COURSE NOTES: Understanding genetics for improving health outcomes

Course Code: CEUGH

Instructor Name: Dr. Rahul Kushwah

Phone Number: (647) 889 6916

Instructor Email: rahulkushwah@anantlife.com

Course Description

The course is intended to provide the tools that holistic nutritionists can use to deliver personalized healthcare to their clients, catered to the genetic makeup of the client. Genetics plays a very robust role in nutrition, detoxification, weightloss and overall health and wellness. Furthermore, genetics can be used not only to improve the health of the clients but also to develop DNA based nutritional plans that can potentially prevent development of chronic diseases. Course participants will gain knowledge on how genetic information can be used to deliver nutritional plans, weightloss strategies, detoxification plans, hormonal balance plans along with nutritional plans to prevent development of chronic disease such as diabetes and cardiovascular diseases.

IHN has partnered with Anantlife Canada Inc., a leader in clinical grade genetic testing for healthcare providers all over the world, to offer a Certified Genetic Testing Provider Certificate upon successful completion of the course. Successful completion of the course implies that the candidates have received the education and training to not only understand genetic concepts pertaining to diet, nutrition, detoxification, fitness, hormonal health and metabolic disorders but have also been trained on interpretation of the genetic testing reports along with development of a DNA based health plan for better health outcomes.

SESSION 1:

INTRODUCTION TO MOLECULAR BIOLOGY, MOLECULAR GENETICS, HUMAN GENETICS AND PREDICTIVE GENETIC TESTING

Molecular biology is the study of all the molecular processes involved in cells and understanding molecular biology is the key to understanding how DNA plays an essential role in regulating overall biology including genetics. Our understanding of molecular biology along with the laboratory techniques utilized have formed the basis for genetic testing. The readings herein are to provide a detailed understanding of molecular biology, which is needed to develop a practical understanding of genetics and genetic testing.

Introduction to Molecular Biology

2

2.1 Introduction

Modern biology has its roots at the work of Gregor Mendel who identified the fundamental rules of hereditary in 1865. The discovery of chromosomes and genes followed later and in 1952 Watson and Crick disclosed the double helix structure of DNA. All living organisms have common characteristics such as replication, nutrition, growing and interaction with their environment. An *organism* is composed of *organs* which perform specific functions. Organs are made of *tissues* which are composed of aggregation of cells that have similar functions. The *cell* is the basic unit of life in all living organisms and it has molecules that have fundamental functions for life. Molecular biology is the study of these molecules in the cell. Two of these molecules called *proteins* and *nucleotides* have fundamental roles to sustain life. Proteins are the key components in everything related to life. DNA is made of nucleotides and parts of DNA called *genes* code for proteins which perform all the fundamental processes for living using biochemical reactions.

Cells synthesize new molecules and break large molecules into smaller ones using complex networks of chemical reactions called *pathways*. Genome is the complete set of DNA of an organism and human genome consists of chromosomes which contain many genes. A gene is the basic physical and functional unit of hereditary that codes for a protein which is a large molecule made from a sequence of amino acids. Three critical molecules of life are deoxyribonucleic acid (DNA), ribonucleic acid (RNA), and proteins. A central paradigm in molecular biology states that biological function is heavily dependent on the biological structure.

In this chapter, we first review the functions performed by the cell and its ingredients. The DNA contained in the nucleus, the proteins, and various other molecules all have important functionalities and we describe these in detail. The central dogma of life is the process of building up a protein from the code in the genes as we will outline. We will also briefly describe biotechnological methods and introduce some of the commonly used databases that store information about DNA, proteins, and other molecules in the cell.

K. Erciyes, *Distributed and Sequential Algorithms for Bioinformatics*,
Computational Biology 23, DOI 10.1007/978-3-319-24966-7_2

2.2 The Cell

Cells are the fundamental building blocks of all living things. The cell serves as a structural building block to form tissues and organs. Each cell is independent and can live on its own. All cells have a metabolism to take in nutrients and convert them into molecules and energy to be used. Another important property of cells is *replication* in which a cell produces another cell that has the same properties as itself. Cells are composed of approximately 70 % water; 7 % small molecules like amino acids, nucleotides, salts, and lipids, and 23 % macromolecules such as proteins and polysaccharids. A cell consists of molecules in a dense liquid surrounded by a membrane as shown in Fig. 2.1.

The *eukaryotic cells* have nuclei containing the genetic material which is separated from the rest of the cell by a membrane and the *prokaryotic cells* do not have nuclei. Prokaryotes include bacteria and archaea; and plants, animals, and fungi are examples of eukaryotes. The tasks performed by the cells include taking nutrients from food, converting these to energy, and performing various special missions. A cell is composed of many parts each with a different purpose. The following are the important parts of an eukaryotic cell with their functions:

- **Nucleus**: Storage of DNA molecules, and RNA and ribosome synthesis.
- **Endoplasmic reticulum**: Synthesis of lipids and proteins
- **Golgi apparatus**: Distribution of proteins and lipids and posttranslational processing of proteins.
- **Mitochondria**: Generation of energy by oxidizing nutrients.
- **Vesicles**: *Transport vesicles* move molecules such as proteins from endoplasmic reticulum to Golgi apparatus, *Secretory vesicles* have material to be excreted from the cell and *lysosomes* provide cellular digestion.

The nucleus is at the center of the cell and is responsible for vital functions such as cell growth, maturation, division, or death. *Cytoplasm* consists of jellylike fluid which surrounds the nucleus and it contains various other structures. *Endoplasmic*

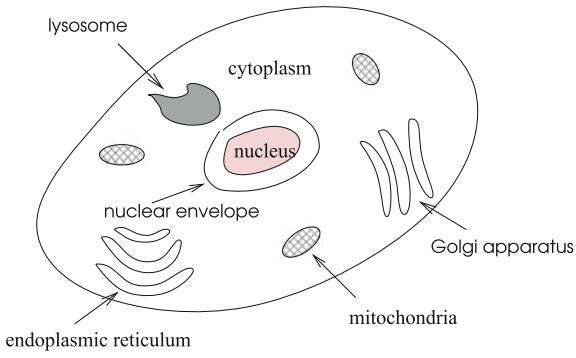

Fig. 2.1 Parts of a cell

reticulum enwraps the nucleus, and processes molecules made by the cell and transports them to their destinations. Conversion of energy from food to a form that can be used by the cell is performed by *mitochondria* which have their own genetic material. These components of the cell are shown in Fig. 2.1. The cell contains various other structures than the ones we have outlined here.

Chemically, cell is composed of few elements only. Carbon (C), hydrogen (H), oxygen (O), and nitrogen (N) are the dominant ones with phosphorus (P) and sulfur (S) appearing in less proportions. These elements combine to form molecules in the cell, using covalent bonds in which electrons in their outer orbits are shared between the atoms. A *nucleotide* is one such molecule in the cell which is a chain of three components: a base B, a sugar S, and a phosphoric acid P. The three basic macromolecules in the cell that are essential for life are the DNA, RNA, and proteins.

2.2.1 DNA

James Watson and Francis Crick discovered the Deoxyribonucleic Acid (DNA) structure in the cell in 1953 using X-ray diffraction patterns which showed that the DNA molecule is long, thin, and has a spiral-like shape [5]. The DNA is contained in the nuclei of eukaryotic cells and is composed of small molecules called *nucleotides*. Each nucleotide consists of a five-carbon sugar, a phosphate group, and a base. The carbon atoms in a sugar molecule are labeled $1'$ to $5'$ and using this notation, DNA molecules start at $5'$ end and finish at $3'$ end as shown in Fig. 2.2. There are four nucleotides in the DNA which are distinguished by the bases they have: Adenine (A), Cytosine (C), Guanine (G), and Thymine (T). We can therefore think of DNA as a string with a four letter alphabet $\sum = \{A,C,G,T\}$. Human DNA consists approximately of three billion bases. Nucleotide A pairs only with T, and C pairs only with G, we can say A and T are complementary and so are G and C as shown in Fig. 2.2.

