UNIT 2  GENETIC EPIDEMIOLOGY AND EPIGENETICS

Contents
2.1  Introduction
2.2  Scope of Genetic Epidemiology
2.3  Methods of Genetic Epidemiology
2.4  Epigenetics
   2.4.1  Epigenetic Mechanisms
   2.4.2  Epigenetics and Inheritance
2.5  DNA Methylation
2.6  The role of DNA Methylation in Mammalian Epigenetics
2.7  The Hap Map Project
   2.7.1  Method of Study
   2.7.2  International Hap Map Project
   2.7.3  Genetic Variation and Use of Hap Map
2.8  Summary

Suggested Reading
Sample Questions

Learning Objectives

After studying this unit, you would be able to:

- explain the concept of Genetic Epidemiology and its scope in areas of medicine, genetics etc;
- discuss the meaning of Epigenetics and the factors responsible for the occurrence of epigenetics;
- elucidate the importance of epigenetics in the understanding of DNA Methylation or noncoding RNA;
- explain the methods employed for the study of Genetic Epidemiology; and
- to understand Hap Map project, the new advancement in Human Genetics.

2.1  INTRODUCTION

The Mendel’s Laws of Heredity discovery led to emergence of new biological field of Genetics which is the study of heredity and variation i.e. how certain characteristics or traits are passed from parents to offspring. His discoveries paved the way for today’s understanding of the role of genetics in human development and in the treatment of genetic disorders. Presently, genetic research has moved from Mendelian genetics to sequence maps for the study of natural human genetic variation at the level of the genome. The inheritance in contemporary populations, uniting genetics with epidemiology, medicine, psychology, and (in other species) with agriculture and with this concept the field of genetic epidemiology emanated.
Genetic Epidemiology is the study of the role of genetic factors in determining health and disease in families and in populations, and the interplay of such genetic factors with environmental factors. Genetic epidemiology was defined by Morton as “a science which deals with the etiology, distribution, and control of disease in groups of relatives and with inherited causes of disease in populations. It is of two types. Classical epidemiology and Molecular Epidemiology. The former deals with disease patterns and factors associated with causation of diseases with the ultimate aim of preventing the disease and the latter measures exposure to specific substances (DNA adducts) and early biological response (somatic mutations), evaluates host characteristics (genotype and phenotype) mediating response to external agents, and use of markers of a specific effect (like gene expression) to refine disease categories (such as heterogeneity, etiology and prognosis). The study methods of Genetic Epidemiology are mentioned below:

- **Familial aggregation studies**: Is there a genetic component to the disease, and what are the relative contributions of genes and environment?
- **Segregation studies**: What is the pattern of inheritance of the disease (e.g. dominant or recessive)?
- **Linkage studies**: On which part of which chromosome is the disease gene located?
- **Association studies**: Which allele of which gene is associated with the disease?

### 2.2 SCOPE OF GENETIC EPIDEMIOLOGY

The goals of genetic epidemiology contrast with those of “traditional” epidemiology and population genetics. “Traditional” epidemiology studies the relationship between the environment and the incidence of a given disease, although it recognizes the significance of the host and his or her genetic makeup. Population genetics, on the other hand, seeks to predict the influences of population structure and selection and mutation on bodily phenotypes and diseases. Finally, genetic epidemiology studies the way environmental risk factors interact with the genetic makeup of a given population.

Further the scope of genetic epidemiology has expanded to include common diseases for which many genes each make a contribution (polygenic, multifactorial or multigenic disorders). Human Genome Project has led to study the association between diseases and genotypes. Advancement in the technology made it feasible to conduct large-scale genome-wide association studies. Genetic epidemiology is relatively new discipline that seeks to unravel the role of genetic factors and their interactions with environmental factors in the etiology of diseases, using population/family study approaches.

### 2.3 METHODS OF GENETIC EPIDEMIOLOGY

Genetic epidemiology uses two types of research strategies: descriptive and analytic. The descriptive strategy, at the population as well as at the family level, is based on the study of time, location, and the individual.
**Family Recurrence Studies**

A fundamental aspect of genetic epidemiology is the study of aggregation (or recurrence) of certain diseases in given families. Is this familial aggregation associated with common environmental exposure, hereditary susceptibility, or cultural inheritance of risk factors? If there is genetic susceptibility, how is it inherited? The existence of familial aggregation can be determined by observing the prevalence of a given disease in family members of the index case (the index case is the affected individual who introduces the family into the study) and of controls (individuals who are not affected). This method is efficient and inexpensive, but one of its limitations is that information about characteristics of family members and controls may give rise to biased.

