spread of disease to distant locations of the body limits its effectiveness. Since chemotherapy is not specific to cancer cells, it is sometimes toxic to healthy tissues. Radiation also damage normal cells and tissues. Therefore, development of novel target specific therapeutics must be necessary for the effective treatment of cancer.

2.2 Biological Aspects of Gene Regulation

Life sciences began with Robert Hooke; who in 1665 discovered cells which are the basic unit of life for all living organisms. There are
different types of cells in our body like brain cells, liver cells, skin cells etc. All these cells have unique characteristics and functions. The nucleus of the cell stores the hereditary material, the genes, in the form of long and thin DNA (deoxyribonucleic acid) strands. The genome of an organism contains necessary information to control of all cellular process like replication of DNA, protein synthesis etc. According to central dogma of molecular biology [19], DNA, RNA and proteins are the three macromolecules essential for all known forms of life. DNA is the carrier of genetic information used in the development and functioning of all organisms. This genetic information is used to encode protein molecules. Three different processes are responsible for the inheritance of genetic information and its conversion from one form to another (see Figure 2.2):

1. Replication: Before a cell divides, its DNA is replicated to give identical copies. It is the basis for biological inheritance. DNA replication is said to be semi conservative since one strand serves as a template for the second strand.

2. Transcription: The process of making single stranded ribonucleic acid (RNA) from DNA template is called transcription. During transcription, a DNA sequence is read by an enzyme called RNA polymerase (RNA pol), which produces a complementary, antiparallel RNA strand. Several types of RNAs are synthesized in the nucleus during transcription. Of particular interest are
   
   - messengerRNA(mRNA)- later used for protein synthesis
• ribosomal RNA (rRNA)-major component of building ribosome, the protein making machinery

• transfer RNA (tRNA)-The molecules that carry amino acids to the growing peptide chain

• micro RNA (miRNA)-tiny RNA molecules that regulate the expression of mRNA

3. Translation: Translation is a process where ribosomes synthesize proteins from the information contained in the mRNA. During translation, the ribosome reads a string of three bases on the RNA (codon) and translates them into one amino acid.

Proteins are further processed in subcellular compartments and transported in-and-out of the cell to carry out different metabolic functions. These highly coordinated activities empower cells to respond to the varying environment with both speed and precision.

Gene expression is a process by which information encoded in a gene is used for the production of gene products such as RNA or proteins. It covers the entire process from transcription through protein synthesis. If the protein is synthesized, a gene is said to be “expressed” and the expression level of gene depends on the amount of mRNA it produced. Different cell types in an organism carry out a range of specialized function depends upon the genes that are expressed only in that cell type. The factors that affect gene expression are the type of tissue, the age of the person, the presence of specific chemical signals etc.
Chapter 2

2.2 Central Dogma of Molecular biology

Genes transcribed in the nucleus are translated into proteins in the cytoplasm. The figure is taken from http://www.accessexcellence.org/

2.2.1 Gene Regulation

Gene regulation refers to the collection of processes that control the amount and timing of appearance of the functional gene product. It is the
basis for diverse biological process including cell growth and development as well as cellular differentiation, versatility and adaptability of any organism. Gene expression is controlled at three possible levels in the production of an active gene product. First and most important mode for regulating eukaryotic gene expression is the transcriptional regulation. Regulation of transcription controls when the gene is transcribed and how much it is transcribed. Different factors that influence transcription regulation are the strength of promoter elements within the DNA sequences of a given gene, the presence or absence of enhancer sequences (which enhance the activity of RNA polymerase and increase transcription), and the interaction among multiple activator proteins and inhibitor proteins. Second is the translational regulation, controls the amount of proteins synthesized from mRNA. Third, post-transcriptional or post-translational regulation mechanisms control the level of active gene products. Active mRNA level can be controlled by addition of poly (A) tail, splicing, silencing by noncoding RNAs (miRNA, siRNA) etc. Some proteins may also undergo modifications such as folding, enzymatic cleavage, bond formation etc. These modifications can play a vital role in the regulation and control of gene expression.

### 2.2.2 Gene Regulatory Network

The interactions among genes, proteins and other cellular components form complex circuits that control all biological functions in a living organism. One type of such circuit is gene regulatory network which
represents the interaction structure of genes. A Gene regulatory network (GRN) models the complex regulatory mechanisms that control the activity of genes in living things and provides the most realistic representation of gene regulation. GRN models can be categorized into two classes, detailed and abstract model, according to the level of complexity in the model. In the detailed GRN model, the true physical interactions between regulatory proteins and their promoters are represented. In such models, regulator nodes are either transcriptional regulator proteins or genes and the target nodes are the mRNA levels for the target genes. The figure 2.3 shows the schematic illustration of a detailed GRN model. For instance, gene1 inhibits gene2 and activates gene3, implies that mRNA1 transcribed from gene 1 is translated to protein1 which in turn inhibits gene 2 and activates gene 3. In abstract GRN model such detailed functional descriptions are not represented explicitly. The abstract GRN model is depicted in figure 2.4. In abstract model; genes are represented as nodes and the regulatory relationships as directed edges. The regulatory relationship can be either an activation (increasing the transcription of other genes) or a repression (inhibiting the transcription level). The absence of link between two nodes implies that there is no relationship between two nodes. The regulation between two genes in a GRN implies direct physical interactions as well as indirect regulations via proteins, metabolites and noncoding RNAs that have not been measured directly [8]. This work focuses on inferring abstract GRN models from high throughput microarray data.
Figure 2.3 An example of a detailed GRN model. Genes can either activate or inhibit themselves or other genes (gene1 inhibit gene2 and activates itself). Often proteins form complex and regulates other genes.

