

Indira Gandhi
National Open University
School of Social Sciences

Block

6

APPLIED AND EMERGING TRENDS IN HUMAN GENETICS

UNIT 1

Introduction to New and Emerging Areas in Human Genetics

5

UNIT 2

Genetic Epidemiology and Epigenetics

26

UNIT 3

Prevention of Genetic Diseases

39

UNIT 4

Genetics and Human Issues

48

Expert Committee

Prof. P. Dash Sharma (Retd.)
Dept. of Anthropology
Ranchi University, Ranchi

Prof. P. Veerraju (Retd.)
Dept. of Human Genetics
Andhra University
Visakhapatnam

Prof. S.M.S. Chahal
Dept. of Human Biology
Punjabi University, Patiala

Prof. A. Papa Rao
Dept. of Anthropology
Sri Venkateswara University
Tirupati

Dr. Roli Mathur
Scientist 'C'
Division of Basic Medical Sciences
Indian Council of Medical Research
New Delhi

Dr. Seema Kalra
Assistant Professor
Dept. of Biochemistry
IGNOU, New Delhi

Faculty of Anthropology SOSS, IGNOU

Dr. Rashmi Sinha, Reader
Discipline of Anthropology
IGNOU, New Delhi

Dr. P. Venkatramana
Assistant Professor
Discipline of Anthropology
IGNOU, New Delhi

Dr. Rukshana Zaman
Assistant Professor
Discipline of Anthropology
IGNOU, New Delhi

Dr. Mitoo Das
Assistant Professor
Discipline of Anthropology
IGNOU, New Delhi

Dr. K. Anil Kumar
Assistant Professor
Discipline of Anthropology
IGNOU, New Delhi

Academic Assistance provided by
Dr. N.K. Mungreiphy, Research Associate
(DBT) for the Expert Committee meeting

Programme Coordinator: Dr. Rashmi Sinha, SOSS, IGNOU, New Delhi

Course Coordinator: Dr. P. Venkatramana, SOSS, IGNOU, New Delhi

Content Editor

Prof. P. Veerraju (Retd.)
Department of Human Genetics
Andhra University, Visakhapatnam

Block Preparation Team

Unit Writers

Dr. S.A.A. Latheef (Unit 1)
Post Doctoral Fellow (CREBB)
School of Life Sciences
University of Hyderabad, Hyderabad

Dr. Kumud Sarin (Unit 2)
Technical Director, Indian Biosciences
and Research Institute, NOIDA

Dr. P. Venkatramana (Unit 3)
Assistant Professor,
Faculty of Anthropology
SOSS, IGNOU, New Delhi

Dr. Roli Mathur (Unit 4)
Scientist 'C'
Division of Basic Medical Sciences
Indian Council of Medical Research
New Delhi

Cover Design

Dr. Mitoo Das
Asstt. Professor
Anthropology
SOSS, IGNOU
New Delhi

Authors are responsible for the academic content of this course as far as the copyright issues are concerned.

Print Production

Mr. Manjit Singh
Section Officer (Pub.), SOSS, IGNOU, New Delhi

November, 2012

© Indira Gandhi National Open University, 2012

ISBN-978-81-266-

All rights reserved. No part of this work may be reproduced in any form, by mimeograph or any other means, without permission in writing from the Indira Gandhi National Open University.

Further information on Indira Gandhi National Open University courses may be obtained from the University's office at Maidan Garhi, New Delhi-110 068.

Printed and published on behalf of the Indira Gandhi National Open University, New Delhi by the Director, School of Social Sciences.

Laser Typeset by : Tessa Media & Computers, C-206, A.F.E.-II, Okhla, New Delhi

Printed at :

BLOCK 6 APPLIED AND EMERGING TRENDS IN HUMAN GENETICS

In recent years, the application of advanced technologies has created a number of new emerging subfields in the field of genetics. Of late, with the availability of latest sequencing technologies leading to the sequencing of genomes of various organisms including human, large scale public-domain genomic databases have been created. To treat this data, the field of Bioinformatics has emerged. Bioinformatics is defined as the computational handling and processing of genetic information. Various tools have been used in this subfield. Further sequencing of the genomes has led to the origin of proteomics dealing with the understanding of post transcriptional and post translational modifications of various proteins. Proteomics has subfields like structural, expression, functional proteomics etc. Various tools like databases, mass spectrophotometry, software tools and protein separation technologies have been used for Proteomics.

Further, a new field of pharmacogenomics has emerged which is concerned with genome wide analysis of genetic determinants of drug efficacy and toxicity. Besides, various factors that influence the drug response like age, sex, race, ethnicity, disease state and genetics also play an important role in drug response.

Further stem cell research has made a great impact on the applied value of genetics. Various types of stem cells have been characterised and the applications of these cells vary from an in vitro model to therapeutic tool.

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 – 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.

Genetic disease is caused by variation or alteration called mutation in gene, or alteration in chromosome. Genetic diseases can be classified into simple and complex diseases. Simple genetic diseases are also called as monogenetic or Mendelian diseases caused by a mutation. Complex diseases have multi-factorial causation. They are caused by the interaction of many genes and gene – environmental interaction. Screening the genetic diseases in the population and genetic counselling to the couples is a prerequisite in avoiding the defective progeny. To prevent the genetic diseases prenatal diagnosis and termination of defective zygote is one of the best ways. Gene therapy is the treating method for genetic diseases, it involves in correcting the default genetic sequence. Currently available method for preventing genetic diseases is to identify the faulty genetic code at an early stage and change the life style. By changing the life style we can prevent some late onset genetic diseases. In curing the genetic diseases future research may bring success in gene therapy.

The last decade has witnessed tremendous advances in science and technology and the prospects of genetic testing, genetic engineering, cloning, stem cell research, prenatal testing, DNA diagnostic etc. all have raised many ethical dilemmas in the minds of scientists and society. There are advantages of such new technologies but there is also the fear of the unknown distant possibilities which needs careful examination. Weighing of risk benefit ratio, protection of privacy and confidentiality and sharing of burdens and benefits are some of the main issues. Close monitoring of research in these areas is important. Education has the potential power to prevent stigmatization and discrimination by emphasising that genetic disorders are not caused by the behaviour of affected persons or families but the fact that most people may carry some recessive lethal mutations and that our offspring or we are all at genetic risk. For all kinds of genetic testing and research respect for persons is important and should include informed consent, right to referral, full disclosure, protection of confidentiality, and respect for minority groups, women and children. The implications of these various ethical, legal and social issues concerning the application of genetic technologies are outlined here.



UNIT 1 INTRODUCTION TO NEW AND EMERGING AREAS IN HUMAN GENETICS

Contents

- 1.1 Introduction
- 1.2 Proteomics
- 1.3 Bioinformatics
- 1.4 Pharmacogenomics
- 1.5 Stem Cell Research
- 1.6 Summary
- Suggested Reading
- Sample Questions

Learning Objectives



After reading this unit, you will be able to:

- appreciate the definition, importance, summary of proteins and their functions, subfields, tools, applications and challenges of proteomics;
- describe scope, historical development, tools, biological databases, applications and challenges of bioinformatics;
- gain insight on scope, classic examples, evolutionary aspects, approaches, in vitro and in vivo models, applications and challenges of pharmacogenomics; and
- discuss scope and evolution of stem cell research, types of stem cells, applications and challenges of stem cell research.

1.1 INTRODUCTION

Like any discipline human genetics is growing by leaps and bounds with the advancement of science and technologies for the noble mission of human welfare and enhancement of quality of life of the people.

Sequencing of the genomes has led to an accumulation of a wealth of sequence data. The functions of a large number of genes are unknown. Deciphering the functions of genes will require a clearer understanding of post transcriptional and post translational modifications. Since defective proteins are the ultimate causes of most diseases, it is important to consider the proteome (set of all proteins) in relation to the genome, the study of which constitutes the subject matter of proteomics.

Advancement of sequencing technologies, and the consequent of sequencing of genomes of various organisms including human, and availability of large-scale public-domain genomic and related databases have necessitated the development of the discipline of Bioinformatics. Bioinformatics has been useful for the

maintenance and analysis of sequences within the reach of scientific community and public in the form of databases. Development of various algorithms have facilitated the comparison of sequences and genomes for predicting the genes and their function, translation of nucleotide sequences into amino acid sequences, predicting the structure of proteins by homology modeling, single nucleotide polymorphism discovery, finding the susceptible biomarkers (or genetic markers) of diseases and identification of drug targets for the better management of diseases.

There is a difference of about 0.1% in the DNA sequences of any two individuals. This variation often resulted in variation in drug response. Studying at genome level for characteristic genetic variants responsible for variation in response to drugs has resulted in the birth of pharmacogenomics. Lack of adequate knowledge of genotype characteristics that contribute to adverse reactions to various drugs is also increasing patient-care costs. Genotype guided management of diseases holds great promise.

Decreasing birth rates, improvement of nutrition and control of infectious diseases have improved the lifespan, and at the same time increased the prevalence of diseases like diabetes, cardiovascular diseases and cancers. With the advancement of medical technology and pharmaceutical research many of these diseases can be managed clinically. Finally, patients of these diseases sometimes face end-stage organ dysfunction requiring organ transplantation. Lack of awareness on organ donation and pending clearances on cadaver organ transplantation have fostered the development of stem cell research efforts. Somatic cell nuclear transfer is one of the tools of stem cell research which helps in developing autologous and immune rejection free cells for therapeutic purposes. Developments in the stem cell research have helped to shape the therapies for the management of some disorders. Stem cell research has raised the hope of suffering patients, although the implementation of the results of this research for patient welfare is still in its infancy.

Let us discuss in detail on the above said areas of Human Genetics.

1.2 PROTEOMICS

The term 'Proteome' was coined by Marc Wilkins in 1994. This term is a linguistic equivalent to genome and deals with large scale analysis of the complete, or at least the major, set of proteins. The Proteome can be defined in terms of the sequence, structure, abundance, localisation, modification, interaction and biochemical function of its components.

Importance of proteomics

- 1) Genes are instruction carriers, while the proteins are the functional molecules of the cells and a true understanding of them can come from the direct study of proteins.
- 2) Unlike the genome whose content with few exceptions remain the same irrespective of cell type or environmental conditions, the proteome is dynamic whose content varies under different conditions due to the regulation of transcription, RNA processing, protein synthesis and protein modification. Study of the proteome can provide the glimpse of the cell in action.

- 3) A good understanding on the structure and function of a protein may provide clues to introduce mutations in order to better understand their function.
- 4) Transcriptome may not represent the true insight on proteome, because not all mRNA in the cell are translated and rates of protein synthesis and protein turnover may vary among transcripts.
- 5) Difference in the stability of mRNA and efficiencies in translation can affect the generation of new proteins. Some transcripts may give rise to multiple proteins. For instance, 22 different forms of alpha-1-antitrypsin were observed in plasma. The individual functions of these proteins can be studied at the protein level only.
- 6) In certain body fluids like serum, cerebrospinal fluid and urine where nucleic acids are not represented, proteins only provide information about determinants of disease progression. In case of degradation and cross linking of nucleic acids in fixed biological specimens, protein may only act as a source material for further study. In many diseases proteins are the drug targets.
- 7) Proteomics attempts to bridge the gap between our understanding of genome sequence and cellular behaviour.

A brief account on proteins is as follows:

Proteins: The term 'protein' is derived from the Greek word 'Proteios', meaning of the first order. J.J.Berzelius in 1938 coined the term 'protein' to describe a class of macromolecules abundant in living organisms. Proteins constitute about 50% of dry weight of the cell. They are made up of amino acids which were earlier 20 in number, but recently two more such as selenocysteine and pyrrolysine were added to the list. Except in marine microorganisms, in which both D and L amino acids were found, proteins contain L-amino acids only. The D and L refers to the property of amino in response plane polarized light. The amino acids form peptide bonds involving carboxylic group of one amino acid with amino group of other amino acids. The involved amino acids are called amino residues. Depending on the number of amino acid involvement in peptide bonds they are called bi, tri, tetra etc. A protein may have one or more polypeptides (a string of amino acids) folded mostly into either globular or fibrous form.

Proteins are synthesized by the translation of mRNA into polypeptides on ribosomes. After translation, they undergo 400 types of reversible and irreversible chemical reactions like glycosylation, phosphorylation, which are collectively called as post translational modifications. At any given time in a cell, the level of protein depends on the rate of transcription of the gene, the efficiency of translation of mRNA into protein and the rate of degradation of the protein. Various agents like oxidants, radiation, chemicals cause modification of proteins which lead to their degradation. Phosphorylation accompanied by conjugation with ubiquitin and lysosomal enzymes also effect degradation of proteins.

Structure: The structure of protein can be explained at four levels.

- a) **Primary:** This contains the linear sequence of amino acids which are bonded by covalent peptide bonds or linkages. This structure determines its function and the composition of amino acids is responsible for physical and chemical properties.

- b) **Secondary:** This is formed by twisting of polypeptide chain leading to spatial arrangement of protein. The basis of secondary structure depends on the pattern of hydrogen bonds between amide and carboxylic groups. α helix and β sheets are the known two main types of secondary structures. α helix has a rigid arrangement of a polypeptide chain in which amino acid side chains extend outward from the central axis. It is stabilized by extensive hydrogen bonding. In case of β sheets, hydrogen bonds are observed between the neighbour segments of polypeptide chains. The arrangement of polypeptide in β sheets is either parallel (same direction) or anti-parallel (opposite direction).
- c) **Tertiary:** This structure provides the stability of the protein. It is the three dimensional structure. In this, hydrophobic side chains are held interior and hydrophilic groups are seen on the surface of the protein.
- d) **Quaternary:** It is the spatial arrangement of subunits (polypeptide chains). These subunits are held together by noncovalent bonds like hydrogen bonds, hydrophobic interactions and ionic bonds. Depending on the number of polypeptides these subunits are known as mono, di, tri or tetramers. If they are identical they are called homo or if unrelated they are known hetero.

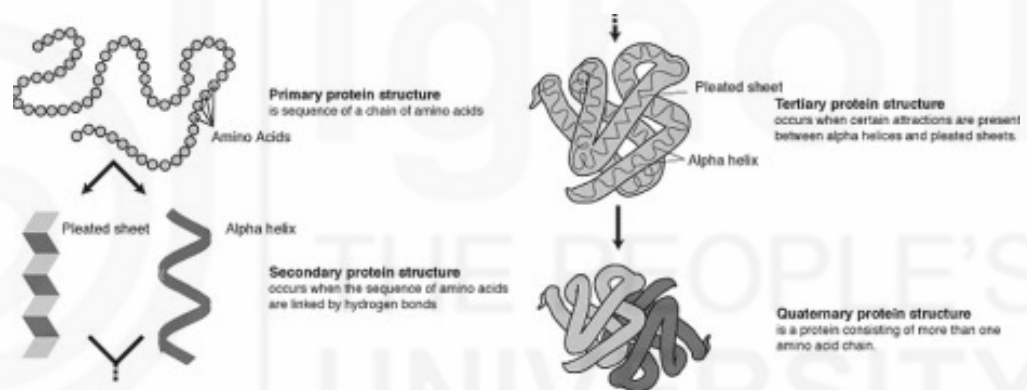


Fig. 1.1: Protein Structure (source: bioweb.wku.edu)

Subfields of Proteomics

- 1) **Sequence and Structural proteomics:** Protein sequences allow the designing of probes or primers which can be used to isolate the cDNA or genomic sequence. It is the protein sequence which acts as a bridge between the activity of a protein and the genetic basis of a particular phenotype. Increasing deposition of protein sequences and consequent development of statistical techniques are facilitating the comparison of proteins. Three primary sequence databases are Genbank, EMBL (European Molecular Biology Laboratory), DDBJ (DNA Data Bank of Japan) that provide translated protein sequences from DNA sequences, whereas, SWISSPROT is a dedicated protein sequence data bank.