Given the sequence S of a DNA strand, we can construct the other strand S' by taking the complement of bases in this strand. If we take the complement of the

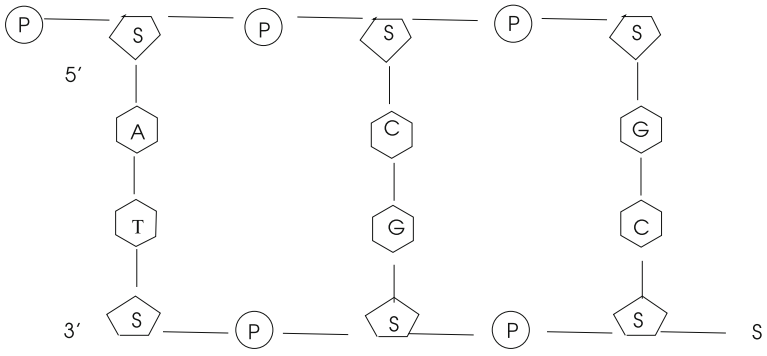

Fig. 2.2 DNA structure

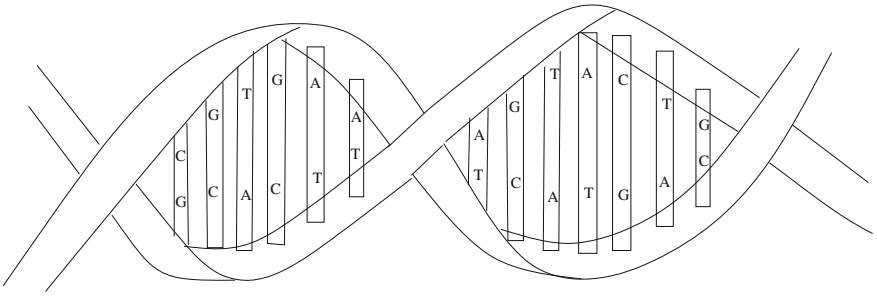

Fig. 2.3 DNA double helix structure

resulting strand we will obtain the original strand. This process is used and essential for protein production. Physically, DNA consists of two strands held together by hydrogen bonds, arranged in a double helix as shown in Fig. 2.3. The *complement* of a DNA sequence consists of complements of its bases. The DNA therefore consists of two complementary strands which bind to each other tightly providing a stable structure. This structure also provides the means to replicate in which the double DNA helix structure is separated into two strands and each of these strands are then used as templates to synthesize their complements.

The DNA molecule is wrapped around proteins called *histones* into complex-walled structures called *chromosomes* in the nucleus of each cell. The number of chromosomes depends on the type of eukaryote species. Each chromosome consists of two *chromatides* which are coil-shaped structures connected near the middle forming an x-like structure. Chromosomes are kept in the nucleus of a cell in a highly packed and hierarchically organized form. A single set of chromosomes in an organism is called *haploid*, two sets of chromosomes is called *diploid*, and more than two sets is called *polyploid*. Humans are diploid where each chromosome is inherited from a parent to have two chromosomes for each of the 23 chromosome set. The sex chromosome is chromosome number 23 which either has two chromosomes shaped X resulting in a female, or has X and Y resulting in a male. The type of chromosome inherited from father determines the sex of the child in this case.

2.2.2 RNA

The ribonucleic acid (RNA) is an important molecule that is used to transfer genetic information. It has a similar structure to DNA but consists of only one strand and does not form a helix structure like DNA. It also has nucleotides which consist of a sugar, phosphate, and a base. The sugar however is a *ribose* instead of deoxyribose and hence the name RNA. Also, DNA base thymine (T) is replaced with uracil (U) in RNA. The fundamental kinds of RNA are the messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA) which perform different functions in the cell. RNA provides a flow of information in the cell. First, DNA is copied to mRNA

in the nucleus and the mRNA is then translated to protein in the cytoplasm. During translation, tRNA and rRNA have important functions. The tRNA is responsible for forming the amino acids which make up the protein, as prescribed in the mRNA; and the rRNA molecules are the fundamental building blocks of the ribosomes which carry out translation of mRNA to protein.

2.2.3 Genes

A *gene* is the basic unit of hereditary in a living organism determining its character as a whole. A gene physically is a sequence of DNA that codes for an RNA (mRNA, tRNA, or rRNA) and the mRNA codes for a protein. The study of genes is called *genetics*. Gregor Mendel in the 1860 s was first to experiment and set principles of passing hereditary information to offsprings.

There are 23 pairs of chromosomes in humans and between 20000–25000 genes are located in these chromosomes. The starting and stopping locations of a gene are identified by specific sequences. The protein coding parts of a gene are called *exons* and the regions between exons with no specific function are called *introns*. Genes have varying lengths and also, exons and introns within a gene have varying lengths. A gene can combine with other genes or can be nested within another gene to yield some functionality, and can be mutated which may change its functionality at varying degrees in some cases leading to diseases. The complete set of genes of an organism is called its *genotype*. Each gene has a specific function in the physiology and morphology of an organism. The physical manifestation or expression of the genotype is the *phenotype* which is the physiology and morphology of an organism cite. A gene may have different varieties called *alleles* resulting in different phenotyping characteristics. Humans are diploid meaning we inherit a chromosome from each parent, therefore we have two alleles of each gene. The genes that code for proteins constitute about 1.5 % of total DNA and the rest contains RNA encoding genes and sequences that are not known to have any function. This part of DNA is called *junk DNA*. There is no relatedness between the size of genome, number of genes, and organism complexity. In fact, some single cell organisms have a larger genome than humans.

2.2.4 Proteins

Proteins are large molecules of the cell and they carry out many important functions. For example, they form the antibodies which bind to foreign particles such as viruses and bacteria. As enzymes, they work as catalysts for various chemical reactions; the messenger proteins transmit signals to coordinate biological processes between different cells, tissues, and organs, also they transport small molecules within the cell and the body. Proteins are made from the information contained in genes. A protein consists of a chain of amino acids connected by *peptide bonds*. Since such a bond releases a water molecule, what we have inside a protein is a chain of amino acid

Table 2.1 Amino acids

Name	Abbrv.	Code	Pol.	Name	Abbrv.	Code	Pol.
Alanine	Ala	A	H	Methionine	Met	M	H
Cysteine	Cys	C	P	Asparagine	Asn	N	P
Aspartic acid	Asp	D	P	Proline	Pro	P	H
Glutamic acid	Glu	E	P	Glutamine	Gln	Q	P
Phenylalanine	Phe	F	H	Arginine	Arg	R	P
Glycine	Gly	G	P	Serine	Ser	S	P
Histidine	His	H	P	Threonine	Thr	T	P
Isoleucine	Ile	I	H	Valine	Val	V	H
Lysine	Lys	K	P	Tryptophan	Trp	W	H
Leucine	Leu	L	H	Tyrosine	Tyr	Y	P

residues. Typically, a protein has about 300 amino acid residues which can reach 5000 in large proteins.The essential 20 amino acids that make up the proteins is shown in Table 2.1 with their abbreviations, codes, and polarities.

Proteins have highly complex structures and can be analyzed at four hierarchical structures. The *primary structure* of a protein is specified by a sequence of amino acids that are linked in a chain and the *secondary structure* is formed by linear regions of amino acids. A *protein domain* is a segment of amino acid sequences in a protein which has independent functions than the rest of the protein. The protein also has a 3D structure called *tertiary structure* which affects its functionality and several protein molecules are arranged in *quaternary structure*. The function of a protein is determined by its four layer structure. A protein has the ability to fold in 3D and its shape formed as such affects its function. Using its 3D shape, it can bind to certain molecules and interact. For example, mad cow disease is believed to be caused by the wrong folding of a protein. For this reason, predicting the folding structure of a protein from its primary sequence and finding the relationship between its 3D structure and its functionality has become one of the main research areas in bioinformatics.

2.3 Central Dogma of Life

The central dogma of molecular biology and hence life was formulated by F. Crick in 1958 and it describes the flow of information between DNA, RNA, and proteins. This flow can be specified as DNA $\rightarrow$ mRNA $\rightarrow$ protein as shown in Fig. 2.4. The forming of mRNA from a DNA strand is called *transcription* and the production of a protein based on the nucleotide sequence of the mRNA is called *translation* as described next.