Figure 2.4 Abstract model of the GRN depicted in figure2.3 is shown. An edge → indicates activation of transcription, whereas an edge –l indicates repression of transcription.

2.3 Role of MicroRNAs in Gene Regulation

MicroRNAs are a class of non-coding RNAs that hybridize to mRNAs and regulate their activities at post transcriptional as well as
translational level [20]. There are at least 800 miRNAs within the human genome, which may target about 60% of mammalian genes [21, 22]. MicroRNAs bind to partially complementary sites in the messenger RNA of other genes and inhibit the translation of these genes. They have been found to regulate a wide range of biological process such as cell differentiation, proliferation, growth, mobility and apoptosis in diverse cancer-related biological processes [23, 24].

MicroRNAs were discovered in 1993 by Rosalind Lee, Rhonda Feinbaum and Victor Ambros during a study of the gene lin-14 in C. elegans development [25]. Since then, over 4000 miRNAs have been identified in almost all metazeon genomes including mammals, flies, worms and plants. In the human genome as many as 700 miRNAs have been identified yet and over 800 more are predicted to exist. The impact of microRNA on the proteome suggests that the microRNA acts as a rheostat, making fine-scale adjustments to protein synthesis from thousands of genes [26, 27].

2.3.1 Biogenesis of miRNAs

MicroRNA biogenesis is a stepwise process that starts in the nucleus and ends in the cytoplasm (see Figure 2.5). Most miRNAs are located in the introns of protein and non-protein coding genes or even in exons of long non-protein coding transcripts [28]. MiRNA genes are usually transcribed by RNA polymerase II (Pol II) in the nucleus [29]. The miRNA sequence and its reverse-complement base pair to form a double stranded RNA hairpin loop called pri-miRNA (primary miRNA structure). The nuclear enzyme Drosha and its cofactor DGCR8/Pasha cleave the base of the hairpin to form pre-miRNA of about 70 nucleotides in length. The pre-miRNA hairpins are transported from the nucleus into the cytoplasm by Exportin 5, a carrier protein.
In cytoplasm, RNase III enzyme Dicer cuts 20-25 nucleotides from the base of the hairpin yielding an imperfect miRNA:miRNA* duplex [30]. The functional strand of the microRNA duplex is then loaded into Argonaute protein within RNA-induced silencing complex (RISC) and becomes mature miRNA, whereas the other strand, miRNA*, is degraded [31, 32]. Finally, the mature miRNA load in RISC is potent for regulating protein production, either by translational repression or mRNA cleavage.

**Figure 2.5 Pathway from microRNA biogenesis to mRNA regulation.** The microRNA gene is transcribed by RNA polymerase II into a double stranded RNA hairpin loop called the primary transcript or pri-microRNA. The nuclear enzyme Drosha cleaves the flanking sequences, resulting in the ~70 nucleotide long pre-microRNA. After the relocation of pre-miRNA into the cytoplasm by exportin-5, Dicer, RNase III enzyme, performs the second cleaving step called ‘dicing’ to produce the microRNA:miRNA* duplex. Subsequently the duplex is separated and one strand gets incorporated into the RISC, while the other strand is degraded. Finally the microRNA loaded RISC is potent for regulating protein production, either by translational repression or mRNA cleavage. The image is taken from http://dna-rna.net
2.3.2 MicroRNA and Cancer

MicroRNAs have diverse expression pattern and might play a key role in various developmental and physiological processes like cell development, proliferation, mobility, differentiation and apoptosis [23, 24]. Accordingly, altered miRNA expression is likely to contribute to a wide range of human diseases, including cancer. The findings that miRNAs have a role in cancer are supported by the fact that many miRNA genes are located at fragile sites in the genome or regions that are commonly amplified or deleted in human cancer [33]. Also, malignant tumors and tumor cell lines contain widespread deregulated miRNA expression compared to the corresponding normal tissues [34,35].

First evidence of involvement of miRNAs in cancer was reported in 1999 [36]. Calin et. al identified that two miRNAs, mir-15 and mir-16, were involved in the pathogenesis of chronic Lymphocytic Leukemia. Later, in 2005, He et. al. [37] demonstrated that miRNAs from mir-17-19 cluster were over expressed in lymphoma cell lines. In the same year, Johnson et. al. [38] experimentally confirmed that loss or reduction of let-7 in lung cancer lead to the over expression of RAS oncogene which in turn results in the increased cell growth and tumorigenesis. The authors suggested let-7 act as tumor suppressor. Recent experiments also show that miRNAs upregulate genes in one condition, but act as a negative regulator in another condition. For example, let7 and the synthetic microRNA miRcxcr4-likewise upreguate target mRNAs upon cell-cycle arrest; yet, they inhibit translation in proliferating cells [39].
In general, changes in the expression pattern of miRNAs can influence carcinogenesis if their mRNA targets are encoded by oncogenes or tumor suppressor genes [17]. Recent functional studies suggest that miRNAs regulate many known oncogenic and tumor suppressor pathways involved in the pathogenesis of Colorectal Cancer [40, 41]. MiRNAs regulate many proteins involved in key signaling pathways of CRC, such as members of the Wnt/β-cat enin pathway, EGFR signaling (KRAS and phosphatidylinositol-3-kinase (PI-3-K) pathways) and p53 pathways [17]. Thus the analysis of such miRNAs is useful for cancer diagnosis, prognosis, treatment and drug target discovery.