Similar sequences may give rise to similar structures and this idea has given birth to a new branch of proteomics known as structural proteomics which paved the way for storage, presentation, comparison, inferring evolutionary relationships and prediction of theoretical protein models, a boon in the absence of crystallographic protein structures, for drug discovery research.

- 2) **Expression proteomics:** It is concerned with the analysis of protein abundance, separation of protein mixtures, the identification of individual components and their systematic quantitative analysis. This sub branch lays emphasis on differences representing alternative states like health and disease and characterisation of post-translational modifications. The key tools used in investigations involve 2D gel electrophoresis, mass spectrometry, multi-dimensional chromatography and protein chips.
- 3) **Interaction proteomics:** It deals with the genetic and physical interactions among proteins as well as interactions between proteins and nucleic acids or small molecules. Study of protein interactions not only provides insight on the function of individual proteins but also how proteins function in pathways, networks, and complexes. It seeks to achieve creation of proteome linkage map based on binary interactions between individual proteins and higher order interactions determined by the systematic analysis of protein complexes. Interactions between proteins and nucleic acids emphasize on processes such as gene regulation, while interaction of proteins with small molecules may enlighten on the interaction of enzymes with substrates and receptors with their ligands and also may play an important role in drug development process. The key approaches used in studies of this kind of interaction are yeast, two hybrid system, mass spectrometry, biochemical assays and X-ray crystallography.
- 4) **Functional Proteomics:** This lays emphasis on testing protein functions on a large scale such as testing expressed proteins for different enzymatic activities.

Tools used in Proteomics

- 1) **Databases:** Protein Expressed sequence tag and complete genome sequence provide information on all the expressed proteins in organisms.
- 2) **Mass Spectrometry:** It provides information on molecular measurement (>100KDa) and sequence analysis of proteins.
- 3) **Soft ware tools:** These tools determine the sequence of a protein with the aid of specialised algorithms and provide automated survey of large amounts of mass spectrometry data for protein sequence matches.
- 4) **Protein Separation technologies:** They resolve complex protein mixtures into individual proteins and permit comparison of differences in protein levels between two samples. The key technologies include 2Dimensional gel electrophoresis, SDS-Poly acrylamide gel electrophoresis, high performance liquid chromatography, capillary electrophoresis, affinity and ion exchange chromatography.

Applications

- 1) Identification and cataloguing of proteins.
- 2) Identification of proteins in a sample of differentiation, developmental state, disease state and exposed to a drug, chemical or physical stimulus.
- 3) Determinate how proteins interact with each other in living systems and characterisation of proteins in more complex networks.
- 4) Mapping of proteins in post-translational modifications.

Challenges: No single technological approach is suitable for every application. Integration and automation of these approaches, using of better materials, advancement in instrument design and methodology for improving sensitivity, resolution and repeatability are the challenges before the proteomics community in order to provide a comprehensive analysis of complex biological system.

1.3 BIOINFORMATICS

Bioinformatics is defined as the computational handling and processing of genetic information. Its goal is to enable the discovery of new biological insight as well as to create a global perspective from which unifying principles can be perceived. It lays emphasis on organizing the data for accessing and generating information, developing tools to analyse the data and interpret the data in a biologically meaningful manner. In analysing the data, two approaches like data mining and knowledge discovery are taken into consideration. These approaches involve generating testing hypothesis regarding the function or structure of a gene or protein of interest by homology searching. Besides deriving meaningful biological information from the data; bioinformatics also serves the scientific community with resources such as databases.

Bioinformatics plays an important role in many areas of biological research like genomics, transcriptomics, proteomics, structural biology, genetics, molecular biology and evolutionary biology. Bioinformatics knowledge is utilised in genomics particularly, genome sequencing, mapping, genome annotation, comparing multiple genomes, calculating evolutionary distance and single nucleotide polymorphism discovery. Transcriptomics applications of bioinformatics includes study of transcribed sequences, both full length cDNA and expressed sequence tags and analysis of gene expression data. In proteomics, bioinformatics is helpful in the analysis of protein sequences, protein abundance and in determination of protein structure either empirically or computationally. In molecular biology, analysis of protein-protein interactions and molecular pathway and in systemic studies of gene regulation bioinformatics has significant contribution. In genetics, bioinformatics is useful in the discovery of new molecular genetic markers such as SNP's and use of these and other markers to dissect the genetic basis of disease and other phenotypes. Bioinformatics is also helpful in the studies of evolution and phylogeny.

Tools used in Bioinformatics: Various tools are used in bioinformatics research like internet, search engines like Google, Scirus, AltaVista, Lycos, HotBot, Northern Light, Dogpile, databases like National centre for biotechnology information (NCBI), PubMed and sequence analysis tools like BLAST, FASTA, multiple sequence alignment (MSA), and visualisation tools such as RasMol, Jmol and Cn3D.

Computational methods help in analysing the data and formulate hypotheses. Sequence data is the most abundant type of biological data available electronically. Pairwise sequence comparison is used in bioinformatics applications for sequence based database searching, building evolutionary trees, identification of characteristic features of protein families, create homology models, compare genomes, explore sequence determinants of protein structure and to connect expression data to genomic information.

Sequence data can be used for sequence analysis to know the sequence characteristics, sequence comparison, multiple sequence alignment, motif discovery and phylogenetic inference. Sequence databases can be found on internet. Databases are searched mostly for similarity search. The tools used for similarity search includes BLAST and FASTA.

BLAST: It stands for basic local alignment search tool. This algorithm is used to perform sequence similarity search. A server at NCBI was established to support BLAST. This server is used widely for sequence database searches. An independent set of BLAST program was developed at Washington University known as WU-BLAST. This BLAST also performs the similarity search like NCBI and produces gapped local alignments. The BLAST requires different statistical methods to evaluate sequence similarity score. BLAST algorithm increases the speed of sequence alignment by searching first for common words or k-tuples in query sequence and each database sequence. It searches the words that are significant. In case of proteins, the significance of word matching is evaluated by BLOSUM 62 amino acid substitution matrix. The word length is 3 for proteins and 11 for nucleic acids in BLAST algorithm. The latest version of BLAST is BLAST 2. It reports the gapped alignment of query and database sequences. BLAST has filtering feature for searching low complexity regions in query sequence (repeats of sequence character) which produces artificial high score alignments.

There are number of variations of BLAST programs like Blastp for comparing an amino acid query sequence against a protein sequence of a database; Blastn for comparing a nucleotide query sequence against nucleotide sequence of a database; Blast x for searching six frame translation product of nucleotide sequence against a protein database; Tblastn for searching a protein sequence against a translated nucleotide sequence of a database; and Tblastx for comparing six frame translations of nucleotide sequence query sequence against a six frame translations of a nucleotide sequence data base. MegaBLAST searches similar sequences that are 300 to 100,000 bp long. A long word is used for searching and the gap penalty is calculated from the match and mismatch scores. RPS BLAST scans conserved domains in a protein sequence. BLASTcl3 is a network client BLAST which is used to access the BLAST server. Standalone BLASTs are executable versions of all the BLAST program for the operating systems Windows, Unix and Macintosh. PSI-BLAST (Position specific iterated BLAST), PHI BLAST (pattern hit initiated BLAST) are used to search for domains in query protein sequence and in database sequences. BEAUTY (BLAST enhanced alignment utility) adds additional features to the BLAST like summarizing the locations of HSPs, PFAM domains and Prosite pattern. BLAST searching with cobbler sequence (consensus) is used to find majority residues in multiple sequence alignment. BLAST2 program is used to align very long sequences.

FASTA: This program is used for aligning pairs of protein and DNA sequences. It searches for matching sequence patterns or words called k-tuples. Patterns contain k consecutive matches of letters. It attempts to build a local alignment on word matches. This program is used for database searches. FASTA compares the query protein or DNA sequences to the target sequences in the database and give the best matched sequence and local alignment of matched sequences. To search for similarity FASTA uses hashing method in which a table of the positions of each word of length k, or k tuple is constructed for each sequence. The position

of each word is calculated by subtracting the position in the first sequence from the position in the second and words having the same offset position show a region of alignment between the two sequences. The number of comparisons increases as the average sequence length. The k tuple length is 1 or 2 for protein and 4-6 for nucleic acid sequences in FASTA program. There are other versions of the FASTA also reported. Among them TFASTA compares the query protein sequence to a six frame translation of DNA sequence of the database; FASTF/TFASTS compares a set of short peptide fragments against a protein sequence database or a DNA sequence database translated in all six reading frames; FASTX and FASTY translate a query DNA sequence in all three reading forward frames and compare all three frames to a protein sequence database; and TFASTX and TFASTY compare a query protein sequence to a DNA sequence database, translating each DNA sequence in all six possible reading frames.

Multiple Sequence alignment (MSA): It is an alignment of three or more sequences and aims to place sequence positions related by function and evolution in the same column of the alignment allowing for mismatches and gaps (deletions or insertions). In msa, both global and local alignments are used. In global alignment, dynamic programming algorithm is used for alignment of three sequences, more than this number, only a small number of relatively short sequences may be analysed. The methods used include progressive methods (ClustalW, ClustalX, MAFFT, MAVID, MSA, MULTIPIPMAKER, POA, PRALINE, T-COFFEE) which start by aligning most alike sequences followed by building an alignment by the addition of more sequences; iterative methods (DIALIGN, PRRP, SAGA) initially align group of sequences and then revise the alignment to achieve a more reasonable result; methods of aligning the sequences based on conserved pattern found in the same order in the sequences; statistical methods generating probabilistic models of the sequences and graph based methods.

Local MSA methods align the most similar regions in sequences. The approaches include profile analysis which identifies highly conserved portion of the alignment and produces a scoring matrix called a profile. A profile includes scores for amino acid substitutions and gaps in each column of the conserved region. In block analysis, blocks (substituted regions without gaps) are searched and used in sequence alignments. Pattern searching or statistical methods scan a localised region of sequence similarity in a set of sequences.

Structure visualisers: Protein structure data is stored as collections of x,y,z coordinates. The connectivity between atoms in proteins has to be taken into account and for the visualisation to be effective a virtual 3D environment which needs to be created. A protein structure visualisation program needs to be able to display use selected subsets of atoms with correct connectivity, draws standard cartoon representations of proteins such as ribbons and cylinders and recolour subsets of a molecule according to a specified parameter.

RasMol: It is a structure visualisation program tool and available for a wide range of operating systems. It reads molecular structure files in the standard PDB format. It comes in three display depths 8, 16 and 32 bit. The molecule can be rotated in window. It has file menu commands for opening molecular structure file, display menu commands for the molecular display style to formats including ball and stick, cartoons and space fill. The colour menu allows colour changes of

the molecule, option menu allows changes of the display style and export menu facilitates writing the displayed images in common electronic image formats such as GIF, PostScript and PPM. Help common allows the creation of own combination of colours and structure display formats.

Cn3D: It allows viewing protein structure files NCBI ASN.1 format. It opens two windows: a colour structure viewer in which a molecule can be rotated, coloured according to different properties and rendered in different display formats; a sequence viewer, which allows you to view sequences and alignments corresponding to the displayed protein and to add graphics to the sequence display to highlight the location of secondary structure features.

Biological Databases: A biological database is a large, organised body of persistent data, usually associated with computerised software designed to update, query and retrieve components of the data stored within the system. These databases are helpful to gain a insight into biological phenomena from the structure of biomolecules and their interactions to the whole metabolism of organism and to understand the evolution of species. Databases are classified as primary databases (DDBJ, EMBL, Gene Bank), protein sequence databases (SWISS-PROT, Protein Information Resource), protein sequence databases (Pfam, PROSITE) protein structure databases (PDB, SCOP), protein-protein interaction databases (BioGRID, STRING), pathway databases (KEGG), microarray databases (Array express, Gene expression omnibus).

DDBJ (DNA Data Bank of Japan): It is run by National Institute of Genetics, Japan. It is the only nucleotide sequence database in Asia. It works in collaboration with EMBL and Gene Bank. It collects experimentally determined sequence data mainly from Japanese researches but also accepts from others as well. The database is a collection of “entry” which is the unit of the data. Each entry includes nucleotide sequence and the information of submitters, references, source organisms, and the biological nature such as gene function and other property of the sequence

EMBL(European Molecular Biology Laboratory): This database is maintained by European Molecular Bioinformatics Institute, Cambridge, U.K. It is Europe primary nucleotide sequence database. The data consists of DNA and RNA sequences drawn from individual researchers, genome sequencing projects and patent applications. As on 30th August, 2012 it contain 252,106,363 sequence entries comprising 450,481,663,919 nucleotides.

Gene Bank: It is a NIH genetic sequence database. As on April 2011, it had a collection of 126,551,501,141 bases in 135,440,924 sequence record. Entrez nucleotide is used for sequence identification and annotations. Entrez nucleotide is divided into core nucleotide (main collection), dbEST (expressed sequenced tags) and dbGSS (genome survey sequences). BLAST programme can be used to align query sequences to Genebank sequences.

SWISS-PROT: It is a curate protein sequence database formed in 1986 and maintained by the Department of Medical Biochemistry of the University of Geneva and the European Bioinformatics Institute (EBI). The characteristic features of SWISS-PROT include availability of high level annotations, a minimal redundancy and integration with other databases.

PIR (Protein Information Resource): It was established in 1984 by the National Biomedical Research Foundation (NBRF) and helps researchers in the identification and interpretation of protein sequence information and provides tools.

Pfam: It is a database of protein families. It has two components namely Pfam-A and Pfam-B. Pfam-A contains large portion of sequences with high quality and manually curated protein families, while, the Pfam-B has a collection of low quality families and useful for identifying functional regions. Pfam also generates clans, a grouping of related protein families.

PROSITE: This database has a collection of protein families and domains. Patterns and profiles of more than thousand protein families and domains data is available in this database.

PDB (Protein Data Bank): This database is of experimentally determined structures of proteins, nucleic acids and complex assemblies. It was established in 1971 at Brookhaven National Laboratories. In 1998 it came under the umbrella of Research colloboratory for structural bioinformatics. It contains information about coordinates, deposited structures and method of structure determination. As on October 2nd 2012, 85 thousand records of structures were available.

SCOP (The Structural Classification of Proteins): It organises and classifies proteins based on their evolutionary and structural relationships. It is organised into four hierarchical levels: family, super family, fold, and classes.

BioGRID (The Biological General Repository for Interaction Datasets): It is a database with collections of genetic and protein interaction data from model organisms and humans. It has holdings of 557,934 interactions scanned from 34,996 publications in the primary literature as on 1st October, 2012.

STRING (Functional Protein Association Network): It is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations. The data is drawn from genomic context, high-throughput experiments, co expression and previous knowledge. This database makes an attempt to integrate interaction data from these sources for a large number of organisms and transfers information between organisms. It has a collection of 5214234 proteins of 1133 organisms.

KEGG: It was initiated in 1995. This database integrates genomic, chemical and systemic functional information. It provides a reference knowledge base for linking genomes to life through the process of PATHWAY mapping to infer systemic behaviours of the cell or the organism. It also links genomes to the environment via BRITE mapping. KEGG BRITE, an ontology database showcase functional hierarchies of various biological objects, including molecules, cells, organisms, diseases and drugs and also relationships among them. KEGG PATHWAY presently has a combined map of about 120 existing pathway maps. Smaller pathway modules are stored in KEGG MODULE. KEGG DRUG contains information about all approved drugs in the US and Japan, and KEGG DISEASE provide a link in to disease genes, pathways, drugs and diagnostic markers.