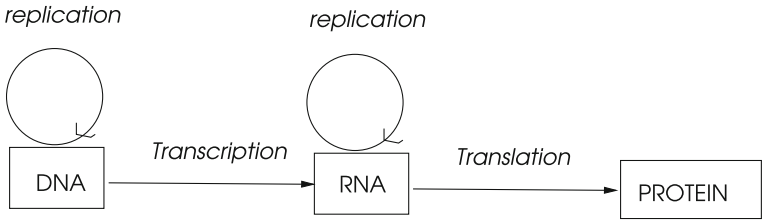

Fig. 2.4 Central dogma of life

2.3.1 Transcription

In the transcription phase of protein coding, a single stranded RNA molecule called mRNA is produced which is complementary to the DNA strand it is transcribed. The transcription process in eukaryotes takes place in the nucleus. The enzyme called *RNA polymerase* starts transcription by first detecting and binding a *promoter* region of a gene. This special pattern of DNA shown in Fig. 2.5 is used by RNA polymerase to find where to begin transcription. The reverse copy of the gene is then synthesized by this enzyme and a terminating signal sequence in DNA results in the ending of this process after which pre-mRNA which contains exons and introns is released. A post-processing called *splicing* involves removing the introns received from the gene and reconnecting the exons to form the mature and much shorter mRNA which is transferred to cytoplasm for the second phase called *translation*. The complete gene contained in the chromosome is called *genomic* DNA and the sequence with exons only is called *complementary* DNA or cDNA [25].

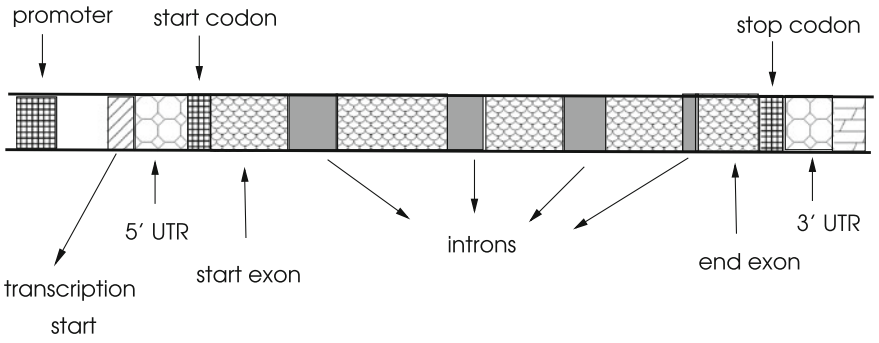

Fig. 2.5 Structure of a gene

Table 2.2 The genetic code

	2nd Letter				
1st L.	U	C	A	G	**3rd L.**
U	UUU Phe UUC UUA Leu UUG	UCU Ser UCC UCA UCG	UAU Tyr UAC UAA **Stop** UAG **Stop**	UGU Cys UGC UGA **Stop** UGG Trp	U C A G
C	CUU Leu CUC CUA CUG	CCU Pro CCC CCA CCG	CAU His CAC CAA Gln CAG	CGU Arg CGC CGA CGG	U C A G
A	AUU Ile AUC AUA AUG **Met**	ACU Thr ACC ACA ACG	AAU Asn AAC AAA Lys AAG	AGU Ser AGC AGA Arg AGG	U C A G
G	GUU Val GUC GUA GUG	GCU Ala GCC GCA GCG	GAU Asp GAC GAA Glu GAG	GGU Gly GGC GGA GGG	U C A G

2.3.2 The Genetic Code

The genetic code provides the mapping between the sequence of nucleotides and the type of amino acids in proteins. This code is in triplets of nucleotide bases called *codons* where each codon encodes one amino acid. Since there are four nucleotide bases, possible total number of codons is $4^3 = 64$. However, proteins are made of 20 amino acids only which means many amino acids are specified by more than one codon. Table 2.2 displays the genetic code.

Such redundancy provides fault tolerance in case of mutations in the nucleotide sequences in DNA or mRNA. For example, a change in the codon UUA may result in UUG in mRNA but the amino acid *leucine* corresponding to each of these sequences is formed in both cases. Similarly, all of the three codons UAA, UAG, and UGA cause termination of the polypeptide sequence and hence a single mutation from A to G or from G to A still causes termination preventing unwanted growth due to mutations. Watson et al. showed that the sequence order of codons in DNA correspond directly to the sequence order of amino acids in proteins [28]. The codon AUG specifies the beginning of a protein amino acid sequence, therefore, the amino acid *methionine* is found as the first amino acid in all proteins.

2.3.3 Translation

The translation phase is the process where a mature mRNA is used as a template to form proteins. It is carried out by the large molecules called *ribosomes* which consist of proteins and the ribosomal RNA (rRNA) [5]. A ribosome uses tRNA to

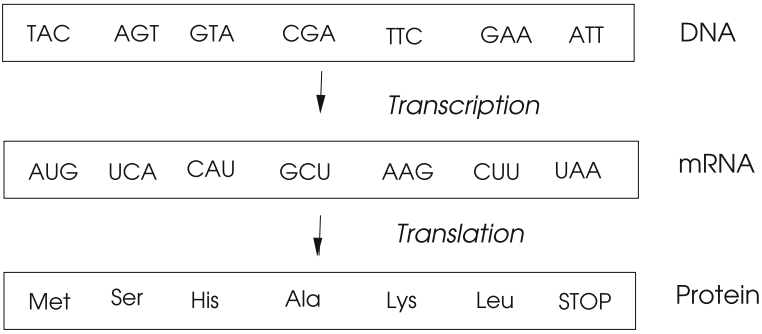

Fig. 2.6 Construction of a protein

first detect the start codon in the mRNA which is the nucleotide base sequence AUG. The tRNA has three bases called *anticodons* which are complementary to the codons it reads. The amino acids as prescribed by the mRNA are then formed and added to the linear protein structure according to the genetic code. Translation to the protein is concluded by detecting one of the three stop codons. Once a protein is formed, a protein may be transferred to the needed location by the signals in the amino acid sequence. The new protein must fold into a 3D structure before it can function [27]. Figure 2.6 displays the transcription and translation phases of a superficial protein made of six amino acids as prescribed by the mRNA.

2.3.4 Mutations

Mutations are changes in genetic code due to a variety of reasons. As an example, a stop codon UAA may be formed instead of an amino acid coding codon UCA (Serine) by a single point mutation of A to C, which will result in a shorter protein that will likely be nonfunctional. An amino acid may be replaced by another one, again by a point mutation such as the forming of CCU (Proline) instead of CAU (Histidine) by the mutation of C to A. These types of mutations may have important or trivial effects depending on the functioning of the mutated amino acid in the protein [5]. Furthermore, a nucleotide may be added or deleted to the sequence, resulting in the shifting of all codons by one nucleotide which will result in a very different sequence.

Mutations can be caused by radiations from various sources such as solar, radioactive materials, X-rays, and UV light. Chemical pollution is also responsible for many cases of mutations and viruses which insert themselves into DNA cause mutations. The inherited mutations are responsible for genetic diseases such as multiple sclerosis and Alzheimer disease. In many cases though, mutations result in better and improved characteristics in an organism such as better eyesight.

2.4 Biotechnological Methods

Biotechnology is defined as any technological application that uses biological systems, living organisms, or derivatives thereof, to make or modify products or processes for specific use, as defined by the Convention on Biological Diversity (CBD) [7]. The main biological methods are the *cloning* and *polymerase chain reaction* to amplify it, and *sequencing* to determine the nucleotide sequence in a DNA segment.

2.4.1 Cloning

DNA needs to be in sufficient quantities to experiment. DNA *cloning* is a method to amplify the DNA segment which could be very small. In this method, the DNA to be amplified called $insert$ is inserted into the genome of an organism which is called the $host$ or the $vector$. The host is then allowed to multiply during which the DNA inserted to it also multiplies. The host can then be disposed of, leaving only the amplified DNA segment. The commonly used organisms for cloning DNA are $plasmids$, $cosmids$, $phages$, and *yeast artificial chromosomes* (YACs) [25]. A plasmid is a circular DNA in bacteria and is used for cloning DNA of sizes up to 15 kbp. Phages are viruses and DNA segment inserted in them gets replicated when the virus infects an organism and multiplies itself. In YAC-based cloning, an artificial chromosome of yeast is constructed by the DNA insert sequence and the yeast chromosome control sections. The yeast multiplies its chromosomes including the YAC and hence multiplying the insert. YAC-based cloning can be used for very large segments of a million base pairs [25].

