Susanne B. Haga

The Book of Genes and Genomes

Springer

The Book of Genes and Genomes

Susanne B. Haga

The Book of Genes and Genomes

Springer

Susanne B. Haga
Duke University Center for Applied Genomics and Precision Medicine (CAGPM)
Durham, NC, USA

ISBN 978-0-387-70915-4 ISBN 978-0-387-70916-1 (eBook)
https://doi.org/10.1007/978-0-387-70916-1

This Springer imprint is published by the registered company Springer Science+Business Media, LLC part of Springer Nature.
The registered company address is: 1 New York Plaza, New York, NY 10004, U.S.A.

Preface

Swirling around in the core of our cells and those of every living creature is a twisted work of art. Physically twisted, that is. In all its simplicity, DNA is the common thread of all life-forms, big and small. If we look closely enough, we can find traces of our past and current life, as well as hints of our future. A personal barcode of sorts, each person's DNA is often regarded as the indisputable proof of identification. So, what does it all mean and why should I care? How does it affect me, my family, the environment, and society?

Skim the daily headlines on your phone or computer, and you are bound to see a story announcing some new discovery about a gene related to a disease, virus, behavior, political preference, or anything else you could imagine. Or maybe it is a story about the Neandertals and their relationship with modern humans, as deduced from genomic sequencing. Or maybe someone is proposing to bring back an extinct species, say the woolly mammoth (think Manny from Ice Age!). If you do not read the news, check out literature, television, and movies, and you will find that these fictional outlets have been keen to incorporate DNA and genetics into storylines: Jurassic Park's recreation of dinosaurs, CSI's and Law and Order's (and all of their spin-offs) reliance on DNA analysis to identify criminals, and genetically engineered (accidental or intentional) action movies (e.g., Minority Report, Wolverine, Jupiter Ascending, Spiderman). You can call it "science fiction," but it appears to have escaped that category, blending fiction and fact to create stories that aim to impress us with the precision and power of genetics and now genomics.

Therefore, it is almost impossible to avoid even if you have tried to—genetics and genomics are literally everywhere. It is actually possible that you have become desensitized to all of the news stories. Can everything possibly be linked to DNA you might wonder? I would argue probably yes, since DNA is present in almost every living organism on Earth (and we are now looking for it beyond Earth). But despite the ubiquity of DNA-related stories, scientists are still very much uncovering the secrets of DNA. It really was not until the

last 100 years that scientists got a handle on what exactly was passed on from generation to generation (DNA), what DNA looked like (a double helix), and finally the sequence of the unique code (the order of the four-letter alphabet). What most people do not realize is that scientists are really trying to read a book written in a language completely unlike any other. Every species has its own genetic "codebook," although there are shared instructions between all living organisms. To further complicate matters, every member of a given group will differ just slightly in their codebook. And, on top of that, the genetic codebook (or parts thereof) can have different meanings if "read" in different environments. Confused yet? Welcome to the world of genetics (and science for that matter).

Ironically, my interest in genetics was sparked by its apparent simplicity and precision (as it was presented to me in ninth grade). And as I continued with my education in human genetics during a time of rapid advances in scientific knowledge and technology, it became very clear that genetics is anything but simple and precise. I believe that this also inspired my passion in education to help others understand genetics, either very broadly as this book attempts to do or about one specific genetic test or application. DNA is a chemical, but not to worry—this is not a chemistry book. However, in order to describe some of the medical and nonmedical applications, I have ventured a little into the science of genetics that you may have avoided in school. By bringing together the wide range of applications based on DNA science and medicine in this book with a little scientific explanation, my ultimate hope is to leave you with some knowledge to make you a more critical reader about genetics and genomics and, if needed, to make informed decisions about your health or that of your family members.

Along with the scientific discoveries and the exciting applications developed thus far, we must also consider the ethical and legal concerns and the potential adverse consequences raised by this newfound knowledge and technology. Genetics has benefitted and suffered from the intense focus of recent decades and the negative history associated with eugenics. Thus, not surprisingly, it draws a range of public responses from fascination to fear and trepidation. With the rapid advances in genetics and this new field called genomics, both practical questions (who has access to new applications, who pays for it) and ethical questions (should we really be doing this) have followed the science, sometimes with unclear answers. We all must continue to ask these important questions. A better understanding of the science and related applications should promote more informed and greater public engagement.