Micro array databases: Microarray is a hybridization of a nucleic acid sample (target) to a very large set of oligonucleotide probes, which are attached to a solid support, to determine sequence or to detect variations in a gene sequence or expression or for gene mapping. Various organisations have created microarray databases like National Center for Biotechnology Information, US-Gene expression Omnibus, European Bioinformatics Institute -Array express and National Institute of Genetics, Japan- Center for information biology gene expression. The purpose of creation of these databases is to store the minimum information about microarray experiments and to allow the researchers to repeat the experiments.

Applications of Bioinformatics (BI):

- is useful for gene and protein prediction, evolutionary distance calculation, active site identification, construction of novel mutations and characterisation of alleles of diseases.
- is applied in the analysis of the organisation of genes and genomes.
- may be helpful in identification of regulatory elements in genes.
- focuses on the development of algorithms to assess relationships among members of large data sets.
- allows the study of evolutionary events like gene duplication, lateral gene transfer.
- is useful in conservation of genomes of endangered species and biodiversity management.
- facilitate matching of data generated by mass spectrometers to the protein sequence databases.
- is helpful in the automation, processing, quantification and analysis of large amount of data from biomedical images.
- provide first hand information on chemical structure, reaction kinetics and synthetic methods, toxic chemical substances in the form of databases.
- can be used to study invention, implementation of structures and algorithms to improve communication, understanding and management of medical information.
- may be used to develop models for simulating intracellular molecular processes to predict dynamic behaviour of living cells.
- tools are used for prediction of function of genes and 3D structure prediction of proteins, selection of suitable ligands, identification of protein motifs and domains.
- is useful to maintain the sequence data and allow analysis of data by making availability of tools.

Challenges: The key challenges to bioinformatics involve the management of current flood of raw data, aggregate information, the diversity of the sources and formats, variable conditions and descriptions of experimentation, the quality of experimental evidence for data, data processing and evolving knowledge arising from the study of the genome and its manifestations.

1.4 PHARMACOGENOMICS

Pharmacogenomics may be defined as the genome wide analysis of genetic determinants of drug efficacy and toxicity. This branch of science began as pharmacogenetics when Vogel gave its name in 1959. The main goal of pharmacogenomics is evaluating the role of genetic variants in drug metabolism and effect, developing innovative ways of minimizing harmful drug effects and optimising care for individual patients. Various tools like gene mapping, gene sequencing, statistical genetics and gene expression are used to derive information.

Various factors influence the drug response like age, genetics, sex, race/ethnicity, disease state, absorption, distribution, metabolism, excretion, body weight, height, receptor sensitivity, organ dysfunction, accompanying medications, smoking, diet, alcohol, stress, pollution, socio-economic status, drug adherence, physiological changes including pregnancy, lactation, unmeasured nucleotide or structural variation, complex methylation/epigenetic mechanisms and gene-environment interactions. Around 30 to 60% of medication response rate for treating many diseases like depression, schizophrenia, rheumatoid arthritis has been observed. It has been reported that genetic factors can account 20 to 95% variation in drug response. Patients are classified into poor, normal and rapid metabolizers depending on the response to the drug intake. When a standard dose is given to the poor metabolizer, the drug is metabolized slowly resulting in an increased risk of toxicity and for the ultra metabolizer, the standard dose may be ineffective.

Genetic variations related to drug response can be classified in to three types namely pharmacokinetic, pharmacodynamic and idiosyncratic, based on their mechanism of action. Variations involving pharmacokinetics are associated with drug transporters and metabolizing enzymes and lead to alterations in the uptake, distribution and elimination of drugs. The Pharmacodynamic type of variations occur in the drug target or a component of the target pathway leading to altered drug efficacy. The target of pharmacodynamics includes receptors, ion channels, enzymes, transducer and regulatory proteins and immune molecules. A third type of variations known as idiosyncratic involved in unintended actions of a drug outside its therapeutic indication.

Inter individual variations in drug response have been reported based on genetic variations in drug metabolizing enzymes, receptors, transporters and pathways. The drug metabolizing enzymes are heterogeneous group of proteins involved in metabolism of drugs. They are two groups namely oxidative drug metabolizing enzymes and conjugative drug metabolizing enzymes. The oxidative drug metabolizing enzymes include cytochrome P450 and Flavin mono oxygenase; these catalyze the introduction of an oxygen atom into substrate molecules leading to hydroxylation or demethylation. The conjugative enzymes catalyze the coupling of endogenous small molecules to xenobiotics that results in the formation of soluble compounds that are more readily excreted. This enzyme family has UDP glycosyltransferases (UGTs), glutathione transferases, sulfotransferases and N acetyltransferases as members.

Classical examples of variation in drug response: Though many examples are available, few are given below.

Warfarin: It is an anticoagulant used to prevent stroke and venous thromboembolism. Its use is limited by narrow therapeutic window, variability in dose–response, interactions with drugs and diet and risk of serious bleeding. The dose is prescribed based on anticoagulation response measured by laboratory assay like internationalised normalised ratio. Clinical factors account for 17-21% of variation and genetic polymorphism in genes such as Cytochrome P450(CYP)2C9 and vitamin K epoxide reductase complex subunit 1(VKORC1) are responsible for 30-35% variation in warfarin dosing. Over coagulation and risk of bleeding was observed in carriers of at least one or more variant alleles of the CYP2C9 genotype. Increase risk of adverse cardiac events were observed in those who possess the variant VKORC1. A 42.6% benefit of warfarin treatment was observed after genotype guided drug regimen.

Clopidogrel: In cardiac patients, who are undergoing percutaneous coronary interventions (PTCA) clopidogrel is a standard medication. Clopidogrel is a prodrug that requires metabolic activation in a reaction catalyzed by cytochrome P450 enzyme CYP2C19 into its active metabolite. It has been reported that around 25% of patients experience subtherapeutic antiplatelet response. A lower capacity to metabolize clopidogrel into its active metabolite and inhibit platelet activation and higher risk of adverse cardiovascular events were observed variant allele carriers when compared to wild type allele of CYP2C19. Another mediator of Clopidogrel platelet effect has been reported and it was Paraoxanase 1(PON1). It was found to drive the conversion of the drug into the active metabolite. A polymorphism in PON1 (PON1Q192R) was observed to affect the platelet response, clopidogrel pharmacokinetics and the risk for thrombosis. Limited platelet inhibition and decreased plasma levels of both active PON1 and clopidogrel metabolites were observed homozygous individuals of PON1QQ192.

Flaxacillin: It is an antibiotic used for the treatment of staphylococcal infections. The usage of this drug has been associated with cholestatic hepatitis in approximately 8.5 cases per 100,000 patients. A single nucleotide polymorphism in HLA-B*5701 has shown strong association with hepatic injury.

Abacavir: It is a nucleoside analogue used to treat patients with HIV type 1 infection. Within six weeks of treatment, approximately 5% of patients developed a hypersensitivity reaction involving multisystem with symptoms of fever, rash and gastrointestinal discomforts which subsided within 72 hours of discontinuation of the drug. Approximately 74% carriers of haplotype HLA-B*5701 showed hypersensitivity when administered with abacavir.

Ribavirin: Chronic hepatitis C is a liver disease characterised by hepatitis C infection. The patients with this infection are treated with pegylated interferon and ribavirin. Approximately 50% of patients depending on ethnic origin show positive response to this treatment and become virus free. Pharmacogenomic studies have shown that CC genotype at the SNP rs12979860, 3kb upstream of the IL28B gene is associated with response to pegylated interferon and ribavirin for patients with chronic genotype 1 infection and natural clearance and the presence of G allele at rs8099917 is associated with non response.

Approaches of Pharmacogenomics

Candidate gene approach: Candidate gene use experimentally derived a priori knowledge about a disease or a drug involving both public and proprietary

databases for identifying candidate genes whose expression may impact drug action or disease pathogenesis. In this approach, genes are identified based on metabolic pathways, molecular targets, biological response pathways and/or disease risk. Based on the perceived likelihood of involvement of drug response the genes are ranked. Though this approach is tested in unrelated subjects, family studies, but population based studies are commonly employed as they detect relative risk as low as 1.5. The main focus of this approach is finding whether there are differences between the case (non responder) and the control (responder) in genetic variation that is assumed to be functional and involved in the observed phenotypic variability. One of the successful examples of this approach is the identification of individual response to the drug 6-mercaptopurine. The drawbacks of this approach is that it fails to consider a potential contribution of other genes particularly those whose function is yet to be understood; it is always not possible to have a information on the functionality of genetic variability or may be unreliable; it relies on the variability of specific point in the whole gene sequence.

Genome wide approach: This approach does not require a prior knowledge of the target gene. It attempts to identify the association between genetic variants and a given disorder by directing marker SNPs and analysing differences between case and control groups. This approach has been successful in mapping rare highly penetrant diseases in family pedigrees and identifying genes for monogenic traits. The strategies in this approach are based on the Linkage Disequilibrium relationships and structuring of haplotype blocks in the genome. It requires thousands of single nucleotide polymorphism (SNP) as genetic markers and it has been estimated that a minimum of 300,000 to 500,000 evenly spaced SNPs needed to find a marker within the range of disequilibrium. One of the successful examples of genome wide approach is identification of genetic variant in the drug transporter gene *SLCO1b1* responsible for statin induced myopathy. The success of genome wide approach depends on study design, sample size, quality control of genotype, collection bias, individual sample data, and ability of high throughput technologies to produce volume of data and ethnicity details. Validating data in an independent cohort in unbiased approach is requisite. The reported odds ratios are relatively small. The practical utility of information generated by using this approach remains controversial. Whole genome studies involve usage of microarrays which is out of reach of routine clinical practice. Complexity in analysis and requirement of additional guarantees in diagnosis make this approach far from applicable. Matching control groups to factors such as underlying disease and ancestry, contributions of genetic variants not detected by current platforms, analysis of gene-gene and gene-environment interactions in determining phenotype, affordable sequencing, storing of sequence data, information management and methods of genome analysis are the challenging issues in this approach.

Applications of Pharmacogenomics

- 1) Pharmacogenomic studies are helpful in developing therapeutic agents suitable for genetically identifiable human sub group populations.
- 2) Pharmacogenomics research can decrease the time and number of subjects needed for clinical trials.
- 3) Pharmacogenomics may facilitate the identification of biomarkers to optimize drug selection, dose and treatment duration and avert adverse drug reactions.

- 4) Pharmacogenomics may be helpful in reducing national health care bills in developing countries by taking into consideration of genomic variations between populations.
- 5) Study of the genotypes of populations with little admixture may be helpful in predicting drug responses without testing each individual.
- 6) Pharmacogenomics may help to improve our understanding of the mechanisms underlying variability in human physiology and its response to drug therapy with a final goal of improving therapy.
- 7) Information generated from pharmacogenomics studies may help health professionals and patients to make informed decisions about treatment options.
- 8) Generation of data on new and existing drugs will help in effective utilisation of scarce resources.
- 9) Pharmacogenomics may be helpful in reducing the costs associated with inappropriate drug treatments or hospitalisations due to serious adverse reactions.
- 10) Pharmacogenomics testing may produce collateral information which may be medically beneficial for ex. polymorphism in dopamine receptor though pharmacogenomics information may help in smoking cessation.
- 11) Long term applications of pharmacogenomics include reducing the burden of disease, improving the economic efficiency of the health care system and reducing some disparities in health care access and health outcomes.
- 12) Pharmacogenomics may be helpful in offering alternatives to the traditional drug development.

Challenges: The challenges of pharmacogenomics include establishing the clinical utility in order to support the value of genotyping; unaffordability of technology for wider application of pharmacogenomics outside the research and development setting; unequal treatment or health disparities due to social and the consequent ethical and legal issues connected to pharmacogenomics testing; measurement challenges such as presence of multiple pathways involved in drug effects, multiple polymorphisms, gene-environment interactions, length of time between testing and clinical outcomes and multiple determinants of clinical outcomes; time and cost intensity; developing technology to find out specific SNP; finding patients who fit into the criteria and possible users of the drugs; defining cut-off within adverse drug response distributions; lack of reproducibility of some gene-drug pairs and the questionable utility of the findings in a large population.

1.5 STEM CELL RESEARCH

Stem cells are precursor undifferentiated cells that are characterised by self renewal and differentiation. The most common definition is based on the properties of hematopoietic stem cells such as multipotency, asymmetric divisions, quiescence, life- long self renewal, niche dependence and long term repopulation ability upon in vivo transplantation. All stem cells may not follow these properties like embryonic stem cells not self renewing in vivo beyond the blastocyst stage;

muscle cells are not multipotent; and mesenchymal stem cells do not transplant robustly. Based on context and organism, stem cell properties of monopotency, transient proliferation, lack of niche and inability to transplant *in vivo* are acceptable. Some stem cells like mesenchymal stem cells have the potential of generating cells not only of their lineage but also other lineages. This phenomenon is known as plasticity or transdifferentiation. Homeostasis is the central mechanism in mammalian cells through which stem cells maintain and preserve organ and tissue integrity by self renewal and multilineage differentiation.

Stem cell populations are heterogeneous. Though stem cells express specific protein markers but accumulating evidence suggests that these markers are transient and dynamic. Stem cells express a vast range of genes at the mRNA levels. Stem cells undergo asymmetric mitotic division and produce two identical daughter cells. Of the two cells, at least one cell retains the stem cell properties, while other differentiates. This process is governed by stem cell niche (environment) and the detachment from this niche would result in differentiation. Stem cells may also follow stochastic differentiation, wherein stem cells make a combination of asymmetric divisions and symmetric ones leading to the formation of either two stem cells (symmetric renewal) or two differentiated cells (symmetric differentiation). Stem cells in olfactory epithelium and muscle follow this kind of cell division.

There is no consensus on the uniformity in the classification of stem cells in the literature. Stem cells were classified as embryonic or adult/non-embryonic/somatic (postnatal to adult)/extra-embryonic stem cells (adult, cancer and induced pluripotent) or hematopoietic and non-hematopoietic (mesenchymal stem cells), based on their origin. In terms of developmental potential, stem cells were categorized as totipotent, multipotent and unipotent. Totipotent stem cells are able to differentiate into all types of embryo tissues in the trophoblast. These cells are found after the first cell divisions in zygote (ex. zygote). Stem cells are pluripotent because of their ability to differentiate into cells of the three germinal layers (ectoderm, mesoderm and endoderm) except trophoblast lineage. Embryonic stem cells belong to this category. Multipotent stem cells like hematopoietic can produce a limited range of differentiated cell lineages. Only one specific cell type is generated by unipotent stem cells such as muscle progenitors.

Types of stem cells

Embryonic stem cells

- 1) These cells are primitive cells and can self renew and differentiate into all cells from all three germ layers such as ectoderm, endoderm and mesoderm.
- 2) Derived from inner cell mass of blastocyst stage at approximately five days of development using an immunosurgical technique.
- 3) Sources are embryos created via in vitro fertilisation/somatic cell nuclear transfer and fetuses obtained through elective abortion.
- 4) These cells do not conform to several cell requirements such as niche dependence, capacity to undergo asymmetrical cell division.
- 5) Mouse not human embryonic stem cells require leukemia inhibitor factor for their propagation.

