INTRODUCTION

Gastrointestinal cancer is among medical and economic burden worldwide ***(Parkin et al., 2001).*** Esophageal and gastric cancers are most common in the non industrialized countries while colorectal cancer is the predominant gastrointestinal malignancy in westernized countries ***(Gerson and Triadafilopoulos, 2002).***

Aetiology of GIT malignancies is combination of both environmental and genetic factors that can be corrected or prevented ***(Parkin et al., 2001).***

Risk factors for GIT malignancies include e.g. tobacco chewing, alcohol consumption, Barrett's esophagus, and GERD in cancer esophagus ***(Wo et al., 2001).*** Helico-bacter pylori infection in case of cancer stomach ***(Hohenberger and Gretsch, 2003).*** Genetic susceptibility and obesity in colorectal cancer ***(Rozen et al., 2002b).***

Prevention of GIT malignancies requires control of diet, alcohol consumption, H pylori eradication ***(Parkin et al., 2001).*** Chemoprevention by (NSAID), and COX-2 inhibitors ***(Rozen et al., 2002b)***. The newer nitric oxide releasing aspirin may have the beneficial effect of aspirin but without its toxicity ***(Fiorucci et al., 2003; Imperial, 2003).***

Screening can be done by tumor markers, self advancing colonoscope, confocal endomicroscopy which offer real time

optical sectioning giving virtual histology with high negative predictive value for abnormalities, capsule endoscopy is established as useful in small bowel polyposis syndrome, now completed therapeutically by double balloon enteroscopy ***(Southey et al., 2005).*** Endoscopic follow up of Barrett's esophagus every 2-3 years with multiple biopsies in case of cancer esophagus ***(Gerson and Triadafilopoulos, 2002).***

In case of cancer stomach, screening can be done by endoscopy followed by endoscopic mucosectomy; this is useful when performed on an asymptomatic population ***(Soetikno et al., 2003; Hohenberger and Gretschel, 2003).*** Colorectal cancer screening of asymptomatic individuals identifies adenomatous polyps and their endoscopic removal prevents colorectal cancer ***(Rozen et al., 2002b).***

AIM OF THE WORK

The aim of this work is to show the molecular biological aspects of malignancy affecting the gastrointestinal tract, to identify different methods of screening for such malignancies, and to identify high risk groups.

The malignancies that will be included in this essay are cancer esophagus, cancer stomach, cancer pancreas, colorectal carcinoma, gastrointestinal stromal tumors (GIST), and gastrointestinal lymphoma.

Clinical picture and Management will be discussed briefly.

OVERVIEW ON MOLECULAR BIOLOGY

Every human began life as a single fertilized egg. This single cell contained all the necessary information to direct development of the various organs and tissues of the body, including germ cells. Understanding the basis of tissue diversity is an ongoing theme of biomedical research. Although many details of this process are unclear, the basic scheme of how tissue-specific functions are established is known. The basics of molecular biology includes the relationship among DNA, RNA, and proteins, as well as the language used in describing molecular and cellular processes ***(Plasterk, 2002).***

Cellular Diversity and Genomic Stability

In general, all cells of an individual possess exactly the same genetic information; this is largely contained in nuclear DNA that is located within 46 discrete chromosomes (22 pairs of autosomes and 1 pair of sex chromosomes) ***(Sapolsky et al., 1999).***

In contrast to the general constancy of DNA across tissues, the structure and function of different tissues are highly variable. As an example, cardiac muscle clearly differs from skin or liver. These differences in function are achieved by selective activation of genes in each cell type. The mechanisms by which this is accomplished are understood at the general

level, but differ in their details in various tissues ***(Freedman et al., 2004).***

Some exceptions to the constancy of genetic information exist. As examples: Immunoglobulin and T cell receptor gene rearrangements occur in normal B and T cells, respectively. The degradation of genetic information, including mutations, chromosomal duplication or loss, and/or rearrangements, is a hallmark of neoplastic disease ***(Marchini et al., 2004).***