2.4.2 Polymerase Chain Reaction

The polymerase chain reaction (PCR) developed by Kary Mullis [3] in 1983, is a biomedical technology used to amplify selected DNA segment over several orders of magnitude. The amplification of DNA is needed for a number of applications including analysis of genes, discovery of DNA motifs, and diagnosis of hereditary diseases. PCR uses *thermal cycling* in which two phases are employed. In the first phase, the DNA is separated into two strands by heat and then, a single strand is enlarged to a double strand by the inclusion of a primer and polymerase processing. DNA polymerase is a type of enzyme that synthesizes new strands of DNA complementary to the target sequence. These two steps are repeated many times resulting in an exponential growth of the initial DNA segment. There are some limitations of PCR processing such as the accumulation of pyrophosphate molecules and the existence of inhibitors of the polymerase in the DNA sample which results in the stopping of the amplification.

2.4.3 DNA Sequencing

The sequence order of bases in DNA is needed to find the genetic information. *DNA sequencing* is the process of obtaining the order of nucleotides in DNA. The obtained sequence data can then be used to analyze DNA for various tasks such as finding evolutionary relationships between organisms and treatment of diseases. The exons are the parts of DNA that contain genes to code for proteins and all exons in a genome is called *exome*. Sequencing exomes is known as *whole exome sequencing*. However, research reveals DNA sequences external to the exons also affect protein coding and health state of an individual. In *whole genome sequencing*, the whole genome of an individual is sequenced. The new generation technologies are developed for both of these processes. A number of methods exist for DNA sequencing and we will briefly describe only the few fundamental ones.

The sequencing technology called *Sanger sequencing* named after Frederick Sanger who developed it [23,24], used deoxynucleotide triphosphates (dNTPs) and di-deoxynucleotide triphosphates (ddNTPs) which are essentially nucleotides with minor modifications. The DNA strand is copied using these altered bases and when these are entered into a sequence, they stop the copying process which results in different lengths of short DNA segments. These segments are ordered by size and the nucleotides are read from the shortest to the longest segment. Sanger method is slow and new technologies are developed. The *shotgun* method of sequencing was used to sequence larger DNA segments. The DNA segment is broken into many overlapping short segments and these segments are then cloned. These short segments are selected at random and sequenced in the next step. The final step of this method involves assembling the short segments in the most likely order to determine the sequence of the long segment, using the overlapping data of the short segments.

Next generation DNA sequencing methods employ massively parallel processing to overcome the problems of the previous sequencing methods. Three platforms are widely used for this purpose: the Roche/454 FLX [21], the Illumina/Solexa Genome Analyzer [4], and the Ion Torrent: Proton/PGM Sequencing [12]. The Roche/454 FLX uses the *pyrosequencing* method in which the input DNA strand is divided into shorter segments which are amplified by the PCR method. Afterward, multiple reads are sequenced in parallel by detecting optical signals as bases are added. The Illumina sequencing uses a similar method, the input sample fragment is cleaved into short segments and each short segment is amplified by PCR. The fragments are located in a slide which is flooded with nucleotides that are labeled with colors and DNA polymerase. By taking images of the slide and adding bases, and repeating this process, bases at each site can be detected to construct the sequence. The Ion proton sequencing makes use of the fact that addition of a dNTP to a DNA polymer releases an H^+ ion. The preparation of the slide is similar to other two methods and the slide is flooded with dNTPs. Since each H^+ decreases pH, the changes in pH level is used to detect nucleotides [8].

2.5 Databases

A database is a collection of structured data and there are hundreds of databases in bioinformatics. Many of these databases are generated by filtering and transforming data from other databases which contain raw data. Some of these databases are privately owned by companies and access is provided with a charge. In most cases however, bioinformatics databases are publicly accessible by anyone. We can classify bioinformatics databases broadly as nucleotide databases which contain DNA/RNA sequences; protein sequence databases with amino acid sequences of proteins, microarray databases storing gene expression data, and pathway databases which provide access to metabolic pathway data.

2.5.1 Nucleotide Databases

The major databases for nucleotides are the GenBank [10], the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) database [19], and the DNA Databank of Japan (DDJB) [26]. GenBank is maintained by the National Center for Biotechnology Information (NCBI), U.S. and contains sequences for various organisms including primates, plants, mammals, and bacteria. It is a fundamental nucleic acid database and genomic data is submitted to GenBank from research projects and laboratories. Searches in this database can be performed by keywords or by sequences. The EMBL-EBI database is based on EMBL Nucleotide Sequence Data Library which was the first nucleotide database in the world and receives contributions from projects and authors. EMBL supports text-based retrieval tools including SRS and BLAST and FASTA for sequence-based retrieval [9].

2.5.2 Protein Sequence Databases

Protein sequence databases provide storage of protein amino acid sequence information. Two commonly used protein databases are the Protein Identification Resource (PIR) [16,31] and the UniProt [15] containing SwissProt [2]. The PIR contains protein amino acid sequences and structures of proteins to support genomic and proteomic research. It was founded by the National Biomedical Research Foundation (NBRF) for the identification and interpretation of protein sequence information, and the Munich Information Center for Protein Sequences (MIPS) [22] in Germany, and the Japan International Protein Information Database later joined this database. SwissProt protein sequence database was established in 1986 and provided protein functions, their hierarchical structures, and diseases related to proteins. The Universal Protein Resource (UniProt) is formed by the collaboration of EMBL-EBI, Swiss Institute of Bioinformatics (SIB) and PIR in 2003 and SwissProt was incorporated into UniProt. PDBj (Protein Data Bank Japan) is a protein database in Japan providing an archive of macromolecular structures and integrated tools [17].

2.6 Human Genome Project

The human genome project (HGP) is an international scientific research project to produce a complete human DNA sequence and identifying genes of human genome as well as other organisms such as mice, bacteria, and flies. This project was planned in 1984, started in 1990 and was finished in 2003. About 20 universities and research centers in United States, Japan, China, France, Germany, and the United Kingdom participated in this project. It aimed to sequence the three billion base pairs in human genome to analyze and search for the genetic causes of diseases to find cure for them, along with analysis of various other problems in molecular biology.

The results of this project are that there are between 20,000–25,000 genes in humans, and the human genome has more repeated DNA segments than other mammalian genomes. The work on results are ongoing but the results started to appear even before the completion of the project. Many companies are offering genetic tests which can show the tendencies of an individual to various illnesses. Comparing human genome with the genomes of other organisms will help our understanding of evolution better. Some ethical, legal, and social issues are questioned as a result of this project. Possible discrimination based on the genetic structure of an individual by the employers is one such concern and may result in unbalance in societies. However, this project has provided data that can be used to find molecular roots of diseases and search for cures.

2.7 Chapter Notes

We have reviewed the basic concepts of molecular biology at introductory level. The processes are evidently much more complex than outlined here. More detailed treatment of this topic can be found [1,20,29,30]. The cell is the basic unit of life in all organisms. The two types of cell are the eurokaryotes which are cells with nuclei and prokaryotes which do not have nuclei. Both of these life forms have the genetic material embedded in their DNA. Human DNA consists of a sequence of smaller molecules called nucleotides which are placed in 23 pairs of structures called chromosomes. A sequence in a chromosome that codes for a protein is called a gene. Genes identify amino acid sequences which form the proteins. The central dogma of life consists of two fundamental steps called transcription and translation. During transcription, a complementary copy of a DNA strand is formed and then the introns are extracted to form mRNA which is carried out of nucleus and the ribosomes form the amino acids prescribed using cRNA and tRNA. The three nucleotides that prescribe an amino acid is called a codon and the genetic code provides the mapping from a codon to an amino acid. Proteins also interact with other proteins forming protein–protein interaction (PPI) networks and their function is very much related to their hierarchical structure and also their position in the PPI networks.