**Twin studies**

Twin studies have typically been used to determine whether genetic factors play a role in the etiology of certain diseases. Such studies consist of comparing the difference in concordance between identical or monozygotic twins (MZ) and fraternal or dizygotic twins (DZ). MZ twins share 100% of their genetic material, whereas DZ twins share, on average, 50% of their genes. If sets of twins are being studied, and the MZ twins are found to be concordant (both have the same disease, for example) with greater frequency than the DZ twins, it is possible to conclude that genetic factors are at least partially involved in the etiology of that disease. It is important to note, however, that genetic differences may exist between MZ twins. They may differ, for example, in the series of T-cell antibodies and receptors, in the number of mitochondrial deoxyribonucleic acid (DNA) molecules, in somatic mutations in general, and in the inactivation pattern of the X chromosome in female twins. It is also well known that MZ twins may differ from DZ twins as a result of environmental factors. One of two calculations is normally made in twin studies, based on the method used to select the twins: (1) pair concordance rate, which describes the proportion of twin pairs where both siblings are affected; and (2) index case concordance rate, which is the proportion of affected individuals among the co-twins of those selected as index cases. Although the pair concordance rate is the simplest method of determining whether genes affect a specific phenotype, it does not measure the magnitude of such an effect. For that purpose, use of the index case concordance rate is preferable.

Twin studies are limited by several factors; in particular those associated with the way participants are selected for the studies. For example, it has been observed that studies that depend exclusively on volunteers have a greater proportion of MZ twins, female pairs, and participants who are concordant for the phenotype under study. Such differences may influence the concordance rate that is calculated, which is why several countries—Sweden is a prime example—have launched population-based twin registries. Another limitation, especially in behavior studies, is that MZ twins tend to share environmental factors more frequently than DZ twins.

**Allelic Association Studies**

1) The allele in question is actually the cause of the phenotype.

2) The allele does not cause the phenotype but is in linkage disequilibrium with the causal allele. Linkage disequilibrium takes place when the causal allele of the phenotype is physically close (or linked) to the allele being studied.
The steps, a genetic epidemiologic research follows, are:

1) Establishing that there is a genetic component of the disorder.

2) Establishing the relative size of that genetic effect in relation to other sources of variation in disease risk (environmental effects such as intrauterine environment, physical and chemical effects as well as behavioral and social aspects).

3) Identifying the gene(s) responsible for the genetic component.

All of these can be achieved either in family studies (segregation, linkage, association) or in population studies (association).

**General methods employed in genetic epidemiology are:**

**Genetic risk studies:** What is the contribution of genetics as opposed to environment to the trait?

**Segregation analyses:** What does the genetic component look like (oligogenic ‘few genes each with a moderate effect’, polygenic ‘many genes each with a small effect’, etc.)? What is the model of transmission of the genetic trait? Segregation analysis requires multi generation family trees preferably with more than one affected member.

**Linkage studies:** What is the location of the disease gene(s)? Linkage studies screen the whole genome and use parametric or nonparametric methods such as allele sharing methods (affected sibling-pairs method) with no assumptions on the mode of inheritance, penetrance or disease allele frequency (the parameters). The underlying principle of linkage studies is the co segregation of two genes (one of which is the disease locus).

**Association studies:** What is the allele associated with the disease susceptibility? The principle is the coexistence of the same marker on the same chromosome in affected individuals (due to linkage disequilibrium). Association studies focus on population frequencies, whereas linkage studies focus on concordant inheritance. Association studies have several practical advantages over linkage studies.