Principles of Molecular Biology

The discovery of the structure of DNA, RNA, and proteins, and of the genetic code provides the conceptual framework by which genetic stability and functional diversity is currently understood ***(Pritchard et al., 1999).***

DNA is the information-storing molecule; each molecule is duplicated during the process of replication that accompanies each cell generation. The process by which this information is transferred into cellular function begins when the genetic information residing in DNA is transcribed into messenger RNA (mRNA) ***(Kalajzic et al., 2002).***

Messenger RNA is then used to direct the translation of genetic information to physiologically active proteins. This is performed by utilizing the information contained within the mRNA sequence to construct a unique polypeptide, which is defined by a linear chain of amino acids. Via the use of

molecular machinery, the specific amino acids are placed within the polypeptide as directed by a template defined by unique triplets of bases (codons) found within the mRNA molecule. This process establishes the correspondence between DNA encoded information and its expression in protein via the universal genetic code ***(Southan, 2004).***

Not all RNA molecules function as mRNA. Some act as components of ribosomes, others are involved in RNA splicing, and still others serve as transfer RNA. Finally, some double stranded RNAs direct targeted degradation of homologous mRNA molecules, inhibiting their translation into proteins. ***(Southan, 2004).***

Structure of DNA and Template-Directed Nucleic Acid Synthesis

DNA is normally present as an antiparallel polymeric double helix composed of four nucleotide subunits. The nucleotide subunits consist of the following bases: Adenine (A) Guanine (G) Thymine (T) Cytosine (C). The two strands of the double helix are held together by specific hydrogen bonds that form between A and T (2 hydrogen bonds) or between G and C (3 bonds) ***(Reich et al., 2001).***

A and G, the larger bases, are purines, while T and C, the smaller bases, are pyrimidines. Double stranded DNA contains equimolar amounts of purine and pyrimidine. In addition, A and T are present in equimolar amounts, as are G and C. The backbone of the DNA molecule is an alternating copolymer of

deoxyribose, a 5 carbon sugar and phosphate groups, linked by phosphodiester bonds to the 5' and 3' carbons of each deoxyribose unit ***(Kwok et al., 2003).***

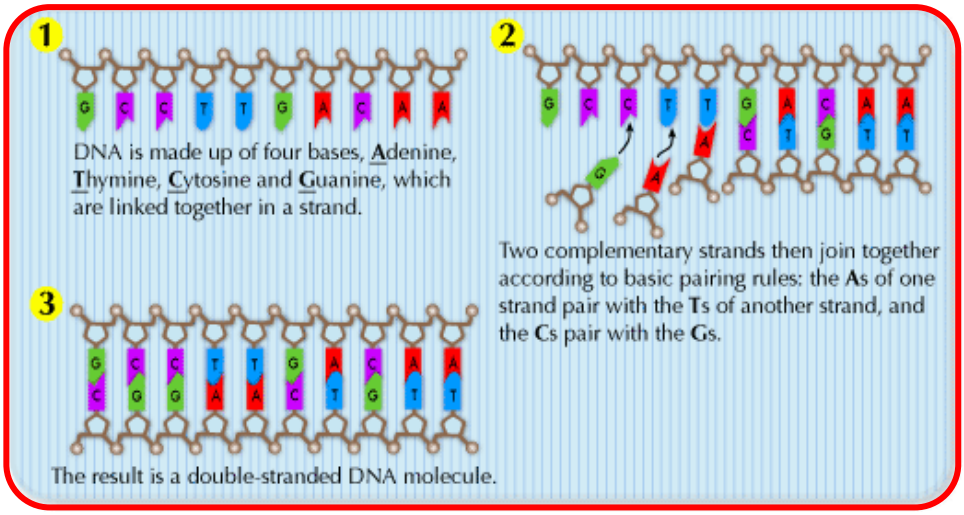

Figure (1): DNA structure ***(Kwok et al., 2003).***

The hydrogen bonding between the complementary base pairs A and T or G and C provides the chemical basis for DNA's function as the storage medium for genetic information. The genetic information is encoded as the sequence of bases along a DNA strand and is read from the 5' to 3' direction ***(Reich et al., 2001).***