We also briefly reviewed the biotechnologies for DNA multiplying, namely cloning and PCR technologies. These techniques are needed to provide sufficient

amount of DNA to experiment in the laboratories. DNA sequencing is the process of obtaining nucleotide sequence of DNA. The databases for DNA and protein sequences contain data obtained by various bioinformatics projects and are presented for public use. DNA microarrays provide snapshots of DNA expression levels of vast number of genes simultaneously and gene expression omnibus (GEO) [11] from NCBI and ArrayExpress [14] from EBI are the two databases for microarray-based gene expression data. There are also pathway databases which provide data for biochemical pathways, reactions, and enzymes. Kyoto Encyclopedia of Genes and Genomes (KEGG) [13,18] and BioCyc [6] are two such databases.

The computer science point of view can be confined to analysis of two levels of data in bioinformatics: the DNA/RNA and protein sequence data and the data of biological networks such as the PPI networks. Our main focus in this book will be the sequential and distributed algorithms for the analysis of these sequence and network data.

Exercises

1. For the DNA base sequence S = AACGTAGGCTAAT, work out the complementary sequence S' and then the complementary of the sequence S'.
2. A superficial gene has the sequence CCGTATCAATTGGCATC. Assuming this gene has exons only, work out the amino acid of the protein to be formed.
3. Discuss the functions of three RNA molecules named tRNA, cRNA, and mRNA.
4. A protein consists of the amino acid sequence A-B-N-V. Find three gene sequences that could have resulted in this protein.
5. Why is DNA multiplying needed? Compare the cloning and PCR methods of multiplying DNA in terms of technology used and their performances.

References

1. Alberts B, Bray D, Lewis J, Raff M, Roberts K (1994) Molecular biology of the cell. Garland Publishing, New York
2. Bairoch A, Apweiler R (1999) The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. Nucleic Acids Res 27(1):49–54
3. Bartlett JMS, Stirling D (2003) A short history of the polymerase chain reaction. PCR Protoc 226:36
4. Bentley DR (2006) Whole-genome resequencing. Curr Opin Genet Dev 16:545–552
5. Brandenberg O, Dhlamini Z, Sensi A, Ghosh K, Sonnino A (2011) Introduction to molecular biology and genetic engineering. Biosafety Resource Book, Food and Agriculture Organization of the United Nations
6. Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Pujar A, Shearer AG, Travers M, Weerasinghe D, Zhang P, Karp PD (2011) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res 40(Database issue):D742D753

7. CBD (Convention on Biological Diversity). 5 June 1992. Rio de Janeiro. United Nations
8. http://www.ebi.ac.uk/training/online/course/ebi-next-generation-sequencing-practical-course
9. http://www.ebi.ac.uk/embl/
10. http://www.ncbi.nlm.nih.gov/genbank
11. http://www.ncbi.nlm.nih.gov/geo/
12. http://www.iontorrent.com/. Ion Torrent official page
13. http://www.genome.jp/kegg/pathway.html
14. http://www.ebi.ac.uk/arrayexpress/. Website of ArrayExpress
15. http://www.uniprot.org/. Official website of PIR at Georgetown University
16. http://pir.georgetown.edu/. Official website of PIR at Georgetown University
17. http://www.pdbj.org/. Official website of Protein Databank Japan
18. Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28(1):2730
19. Kneale G, Kennard O (1984) The EMBL nucleotide sequence data library. Biochem Soc Trans 12(6):1011–1014
20. Lewin B (1994) Genes V. Oxford University Press, Oxford
21. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS et al (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376–380
22. Mewes HW, Andreas R, Fabian T, Thomas R, Mathias W, Dmitrij F, Karsten S, Manuel S, Mayer KFX, Stmpflen V, Antonov A (2011) MIPS: curated databases and comprehensive secondary data resources in 2010. Nucleic Acids Res (England) 39
23. Sanger F, Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 94(3):441–448
24. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74(12):5463–5467
25. Setubal JC, Meidanis J (1997) Introduction to computational molecular biology. PWS Publishing Company, Boston
26. Tateno Y, Imanishi T, Miyazaki S, Fukami-Kobayashi K, Saitou N, Sugawara H et al (2002) DNA data bank of Japan (DDBJ) for genome scale research in life science. Nucleic Acids Res 30(1):27–30
27. Voet D, Voet JG (2004) Biochemistry, 3rd edn. Wiley
28. Watson J, Baker T, Bell S, Gann A, Levine M, Losick R (2008) Molecular biology of the gene. Addison-Wesley Longman, Amsterdam
29. Watson JD (1986) Molecular biology of the gene, vol 1. Benjamin/Cummings, Redwood City
30. Watson JD, Hopkins NH, Roberts JW, Steitz JA, Weiner AM (1987) Molecular biology of the gene, vol 2. Benjamin/Cummings, Redwood City
31. Wu C, Nebert DW (2004) Update on genome completion and annotations: protein information resource. Hum Genomics 1(3):229–233

http://www.springer.com/978-3-319-24964-3

Distributed and Sequential Algorithms for Bioinformatics

Erciyes, K.

2015, XVII, 367 p. 157 illus. in color., Hardcover
ISBN: 978-3-319-24964-3

Chapter 2
Predictive Genetic Testing

The second chapter gives an introduction to and analysis of predictive genetic testing (PGT). This chapter explains PGT to identify the distinguishing characteristics of PGT that shape the revised model of consent in future chapters.

In order to fully understand the distinguishing characteristics, the beginning of the chapter will focus on the science of PGT. Having an understanding of the science behind this testing will help to elicit and further develop the three distinctive characteristics for this and future chapters. This chapter will introduce each of the three characteristics that arise from the science of PGT. As a result, this chapter establishes the foundation of the analysis for all the subsequent chapters.

The first part of Chap. 2 focuses on the background and science of PGT in order to get a better idea of the technical aspects behind genetic testing. The second part deals with the first distinguishing characteristic of PGT, understanding genetic risks and probabilities. Because genetic risks and probabilities are typically unique to PGT, this section will look at possible misunderstandings that can arise from not fully understanding genetic information. The third part looks at treatment options for diagnosed genetic traits. PGT can be different from other medical tests in that this testing can return results for diseases that have no treatment or preventative measures. This section will analyze that fact and identify some possible treatments that might help with PGT. The fourth area analyzes family-related genetic information. While some medical tests can involve family information, PGT has significant implications for family members of the individuals being tested. This section will discuss different styles of communication and disclosure of test results, and it will also look at the different concerns of family-related information for genetics. Then this fourth area will analyze the ideas of genetic exceptionalism versus normal medical information in regards to family-related genetic information. The fifth section will conclude with a summary of the implications and the distinguishing characteristics of PGT. By identifying the distinguishing characteristics of genetics, the chapter will highlight the power and limitations of genetic information and testing. The conclusion will bring together all the aspects of PGT that can influence the revised model of informed consent.

J. Minor, *Informed Consent in Predictive Genetic Testing*,
DOI 10.1007/978-3-319-17416-7_2

A. The Science Behind PGT

This section will look at PGT and the science behind it including issues of new genetic technologies and genetic information associated with specific diseases. Predictive genetic testing (PGT) looks at those at risk, the asymptomatic people.[1] One of the differences between PGT and a typical medical diagnostic test is the fact that PGT looks at the future while a diagnostic test gives information concerning the present.[2]

An article by Philip Mitchell, Bettina Meiser, Alex Wilde, et al. defines genetic testing as a test "used to identify a particular genotype (or set of genotypes) for a particular disease in a particular population for a particular purpose."[3] Philip Mitchell, Bettina Meiser, Alex Wilde, et al. go on to say that population is important because of the positive predictive value (PPV). The predictive value of a test is influenced by how often that disease occurs in a specific population. Generally there are three concepts used to evaluate a test: analytical validity, clinical validity, and clinical utility. The article by Philip Mitchell, Bettina Meiser, Alex Wilde, et al. says that analytical validity is related to the reliability and accuracy of a specific test. In genetic testing this is the ability of the test to identify the specific genotype. Analytical validity answers the question of does the genetic test actually identify what it is supposed to identify such as a Huntington's mutation or the BRCA mutation. Clinical validity is "determined by: (1) the strength of evidence for the link between genotype and disease; and (2) test performance characteristics such as sensitivity, specificity, positive and negative predictive values, and likelihood ratios."[4] Clinical validity looks at the accuracy and consistency of test performance. Clinical utility looks at the actual value of the test. This area answers the question of whether or not the test provides information that could be useful to the patient. There are 8 areas that should be analyzed in clinical utility. The testing purpose should look at areas of legitimacy, efficacy, effectiveness, and appropriateness. The possibility of testing should analyze areas of acceptability, efficiency of economic evaluation, optimality of economic evaluation, and equity of resources.[5]

This science behind PGT will look at four different areas including the science, utility, benefits, and risks of testing. The science will look at the technical aspects of testing, the utility will provide the basic value and purpose of testing, and then the benefits and risks of testing will be analyzed.