- 6) These cells are grown on feeder layers.
- 7) These cells can be maintained in undifferentiated state for at least 80 passages.
- 8) They can form embryoid bodies, the cell aggregations containing all three embryonic germ layers.
- 9) These cells have the potential of forming teratomas *in vivo*.
- 10) These cells are considered as the optimal stem cell source for regenerative medicine applications in view of their potential to form any tissue in the body.
- 11) Application of these cells are limited by ethical, political, biologic and regulatory hurdles.

Epiblast-stem cells (EpiSC)

- 1) Cells are derived from pre- gastrula embryo.
- 2) Express surface markers like mouse embryonic stem cells, Oct4, Nanog and Ssea-1, and form teratoma.
- 3) These cells do not require feeders or leukemia inhibitory factor (Lif) for culture but can be expanded in the presence of Activin/nodal signaling.
- 4) These cells are unable to contribute to somatic cells and the germ line following injection into blastocyst or following morula aggregation.

XEN Stem cells

- 1) These cells are isolated from extra embryonic endoderm.
- 2) Express markers like Sox7, Hnf4, Gata4 and Foxa2 but lack expression of Oct4 and Nanog.
- 3) In a chimera assay, these cells contributed to the parietal endoderm and to the parietal yolk sac at later stages during embryo development.

XEN P Stem cells

- 1) Express genes Oct4, Gata 6 and Ssea-1 and isolated from rat blastocysts.
- 2) These cells require leukemia inhibitory factor (Lif) for exvivo maintenance.
- 3) Upon morula aggregation or injection in the blastocyst, these cells contribute to primitive/visceral and parietal extra embryonic endodermal lineages but not the embryo proper.
- 4) They form tumors when injected postnatal.

Hematopoietic stem cells (HSCs)

- 1) These are best characterised stem cells.
- 2) The size of the total pool of HSCs remains roughly the same in the absence of injury, about half of all HSC divisions must, at the population level, be self-renewing.
- 3) In the steady state, HSCs redistribute via the bloodstream among distinct anatomical locations and therefore are likely to be found in all tissues of the body.

- 4) These cells are multipotent and can differentiate into all myeloid and lymphoid blood lineages.
- 5) Through fusion HSC contribute to other tissues.
- 6) Sources of these cells are bone marrow, peripheral blood and umbilical cord blood.
- 7) These stem cell transplantations are performed using HLA matched siblings, parents or donors.
- 8) These cells represent less than 0.05% of the total bone marrow.
- 9) Enriched using a complement of cell surface antigens.

Mesenchymal Stem cells

- 1) These are adherent cells with fibroblast like morphology and are capable of self replication through many passages.
- 2) These cells are pluripotent and are capable to differentiate into multiple types of tissue like bone, cartilage, muscle, neuron, cardiomyocyte and hepatocytes.
- 3) These cells have proangiogenic and immunomodulatory effects.
- 4) These cells have the potential utility for treating a variety of diseases and disorders like graft versus host disease, organ transplantation, cardiovascular disease, brain and spinal cord injury, lung, liver and kidney diseases and skeletal injuries.
- 5) They have conserved long telomere lengths and do not form teratomas *in vivo*.
- 6) These cells are isolated from a number of tissues like bone marrow, adipose tissue, umbilical cord blood, placenta, amniotic fluid, amniotic membrane, gingival, circulating blood, synovium, trabecular bone, dermis, dental pulp and lung and have the capacity of expansion in vitro on a clinical scale.
- 7) These cells participate in maintaining essential environment to support the hematopoietic stem cells in the bone marrow.

Pancreatic stem cells (PSC)

- 1) They provide an alternative renewable source of surrogate β cells.
- 2) They have the potential for lineage-restricted differentiation and are capable of developing into a pancreatic phenotype.
- 3) The adult pancreatic stem or progenitor cells are found in duct cells, exocrine tissue, nestin positive islet-derived progenitor cells, neurogenin-3-positive cells, pancreas-derived multipotent precursors and mature β cells.
- 4) In the present state of knowledge, the expansion potential of PSC is limited.
- 5) Pdx-1, nestin and Ngn-3 markers have been shown to be expressed by these cells.

Cancer stem cells (CSCs)

- 1) These are small population of cancer cells that have the ability of unlimited growth, self renewal, as well as differentiation into more specialised cancer cell types.
- 2) CSCs have been identified and isolated in various hematological as well as solid malignancies.
- 3) CSC may form new tumour tissue when transferred into immunodeficient animal models and have been shown to survive and regenerate tumours tissue even after large percentage of tissues has been destroyed by chemotherapy.
- 4) Standard pathways for self-renewal of normal stem cells, such as Wnt, Notch and Hedgehog signaling, are also present in CSCs and have an important role in their function.
- 5) These cells produce higher levels of proangiogenic factors than their differentiated counterparts, and exhibit more potent proangiogenic capability.

Applications

- 1) Stem cells provide an opportunity to study the growth and differentiation of cells into tissues.
- 2) Stem cells can be used to produce large amounts of one cell type.
- 3) These cells can be used to test new drugs for effectiveness and chemicals for toxicity.
- 4) The damaging side effects of medical treatments might be repaired with stem cell treatment.
- 5) Somatic cell nucleus transfer technique stem cells created by using patient cell would avoid any tissue rejection problems that could be encountered in other stem cell therapeutic approaches.
- 6) In view of their migratory properties these cells can be used to target organs ex. tumours.
- 7) These cells can be used as vehicles to carry therapeutic molecule which they excrete spontaneously.
- 8) Stem cell therapies, in future, may circumvent the traditional use of chemicals as therapeutic drugs.
- 9) Stem cell transplantations are used to treat or greatly ameliorate a variety of genetic diseases ranging from inherent defects of hematopoietic cell production (thalassaemia) or function to metabolic diseases (lysosomal storage diseases) mostly affecting solid organs.
- 10) In acute myeloid leukemia and high grade lymphoma, hematopoietic cell therapy is used as adjuvant therapy.
- 11) Stem cells therapies are promising in the management of variety of disease conditions like cardiovascular diseases, neurological diseases (parkinson's disease, Amyotrophic lateral sclerosis, Huntington's diseases or Alzheimer's disease, Duchenne muscular dystrophy), diabetes, eye diseases and bone diseases.

- 12) Mesenchymal stem cells have been shown to be helpful in rapid engraftment of allogenic bone marrow transplantations. These cells were found inhibiting T cell growth, reducing graft versus host disease and effective in a large number of steroid resistant patients.
- 13) Early wound healing, delayed progression in human multiple system atrophy and beneficial effects in patients with hemorrhagic cystitis, penumediastinum and perforated colon due to the transplantation of mesenchymal stem cells.
- 14) In experimental studies entire organs were generated from stem cells and this raises the hope of using stemcells/progenitors for tissue engineering applications to generate organs for therapeutic purposes.

Challenges: Overcoming the immunological barriers, understanding of tissue restrictive signals, betterment of methodologies for stem cell isolation, in vitro propagation and transplantation either allo or xeno conditions by awakening resident stem cells, lineage tracing and long term real time follow-up of single cells including the real time assessments of gene and protein expression may go a long way in better utilisation of stem cell therapies for improving the quality of life in people afflicted with various disease conditions.

1.6 SUMMARY

The Proteome can be defined in terms of the sequence, structure, abundance, localisation, modification, interaction and biochemical function of its components. Proteomics has subfields like sequence and structural proteomics, expression proteomics, interactional and functional proteomics. Various tools are used in proteomics like databases, mass spectrometry, software tools and protein separation technologies.

Bioinformatics is defined as the computational handling and processing of genetic information. The goal of this field is to enable the discovery of new biological insight as well as to create a global perspective from which unifying principals can be perceived. Various tools like multiple sequence alignment, BLAST, FASTA, visualisation tools like RasMol, CnD3 are used in this subfield. Biological databases are helpful to gain a insight into biological phenomena from the structure of biomolecules and their interactions to the whole metabolism of organism and to understand the evolution of species. Databases are classified as primary databases (DDBJ, EMBL, Gene Bank), protein sequence databases (SWISS-PROT, Protein Information Resource), protein sequence databases (Pfam, PROSITE) protein structure database (PDB, SCOP), protein-protein interaction databases (BioGRID, STRING), pathway databases (KEGG), microarray databases (Array express, Gene expression omnibus).

Pharmacogenomics may be defined as the genome wide analysis of genetic determinants of drug efficacy and toxicity. Besides, various factors that influence the drug response like age, sex, race/ethnicity, disease state, genetics also play an important role in drug response. Genetic variations related to drug response can be classified in to three types namely pharmacokinetic, pharmacodynamic and idiosyncratic, based on their mechanism of action. Candidate gene, genome wide and risk of disease approach are used for the study of pharmacogenomics. Randomized clinical trials, case-control and cohort designs are used in this

discipline. Lymphoblastoid cell lines, NC160 cell line panel, various blood cells as in vitro models and animal models in pharmacogenomics.

Stem cells are precursor undifferentiated cells that are characterised by self renewal and differentiation. Stem cell populations are heterogeneous in nature. Stem cells were classified as embryonic or adult/non-embryonic/somatic (postnatal to adult)/extra- embryonic stem cells (adult, cancer and induced pluripotent) or hematopoietic and non-hematopoietic (mesenchymal stem cells), based on their origin. In terms of developmental potential, stem cells were categorized as totipotent, multipotent and unipotent. Applications of stem cells vary from a invitro model to therapeutic tool.

Suggested Reading

Dov Zipori (ed). 2009. Biology of Stem cells and the Molecular basis of the stem state. New York: Humana Press.

Kalow W. 2005. Pharmacogenomics: Historical perspective and status. In: Pharmacogenomics, Methods in Molecular Biology Innocenti, F (Ed). Totowa NJ: Humana Press Inc. Vol 311.

Krishna Rao, A and Raghu KA. 2011. Stem cells and regenerative medicine. From molecular embryology to tissue engineering. New York: Humana Press.

Liebler, D.C. 2002. U.K. Introduction to Proteomics: Tools for the New Biology. New Jersey: Humana Press, Totowa.

Mount, DW. 2004. Bioinformatics. Sequence and Genome Analysis. 2nd edition, New Delhi: CBS publishers and distributors.

Ouzounis, CA, Valencia, A. 2003. Early bioinformatics: the birth of a discipline-a personal view. Bioinformatics, 17: 2176-2190.

Richard JS, 2003. Proteins and proteomics: A Laboratory manual. New York: cold spring harbor laboratory press.

Searls, DB. 2010. The roots of bioinformatics. PLOS Computational Biology. 6(6):e1000809.

Trifonov, EN. 2000. Earliest pages of bioinformatics. Bioinformatics, 16:5-9.

Twyman, R.M. 2004. Principles of Proteomics. Oxon: Garland Science/Bios Scientific Publishers.

Sample Questions

- 1) What is bioinformatics? Mention in brief the important sequence analysis tools.
- 2) What are the different subfields of proteomics?
- 3) Give a brief note on genome wide approach.
- 4) What are the different types of stem cells?
- 5) What are the tools employed in proteomic studies?
- 6) What is Pharmacogenomics? Discuss with examples.

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.

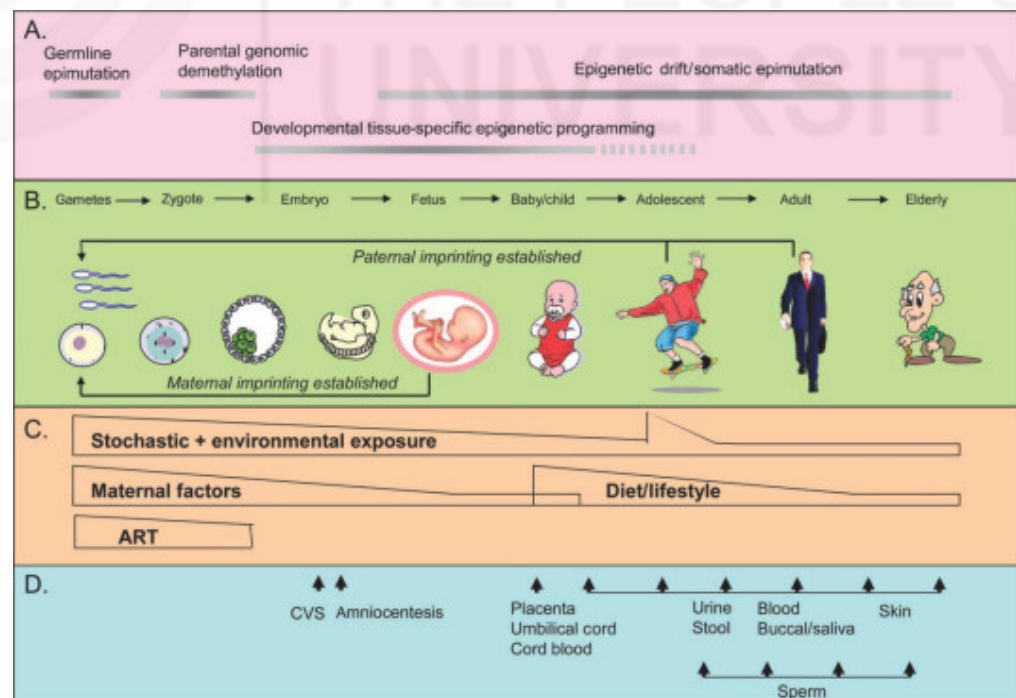


Fig. 2.1: Epigenetics, environment, and development. A. Dynamic epigenetic profile throughout the life course, with both genome-wide remodeling and tissue-specific programming events.

(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

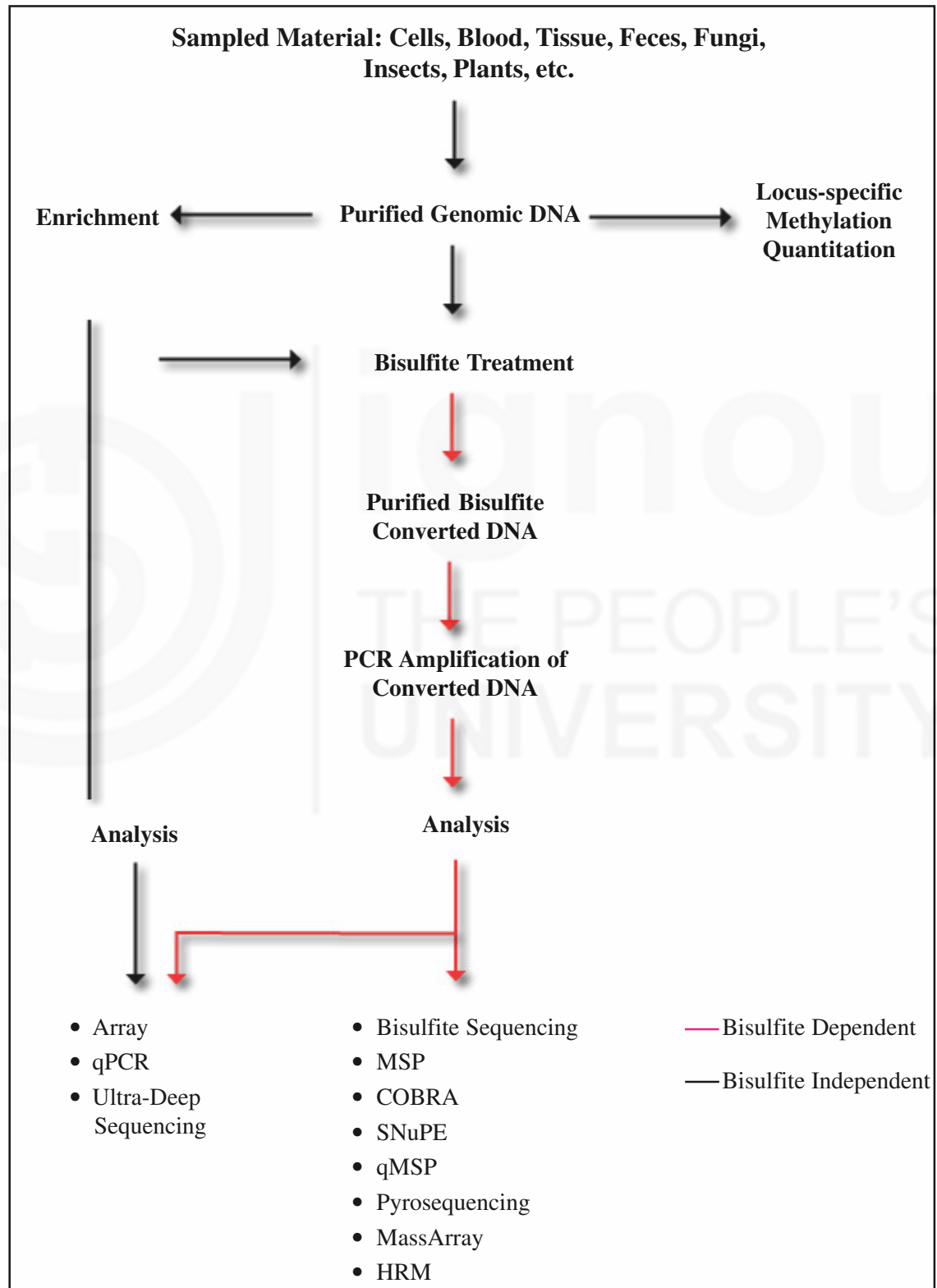


Fig. 2.2: DNA Methylation Analysis Flowchart

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.