The weakness of hydrogen bonds, which each possess a strength of approximately 2 kcal/mol, is an important feature with regard to nucleic acid function. This weakness allows denaturation, or separation of the DNA strands, to occur at physiologic temperatures. Separation of the complementary DNA strands and synthesis of new DNA strands by sequences

directed by the templates of the original strands allows for accurate copying of sequence information ***(Reich et al., 2001).***

In general, DNA replication is semi-conservative, in that each daughter molecule contains one old and one newly synthesized strand ***(Gut, 2004).***

During the S phase of each cell cycle, DNA is replicated by DNA polymerases to provide each daughter cell with a complete genome. The genome is the total genetic complement of an organism ***(Gut, 2004).***

There are several important structural differences between DNA and RNA. In general, DNA is double-stranded and RNA is single stranded. In DNA, the sugar is deoxyribose, while it is ribose in RNA. In DNA, thymine is the pyrimidine complementary to adenine, but uracil replaces thymine in RNA ***(Gut, 2004).***

Transcription

Template-directed synthesis, in which one strand of DNA provides sequence information, is used in both DNA replication and transcription of DNA to form mRNA. However, some DNA sequences do not encode protein: Some DNA sequence elements provide control information; they specify the location of an active gene or allow the binding of transcription factors that modulate the rate at which a gene is transcribed. Coding regions of a gene's DNA sequence are characteristically

interrupted by introns or noncoding intervening sequences that are spliced (or removed) out of mature mRNA. Intron boundaries are marked by splice donor and splice acceptor sites, which provide sequence recognition sites for the spliceosomes; spliceosomes are the enzymatic ribonucleoprotein complex that removes introns from the primary transcript to produce mature mRNA. The 3' end of an mRNA molecule is a tail of adenines that are not present in the DNA, but are added when the transcriptional machinery recognizes a polyadenylation site ***(Dearlove, 2002).***

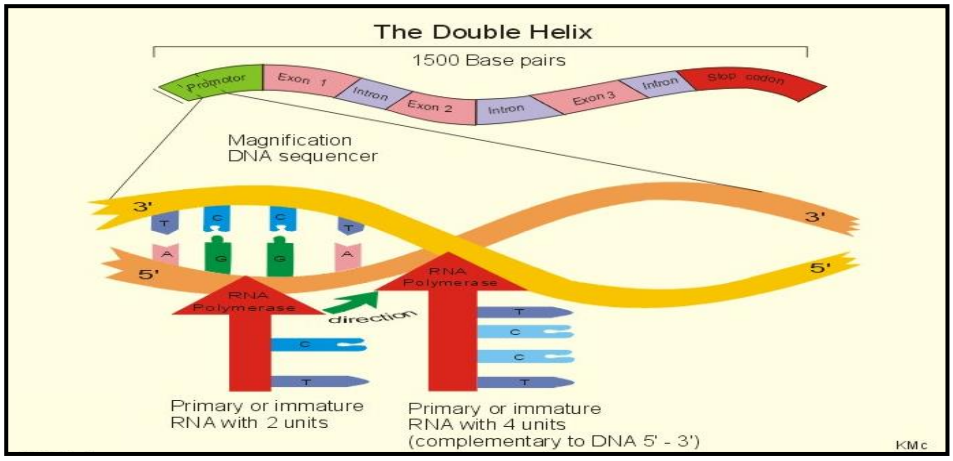

Figure (2): DNA transcription ***(Dearlove, 2002).***

The transcription initiation complex and various transcription factors recognize sequence signals present in the DNA to identify the presence of an active gene. Local denaturation of the DNA allows RNA polymerase to synthesize an mRNA molecule using the coding strand of the DNA as a template. The primary transcript synthesized in this step is

spliced and polyadenylated to yield a mature mRNA molecule ***(Dearlove, 2002).***