### 2.4 EPIGENETICS

The cells in a multicellular organism have normally identical DNA sequences (and therefore the same genetic instruction sets), yet maintain different terminal phenotypes. This nongenetic cellular memory, which records developmental and environmental cues (and alternative cell states in unicellular organisms), is the basis of epi-(above)–genetics. The lack of identified genetic determinants that fully explain the heritability of complex traits, and the inability to pinpoint causative genetic effects in some complex diseases, suggest possible epigenetic explanations for this missing information. The desire to understand the “deprogramming” of differentiated cells into pluripotent/totipotent states, has led to “epigenetic” becoming shorthand for many regulatory systems involving DNA methylation, histone modification, nucleosome location, or non-coding RNA. Epigenetics was coined by Waddington (1942) to refer to the study of the “causal mechanisms” by which “the genes of the genotype bring about phenotypic effects.”
An epigenetic system is heritable, self-perpetuating, and reversible. Whether histone modifications (and many non-coding RNAs) are epigenetic is debated; it is likely that relatively few of these modifications or RNAs will be self-perpetuating and inherited.

Robin Holliday defined epigenetics as “the study of the mechanisms of temporal and spatial control of gene activity during the development of complex organisms” The development and maintenance of an organism is orchestrated by a set of chemical reactions that switch parts of the genome off and on at strategic times and locations. Epigenetics is the study of these reactions and the factors that influence them.

Epigenetics has different meanings for different scientists. Molecular biologists define epigenetics as “the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence” (Riggs et al. 1996). Functional morphologists Herring (1993), for whom epigenetics refers to “the entire series of interactions among cells and cell products which leads to morphogenesis and differentiation” and further it was stated that “epigenetic influences range from hormones and growth factors to ambient temperature and orientation in a gravitational field.”

2.4.1 Epigenetic Mechanisms

Several epigenetic mechanisms regulate genes i.e. DNA methylation and changes to histone proteins, around which DNA is packed, or involve functional non-coding RNAs. In general, low levels of DNA methylation (hypomethylation) are associated with higher gene activity and high levels of methylation with gene silencing. Repeat associated CpG islands and non regulatory CpG sites generally exist in a methylated state.

![Epigenetics, environment, and development](aje.oxfordjournals.org)

Fig. 2.1: Epigenetics, environment, and development. A. Dynamic epigenetic profile throughout the life course, with both genome-wide remodeling and tissue-specific programming events.

(Source:aje.oxfordjournals.org)
2.4.2 Epigenetics and Inheritance

Epigenetic inheritance is an unconventional finding. It goes against the idea that inheritance happens only through the DNA code that passes from parent to offspring. It means that a parent’s experiences, in the form of epigenetic tags, can be passed down to future generations.

The Challenges of Proving Epigenetic Inheritance

Proving epigenetic inheritance is not always straightforward. To provide a watertight case for epigenetic inheritance following postulates:

- **Rule out the possibility of genetic changes.** In organisms with larger genomes, a single mutation can hide like a needle in a haystack.
- **Show that the epigenetic effect can pass through enough generations to rule out the possibility of direct exposure.**

In a pregnant mother; three generations are directly exposed to the same environmental conditions at the same time. An epigenetic effect that continues into the 4th generation could be inherited and not due to direct exposure.

The Human Epigenome Project

Research into epigenetics has provided new and exciting advances in plant technology, potential cancer treatments and new tools for researchers trying to identify the function of genes. In recognition of the importance of DNA methylation in epigenetics, it is now the subject of the multi-million dollar Human Epigenome Project.

2.5 DNA METHYLATION

In DNA, methylation occurs in the CpG islands, a CG rich region, upstream of the promoter region. The letter “p” here signifies that the C and G are connected by a phosphodiester bond. In humans, DNA methylation is carried out by a group of enzymes called DNA methyltransferases. These enzymes not only determine the DNA methylation patterns during the early development, but are also responsible for copying these patterns to the strands generated from DNA replication. DNA methylation is a biochemical process that is important for normal development in higher organisms. It involves the addition of a methyl group to the 5’ position of the cytosine pyrimidine ring or the number 6 nitrogen of the adenine purine ring and the modification is inherited through cell division.

DNA methylation stably alters the gene expression pattern in cells such that cells can “remember where they have been” or decrease gene expression; for example, cells programmed to be pancreatic islets during embryonic development remain pancreatic islets throughout the life of the organism without continuing signals telling them that they need to remain islets. DNA methylation is typically removed during zygote formation and re-established through successive cell divisions during development. However, the latest research shows that hydroxylation of methyl group occurs rather than complete removal of methyl groups in zygote. Some methylation modifications that regulate gene expression are inheritable and are referred to as epigenetic regulation. DNA methylation is essential for normal development and is associated with a number of key processes including genomic imprinting, X-chromosome inactivation, suppression of
repetitive elements, and carcinogenesis. Alterations of DNA methylation have been recognised as an important component of cancer development. Hypomethylation, in general, arises earlier and is linked to chromosomal instability, loss of imprinting, whereas hypermethylation is associated with promoters that can arise secondary to gene (oncogene suppressor) silencing, but might be a target for epigenetic therapy.