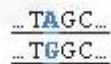
To find the genetic factors involved in a particular disease, one can proceed as follows. First a certain region of interest in the genome is identified, possibly from earlier inheritance studies. In this region one then locates a set of tag SNPs

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..C..A..T..G..T..
..A..C..C..G..C..T..
..G..T..C..G..G..A..

```

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

Andreas Ziegler , Inke R. König. 2010. A Statistical Approach to Genetic Epidemiology. Concepts and Applications.2nd Edition.,Weinheim: Wiley-VCH Verlag GmbH and Co.KGaA.

Dawn Teare (ed). 2011. Genetic epidemiology. Methods in molecular biology.Vol.713, Newyork: Springer.

Francis CR. 2011. Epigenetics: The Ultimate mystery of Inheritance. Newyork: W.W.Norton & Company.

Francis CR. Epigenetics: How environment shapes our genes.W.W.Norton and Company, 2012.

Lyle J. Palmer, Paul Burton and George Davey Smith (Editors). 2011. An Introduction to Genetic Epidemiology (Health & Society Series). First edition. Bristol, U.K: The Policy Press.

Morton, N and Chin Sik Chung. Genetic Epidemiology. London: Academic press.

Trygve Tollefsbol. (eds). 2010. Handbook of epigenetics: The new molecular and medical genetics. London:Academic Press.

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.

UNIT 3 PREVENTION OF GENETIC DISEASES

Contents

- 3.1 Introduction
- 3.2 Genetic Screening
- 3.3 Genetic Counselling
- 3.4 Prenatal Diagnosis
- 3.5 Gene Therapy
- 3.6 Summary
- Further Reading
- Sample Questions

Learning Objectives



It is expected that after reading this unit, you would be able to:

- explain what is genetic disease and types of genetic diseases;
- discuss the processes of genetic screening and genetic counselling;
- describe on prenatal diagnosis; and
- explain gene therapy and its importance.

3.1 INTRODUCTION

Genetic diseases are those pathological conditions that are caused or influenced by abnormalities in the genetic material (genome) of a person and these are caused by genemutations and chromosomal abnormalities. Genetic diseases place a substantial health, emotional, and economic load on affected people, their families and on the society. There are a number of diverse types of genetic disorders. Geneticists group genetic disorders into four categories: single gene/monogenic genetic diseases, multifactorial/polygenic genetic diseases, chromosomal diseases and mitochondrial DNA disorders.

Single Gene Genetic diseases: These are also called as monogenic genetic diseases or Mendelian genetic disorders. These type of disorders are caused by changes or mutations that occur in the DNA sequence of a single gene. Single gene disorders are passed on between generations from parents to offspring. Four patterns of inheritance occur: autosomal dominant, autosomal recessive, X-linked dominant and X-linked recessive. Some of the diseases of single gene disorders are sickle cell anaemia, cystic fibrosis, Aicardi Syndrome, Huntington's disease etc.

Polygenic genetic diseases: These diseases are also called multifactorial genetic disorders. These diseases are influenced by the interaction of multiple genes and interaction between genes and environmental factors. Heart disease, diabetes, hypertension, Alzheimer's disease are some of the examples of polygenic diseases.

Chromosomal diseases: These diseases are caused by the loss or gain of one or more chromosomes or by alterations in chromosome structure. Down syndrome, Turner syndrome and Klinefelter syndrome are some of the examples for chromosomal diseases.

Mitochondrial DNA disorders: involve the defects in mitochondrial DNA not nuclear DNA. Ex: Some rare neurological and skeletal disorders.

So far we have studied what are genetic disease and their types. Now let us discuss how to prevent them. The desire of all parents is to have normal and healthy children. Hence many parents think ways to prevent or lower the risk of passing the genetic diseases to their offspring by preventing the genetic diseases. To prevent genetic diseases, parents have to undergo certain preventive measures such as Genetic screening, Genetic counselling, Prenatal diagnosis and Gene therapy. A brief description of the above mentioned preventive measures is described below.

3.2 GENETIC SCREENING

Genetic screening is a routine diagnostic procedure devised to detect those who are carriers of, or who are themselves affected by, a hereditary disease. Genetic screening applies to populations rather than to individuals.

The most-widespread application of genetic screening in the United State is for phenylketonuria (PKU). All hospitals in the United States screen newborn babies for PKU by a blood test called Guthrie test.

After genetic screening if both the parents are heterozygous for a genetic disease and the genotypes of both the prospective parents become known and they can decide to produce a child or not. It is simple matter to work out the probability of their child inheriting the disease. This can be done through appropriate tests around 2-3 months after conception. This is achieved through amniocentesis or through chorionic villus biopsy; the cultured fetal cells may be used for determining their karyotype, levels of the critical enzymes and restriction patterns of DNA. Such an antenatal diagnosis is now available for several genetic diseases and for a variety of chromosomal defects. The purpose of such a diagnosis is premature termination of abnormal fetuses.

Genetic counselling and antenatal diagnosis provide definite relief to the prospective parents and reduce the frequency of genetically defective individuals in the population. However, it is unlikely that these measures would eliminate the deleterious alleles from a population. This is so because most genetic defects are recessive and heterozygotes for such alleles. Thus even after a total ban on reproduction by the homozygotes for such recessive alleles they would remain in the population through the heterozygotes, therefore even such an extreme selection would lead to only a slow decline in their frequency. Further, it is not likely that all the couples in any society will willingly submit themselves, at least in the foreseeable near future, to these procedures. But genetic counselling has become a routine aspect of medical practice in most developed countries.

It has been advocated that defective genes may be corrected through sophisticated genetic techniques either during the early stages of embryo development (embryo

therapy) or in specific tissues of the adult patient (patient therapy); such an approach is referred to as genetic surgery. Embryo therapy would involve:

- 1) In vitro fertilization of egg.
- 2) Production of several copies of the normal allele of the defective gene.
- 3) Introduction of this DNA into the zygote or in the cells of the developing embryo.
- 4) Integration of DNA, preferably in place of the defective allele, so that it may function normally.

The aim of patient therapy is to introduce the normal gene into the critical tissue of the patient that is affected by a genetic disease, i.e., the tissue where the concerned gene is required to express itself the most e.g., pancreas in the case of diabetes. The steps involved in patient therapy would be similar to those of embryo therapy. But in this case, cells from the concerned tissues may have to be exercised and treated in vitro to correct their genetic defects and then reintroduced into the tissue where they may function normally. Techniques for isolation, identification and multiplication of many human genes are now available, and for many others they are likely to be developed soon. The techniques for gene transfer in eukaryotes are being refined and it may not remain a great problem in the near future.

A suggestion has also been made to use highly specific chemical mutagens that will correct the defect in the concerned gene. Such a directed mutagenesis however is a dream that may be more difficult to fulfil the patient and embryo therapies through DNA mediated genetic modifications. Genetic screening and counselling may also lead to certain problems. The cases of mistaken paternity, the problem of confidentiality, delayed counselling are important among them.

3.3 GENETIC COUNSELLING

Genetic counselling is essentially a communication process that informs prospective parents about the nature of genetic disorders, about the risk of their having a genetically defective child, and about the options available to them in dealing with that risk. Or it can help them cope with the care of an existing genetically handicapped child. The advanced knowledge of the applied aspect of human genetics has gained importance. Therefore genetic counseling is an important area of applied human genetics. The counselor should have adequate knowledge in the counseling process. The counselor is preferably a medical doctor or a profession who is having knowledge in human genetics. The duty of a counselor is to offer the necessary genetic information and also the information about social, economic and psychological aspects related through the case. The persons receiving genetic counseling are called the consultands.

Genetic Counselling, as defined by Harper (1984), is “the process by which patients or relatives at risk of a disorder that may be hereditary are advised of the consequences of the disorder, the probability of developing and transmitting it, and the ways in which this may be prevented or ameliorated”. However the American Society of Human genetics (1975) formulated the definition as “Genetic counselling is a communication process which deals with the human problems associated with the risk of occurrence of a genetic disorder in a family”. This process involves an attempt by one or more appropriately trained persons to help

the individual or family to: (1) comprehend the medical facts including the diagnosis, probable course of the disorder, and the available management, (2) appreciate the way hereditary contributes to the disorder and the risk of recurrence in specified relatives, (3) understand the alternatives for dealing with the risk of recurrence, (4) choose a course of action which seems to them appropriate in their view of their risk, their family goals, and their ethical and religious standards and act in accordance with that decision, and (5) to make the best possible adjustment to the disorder in an affected family member and/or to the risk of recurrence of that disorder.

Genetic counselling involves the following steps:

The prospective parents either suffering from or suspected to be heterozygous for some genetic disease on the risk of their children suffering from the same disease should be educated. By creating a suitable social environment such parents may be encouraged to voluntarily abstain from producing children. To identify people suffering from a genetic disease is relatively easy for a trained clinician. It is difficult to identify a carrier for genetic disease and in most cases not possible. Information on the likelihood of an individual being a carrier for a genetic disease can be obtained by the analysis of family pedigree in some conditions or possible in case of certain conditions by biochemical and molecular tests.

The process of Genetic counselling

- 1) After genetic screening of a disorder, patients or their parents should be informed about the genetic or medical implications of the disease;
- 2) The likelihood of occurrence of such genetic disorder in the family should be calculated; and
- 3) To suggest ways in which the occurrence of the genetic disorder can be controlled.

Pedigree Analysis

Pedigree is a family history of hereditary condition or diagram of a family history indicating the family members, their relationships to proband (the affected individual that brings family to medical attention), and their status with respect to a particular hereditary condition.

Pedigree analysis assists the counsellor to decide if a trait is Mendelian. It also helps to establish the mode of inheritance. The first step in pedigree analysis is to observe the number and relationships of all individuals who express the same or similar clinical features. From this, it should be possible to determine if the disorder is dominant or recessive, autosomal or X-linked or Y-linked by looking for the typical patterns of inheritance. For example, a Y-linked disease can usually be distinguished by seeing male-to-male transmission of mutation, but since males pass only the Y chromosome to their sons, should never be father to son transmission of an X-linked gene. Males will be most commonly affected in an X-linked disease, whereas males and females should be equally affected in autosomal disorders. In general, a dominant disease will be seen in approximately half of the individuals in each generation, but recessives occur very rarely. If the mutation is in the mitochondrial genome, affected mothers will pass the trait to all of their children, but none of the offspring of an affected male should have the disease.

Once the inheritance pattern of the disorder is determined, the status of family members in the pedigree can be evaluated. By carefully observing the position of affected individuals, mutation carriers may be identified. From this data, the risk of carrier status for other family members or the chance that a couple may have an affected child can be estimated.

3.4 PRENATAL DIAGNOSIS

Prenatal diagnosis is the detection or exclusion of abnormality in the foetus prior to birth. In the process of having healthy progeny, prenatal diagnosis useful in detecting health of a baby to be born. It is useful to detect genetic disorders where the couple or populations are at high risk of inheriting genetic disorders. If an abnormality is detected in the pregnancy, the parents can take decision as to whether to continue with the abnormal foetus or to go for abortion. Prenatal diagnosis enables: (1) Timely medical or surgical treatment of a condition before or after birth, (2) parents to abort a foetus with the diagnosed condition, and (3) parents be prepared psychologically, socially, financially and medically for a baby with a health problem or disability or for likelihood of a still birth.

Prenatal diagnosis for genetic diseases is now commonly available for pregnancies at risk. Prior to prenatal diagnosis genetic counselling could only give likelihood of recurrence, based on Mendelian laws or empirical data. If an abnormality is detected in the pregnancy the medical people have to convey the information to parents and allow the parents to take a decision as to whether to continue with the abnormal foetus or to go for abortion. The couples at high risk of having a child with a genetic disorder has to choose between taking the risk or considering other reproductive options such as long term conception, sterilization and termination of pregnancies. Most parents would terminate an affected pregnancy but some choose to use the information to prepare for the birth of an affected child. When the foetus is found to be normal it allows the parents to continue the pregnancy with mental relief.

Many factors are to be considered before undertaking prenatal diagnosis, like risk of inheriting a genetic disorder, severity of genetic disease etc. Some genetic diseases lead to death in utero, or infancy or childhood. Some diseases are compatible with survival for many years with severe handicaps. Many genetic diseases cannot be cured even if the basis for the diseases is known. In some cases gene therapy may help. Currently prenatal diagnosis tests are available for more than 200 genetic disorders. Before going to a test one should be careful in finding the reliability of the test.

Indications for Prenatal Diagnosis

There are many indications for offering prenatal diagnosis. While undertaking prenatal diagnosis the risk of an abnormal foetus is at least as great as risk of the procedure itself. The following are some of the indications:

Advanced maternal age: Women who are pregnant at 35 or more have higher risk of giving birth to a baby with the defect in chromosomes. Down syndrome (trisomy 21) is the common chromosomal defect. A child with Down syndrome can be born to parents of any age; however, the risk rises steeply once the mother's age is over 35 years.

Previous child with a chromosome abnormality: Parents who have had a child with trisomy 21 (Down syndrome) or any other trisomy are at increased risk of having a trisomy in subsequent pregnancies.

Family history of a chromosome defect: A pregnant woman with a family history of chromosome defect are often advised for prenatal diagnosis. The most common abnormality is Down syndrome (Trisomy 21).

Family history of a neural tube defect: If a first-degree relative of the foetus (parent or sib) has a neural tube defect, the risk to the foetus is in the 2 to 5 percent range. If more remote relatives are affected, the risk of the foetus is less but may still be above the population risk.