Growing up in the pre-genomic era, before fancy sequencing machines could whip out DNA codes, I was in awe of the power of DNA. My

amazement continues to grow as I have witnessed how the knowledge of genetics and genomics has transformed medicine and society in such a relatively short period. It is my goal to share some of this excitement to the readers of this book and possibly create a lifelong follower of the field while imparting a bit of knowledge. Undoubtedly, something has been discovered or developed after I finished writing—that is the difficult part about writing a book on this field. There continues to be so much to learn, and I hope that this is just the start for you.

I am indebted to all of the wonderful teachers, mentors, and collaborators that I have learned from and been inspired by. May I continue to pay it forward.

Durham, NC, USA Susanne B. Haga

Contents

1

From Genes to Genomes in All Living Things

Although this book aims to present several applications of genetics and genomics, if I actually started off with that, I would have to keep stopping to explain the science a bit. Thus, this first chapter aims to provide readers with a short overview of the history of genetics and genomics, starting at a time when the words *genetics* and *genomics* did not yet exist to the present.

Most people are probably quite familiar with the words "gene" or "genetics." A word association game may yield words like family, health/disease, and identification (e.g., paternity, forensics). In contrast, the word "genome" (pronounced jee´nōme) is much newer to our lexicon and thus likely to be quite unfamiliar to many. The genome refers to the entire DNA content found in a given cell (as opposed to a gene, which is one very small part of the genome). Although the term genome was coined in 1920, from the words *gene* and chromo*some* (a condensed form of DNA to be explained more later), it really did not garner much attention by the scientific or medical community until fairly recently, beginning in the 1980s and 1990s, as will explain why later.

In the early part of the century, the knowledge and technology were not available to enable scientists to fully understand the human genome, let alone a single gene. The stepwise process of scientific research can seem painfully slow, but lots of information was being learned about basic cellular processes that are taken for granted today. Now we are able to analyze an unknown sample of DNA extremely rapidly and determine from which species the DNA was from, and potentially determine the exact individual human or animal it came from. But despite the huge advances afforded by new scientific technologies and the generation of a lot of data, scientists are still looking for

S. B. Haga, *The Book of Genes and Genomes*, https://doi.org/10.1007/978-0-387-70916-1_1

answers to questions regarding human health, environment, and other fields that may reside in genetic material (or DNA). Imagine trying to put a 3-billion-piece rocket ship together that you do not have the instruction manual for and then try to figure out how it is supposed to work. Now you will begin to get some inkling of the challenges faced by geneticists and genome scientists in trying to uncover the secrets stored and what each "part" does within a genome.

The 1800s

Flashback, if you will, to a time of intellectual curiosity, and of relative peace and prosperity, and the quiet sanctity of an Augustinian monastery. During the mid-1800s, the town of Brünn, Austria (now Brno, Czechoslovakia), was part of the Austrian Empire and later the Austro-Hungarian Empire. Brünn was a hub for textile industries and agriculture—particularly wool and fruit. In 1850, the town's population had reached about 47,000. Perhaps an unlikely place to begin a chapter on genetics and genomics, but this has been referred to as the origin of the theories of inheritance and genetics.

In 1822, Johann Mendel was one of five children born to a peasant family. His father was a farmer, managing his own crop of fruit trees as well as tending to the fields of the feudal lord whom he worked for 3 days a week. Mendel excelled in school but was shy by nature and often needed to return home due to illness caused by stress. During his pre-university studies, Mendel focused on physics and math. Due to a lack of funds though, he did not immediately continue his education at the university level. His physics teacher recommended Mendel to the Abbot at the Brno monastery for the novitiate (a sort of mentorship program for prospective candidates of a religious order who have not yet been admitted).

In 1843, he was admitted to the Augustinian monastery and was ordained in 1848. Although Mendel was not of deep religious faith, entry into the monastery was a way to continue his education and training in science. He took the name Gregor after entering the monastery. He first served in a role similar to a parish priest to a parish affiliated with the monastery, which included responsibilities like tending to the sick at a nearby hospital. However, the stress due to the constant suffering and pain he witnessed took an emotional tool and he became ill and depressed himself.

Recognizing his struggles in providing comfort to the ill, the Abbot reassigned him to teach math and science to seventh graders. In 1850, as required by law, he took the exam to be a teacher of natural history and physics, but

failed. In 1852, to address his apparent deficit of knowledge in the sciences, he attended the University of Vienna and learned from several well-known research scientists of the time. He recalled that the Abbot had mentioned that the mystery of heredity would only be solved through rigorous experimentation. Upon his return to the monastery in 1853, Mendel began studying pea weevil plants. He was familiar with the techniques of artificial fertilization learned during his childhood experiences with fruit trees.