DNA Methylation Analysis Flowchart
Epigenetics

Methylation contributing to epigenetic inheritance can occur through either DNA methylation or protein methylation.

DNA methylation in vertebrates typically occurs at CpG sites (cytosine-phosphate-guanine sites, that is, where a cytosine is directly followed by a guanine in the DNA sequence) and the methylation results in the conversion of the cytosine to 5-methylcytosine. The formation of Me-CpG is catalyzed by the enzyme DNA methyltransferase. Human DNA has about 80%-90% of CpG sites methylated, but there are certain areas, known as CpG islands, that are GC-rich (made up of about 65% CG residues), wherein none are methylated. These are associated with the promoters of 56% of mammalian genes, including all ubiquitously expressed genes. One to two percent of the human genome is CpG clusters, where there is an inverse relationship between CpG methylation and transcriptional activity.

Protein Methylation typically takes place on arginine or lysine amino acid residues in the protein sequence. Arginine can be methylated once (monomethylated arginine) or twice, with either both methyl groups on one terminal nitrogen (asymmetric dimethylated arginine) or one on both nitrogens (symmetric dimethylated arginine) by peptidyl arginine methyltransferases (PRMTs). Lysine can be methylated once, twice or three times by lysine methyltransferases. Protein methylation has been most-studied in the histones. The transfer of methyl groups from S-adenosyl methionine to histones is catalyzed by enzymes known as histone methyltransferases. Histones that are methylated on certain residues can act epigenetically to repress or activate gene expression. Protein methylation is one type of post-translational modification.

2.6 THE ROLE OF DNA METHYLATION IN MAMMALIAN EPIGENETICS

Genes constitute only a small proportion of the total mammalian genome, and the precise control of their expression in the presence of an overwhelming background of noncoding DNA presents a substantial problem for their regulation. Noncoding DNA, containing introns, repetitive elements, and potentially active transposable elements, requires effective mechanisms for its long-term silencing. Genes are transcribed from methylation-free promoters even though adjacent transcribed and nontranscribed regions are extensively methylated. Gene promoters are used and regulated while keeping noncoding DNA, including transposable elements, suppressed.

Role of DNA methylation in Regulating Gene Activity

DNA methylation prevents the expression of genes by altering the amount of messenger RNA. Enzymes attach chemical tags called methyl groups to the bases from which DNA is made.

DNA Methylation in Plants

DNA methylation in plants is more diverse than in animals. In addition to methylating CpGs, plants also methylate the cytosine at CpNpG and CpNpNp sequences, where N can be any base. Plants also have a greater variety of enzymes
involved in methylating DNA than animals. Methylation of plant DNA occurs in transposon sequences, regions of repeated DNA sequences and in the coding region of genes.

**DNA Methylation Patterns are Heritable**

Once a gene has been methylated, all the daughter cells from that cell retain the methylation, making it a heritable change. Some genetic conditions are caused by inappropriate over or under methylation of the same region of DNA, such as Prader-Willi and Angelman’s syndromes.

**The link between DNA Methylation and Cancer**

Cancer is now recognised as both a genetic and epigenetic disease. While some types of cancers can be inherited, other cancers result from changes to DNA that accumulates throughout life. There are three types of cancer-causing genes: oncogenes, tumour suppressor genes and DNA repair genes.

**Age Related Cancers**

DNA methylation is a dynamic process, with the enzymes involved constantly working to methylate and demethylate CpG sites throughout the genome. Inappropriate methylation patterns leads to inactivation of genes that should be expressed, which poses a particular problem when those genes are tumour suppressor genes vital for controlling normal cell growth.