Other indications for prenatal diagnosis include structural aberrations in a family or previous pregnancy, single gene genetic disorders, parental consanguinity, mental retardation etc.

Techniques of prenatal diagnosis: There are several techniques available for prenatal diagnosis. The following are some of the methods.

Amniocentesis: In this technique amniotic fluid would be aspirated by inserting a needle through the abdomen into the mother's amniotic cavity (Fig. 3.1). This procedure is to be performed under ultrasound guidance. Amniocentesis is usually carried out between 14-16 weeks of gestation and 10-20ml of amniotic fluid is aspirated. The cells have to be cultured for an average of 8-14 days to obtain a foetal karyotype. The risk of the method appears very small; however there is a small risk of inducing miscarriage. With this method Karyotyping and some biochemical disorders can be detected.

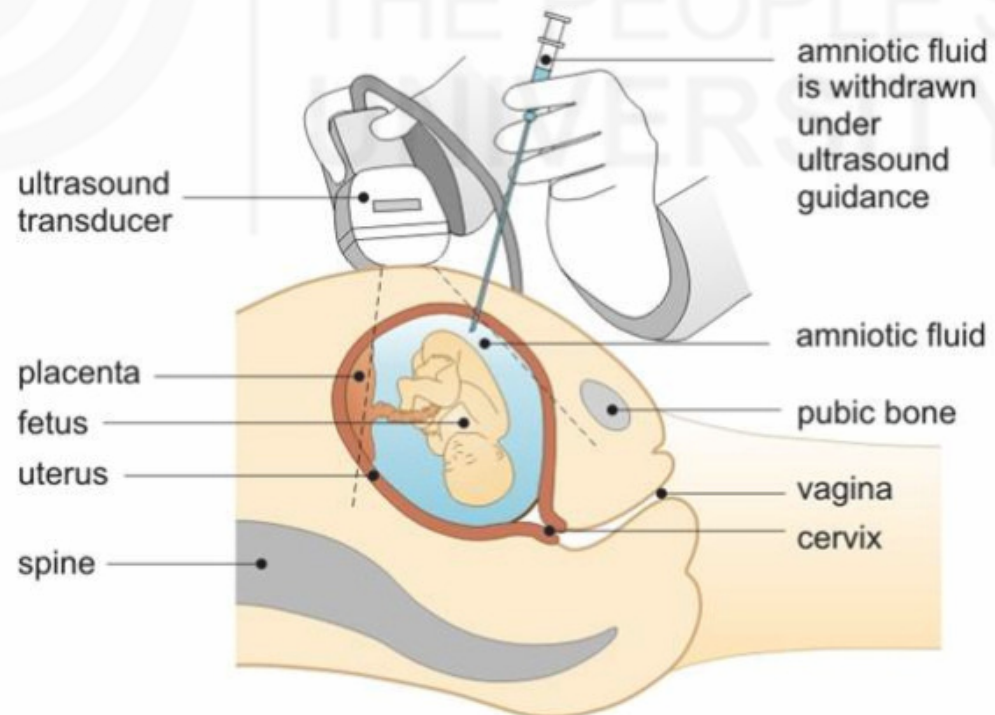


Fig. 3.1: Aspiration of amniotic fluid (source:gi.net.au)

Chorionic villus sampling: This technique is carried out between 9 and 12 weeks of gestation and involves the removal of small amount of placenta. This

technique is suitable for chromosome analysis following the determination of a high risk of chromosome abnormality. The advantage of this technique is early and definitive chromosomal analysis. This test bears a risk of pregnancy loss varying from 1 to 3 percent. In this technique karyotyping, DNA and biochemical disorders can be detected.

Foetal blood sampling: This technique is also called cordocentesis. It involves the removal of small amount of foetal blood by inserting a small needle into the site where the cord joins the placenta. In this technique rapid karyotyping can be done.

3.5 GENE THERAPY

Gene therapy can be defined as the insertion of genetic material into cells for the treatment of disease. Although gene therapy was envisaged initially as a means of treating inherited genetic diseases, disciplines as diverse as oncology, cardiology, endocrinology, and infectious diseases are currently developing novel therapeutic strategies that have in common gene transfer into cells.

Gene therapy naturally intends to supplement a defective mutant allele with a functional one. In most gene therapy studies, a normal gene is inserted into the genome to supplement an abnormal disease causing gene. To deliver the therapeutic gene to the patient's target cells, a carrier called a vector have to be used. Currently, the most common vector is a virus that has been genetically altered to carry normal human DNA. The vector unloads its genetic material containing the therapeutic human gene into the target cell. The generation of a functional protein product from the therapeutic gene restores the target cell to a normal state.

Types of Gene Therapy

Gene Therapy can be classified into two types: Somatic gene therapy and Germ line gene therapy.

Somatic gene therapy

In this gene therapy, the therapeutic genes are moved into the somatic cells or body of a patient. Due to somatic gene therapy the modifications that occur will be restricted to the individual patient only. But these modifications will not be inherited by the patient's offspring or later generations.

Germ line gene therapy

In this type, germ cells such as sperm or eggs are modified by the introduction of genes which are in function and these functional genes are included into their genomes. In this type the modifications would be heritable to and passed on to later generations. But this type is highly controversial for a variety of technical and ethical reasons.

The gene therapy can be categorised into two types: *in vivo* therapy and *ex vivo* therapy. In the case of *in vivo* therapy the therapeutic genes are directly introduced in to the patient's body (fig. 3.2) and in *ex vivo* category the cells are removed from the patient and after manipulation they are returned to the patient (fig. 3.3).

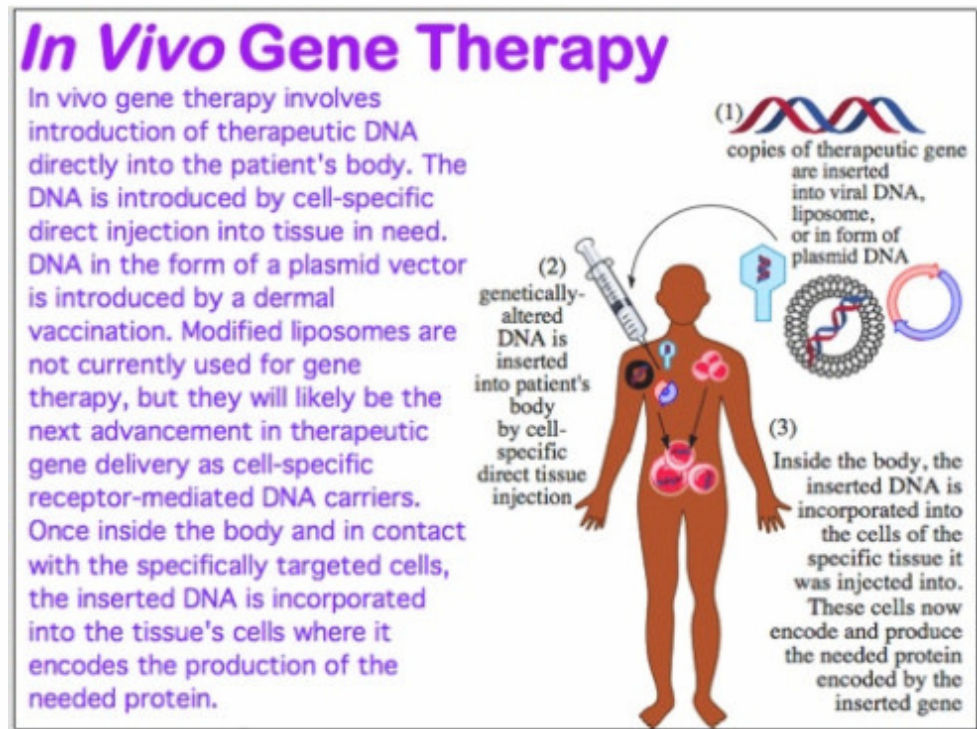


Fig. 3.2: Schematic representation of *in vivo* Gene therapy
(source: gene-therapy.yolasite.com)

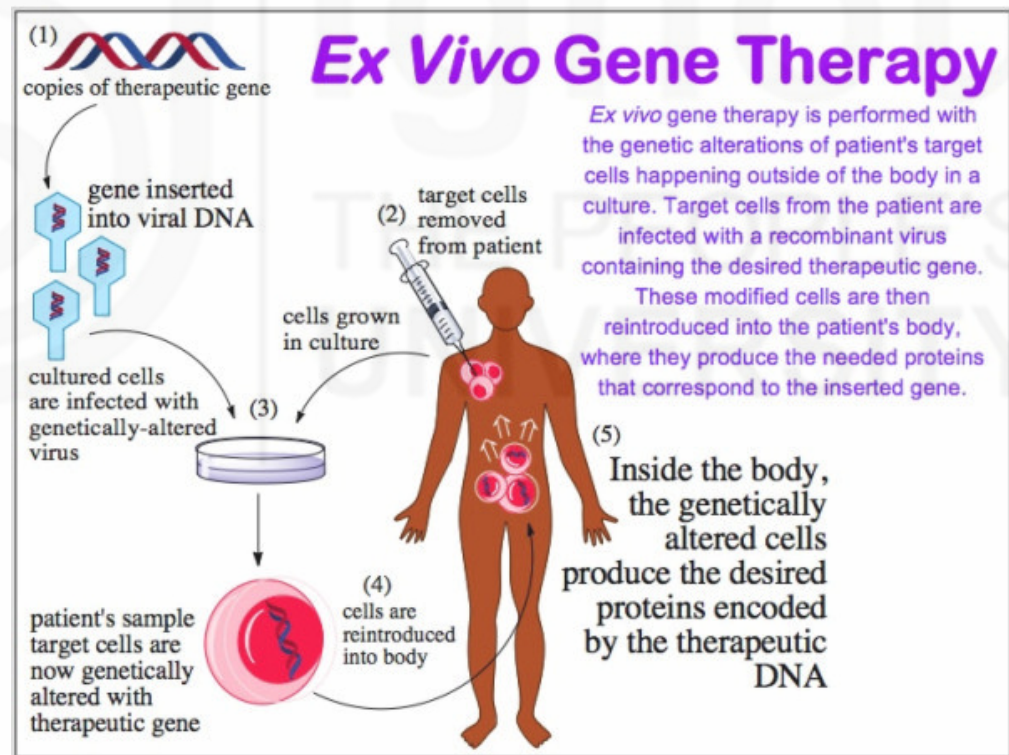


Fig. 3.3: Schematic representation of *ex vivo* Gene therapy (source: en.wikibooks.org)

3.6 SUMMARY

Genetic disease is caused by variation or alteration called mutation in gene, or alteration in chromosome. Genetic diseases can be classified into simple and complex diseases. Simple genetic diseases are also called as monogenetic or Mendelian diseases caused by a mutation. Complex diseases have multi-factorial causation. They are caused by the interaction of many genes and gene -

environmental interaction. Screening the genetic diseases in the population and genetic counselling to the couples is a prerequisite in avoiding the defective progeny. To prevent the genetic diseases, prenatal diagnosis and termination of defective zygote is one of the best ways. Gene therapy is the treating method for genetic diseases, it involves in correcting the default genetic sequence. Currently available method for preventing genetic diseases is to identify the faulty genetic code at an early stage and change the life style. By changing the life style we can prevent some late onset genetic diseases. In curing the genetic diseases future research may bring success in gene therapy.

Further Reading

Gardner, A, Howell, R.T and Davies, T. 2008. Human Genetics. New Delhi: Viva Books private Ltd.

Hartl, D.L. 1991. Basic Genetics. Boston: Jones and Bartlett Publishers.

Mueller, RF and Young, ID. 1995. Emery's Elements of medical genetics. New York and London: Churchill Livingstone.

Thompson, JS and Thompson, MW. 2005. Genetics in Medicine. Philadelphia and London: WB Saunders Company.

Uhlmann, WR, Schuette, JL and Yashar, BM. 2009. A guide to Genetic counseling. New Jersey: Wiley-Blackwell.

Verma, PS and Agarwal, VK. 1999. Cell Biology, Genetics, Molecular Biology, Evolution and Ecology. New Delhi: S. Chand & company Ltd.

Sample Questions

- 1) What is a genetic disease? Briefly explain on various genetic diseases.
- 2) Discuss briefly about Prenatal diagnosis.
- 3) What is genetic screening? Explain its processes.
- 4) Write short notes on the following:
 - a) Gene therapy
 - b) Genetic counselling

UNIT 4 GENETICS AND HUMAN ISSUES

Contents

4.1 Introduction

- 4.1.1 Human Genome Project
- 4.1.2 What is Ethics?
- 4.1.3 Ethical, Legal, Social Issues in Genetics

4.2 Genetic Testing and Research

- 4.2.1 General Concerns
- 4.2.2 Ethical Principles in Genetics Research
 - 4.2.2.1 Autonomy or Respect for Persons
 - 4.2.2.2 Beneficence and Non Maleficence
 - 4.2.2.3 Justice
- 4.2.3 Special Ethical Issues in Genetics Studies
 - 4.2.3.1 Informed Consent
 - 4.2.3.2 Vulnerable Participants
 - 4.2.3.3 Genetic Counselling
 - 4.2.3.4 Risks and Benefits
 - 4.2.3.5 Pedigree Studies
 - 4.2.3.6 Privacy and Confidentiality
 - 4.2.3.7 Inducement vs Compensation
 - 4.2.3.8 Genetic Screening
 - 4.2.3.9 DNA Testing
 - 4.2.3.10 Prenatal Testing
 - 4.2.3.11 Presymptomatic and Susceptibility Testing
 - 4.2.3.12 Gene Therapy
 - 4.2.3.13 DNA and Cell-line Banking/Repository
 - 4.2.3.14 Cloning
 - 4.2.3.15 Stem Cell Research and Therapy
- 4.2.4 Other Concerns in Genetics Research
 - 4.2.4.1 Commercial Issues
 - 4.2.4.2 Patenting
 - 4.2.4.3 Research in International Collaboration

4.3 Summary

Further Reading and References

Sample Questions

Learning Objectives



Once you have studied this unit, you should be able to:

- describe how ethical issues in human genetics have emerged and why it is important to take care of them;
- understand about the various ethical principles and their importance; and
- discuss the emerging ethical, legal, social concerns in genetic testing and research.

4.1 INTRODUCTION

Hereditary conditions affect millions of families throughout the world. About 5% of all pregnancies result in the birth of a child with a significant genetic disorder, congenital malformation or disability. The burden of genetic disorders and birth defects is yet to be recognised. It has been estimated that there is a high prevalence of genetic disorders in India with about 495,000 infants with congenital malformations, 390,000 with G6PD deficiency, 21,400 with Down syndrome, 9000 with Beta thalassaemia, 5,200 with sickle cell disease and 9760 with amino acid disorders born each year (Verma and Bijarnia, 2002). A high prevalence of inborn errors of metabolism of 1 in 1000 live births has been reported from India (Rama Devi and Naushad, 2004). It is estimated that 8000-10000 children are born with thalassaemia every year and there are about 20 million carriers of β -thalassaemia in the country (Mohanty et al, 2002). Most non-infectious diseases, may have a genetic component. Therefore, continued efforts are required to develop effective treatments and make them available in India for diseases that are important to the health of communities as well as individuals and families.

4.1.1 Human Genome Project (HGP)

The human genome project (HGP) was an international research effort, with the aim to sequence the entire human DNA and also determine the location of all identified genes which was carried out from the year 1990 to 2003. This project was coordinated by the U.S. Department of Energy and the National Institutes of Health. The project was initiated with a hope to develop new ways to diagnose, treat, and someday prevent the thousands of disorders that affect us.

Goals of Human Genome Project

identify all the approximately 20,000-25,000 genes in human DNA.

determine the sequences of the 3 billion chemical base pairs that make up human DNA.

store this information in databases.

improve tools for data analysis.

transfer related technologies to the private sector.

address the ethical, legal, and social issues (ELSI) that may arise from the project.

Upon completion of this project enormous amount of genetic information has been made available. All adults have a right to know their genetic makeup and also implications for the health of their potential to next generation. HGP has raised some concerns in regard to issues such as access to genetic information, diagnostic services, privacy and confidentiality, disclosure to family members, freedom of reproductive choices, misuse of genetic information, stigmatization/discrimination of individuals based on their genetic makeup etc. Therefore though there is potential for advances through better diagnostic and preventive methods however on the other hand this knowledge has increased concerns about genetics in the public. It is therefore important to improve public knowledge and create better understanding about genetics. To remove the fear from the minds of people as well as help to protect families with genetic disabilities, it is important to understand societal and the ethical issues. The following are some examples of what could be the implications of genetic knowledge to various people:

- individuals and families – whether to participate in testing, with whom to share the results, implications for individual and other family members;
- health professionals – when to offer genetic testing, obtaining informed consent, how to ensure quality, how to interpret results, counsel patients, whom to disclose information;
- employers, insurers, the courts, other social institutions – the value of genetic information to the decision they must make about individuals;
- governments – how to regulate the use of genetic tests and information they provide and how to provide access to testing and counselling services for society; and
- society – how to improve public understanding of science and its social implications and increase participation of the public in science policy making.

The scientific community should address the above mentioned questions before use of this knowledge could be considered as ethically valid. As the challenges are conquered we will begin to harness genomics for the benefit of our populations (Hardy, 2008, Seguin, 2008).

4.1.2 What is Ethics?

Ethics, as a field in philosophy or religion, is concerned with systematic reflection on the moral life and its conflicts. “Ethics” is a generic term for various ways of understanding and examining the moral life and for resolving ethical problems (Beauchamp and Childress, 1994). Biomedical ethics (or bioethics) is an interdisciplinary field for the systematic study of ethical issues that arise in research, medicine and society. Medical ethics is a system of moral principles that apply values and judgments to the practice of medicine or its practical application in clinical settings. To be ethically correct these principles should be based on the community that is served and in view of their religious, social or cultural value systems so as to adequately protect their rights, welfare and safety. Under the principles of ethics there is an obligation to confer benefits, to prevent and remove harms, and to weigh and balance the possible goods against possible harms of an action.

4.1.3 Ethical Legal and Social Issues (ELSI) in Genetics

As you know during the last 1-2 decades there has been explosion in knowledge in human genetics, and several new technologies have been developed for example gene therapy, genetic engineering, stem cell therapy, cloning, prenatal diagnosis, preimplantation diagnosis, creating designer babies, genetic screening etc. In addition new issues have also emerged, for example those relating to patenting of DNA sequences, Intellectual Property Rights (IPR) issues and potential for bio-terrorism. In this rapidly evolving field there is a need to continuously monitor developments and respond to emerging ethical issues promptly and judiciously since improper use of these can have far reaching implications for human race.

You may be surprised to know that in no other area of biomedical research there has been a greater concern for ethical issues than in the field of human genetics. With the successful completion of Human Genome sequencing in April, 2003, number of guidelines were laid identifying the ethical, legal and social issues and measures that may be taken for improving awareness and understanding of human genetic disorders.

The integration of genetic technologies into the public health raises significant challenges. How can we assure the quality and accuracy? What impact will these new technologies have on existing programs? Issues of autonomy, privacy, confidentiality will require special consideration. Will it be possible to obtain fully informed consent from large populations? How to protect privacy in genetic registries and databases? How can group stigmatization and discrimination be prevented? None of these questions have easy answers, but they raise important challenges for the community. These are the various ethical, legal and social issues that address the societal concerns emerging from the new genomic information and technologies.

4.2 GENETIC TESTING AND RESEARCH

4.2.1 General Concerns

It is important for you to note that in human genetics the concerns may not only be to the individual but also the family, community or society from which s/he has been drawn. Kindly note some of these concerns.

What are the Concerns?

- Genetic information is sensitive and private and needs adequate protection. In genetic research the harm may not only be physical, but also psychosocial which may produce anxiety and depression or damage familial relationship.
- There is a likelihood of social stigmatization and discrimination in schooling, employment, health and general insurance, which requires much great care.
- When involving participants in research it is important to obtain written informed consent and maintain confidentiality of their personal or other details or data collected from them.
- There is great importance of spoken word in medical genetics, since genetic counselling is akin to therapy in other fields. Which means that the 'word' is equivalent to drug/ intervention in other fields of medicine. Appropriate communication skills are necessary for genetic counselling.
- Genetic manipulations have consequences for the future, some of which are unknown and long term. Hence, greater care towards potential dangers is necessary.
- Any biomedical research conducted by student, or faculty on human participants should be reviewed by an independent committee known as the "Institutional Ethics Committees" or IEC to ensure protection of participants in research.

4.2.2 Ethical Principles in Genetics Research

There are three main ethical principles for all kinds of biomedical research including genetics research in human participants :

4.2.2.1 Autonomy or Respect for Persons

This principle refers to respecting the self-determination of individuals and protecting those persons with diminished autonomy. Informed consent is one