In 1855, he began some experiments that would support his now famous work on the theories of inheritance, which he would publish 11 years later in the proceedings of the Brunn Society for Natural History. So what exactly did Mendel figure out with his simple pea plant experiments? He presented his findings in 1865, opening with the following introductory remarks:

> Experience of artificial fertilization, such as is effected with ornamental plants in order to obtain new variations in color, has led to the experiments which will here be discussed. The striking regularity with which the same hybrid forms always reappeared whenever fertilization took place between the same species induced further experiments to be undertaken, the object of which was to follow up the developments of the hybrids in their progeny.

To understand the significance of his work, it is helpful to backtrack a moment to provide some context of Mendel's work. For centuries before Mendel, it was well known, by farmers and ranchers in particular, that the qualities of the next generation's crop were dictated in part by the parents. Two parents with certain desirable traits would be mated to produce the next generation with the same desirable traits, giving rise to "pure-bred" strains. Experiments in the 1700s on plant hybrids were beginning to shed light on the inheritance of traits that were believed to have informed Mendel's thinking. Oftentimes (as was expected), the plant hybrids represented a mixture (blending) of the parental traits, but some would occasionally appear more like one parent than the other. The understanding of whether plants sexually reproduced still was not clear (as was readily obvious in animals).

More than 100 years prior to Mendel, the observation of predictable transmission patterns of human diseases, particularly those that affected only one sex, had been recorded by a number of scientists. In 1794, the English chemist John Dalton noted that he and several of his male relatives were affected with color blindness, a condition now understood to predominately affect males.

Mendel's success was due in part to his choice of organisms (the pea plant) and his selective study of traits (those with only two possible outcomes) rather

than more complex traits with multiple potential outcomes. Traits such as pea color (yellow or green) and texture (smooth or wrinkled) produced unambiguous results (Fig. 1.1). Based on repeated observations and tracking of multiple generations of peas for selected traits, Mendel deduced that each trait was due to the combination of two versions (later defined as an "allele") of a gene in each plant. He did not use the term "gene" since it had not been coined until the early 1900s. Instead, Mendel actually used the word "factor" to describe a unit of inheritance, and deduced that one factor was derived (or inherited) from each parent. During reproduction, these two copies would separate and each parent would donate one or the other copy to the offspring (Fig. 1.2).

The breeding of peas for certain traits for several generations revealed that some traits were dominant compared to others. In other words, some traits that appeared in the first generation of offspring were referred to as *dominant*, and those that appeared in the second generation as *recessive*. For example, if a round pea was bred (or crossed) with a wrinkled pea, all of the offspring in the next generation were observed to be round (and thus, the trait roundness was determined to be dominant). However, if two round peas from this first generation were crossbred, the offspring were a mix of round and wrinkled (defined to be recessive). Although Mendel was not the first person to describe

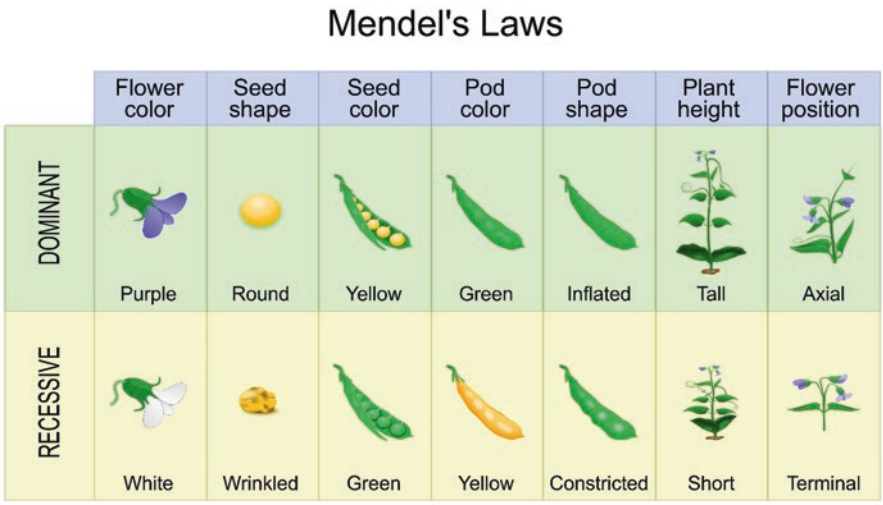

Fig. 1.1 Peas were an ideal organism to study the passage of traits from generation to generation given the wide array of distinct features and multiple combinations possible, e.g., seed form and flower color (source: Adobe Stock)