### 2.7 THE HAP MAP PROJECT

Linkage studies played an important role to understand pattern of inheritance in the human populations particularly in genetically isolated subpopulations, i.e. a group of alleles for neighboring genes on a segment of a chromosome are inherited together. Such a combination of linked alleles is known as a haplotype. When a new mutation occurs in a single individual and is passed down to his or her descendants, and is carried on a specific chromosome. Therefore, every person has a unique combination of alleles of all genes which are not inherited on a completely random basis but come in bunches, that is, haplotype. The successful completion of Human Genome Project led to characterise and sequence the entire genomes of several other organisms, many of which are used extensively in biological research. Identification of the sequence or function of genes in a model organism is an important approach in finding and elucidating the function of human genes. The next key step of the Human Genome Project (HGP) (following the creation of the genetic, physical, sequence and SNP maps) is the generation of a “haplotype” map of the human genome which consists of a high density of SNPs defining the small number of ancestral haplotype (blocks of tightly correlated genetic variants) in each region of the human genome. Knowledge of these haplotype will open a new path for biomedical research.

The variations have an effect on individual’s disease risk and the sites in the DNA sequence where individuals differ at a single DNA base are called single nucleotide polymorphisms (SNPs). Sets of nearby SNPs on the same chromosome are inherited in blocks. This pattern of SNPs on a block is a haplotype. The Hap Map is a map of these haplotype blocks and the specific SNPs that identify the haplotype are called tag SNPs. The “Hap Map,” is a tool to find genes and genetic variations that affect health and disease. The Hap Map is valuable by reducing
the number of SNPs required to examine the entire genome for association with a phenotype from the 10 million SNPs that exist to roughly 500,000 tag SNPs. The HAPMAP would help in genome scan approaches to finding regions with genes that affect diseases much more efficient and comprehensive, since effort is not wasted in typing more SNPs than necessary and all regions of the genome are included.

In addition to its use in studying genetic associations with disease, the Hap Map is a powerful resource for studying the genetic factors contributing to variation in response to environmental factors, susceptibility to infection, and to study the adverse responses to drugs and vaccines.

2.7.1 Method of Study

HapMap results would help to study based on the expectation that there are higher frequencies of the contributing genetic components in a group of people with a disease or particular response to a drug, vaccine, pathogen, or environmental factor than in a group of similar people without the disease or response. Using just the tag SNPs, researchers are able to find chromosome different haplotype distributions in the two groups of people, those with a disease or response and those without. Each region is then studied in more detail to discover which variants in which genes in the region contribute to the disease or response, leading to more effective interventions. These investigations will lead to the development of tests to predict which drugs or vaccines would be most effective in individuals with particular genotypes for genes affecting drug metabolism. Thus the haplotype map should be generated rapidly and made freely available to researchers worldwide.

2.7.2 International Hap Map Project

The International Hap Map Project is an organisation that aims to develop a haplotype map (Hap Map) of the human genome, which describes the common patterns of human genetic variation. Hap Map is a key resource for researchers to find genetic variants affecting health, disease and responses to drugs and environmental factors. The information produced by the project will be made freely available to researchers around the world. The International Hap Map Project is collaboration among researchers at academic centers, non-profit biomedical research groups and private companies in Canada, China, Japan, Nigeria, the United Kingdom, and the United States.

The Hap Map project proposes a shortcut. Although any two unrelated people share about 99.5% of their DNA sequence, some people may have an A at a particular site on a chromosome while others have a G instead. Such a site is known as a single nucleotide polymorphism (SNP), and each of the two possibilities is called an allele. The Hap Map project focuses only on common SNPs, those where each allele occurs in at least 1% of the population. Each person has two copies of all chromosomes, except the sex chromosomes in males. For each SNP, the combination of alleles a person has is called a genotype. Genotyping refers to uncovering what genotype a person has at a particular site.

To find the genetic factors involved in a particular disease, one can proceed as follows. First a certain region of interest in the genome is identified, possibly from earlier inheritance studies. In this region one then locates a set of tag SNPs
from the Hap Map data; these are SNPs that are very well correlated with all the other SNPs in the region, so that knowledge of the alleles of the tag SNPs in an individual will determine the individual’s haplotype with high probability. Next, one determines the genotype for these tag SNPs in several individuals, some with the disease and some without. By comparing the two groups, one can then determine the likely locations and haplotype that are involved in the disease.