---

[1] Anita Silvers and Michael Ashley Stein, "An Equality Paradigm for Preventing Genetic Discrimination," 1342; Michael J. Green and Jeffrey R. Botkin, "'Genetic Exceptionalism" in Medicine: Clarifying the Differences between Genetic and Nongenetic Tests," 571.

[2] James Evans, Cecile Skrzynia, and Wylie Burke, "The Complexities of Predictive Genetic Testing," 1053.

[3] Philip Mitchell, Bettina Meiser, Alex Wilde, et al., "Predictive and Diagnostic Genetic Testing in Psychiatry," 227–228.

[4] Philip Mitchell, Bettina Meiser, Alex Wilde, et al., "Predictive and Diagnostic Genetic Testing in Psychiatry," 227–228.

[5] Philip Mitchell, Bettina Meiser, Alex Wilde, et al., "Predictive and Diagnostic Genetic Testing in Psychiatry," 227–228.

1. Science

The purpose of PGT is to assess an individual's risk of developing a specific disease. PGT analyzes DNA and genetics to identify whether a person will develop or already has an at-risk gene that causes them to have a higher or lower than an average risk of developing a specific disease. In order to make predictions about a person's risk for developing a specific disease, scientists analyze areas in a person's genes and chromosomes for genetic variants and mutations.[6] PGT looks at "polymorphisms that increase the probability of disease development."[7] Mutations can be identified by looking at segments of genes and chromosomes in order to identify genes that are different from the normal size and shape of a specific gene.[8] Currently, studies have identified the genetic variants that might be linked to a higher risk of at least 40 diseases.[9] Also studies have identified over 1000 genetic variants associated with a risk of disease.[10] The genetic variants make up what is called a single nucleotide polymorphism (SNP).Generally companies predict genetic risks by "calculating how often that condition occurs among people of the customer's general age, sex and ethnicity, then factor in the presence or absence of the relevant SNP."[11] It is important to note that the results depend on what genetic variants the laboratory decides to use when analyzing the predicted risk of disease.[12] One company could be analyzing 9 variants, while the other company is analyzing 13 variants. The number of variants analyzed can make a difference in the results as well. Once a genetic variant is associated with a risk of a particular disease, tests can be run to look at the SNPs in a specific sample in order to determine risk. If the individual does have the mutation, then it means that that individual has a higher chance of developing that specific disease or cancer than a person without that mutation.[13]

Diseases can be caused by a number of genetic and environmental factors. An important aspect to predicting disease is the fact that there can be several variants in a gene and those variants can be linked to several different diseases.[14] SNPs can be associated with many different areas. One SNP could be more common in a person

---

[6] Anita Silvers and Michael Ashley Stein, "An Equality Paradigm for Preventing Genetic Discrimination," 1346–1347.

[7] Michael J. Green and Jeffrey R. Botkin, "'Genetic Exceptionalism" in Medicine: Clarifying the Differences between Genetic and Nongenetic Tests," 571.

[8] Abigail L. Rose, Nikki Peters, Judy A. Shea, et al., "Attitudes and Misconceptions about Predictive Genetic Testing for Cancer Risk," 146.

[9] Peter Kraft, Ph.D. and David Hunter, "Genetic Risk Prediction—Are We There Yet?" 1701.

[10] Pauline C. Ng, Sarah S. Murray, Samuel Levy, et al., "An Agenda for Personalized Medicine," 724–725.

[11] Chris Berdik, "Genetic Tests Give Consumers Hints About Disease Risk; Critics Have Misgivings," 2.

[12] Francis Collins, *The Language of Life,* xxi–xxii.

[13] Abigail L. Rose, Nikki Peters, Judy A. Shea, et al., "Attitudes and Misconceptions about Predictive Genetic Testing for Cancer Risk," 146.

[14] Anita Silvers and Michael Ashley Stein, "An Equality Paradigm for Preventing Genetic Discrimination," 1346, 1385.

with breast cancer, Parkinson's disease, or another type of cancer.[15] The BRCA 1/2 gene mutations were identified in 1994 and 1995.[16] These mutations increase the likelihood of developing breast and/or ovarian cancer. Generally many different genes are involved in the development of a disease. For example, Anita Silvers and Michael Ashley Stein in the *Vanderbilt Law Review* say that researchers have found over eight hundred gene mutations that are linked to cystic fibrosis. Since there are so many different variations and mutations for this disease, the predictability of developing cystic fibrosis can vary greatly from variation to variation. One genetic mutation can impact the severity of the disease while another mutation might not impact the disease at all. Anita Silvers and Michael Ashley Stein in the *Vanderbilt Law Review* state that "identical mutations in such genes will affect individuals from different populations to different degrees because of variations in environmental factors."[17] Each mutation does not carry the same weight.

Huntington's disease is more of a unique disease for PGT in that the test has a higher degree of certainty for predicting. In 1993, the genetic variation for Huntington's disease was identified, and as a result PGT was offered for at-risk individuals. Huntington's disease is an incurable neurodegenerative disease with no medical treatments available to slow the progression.[18] There are approximately 30,000 individuals with Huntington's disease, but it is suggested that there are another 200,000 that have not been tested yet and are at risk in the United States. This disease is generally categorized as a late-onset condition, because typically the disease manifests itself around 40 years old. The test for Huntington's disease has 100% penetrance, which means if the gene is present, then the individual will develop Huntington's disease at some point in his or her life. On the other side, if the gene is not present, then the individual is not at risk for Huntington's disease. With this disease, either the Huntington's gene is present or it is absent. Because of the certainty and the predictive value of testing for Huntington's disease, the results for PGT are more clear cut. The science behind PGT for Huntington's disease is not as complicated as other diseases. However, there are other issues to consider with this testing such as the non-existent treatment options and the difficult family implications of this predictive testing, and both of these issues will be discussed in later sections.[19]

The confidence in predictive genetic testing can be both over- and underestimated at times. Most genetic tests will not predict with certainty the likelihood of develop-

[15] Chris Berdik, "Genetic Tests Give Consumers Hints About Disease Risk; Critics Have Misgivings," 2; Marion Harris, Ingrid Winship, and Merle Spriggs, "Controversies and Ethical Issues in Cancer-Genetics Clinics," 301.

[16] Abigail L. Rose, Nikki Peters, and Judy A. Shea, et al., "Attitudes and Misconceptions about Predictive Genetic Testing for Cancer Risk," 145.

[17] Anita Silvers and Michael Ashley Stein, "An Equality Paradigm for Preventing Genetic Discrimination," 1385.

[18] Susan M. Cox, "Stories in Decisions: How At-Risk Individuals Decide to Request Predictive Testing for Huntington Disease," 258.

[19] Kathryn Holt, "What Do We Tell the Children? Contrasting the Disclosure Choices of Two HD Families Regarding Risk Status and Predictive Genetic Testing," 254–255.

ing a specific disease, because most diseases have a number of genetic and environmental aspects to them. Huntington's disease, however, has a genetic test that will predict with a high degree of certainty that a person will develop the disease in the future. But even with a disease like Huntington's, there is no way to predict how the disease will affect an individual person. Many of the tests are subject to uncertainty, false positives or negatives, and possible misinterpretation. Anita Silvers and Michael Ashley Stein in the *Vanderbilt Law Review* say "neither now nor in the future will someone's genetic makeup forecast that person's future health condition with certainty."[20] There are many factors that can influence the development and severity of the disease or illness. Some of the factors that influence the predictive value of the test include the differences of gene expression, accuracy of the specific test, and reliability of the research. But on the other side, Anita Silvers and Michael Ashley Stein in the *Vanderbilt Law Review* say "it is equally misleading to say that basing health predictions on genetic testing is 'little more than medical speculation.'"[21] So while PGT cannot attest to the severity and/or certainty of a specific disease, this testing does have a certain value and legitimacy for medical care.[22]

Also included within the science of PGT are the inheritance patterns for disease. If one parent is homozygous for a specific disease, then he or she has two copies of the mutation. If a parent is heterozygous for a disease, then he or she has one copy of the mutation and one copy of the normal gene. Because there are different patterns of inheritance, not all disease inheritance is the same. For example, Huntington's disease is an autosomal dominant disease. The inheritance pattern for Huntington's disease is easy, because it is only concerned with one gene. Because it is dominant, the disease will be passed down to the children if any of the children inherit one mutated gene. The Punnett square below further illustrates this concept. The capital "H" is the mutation causing Huntington's disease, while the lower case "h" represents the normal gene (Table 2.1).