way of protecting this principle i.e., rights and voluntary decisions of individuals. For all biomedical research the informed consent of the individual, or legal guardian should be obtained. It protects the individual's freedom of choice to voluntarily agree/ disagree to participate in research or give samples for genetic testing. Adequate information about the testing is given in a simple and easily understandable language in a document known as the Informed Consent Form with Participant/ Patient Information Sheet.

4.2.2.2 Beneficence and Non Maleficence

In simple terms it just means doing good and protection from harm. Beneficence involves giving importance to the welfare of persons and providing maximum benefits. On the other hand Non-maleficence refers to preventing from harm or trying to reduce harm and is derived from the traditional medical norm of “do no harm”. The principle of beneficence involves giving highest loyalty to the welfare of people and families with the goal to improve the health of population.

4.2.2.3 Justice

This principle requires treating persons with fairness. It involves distributing benefits and burdens of health care as fairly as possible in society. Therefore persons should be treated fairly, giving them what they deserve, or are entitled to. Distributing the benefits and the burdens of health care ought to be governed by ethically justified rules such as: to each according to need, to each according to an equal share or opportunity, etc.

Table 4.1: Examples highlighting use of ethical principles in genetic testing

- Involve voluntary approach in testing and treatment; avoid coercion (force) by society, or health professionals (autonomy).
- Respect for human diversity and for those whose views are in the minority and respect for people's basic intelligence, regardless of their knowledge (autonomy).
- Education about genetics for the public, medical and other health professionals, teachers, and other persons (beneficence).
- Protecting privacy & confidentiality (autonomy)
- Prevention of unfair discrimination in employment, insurance, or schooling based on genetic information (non-maleficence).
- Teamwork with other professionals through a network of referrals to provide better tests, care and treatment (beneficence).
- Timely provision of indicated services or follow-up treatment (non-maleficence).
- Refraining from providing tests or procedures not medically required (non-maleficence).
- Providing good quality of services, including lab procedures (non-maleficence).
- Sharing results as well as benefits of research with communities (Beneficence, justice).
- Equitable selection of participants from community (justice).
- Protection of poor, vulnerable groups (justice).

4.2.3 Specific Ethical Issues in Genetic Studies

4.2.3.1 Informed Consent

A well informed consent and understood consent is essential part of all genetic testing. The central issue of human research in respect to the individual's autonomy is achieved by obtaining "Informed consent" from adult individuals who are capable of giving a valid informed consent after they have been provided with complete information regarding the condition and the tests. Those persons who are tested are entitled to receive all information in a way that they can understand what is proposed to be done, they must be made aware of any risk, they must be given time to decide whether or not they would like to participate or withdraw from genetics testing/screening. In addition to this the details about the disorder to be screened and its inheritance pattern, reliability of the screening test and what will be done with the samples should also be explained. Information about the implications of a positive screening test (abnormal) should also be explained. In the case of an individual who is not capable of giving informed consent (for example individuals such as children, mentally challenged persons or prisoners etc), the consent of a legal guardian should be taken. Adequate information about the research given in a simple and easily understandable language in a document known as the Informed Consent Form with Participant/ Patient Information Sheet. Table 2 gives the details of elements that should be given in an informed consent form/ patient information sheet (ICMR Guidelines, 2006).

Table 4.2: Elements of Informed Consent Form

1) Nature and purpose of study stating it as research.
2) Duration of participation with number of participants.
3) Procedures to be followed.
4) Investigations, if any, to be performed.
5) Foreseeable risks and discomforts adequately described and whether project involves more than minimal risk.
6) Benefits to participant, community or medical profession as may be applicable.
7) Policy on compensation.
8) Availability of medical treatment for such injuries or risk management.
9) Alternative treatments if available.
10) Steps taken for ensuring confidentiality.
11) No loss of benefits on withdrawal (from participation).
12) Benefit sharing in the event of commercialization.
13) Contact details of PI or local PI/Co-PI in multicentric studies for asking more information related to the research or in case of injury.
14) Contact details of Chairman of the IEC for appeal against violation of rights.
15) Voluntary participation.
16) If test for genetics and HIV is to be done, counseling for consent for testing must be given as per national guidelines.
17) Storage period of biological sample and related data with choice offered to participant regarding future use of sample, refusal for storage and receipt of its results.

A copy of the participant/patient information sheet should be given to the participant for her/ his record. When the written consent (signature or thumb impression) is not possible due to sensitive nature of the project or the participant is unable to write, then oral consent can be taken after ensuring its documentation by an unrelated witness. However the approval for oral consent has to be taken from an ethics committee beforehand. When human biological material or samples are collected, whether in a research or clinical setting, it is appropriate to ask persons for their consent to future storage and use of their samples.

4.2.3.2 Vulnerable Participants

Vulnerable individuals are those persons who are unable to give valid and understood consent for e.g., minors, mentally disadvantaged, prisoners, institutionalised individuals, terminally ill, students, subordinates and people who do not speak the language etc. Further there are influences of culture or customs which can make certain communities vulnerable. Even adult women may be incapable of giving informed consent without consulting family members like mother in law or husband. In community research where even if consent of the community leaders or head of family is taken, individual consent of every person is a must. In telling about test results (e.g., carrier status for a recessive disorder; which parent carries a balanced autosomal translocation that has caused a disorder in their child; incidental discovery of non-paternity) and in assisting couples to reach reproductive decisions, professionals should be careful to protect the interests of those who may be vulnerable to harm from a hostile environment. Similarly children, persons of diminished mental capacity, and some other groups may not be in a position to make a decision due to their limited capacities. Such people may be vulnerable to harm because of their position in society and need protection from any potential adverse effects of genetic testing. Minority groups are entitled to respect, to their beliefs, customs or way of thinking even if the medical geneticist disagrees with these views. They should be treated equally with persons whose views are in the majority. They should be provided with due information and be given opportunity to take their own decisions.

4.2.3.3 Genetic Counselling

Every individual should be provided with complete information and opportunity to make choices regarding one's health. This is important to the person's integrity and contributes to psychological well-being. All genetic testing should be accompanied with pre-test counselling (to tell about the nature of tests, what is expected etc) as well as a post test counselling (to explain the results of the testing and its implications). Pretest and post test counseling is an integral part of prenatal diagnosis (Kabra, 2003). In order to help a person to make a valid choice, he/she should be provided with more than one alternative. Only those persons who are qualified and experienced in communicating the meaning of genetic information should undertake genetic counselling.

Purpose of genetic counselling

- provide detailed information about the genetic disease in question.
- help individuals/couples understand options and present state of medical knowledge so they can make informed decisions.
- help individuals/couples adjust to and cope with their genetic problems.
- the removal or lessening of guilt or anxiety or fear that they may have.
- helping individuals/couples achieve their parenting goals.

4.2.3.4 Risks and Benefits

Any research procedure that exposes a person to any risk should be done only if there is proper justification for doing so and there should be expected benefits of the research. All expected or potential risks as well as benefits should be discussed in detail before participation of the person. Also it is important for you to note that in genetic research, the primary risk may not be physical but rather may be psychosocial and can have effects or implications on other members of the family as well. Genetic testing therefore should be offered when the probability of harm is less and the expected benefits are more. Adequate counseling should be given to participants on the meaning of genetic information they receive.

4.2.3.5 Pedigree Studies

Pedigrees involve obtaining history of other members of the family of the patient or the proband under investigation. It is an important tool in genetics however it may reveal information about the likelihood of other members of the family being either carriers of genetic diseases or being affected by the disease. Special privacy and confidentiality concerns arise in genetic family studies because of relationship between the participants. It should be kept in mind that within families each person is an individual who has the right to keep the information about himself or herself confidential. Family members are not entitled to know each other's diagnosis. Other members are secondary subjects and they are not entitled to know about the genetic diseases of other members without appropriate informed consent. Before revealing medical or personal information about individuals to other family members, investigator must obtain consent of the individual to do so.

To explain this further I will give you an example:

In view of the cultural background of our country where woman is still a vulnerable and exploited participant, revealing information to the husband that his wife is the carrier of balanced chromosomal translocation (leading to recurrent abortions or a genetic syndrome in her child) or that she is a carrier of a single gene causing 'X' linked or recessive disease, may lead to grounds for a divorce despite the fact that the husband himself is a carrier of the autosomal recessive disorder. While general principles of counseling require presence of both the spouses, necessary care must be taken not to end up in breaking the families. In view of above concerns appropriate caution may be exercised. We must understand that revealing who else in the family has agreed to participate may lead to breach of confidentiality. Also if there is a patient, out of personal interest s/he may put undue pressure on relatives to enroll in the study which may not be acceptable to other members.

4.2.3.6 Privacy and Confidentiality

Secure safeguards should be in place for protecting private 'identifying information' of participants and maintaining confidentiality of research data. Identifying information comprises of all details like name, date of birth, contact details, and other personal information that can be used to identify an individual and therefore needs to be protected. Adequate care should be taken in keeping genetic records, however participants should be told beforehand regarding the limits to safeguard confidentiality and of the anticipated consequences of breach of confidentiality. If the result of the research is of benefit to the health of the

participant, then they should be communicated. Once study is over genetic data should be coded or delinked with clinical details to maintain confidentiality (anonymised). Adequate measures should be in place to restrict access of records to unauthorised persons. It needs to be emphasized that consent for screening or a subsequent confirmatory test does not imply consent to any specific treatment or termination of the pregnancy. Specific consent is required from the affected patient to share his/ her genetic information with family members who may be benefited from it.

4.2.3.7 Inducement vs Compensation

Individuals who participate in research may need to be duly compensated for their participation. Provisions should be made for adequate compensation to reimburse extra travel expenses, loss of wages, compensation and medical care in case of any injury. However the amounts should be reasonable and should not become undue inducement, i.e., so huge that it becomes the only reason for participation as an extra income and motivation factor for a person to give samples for research. Inducement can be in cash as well as in kind. Making promises for extra services or facilities may induce a person to agree to participate in testing even if it is not of any direct benefit to the person. Such promises for extra payments in cash or kind and the amount of compensation must be seen reviewed and approved by the ethics committees before it is provided to individuals. In case of any injury happens due to participation of a persons in research it is the obligation of the researcher/ physician to provide adequate treatment or compensation to cover the extra costs incurred due to this.

4.2.3.8 Genetic Screening

Genetic screening implies search in population or individuals who have, or are susceptible to have a serious genetic disease. It also includes persons who, though not at risk themselves, are carriers and therefore at risk of having children with the particular genetic disease. It is essential that screening must be done with specific purpose using validated tests. Genetic screening should be done for conditions if it can be ensured that a suitable treatment/ management of the disease is possible. Depending on nature of the genetic defect that is identified and its pattern of inheritance, siblings and other blood relations as well as existing and future generation/offspring may be affected. Screening of newborns is permissible to detect those genetic diseases like phenylketonuria where serious effects of the disease could be prevented by early intervention such as special diet or treatment. It should not be done when there is no cure or for diseases which manifest later in life. Screening of children should be deferred till the time they are old enough to understand and are able to participate in the decision making process, unless the intervention based on result of the test is likely to be of direct benefit to them at a younger age. Anonymous screening may be conducted on general population in order to establish prevalence of genetic traits/diseases. Anonymous samples means all identifying information about person who has contributed sample is removed and there is no way to know whose sample is it. Left over anonymous blood spots collected for screening newborns for treatable disorders could also be used for this purpose.

4.2.3.9 DNA Testing

For all kinds of DNA diagnosis the general principles of informed consent, confidentiality and other criteria used for any investigation in genetics should be

followed. Since the knowledge in this field is new, and relatively complex, a DNA test must be preceded and followed by appropriate genetic counselling. The laboratories carrying out DNA diagnosis should make adequate provisions to explain fully the nature of testing, implications of results to families. Preferably they should prepare brochures in simple language which can be understood by the persons. Differentiating between clinical practice and research can be difficult in genetics as genetic investigations may extend to other individuals as well as families (Parker, 2004).

4.2.3.10 Prenatal Testing