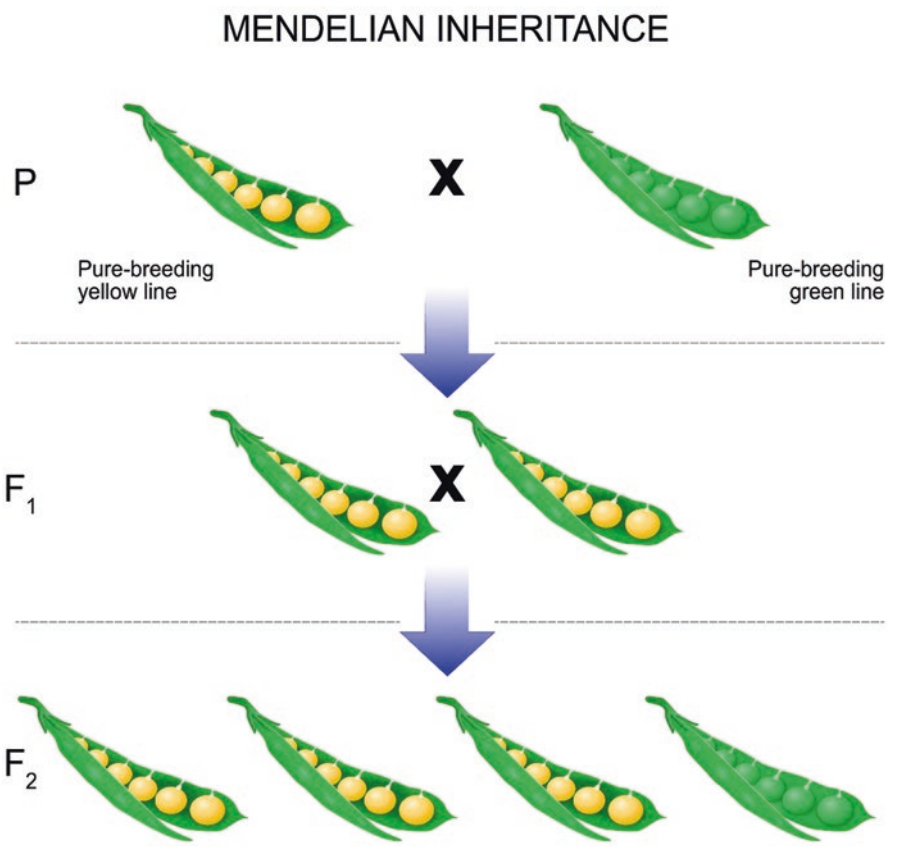

Fig. 1.2 Illustration showing the passage of traits from parents to offspring (P = parents; F1 = first generation; F2 = second generation) (source: Adobe Stock)

the idea of dominant and recessive traits, his experiments unequivocally proved the concepts.

We now know that genes were these "factors" that Mendel described and that they contain the instructions to make proteins, the molecules that actually do the work required of our cells and bodies that would give rise to certain traits or characteristics. If, by chance, one of the genes is mutated or altered, the protein produced from that gene may not work properly, correctly, and/or as efficiently. If the bodily function in which that gene participates requires the normal dose from two copies of the gene to perform its function, the absence of one copy due to a mutation (alteration in the gene) may lead to disease. In this situation, the disease is said to be dominant as only a single mutated copy of a gene will give rise to it. On the other hand, if both copies need to be mutated in order to give rise to a disease, it is referred to as

recessive. In this situation, both parents would likely be carriers of a mutated gene, but neither would be affected. Each parent would have a 50% chance of passing on the mutated gene to the offspring, for a 1 in 4 chance of having an offspring with the recessive disease (inheriting two copies of the mutated gene).

When Mendel studied two or more traits together (e.g., round and yellow peas vs. green and wrinkled peas), he also observed that each individual trait was passed on independently of the other trait and the outcome of one trait was not linked with the outcome of the other traits. All combinations were possible (round/yellow, round/green, wrinkled/yellow, wrinkled/green), although as with the single traits, each combination would appear with different frequencies in every generation. After collecting data on thousands of pea plants and using his assumptions about dominant and recessive alleles, genetic segregation, and independent assortment, Mendel eventually could calculate with precision what ratios could be expected in a given generation for a given trait.