The box represents one parent with Huntington's disease (Hh in bold) and one parent without Huntington's disease (hh in italic). When children are born, each child will have a 50% chance of inheriting Huntington's disease (Hh) and a 50% chance of not having Huntington's disease (hh). This is a basic representation of genetic inheritance patterns. While this seems fairly easy, the inheritance pattern for

Table 2.1 Punnett square for Huntington's disease

	H	**h**
h	Hh	hh
h	Hh	hh

[20] Anita Silvers and Michael Ashley Stein, "An Equality Paradigm for Preventing Genetic Discrimination," 1347–1348.

[21] Anita Silvers and Michael Ashley Stein, "An Equality Paradigm for Preventing Genetic Discrimination," 1347–1348.

[22] Susan Wolf and Jeffrey Kahn, "Genetic Testing and the Future of Disability Insurance: Ethics, Law & Policy," 8; Anita Silvers and Michael Ashley Stein, "An Equality Paradigm for Preventing Genetic Discrimination," 1347–1348.

Huntington's disease is probably one of the easiest. Most other diseases that can be predicted with PGT are multifactorial diseases. There are several factors that influence the inheritance of the BRCA mutation. Even if a person has the mutation, it is not necessarily a positive diagnosis since there are other influencing factors such as the environment and the individual's lifestyle.[23]

2. Utility

Sometimes PGT is recommended based on the utility of the testing, and other times it is not recommended because of a lack of insufficient evidence of testing benefit. Some instances of increased utility include the following: "high morbidity and mortality of disease, effective but imperfect treatment, high predictive power of the genetic test (high penetrance), high cost or onerous nature of screening and surveillance methods, and preventive measures that are expensive or associated with adverse effects."[24] On the other hand, decreased utility for predictive genetic testing include: "low morbidity and mortality of disease, highly effective and acceptable treatment, poor predictive power of the genetic test (low penetrance), availability of inexpensive, acceptable, and effective screening and surveillance methods, and preventive measures that are inexpensive, efficacious, and high acceptable—for example, vaccination."[25] James Evans, Cecile Skrzynia, and Wylie Burke in "The Complexities of Predictive Genetic Testing," from the *British Medical Journal* say that the usefulness of the test can decrease if the disease is curable. For example, James Evans, Cecile Skrzynia, and Wylie Burke suggest that when breast and colon cancer are able to be cured or treated by effective and safe measures, then the benefit of testing is reduced. However, if the disease is curable and is identified earlier, then the disease might be able to be cured earlier rather than later. Evans, Skrzynia, and Burke also suggest that if there are successful and economical screening tools in place for certain diseases, then the utility of PGT will decrease. One example given is of hypertension. Since there are acceptable screening methods that are not expensive, there is not a need to participate in PGT for hypertension. James Evans, Cecile Skrzynia, and Wylie Burke suggest that if the cost of screening is much higher, then PGT will be more economical and attractive to individuals. Evans, Skrzynia, and Burke also suggest that in order for PGT to have higher utility, the preventive measures would generally have some problems and/or be fairly costly. For example, the utility of PGT typically can increase when a person is at risk for breast cancer and is considering a prophylactic mastectomy. James Evans, Cecile

[23] Heidi Chial, "Mendelian Genetics: Patterns of Inheritance and Single-Gene Disorders;" Susan Wolf and Jeffrey Kahn, "Genetic Testing and the Future of Disability Insurance: Ethics, Law & Policy," 8.

[24] James Evans, Cecile Skrzynia, and Wylie Burke, "The Complexities of Predictive Genetic Testing," 1054.

[25] James Evans, Cecile Skrzynia, and Wylie Burke, "The Complexities of Predictive Genetic Testing," 1054–1055.

Skrzynia, and Wylie Burke say "when prevention is simple, however, the value of testing decreases," and the example of a vaccination is given.[26] Since vaccinations are so easy to prevent diseases, there is no need for PGT of diseases like measles, mumps, and rubella. The utility of PGT can play an important role in the utilization of testing and the risks and benefits of testing.

3. Benefits

Reasons to undergo testing include motivational and emotional. Motivational reasons include "early detection, prevention, and control."[27] The goal of PGT is to identify which people have the mutation so that additional monitoring can take place for those at risk. Identification and monitoring of at-risk individuals will hopefully "lead to reduced morbidity and mortality through targeted screening, surveillance, and prevention."[28] If there is additional monitoring, then the hope is that there can be early diagnosis of the disease. PGT can help to monitor those at increased risk and decrease the amount of screening for those that are not at risk or are low risk. Possible prevention and treatment plans are another benefit of testing. Sometimes there can be surgeries or chemotherapy that can help to decrease a person's risk for a specific disease.[29] Also PGT could help with future plans and "may lead individuals to alter their diet or avoid exposure to certain chemicals in an attempt to avoid future disease."[30] The hope is that people will avoid "risk-inducing behaviors."[31] Shoshana Shiloh and Shiri Ilan, the authors of "To Test or Not To Test? Moderators of the Relationship Between Risk Perceptions and Interest in Predictive Genetic Testing," described a study about risk perceptions and testing utilization. The study concluded that the high interest in the test was associated with "both motivations and especially with emotional-reassurance motivation, but not with risk perceptions, health/illness orientations, and cancer anxiety."[32] The study demonstrated that risk perceptions did not necessarily lead to increased test utilization or interest in PGT. Shoshana Shiloh and Shiri Ilan say that understanding the perceived risk is crucial but that is not enough to alter behaviors including fitness modifications. In order to change behaviors and goals, there needs to be psychological changes as well. The

---

[26] James Evans, Cecile Skrzynia, and Wylie Burke, "The Complexities of Predictive Genetic Testing," 1054–1055.

[27] Clara L. Gaff, Veronica Collins, Tiffany Symes, et al., "Facilitating Family Communication about Predictive Genetic Testing: Probans' Perceptions," 133.

[28] James Evans, Cecile Skrzynia, and Wylie Burke, "The Complexities of Predictive Genetic Testing," 1052.

[29] Neil Sharpe and Ronald Carter, *Genetic Testing: Care, Consent, and Liability,* 269–70.

[30] Susan Wolf and Jeffrey Kahn, "Genetic Testing and the Future of Disability Insurance: Ethics, Law & Policy," 8.

[31] Neil Sharpe and Ronald Carter, *Genetic Testing: Care, Consent, and Liability,* 269–70.

[32] Shoshana Shiloh and Shiri Ilan, "To Test or Not To Test? Moderators of the Relationship Between Risk Perceptions and Interest in Predictive Genetic Testing," 471.

article by Shoshana Shiloh and Shiri Ilan concludes that in order to have informed consent and decision making, there needs to be objective and reasonable information about the possible risks and benefits. Also there can be psychological and emotional benefits of testing. Emotional motivations include eliminating uncertainty, gaining support or hope, and preparing emotionally.[33] Finding out that a person is not at risk or is at a very low risk of developing a certain disease can often decrease his or her anxiety levels. This can also lead to a greater "self-perception." Knowing a person's risk status can help to alter or reinforce their view of themselves. Sometimes PGT can result in a greater "sense of control," because the person at risk can follow certain procedures and/or treatments that could potentially decrease their risk.[34] In some people's minds, PGT can help to gain control by knowing their risk status and/or organizing events in their life and future.[35] Motivational and emotional reasons for testing can be benefits of PGT.