There are three major RNA polymerases present in mammalian cells: The outline of transcription given above is applicable to RNA polymerase 2, the polymerase that is responsible for the expression of most genes. RNA polymerase 1 functions primarily to transcribe ribosomal RNA RNA polymerase 3 functions primarily to transcribe a variety of small RNAs, such as tRNAs (transfer RNAs) and the RNA components of the spliceosomes ***(Li, 2004).***

Genetic Code and Translation

Mature mRNA leaves the nucleus and reaches the ribosomes, where its sequence is recognized and used to direct the synthesis of a polypeptide chain. The ribosomes are complex ribonucleoprotein structures that include the enzymatic machinery for protein synthesis. Protein synthesis is template directed, with the mRNA's sequence information being used to specify the protein's amino acid sequence ***(Griffiths et al., 2004).***

There is a complication concerning the relationship between mRNA and proteins: RNA contains 4 bases while proteins may contain up to 20 amino acids (if amino acid modification is excluded). To overcome this numerical difference, the genetic code establishes a correspondence between specific triplets of bases (codons) and specific amino

acids. However, since there are 64 ways to combine three bases, the code is redundant: some amino acids are encoded by more than one codon. In addition, there are three codons (UAG, UGA, UAA) that do not encode amino acids; instead, they specify the end of a polypeptide chain ***(Griffiths et al., 2004).***

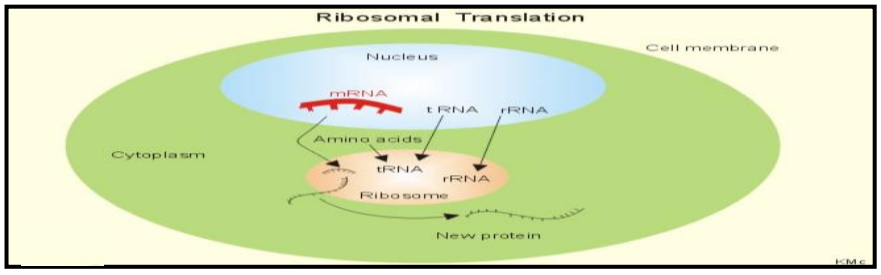

Figure (3): DNA translation ***(Griffiths et al., 2004).***

By convention, the codons are given as the sequence of mRNA, not the sequence of the complementary DNA strand. The nucleic acid sequence is given 5' to 3' and the protein sequence is given N-terminal- to C-terminal. These directions correspond to the direction of synthesis ***(Hartl et al., 2004).***

To synthesize proteins, the following processes must occur sequentially. Beginning with the start codon, each codon is held at the synthetic site in the ribosome; at this site, an amino-acid-charged transfer RNA (tRNA) molecule containing a complementary anticodon base pairs with the mRNA. The carried amino acid is subsequently added to the nascent polypeptide chain. Once the amino acid is added, the ribosome moves processively codon by codon along the mRNA strand, adding an additional amino acid to the polypeptide at each step as dictated by the unique codon. When the ribosome encounters a chain

termination, or nonsense, codon it releases both the mRNA and the newly synthesized protein ***(Hartl et al., 2004).***

The processive movement of the ribosome along the mRNA molecule allows multiple ribosomes to simultaneously synthesize multiple copies of a protein from a single mRNA molecule. This is observed microscopically by the presence of polyribosomes or polysomes, which is a tight spatial array of ribosomes translating a single mRNA molecule ***(EBI, 2004).***

Implications to Medicine

The importance of specific base pairing for the development of a normal organism and/or the maintenance of health cannot be overstated. The mechanisms of template-directed replication and transcription allow preservation of genetic information and its use to encode functional proteins ***(Nadeau et al., 2000).***

Errors in these processes as well as intrinsic properties of this molecular machinery have direct implications for the practice of medicine. As examples: Errors in replication account for mutations that cause a wide array of diseases, including inherited disorders and malignancies. Divergence between humans and bacteria in the enzymatic machinery that carry out the functions of transcription and translation provides the molecular targets for an array of antibiotics that are lethal to bacteria but harmless to humans ***(Risch, 2000).***

Cancer Aetiology: Inheritance or Environment?