Hap Map resource is to guide the design and analysis of genetic association studies, shed light on structural variation and recombination identifies loci that may have been subject to natural selection during human evolution. Despite the ever-accelerating pace of biomedical research, the root causes of common human diseases remain largely unknown, preventative measures are generally inadequate and available treatments are seldom curative. Family history is one of the strongest risk factors for nearly all diseases—including cardiovascular disease, cancer, diabetes, autoimmunity, psychiatric illnesses and many others—providing the tantalizing but elusive clue that inherited genetic variation has an important role in the pathogenesis of disease. Identifying the causal genes and variants would represent an important step in the path towards improved prevention, diagnosis and treatment of disease. More than a thousand genes for rare, highly heritable diseases have been identified so far.

2.7.3 Genetic Variation and Use of Hap Map

Most common diseases, such as diabetes, cancer, stroke, heart disease, depression, and asthma, are affected by many genes and environmental factors. Discovering the DNA sequence variants that contribute to common disease risk offers one of the best opportunities for understanding the complex causes of disease in humans. Sites in the genome where the DNA sequences of many individuals differ by a single base are called single nucleotide polymorphisms (SNPs). For example, some people may have a chromosome with an A at a particular site where others have a chromosome with a G. Each form is called an allele.

```
   TAGC
   TGGC

A part of two chromosomes showing a SNP. Both the A and G alleles are shown.
```

Each person has two copies of all chromosomes except the sex chromosomes. The set of alleles that a person has is called a genotype. For this SNP a person could have the genotype AA, AG, or GG (for basic genetics information). The term genotype can refer to the SNP alleles that a person has at a particular SNP or for many SNPs across the genome. A method that discovers what genotype a person has is called genotyping. About 10 million SNPs exist in human populations, where the rarer SNP allele has a frequency of at least 1%. Alleles of SNPs that are close together tend to be inherited together. A set of associated SNP alleles in a region of a chromosome is called a “haplotype”. Most chromosome regions have only a few common haplotype (each with a frequency of at least 5%), which account for most of the variation from person to person in a population. A chromosome region may contain many SNPs, but only a few “tag” SNPs can provide most of the information on the pattern of genetic variation in the region.
A chromosome region with only the SNPs shown. Three haplotype are shown. The two SNPs in color are sufficient to identify (tag) each of the three haplotype. For example, if a chromosome has alleles A and T at these two tag SNPs, and then it has the first haplotype.

The Hap Map describes the common patterns of genetic variation in humans.

Population and Sample
Most of the common haplotype occur in all human populations; however, their frequencies differ among populations. Therefore, data from several populations are needed to choose tag SNPs.

Impact on Biomedical Research
The availability of a haplotype map of the human genome will have a substantial impact on human genetic studies. Specifically, these studies include:

- Genome-wide association studies.
- Human population structure and history.

Numbers of SNPs to be genotyped
Haplotype could be used to greatly simplify the potential task of genetic testing to determine the genetic disease risks in the context of hundreds of common haplotype rather than considering all of the individual interactions of tens of thousands of genes, each with their own unique distributions among various subpopulations. There are many other clinical manifestations of human genetic variation. In fact, all disease has a genetic component, and all therapies should consider genetic variations (perhaps that can become the motto for the new era of genomic medicine). The physician should be aware of the genetic components of susceptibility versus resistance to various pathogens, variations in disease severity or symptoms, reactions to drugs (pharmacogenomics), and the variable disease course and prognosis that emerges as a synthesis of all of these factors.

2.8 SUMMARY
Genetic epidemiology aims at diseases that are caused by genetic alterations among closely related individuals or in a Mendelian population. It measures the frequency and prevalence of a genetic disease, gene - gene interaction, gene - environment interaction, genotype and phenotype correlation, the extent of expression of default gene in families or in a population. When there is no fault in the structure of a gene, a disease may occur in the faulty expression of a gene. Such changes are due to DNA methylation and histone modification. It is called as epigenetics. Segregation, association and linkage analysis are important methods used in Genetic epidemiological studies. Using these methods internationally, a Haplotype map of humans (HapMap) was constructed to understand common genetic variation and its association with diseases.
Suggested Reading


Sample Questions

1) Define Genetic Epidemiology?

2) Give a brief account on the scope of genetic epidemiology.

3) Explain the term epigenetics. Epigenetics concept has given new ideas to Human Genetics. Justify the statement.

4) DNA Methylation is an important process to understand gene-expression. Explain DNA Methylation process.

5) HAP MAP Project is an essential innovation for biomedical research. State the aims of the project.