There can also be future benefits. Anita Silvers and Michael Ashley Stein in the *Vanderbilt Law Review* suggest that some genetic variants and mutations could have different functions and outcomes than originally thought. "In the future, scientists could discover that having a particular breast cancer gene mutation correlates with immunity from AIDS (as sickle-cell trait correlates with heightened immunity to malaria)."[36] Even with emotional, motivational, and future benefits, sometimes there can also be anxiety, worry, and discrimination that can impact themselves and their family.[37]

4. Risks

Shoshana Shiloh and Shiri Ilan in their article, "To Test or Not To Test? Moderators of the Relationship Between Risk Perceptions and Interest in Predictive Genetic Testing," say that while PGT can promote encouragement and hope, it can also "cause considerable distress to others from premature knowledge of likely illness."[38] Discrimination and psychological harms are often cited as the main harms, but there can be others associated with PGT. Sometimes there can be false assurances which can cause problems for future treatment. False assurances come from getting a lower risk than what is actually true. On the other hand, there can be problems from getting a higher risk than what is actually true. If a woman is told she

[33] Shoshana Shiloh and Shiri Ilan, "To Test or Not To Test? Moderators of the Relationship Between Risk Perceptions and Interest in Predictive Genetic Testing," 471–472, 476–477.

[34] Neil Sharpe and Ronald Carter, *Genetic Testing: Care, Consent, and Liability,* 269–70.

[35] Abigail L. Rose, Nikki Peters, Judy A. Shea, et al., "Attitudes and Misconceptions about Predictive Genetic Testing for Cancer Risk," 148.

[36] Anita Silvers and Michael Ashley Stein, "An Equality Paradigm for Preventing Genetic Discrimination," 1385.

[37] Abigail L. Rose, Nikki Peters, Judy A. Shea, et al., "Attitudes and Misconceptions about Predictive Genetic Testing for Cancer Risk," 148.

[38] Shoshana Shiloh and Shiri Ilan, "To Test or Not To Test? Moderators of the Relationship Between Risk Perceptions and Interest in Predictive Genetic Testing," 469.

has a high risk for developing breast cancer, then she will make decisions based on that information such as having a prophylactic double mastectomy. Problems arise when people adopt "irreversible, risk-inducing or expensive risk prevention strategies based upon incorrectly high estimations of risk."[39] Also sometimes people can have feelings of powerlessness. People often have no control over whether or not they develop a certain disease, and this can cause additional problems emotionally and even physically by not following recommended protocols. The possibility of discrimination and stigmatization can also be a risk of this information.[40]

Discrimination can occur in many areas including employment, insurance, and social situations. Genetic discrimination "arises when individuals with no symptoms or signs receive less favorable or adverse treatment because of their genotype."[41] If the concern about discrimination is high, then sometimes people might not participate in PGT, because he or she is concerned about his family's insurance premiums. If the tests are not conducted, then the individual could be missing out on possible treatment options. In a study with 163 cancer geneticists, about 68 % of the geneticists said that if a person underwent testing for BRCA1 or 2 or hereditary non-polyposis colorectal cancer (HNPCC) they would not bill insurance so that there would be no discrimination. Also with this study, 26 % of the geneticists said that they were in favor of using an alias for testing so as not to cause potential discrimination.[42] Another study conducted by phone found that people were also discriminated against because a family member had a hereditary genetic disease. Anita Silvers and Michael Ashley Stein in the *Vanderbilt Law Review* say that having a negative test result could result in other prospects. Anita Silvers and Michael Ashley Stein continue to say that "Proof that they are not at risk will reassure them of their ability to succeed in endeavors aversive for people who develop the disease."[43] Sometimes people who are at-risk will not participate in certain events or careers in life, because the individuals assume there is nothing he or she can do. Other times, people do not allow individuals to participate, because people consider those individuals to be at risk.[44] Discrimination can occur in many different areas and activities. At the beginning, individual and family discrimination played a significant part in PGT, but in 2003, the Genetic Information Non Discrimination Act (GINA) was established. This act tries to eliminate employment and insurance discrimination.[45] How-

---

[39] Neil Sharpe and Ronald Carter, *Genetic Testing: Care, Consent, and Liability,* 270.

[40] Douglas Martin and Heather Greenwood, "Public Perceptions of Ethical Issues Regarding Adult Predictive Genetic Testing," 107.

[41] Marion Harris, Ingrid Winship, Merle Spriggs, "Controversies and Ethical Issues in Cancer-Genetics Clinics," 304–305.

[42] Marion Harris, Ingrid Winship, Merle Spriggs, "Controversies and Ethical Issues in Cancer-Genetics Clinics," 304–305.

[43] Anita Silvers and Michael Ashley Stein, "An Equality Paradigm for Preventing Genetic Discrimination," 1349.

[44] Anita Silvers and Michael Ashley Stein, "An Equality Paradigm for Preventing Genetic Discrimination," 1349.

[45] Marion Harris, Ingrid Winship, Merle Spriggs, "Controversies and Ethical Issues in Cancer-Genetics Clinics," 305; Cynthia Marietta and Amy McGuire, "Direct-to-Consumer Genetic Testing: Is It the Practice of Medicine?" 370.

ever, Susan Wolf and Jeffrey Kahn in "Genetic Testing and the Future of Disability Insurance: Ethics, Law & Policy," say that the "fear of discrimination is important, as individuals may decide to forego genetic testing (even when it might prove medically useful) in order to protect themselves against insurance discrimination."[46]

Another possible concern about knowing a person is at risk is psychological harms.[47] Marita Broadstock, Susan Michie, and Theresa Marteau in "Psychological Consequences of PGT: A Systematic Review" from the *European Journal of Human Genetics*, suggest that there were no significant changes in the emotional suffering for the carriers and non-carriers in a period of 3 years after PGT. The authors suggest a couple of reasons for this finding. In this study, there could be general psychological defense methods that had already been started. Broadstock, Michie, and Marteau say that research suggests that the people participating in testing are often stronger emotionally and are more capable of handling information. By already deciding to take a PGT and coming forward for testing, people tend to have thought about the testing in advance. Another study found that people getting tested for Huntington's disease (HD) generally "had higher ego strength, were more socially extroverted, and had more positive coping strategies than the general population."[48] Often the people considering getting tested for HD have increased experiential knowledge about this disease. Another study with HD looked at those who decided not to undergo testing and then compared them to those who came forward for testing. The study suggested that those who did not undergo HD testing were more pessimistic about the future. The most common attitudes presented in those who participated in the study were denial of the results or test, elimination of uncertainty, or both. Also sometimes people from at-risk families already had strong coping methods. Broadstock, Michie, and Marteau concluded that as genetic testing is increasingly brought into routine medical care, "some of the protective factors associated with the research environment are likely to be reduced," and likely there will be more psychological harms.[49]

However on the other hand, in "Predictive Genetic Testing in Children and Adults: A Study of Emotional Impact" from the *Journal of Medical Genetics,* S. Michie, M. Bobrow, and T. M. Marteau suggest that there is a significant level of anxiety after getting a positive test result. S. Michie, M. Bobrow, and T. M. Marteau say that this conclusion is especially important in adults with not as many psychological resources.[50] One article by Regina E. Ensenauer, Virginia V. Michels, and Shanda S. Reinke, "Genetic Testing: Practical, Ethical, and Counseling Consider-

[46] Susan Wolf and Jeffrey Kahn, "Genetic Testing and the Future of Disability Insurance: Ethics, Law & Policy," 11.

[47] Neil Sharpe and Ronald Carter, *Genetic Testing: Care, Consent, and Liability,* 270.

[48] Marita Broadstock, Susan Michie, Theresa Marteau, "Psychological Consequences of Predictive Genetic Testing: A Systematic Review," 735–736.

[49] Marita Broadstock, Susan Michie, Theresa Marteau, "Psychological Consequences of Predictive Genetic Testing: A Systematic Review," 735–736.

[50] S. Michie, M. Bobrow, and T. M. Marteau, on behalf of the FAP Collaborative Research Group, "Predictive Genetic Testing in Children and Adults: A Study of Emotional Impact," 520, 526.

http://www.springer.com/978-3-319-17415-0

Informed Consent in Predictive Genetic Testing
A Revised Model
Minor, J.
2015, VII, 232 p., Hardcover
ISBN: 978-3-319-17415-0