It is clear that a mutation causing the dysfunction of a single oncogene or tumour suppressor gene is not sufficient to induce unregulated clonal expansion. The Knudson multi-hit hypothesis elegantly unites the genetics of familial and sporadic tumour development: an inherited mutation in one gene allele may be insufficient to cause a tumour but will cause a significant susceptibility to the development of a particular cancer. Subsequent lifetime exposure to environmental carcinogens (viral, chemical, radiation), along with simple mistakes during cell division, may deregulate the normal allele. Other mutations which accumulate in a similar manner then lead to tumour development. Research has clearly shown that germline mutations in particular genes, such as *p53* and *RB,* have a much stronger influence on the chance of subsequent tumour development than others ***(Carmeliet, 2004).***

Molecular Biological Approach to GIT Cancers

Introduction

Colorectal cancer represents an ideal model for the study of molecular pathogenesis of cancer due to the accessibility of tissue for biopsy, and the clear progression from normal colonic epithelium to invasive cancer via an intermediate precursor, the adenomatous polyp. A multistep process of specific genetic changes is thought to drive the transformation from normal colonic epithelium to an invasive cancer. Single, specific germline mutations underlie the common inherited syndromes

[eg,Adenomatous polyposis coli (APC), Hereditary nonpolyposis colorectal cancer (HNPCC)], while sporadic cancers result from the stepwise accumulation of multiple somatic mutations. Mutations in the APC gene occur early, while others, such as mutations of the p53 suppressor gene, generally occur late in the process ***(Jemal et al., 2007).***

Microsatellite instability

Microsatellites are short (50-300 bp) sequences composed of randomly repeated segments of DNA two to five nucleotides in length (dinucleotide/ trinucleotide/ tetranucleotide repeats). Scattered throughout the genome in the non-coding regions between genes or within genes (introns), many of these microsatellites are highly polymorphic ***(Carraro et al., 2001).*** Often used as markers for linkage analysis because of high variability in repeat number between individuals, these regions are inherently unstable and susceptible to mutations. Somatic microsatellite instability (MSI) has been detected in a number of tumours. Detecting MSI involves comparing the length of microsatellite alleles amplified from tumour DNA with the corresponding allele in normal tissue from the same individual ***(Jemal et al., 2007).***

Recent studies indicate that MSI can be detected in approximately 90% of tumours from individuals with hereditary non-polyposis colorectal cancer (HNPCC). The presence of these additional microsatellite alleles (repeated segments) in tumour cells, results from the inherent susceptibility of these

areas to such alterations and from mutations in the DNA mismatch repair mechanism that would normally correct these errors ***(Chen et al., 2006).***

The data suggest that most colorectal neoplasias begin with inactivation (through a germline or sporadic mutation) of the APC gene, whether or not MSI is present. However, the genetic events may thereafter diverge, depending upon the mechanism underlying genetic instability (ie, chromosomal instability or MSI). Compared to patients with chromosomal instability [which includes those with familial adenomatous polyposis [FAP] and the majority of sporadic colorectal cancer (CRCs)], the MSI or "mutator" pathway involves a unique destabilizing mechanism and inactivation of a different set of genes. Regardless of the molecular pathway involved, the final result is the same ***(Popat et al., 2005).***

In addition to providing insight as to the biology of all CRCs, the identification of specific genetic mutations responsible for colorectal tumorigenesis has had a direct influence on clinical care. Patients at highest risk for developing CRC can be identified via genetic testing for specific germline mutations, and new molecular screening methods for early detection of CRC by detection of mutations in fecal material are under study. In addition, these mutations are also being examined as prognostic markers, and as potential therapeutic targets ***(Ricciardiello et al., 2003).***