It is aimed at detecting presence of abnormalities in the foetus. The foetal sample for examination may be obtained through amniocentesis, chorionic villi sampling, cordocentesis or other biopsies. Foetal cells in maternal circulation can also be used for prenatal testing. Non-invasive methods like ultrasound should be preferred whenever available. Prenatal diagnosis should be performed only for reasons relevant to the health of the foetus or the mother and not to select the sex of the child (in the absence of an X-linked disorder). Prenatal diagnosis can be used to prepare parents for the birth of a child with a disability or they have a choice to go for abortion of affected foetus before 22 weeks of pregnancy. Professionals should recognise the human and economic costs involved in prenatal diagnosis and should limit its use to situations where there is a clear benefit. Many of the centers in tertiary care hospitals or urban centers have now initiated carrier screening for some common genetic ailments. Thalassaemia screening an effective control program is being carried out at the Institute of Immunohaematology, Mumbai, India. More than 14,000 women were screened and carriers identified, followed up and offered prenatal diagnosis (Mohanty et al, 2002).

Sex selection, whether for male or female, is very harmful to society since it creates an imbalance in sex ratios. For example you must be aware about the change in the sex ratio in India as commonly pregnancies with girl foetuses are being terminated. The situation is very bad in several states of the country including Delhi, Punjab, Haryana, Rajasthan, Bihar where sex ratio is greatly altered. Concerned with the misuse of genetic tests, particularly for the pre-selection of sex, the Government of India has enacted a law known as “The Prenatal Diagnostic Techniques (Regulation & Prevention of Misuse) Act 1994” amended in 2003 to include the Preconceptual diagnostic techniques also. Pre-implantation DNA diagnosis is also a type of prenatal diagnosis. Same precautions and safeguards should be adopted for this purpose also.

4.2.3.11 Presymptomatic and Susceptibility Testing

Presymptomatic testing (e.g., for Huntington disease) identifies individuals who will develop a genetic disorder later in life. Susceptibility testing (often referred to as ‘predictive testing’) identifies persons who are at increased risk for developing common diseases, such as heart disease, but who may never develop the disease in question. In some cases, presymptomatic testing (e.g., for familial polyposis coli) can lead to prevention of the disorder’s most serious effects (e.g., by colon surgery to prevent cancer). Susceptibility testing can lead to preventive programmes for heart disease or regular examinations to make possible early diagnosis and treatment (e.g., for breast cancer). In other cases, where successful prevention or treatment are not possible, as in Huntington disease, the major benefit of presymptomatic testing is to provide information for planning one’s

life and for deciding whether or not to have children. Pre-morbid diagnosis (diagnosing before appearance of adverse symptoms) in children should not be done for which there is no available intervention. Pre-morbid diagnosis in adults may be carried out with informed consent and appropriate genetic counselling.

4.2.3.12 Gene Therapy

Somatic cell gene therapy is permissible for the purpose of preventing or treating a serious disease when it is the only therapeutic option. It should be restricted to alleviation of life threatening or seriously disabling genetic disease in individual patients and should not be permitted to change normal human traits. The guidelines and clearance for gene therapy is regulated by the National Bioethics Committee under Department of Biotechnology (DBT) and clearance from the local Institutional ethics committee should be obtained. Safety should be ensured especially because of the possibility of unpredicted consequences of gene insertion. All gene therapy trials should have the provision for long term surveillance. Germ Line Therapy is prohibited at present. Gene Therapy for enhancement of genetic characteristics (so called designer babies) should not be attempted, as long term effects are not understood regarding alterations to genetic machinery of humans. Similarly it would be unethical to use genetic engineering for improvement of intelligence, memory, physical abilities etc. even if specific gene/genes are identified in future. Eugenic Genetic Engineering for selection against personality, character, formation of body organs, fertility, intelligence and physical, mental and emotional characteristics is prohibited (ICMR Ethical Guidelines, 2006).

4.2.3.13 DNA and Cell-line Banking/ Repository

Human genetic material can be enormously precious and especially if it is regarding rare disorders or is derived from special communities or population groups. This material needs adequate preservation and protection for future use and from potential exploitation. Biobanks/Repositories collect, store, and distribute human biological materials for research purposes. Human biological samples in biobanks include organs, tissues, cells, body fluids or samples like serum, buffy coat, DNA, hair, nails, excreta, sweat, buccal scrapings *etc.* The samples may be anonymously stored with or without its corresponding clinical information in a database. Appropriate informed consent is needed for long term storage of samples for future testing and research. Issues related to use of samples, their control and ownership, and the benefit sharing to the individuals or community also need discussion. To prevent any exploitation and protect the rights of participants, the three main requirements are individual informed consent for future research, approval of the IEC and the Repository Ethics Committee, wherever applicable.

4.2.3.14 Cloning

Research using human stem cells to grow new tissues (in order to repair or replace those damaged by disease) holds potential promise. Some of this research may involve nuclear fusion of an adult individual's cell with an enucleated egg, a first step toward potential human cloning. The possible benefits of research using nuclear fusion to produce tissues for the treatment of disease are recognized, provided that there would be no attempt to reproduce an entire human being. At the present time, "reproductive human cloning" is unsafe and should not be attempted.

4.2.3.15 Stem Cell Research and Therapy

The stem cell research holds a great promise for improving human health by control of degenerative diseases and restoration of damage to organs by various injuries. At the same time it also raises several ethical and social issues such as destruction of human embryos to create human embryonic stem (HES) cell lines, potential for commodification in human tissues and organs, possible use of technology for germ-line engineering and reproductive cloning etc. The research in this field, therefore, needs careful consideration. It is important to protect safety and rights of those donating gametes/ blastocysts/ somatic cells for derivation of stem cells; or fetal tissues/umbilical cord cells/ adult tissue (or cells) for use as stem cells. Safeguards are also needed to protect research participants receiving stem cell transplants, and patients at large from unproven therapies/remedies. Most of the stem cell procedures are still in the research or experimental phase and long term effects are unknown therefore stem cell studies should be done as a clinical trial and not offered as proven therapy to patients.

4.2.4 Other Concerns in Genetics Research

4.2.4.1 Commercial Issues, Stigmatisation and Discriminations Based on Genetic Characteristics

When commercial companies are involved in research, it is necessary to protect researchers and participants from possible coercion or inducement to participate in the study. Academic institutions conducting research in alliance with industries or commercial companies require a strong review to probe possible conflicts of interest between scientific responsibilities of researchers and business interests (e.g. ownership or part-ownership of the investigator in the company developing a new product). Prospective participants in research should also be informed of the sponsorship, so that they can be aware of the potential for conflicts of interest and commercial aspects of the research.

The other concern is regarding insurance issues where knowing about certain types of genetic information can lead to discrimination by insurance companies who may refuse to provide insurance or charge higher premium. Similarly there is a potential threat that employers may not provide jobs to individuals with certain type of genotypes or school may not admit children based on their genetic diagnosis. There are also major concerns regarding stigmatisation of certain population groups or certain communities based on the genetic information. Many of these examples may sound theoretical today however with emerging new tools for genetic diagnosis there is a need to make provisions in advance for protecting populations or groups from such potential harm. It is need for us to be aware of these scenarios and protect our rich heritage.

4.2.4.2 Patenting

Biomedical research in human genetics can lead to the development of diagnostic and pharmaceutical products. Patents may be necessary to raise funding to develop such products commercially, but gene sequences without proven utility should not be granted patents. Patenting has the potential to impede international collaboration, especially between developing and developed countries, to the ultimate detriment of service delivery to those with genetic disorders. Genetics differs from many areas of research in that important new knowledge can come from a family, or an ethnic group, with a particular genetic variant. If this leads

to the development of a diagnostic test or new therapies, equity requires that the donors, or the community generally, should receive some benefit.

4.2.4.3 Research in International Collaboration

India with its rich biodiversity and genetic resources, conserved gene pools, indigenous population groups is often an attractive destination for researchers from other countries. However it should not only be provider of samples but be an equal collaborating partner in research to ensure development in science. Such collaborative partnerships should lead to development of capacity within the country for better genetic research rather than mere transfer of samples abroad for diagnostic purposes. At present a lot of genetic research is carried out in International collaboration. This requires adequate care so that the rights of the communities/ persons from whom samples are derived from are adequately protected if there is transfer of biological material across borders. On one hand, collaboration in genetics helps build capacity and is useful for betterment of scientific knowledge but on the other it should not give the impression of experimentation on the population of one country by another. There is a Govt of India notification on “Exchange of Human Biological Material for Biomedical Research” issued on 19.11.97 by Ministry of Health and Family Welfare. Appropriate regulatory clearances are needed for international collaboration and Ethics committee approval has to be taken before the initiation of research. Some special concerns are given below in Table 4.3.

Table 4.3: Safeguards for Research in International Collaboration

- The collaborating investigators, institutions and countries can function as equal partners by building appropriate safeguards.
- Careful consideration should be given to protect the dignity, safety and welfare of the individuals if there is any possibility of exploitation of participants/ populations.
- Choice of study populations for genetic study should be justified in scientific and ethical terms.
- Nature, magnitude, and probability of all foreseeable harms resulting from participation in a collaborative research programme should be explained to participants.
- Research protocol should outline the benefits that persons / communities / countries should experience as a result of their participation.
- The burden and the benefit should be equally borne by the collaborating countries.
- Guidelines, rules, regulations and cultural sensitivities of countries participating in collaborative research should be respected.
- Issues such as intellectual property rights (IPR), exchange of biological materials, data transfer should be carried out as per signed Memorandum of understanding (MoU).
- Collaborative research requires clearance from ethics committees as well as from Health Ministry Screening Committee (HMSC) as per regulatory requirements.

4.3 SUMMARY

The last decade has witnessed tremendous advances in science and technology and the prospects of genetic testing, genetic engineering, cloning, stem cell research, prenatal testing, DNA diagnostic etc. all have raised many ethical dilemmas in the minds of scientists and society. There are advantages of such new technologies but there is also the fear of the unknown distant possibilities which needs careful examination. Weighing of risk benefit ratio, protection of privacy and confidentiality and sharing of burdens and benefits are some of the main issues. Close monitoring of research in these areas is important. Education has the potential power to prevent stigmatization and discrimination by emphasizing that genetic disorders are not caused by the behaviour of affected persons or families but the fact that most people may carry some recessive lethal mutations and that our offspring or we are all at genetic risk. For all kinds of genetic testing and research, respect for persons is important and should include informed consent, right to referral, full disclosure, protection of confidentiality, and respect for minority groups, women and children. The following statements summarize the issues:

- Genetic services should be voluntary but should be made available to those interested in them.
- Genetic counselling should be non-directive to let the decision making be done by individuals without any force or coercion.
- All clinically relevant information that may affect the health of an individual or fetus should be disclosed and various options explained.
- Confidentiality of genetic information should be maintained. When there is a high risk of serious harm to family members at genetic risk, the information should be used to avert this harm.
- Individual privacy should be protected from third parties, such as employers, insurers, schools, commercial entities etc.
- Prenatal diagnosis should be performed only for reasons relevant to the health and only to detect genetic conditions or malformations.
- Choices about counselling, screening, testing and abortion following prenatal diagnosis should be available on a voluntary basis.
- Optimum support and education should be provided for children and families with genetic conditions.
- Research protocols should follow established procedures for ethical review and informed consent.

Further Reading and References

Ethical Guidelines for Biomedical Research on Human Participants, 2006.

Guidelines for Exchange of Human Biological Material for Biomedical Research Purposes, 1997.

Hardy BJ, Séguin B, Goodsaid F, Jimenez-Sanchez G, Singer PA, Daar AS. 2008. The next steps for genomic medicine: challenges and opportunities for the developing world. *Nat Rev Genet.* Oct;9 Suppl 1:S23-7.

Kabra, M. 2003. Prenatal diagnosis. *Indian J Pediatr.* 70:81-5.

Mohanty D, Colah RB, Gorakshakar AH, Nadkarni AH, Phanasgaonkar SP, Shetty S, Ghosh K, Mukherjee MB. 2002. Genetic Disorders in Haematological Practice in India. *Community Genetics*, 5:197-200.

National Guidelines for Stem Cell Research and Therapy, 2007.

Nuffield Council on Bioethics. 1993. Genetic screening and public policy. In *Genetic Screening, Ethical Issues*. London: Nuffield Council on Bioethics, 75–81.

Parker, M Ashcroft R, AOM Wilkie, A Kent, 2004. Ethical review of research into rare genetic disorders. *British Medical Journal*, 329:288–9.

Rama Devi AR, Naushad SM. 2004. Newborn screening in India. *Indian J Pediatr*; 71: 157-60.

Séguin B, Hardy BJ, Singer PA, Daar AS. 2008. Genomic medicine and developing countries: creating a room of their own. *Nat Rev Genet.* 9 (6):487-93.

The Universal Declaration on the Human Genome and Human Rights, 29th session General Conference UNESCO, 11 November 1997.

The ethics of patenting DNA: a discussion paper. Nuffield Council on Bioethics July 2002.

Tom L. Beauchamp, James F. Childress. 1994. *Principles of Biomedical Ethics*, Oxford University Press.

Verma IC, Bijarnia S. 2002. The burden of genetic disorders in India and a framework for community control. *Community Genetics*, 5:192-196.

Sample Questions

- 1) Write a short note on Human Genome Project and its implications.
- 2) What do you understand by Genetic Counselling and why is it important?
- 3) What are the ethical, legal and social issues? Discuss.
- 4) Write a note on the Ethical Principles in Genetic research.
- 5) What is informed consent in research? Give a few main elements of informed consent.
- 6) What do you understand by vulnerable persons/ populations?
- 7) Why should there be adequate provisions to protect the privacy and confidentiality of persons in genetic research?
- 8) Describe any three specific concerns in genetics research.
- 9) Write a short note on biobanks.
10. What safeguards should be in place for genetic research in International Collaboration?