Shortly after his publication, in 1868, Mendel was appointed as the sixth Abbot of the Brunn monastery. With much of his time now devoted primarily to administrative duties, he had far less time to spend on scientific experiments and observation, which appeared to have ceased in 1871. At the age of 61, Mendel died in 1884 without any recognition of the significance of his work by his peers.

More than three decades passed before the significance of Mendel's work was realized. Why did it take so long? One reason was that scientists were focused on other pressing issues of the time—namely that of Charles Darwin and his theories of evolution. It was a challenge, to say the least, to make sense of how Darwin's theories of natural variation and Mendel's theories of inheritance fit together, especially as each was still unclear on its own. More theories on heredity were developed between Mendel's publication and its discovery in the early 1900s, which greatly benefited from advancements at the time in laboratory experimentation, microscopy, and evidence that particular traits or diseases run in families; that is, multiple family members in multiple family generations were affected with the same condition. Eventually, Mendel's "factors" and theories were rediscovered, further advancing the young field of genetics and understanding of inheritance (Fig. 1.3).

The 1900s

As in any field of science, discovering the answer to one question only leads to more questions. Understanding the process of inheritance was only part of the

Fig. 1.3 Commemorative stamp in Czechoslovakia of Gregor Mendel (source: Adobe Stock)

puzzle to be solved. The 1900s were a time of rapid discovery (and rediscovery), when the puzzle pieces seemed to finally fall into place one by one.

The year 1900 marked the recognition of Mendel's work, when three scientists independently noted the significance of Mendel's work. During that year, the British zoologist William Bateson mentioned Mendel's work during a scientific presentation in London. In 1904, Bateson even visited Brno to learn more about Mendel, but no one was able to tell him much about the quiet man and his scientific experiments. In 1902, the British physician Archibald Garrod published his work on the biochemical disease known as alkaptonuria, and determined that this disease was inherited in a "Mendelian" recessive fashion. In 1906, Bateson coined the term "genetics" to describe the study of

heredity. In 1909, Wilhelm Johannsen introduced the word "gene" to denote the unit (the ambiguous term previously used by Mendel) that was passed from parent to offspring.

Confident in their understanding that something was being passed on from generation to generation in predictable patterns, the next big question facing scientists was to determine what exactly was being passed on from parent to offspring—i.e., what exactly was a gene made of? While the European and British scientists had made great contributions in establishing the foundation for the field of genetics, US scientists began making up for lost time with a series of experiments that would confirm that genes were made of DNA. In 1901–1902, separate experiments proved that half of the chromosomes were passed on from the mother and half from the father. Subsequent experiments conducted by US scientists in bacteria and viruses proved that the hereditary material was indeed DNA.

The next question was what exactly did DNA look like—or what was its structure? This was important to learn, as by understanding its structure, scientists might then be able to deduce how it actually functioned, replicated, and was passed on from parent to offspring. It was known that DNA was comprised of four chemical units, abbreviated A, T, C, and G. And it was also known that the number of As and Ts was equal to the number of Cs and Gs in a given sample of DNA. However, the AT/CG ratio differed between species. But how these chemical units were assembled was unknown.

In 1953, the stunningly simple structure of DNA was revealed by the American scientist James Watson and British scientist Francis Crick working together in Cambridge, England. Clues from a special type of X-ray photograph of DNA led them to hypothesize that the structure of DNA was some type of helix. After several attempts to arrange the As, Ts, Cs, and Gs through chemical models, they realized that DNA resembled a twisted ladder, whereby the chemical units were located on the "rungs." Based on the earlier observation about the proportion of As and Ts and Cs and Gs, Watson and Crick predicted that the A units paired only with the T units (A–T) and the Cs paired only with the Gs (C–T) (Fig. 1.4). When cells divide, the DNA has to make a faithful copy of itself to pass on to the next generation or "daughter" cells. Based on this double-helical structure, it was proposed that the DNA strands come apart like a zipper.

These "parent" single strands then serve as the template for two new strands. Thus, half of each new DNA molecule was composed of the original parent strand and a new strand. Because of the understanding that As only connect to Ts and Gs to Cs on the rungs of the DNA ladder, the newly formed DNA

Fig. 1.4 The pairing of DNA subunits (source: BioRender)

was the exact same code or sequence of chemical units present in the parent DNA.

In the years following the discovery of the structure of DNA, much research was done to understand how the DNA code was "read" by the cell's machinery, giving rise to the production of proteins, the molecules that are the "workhorses" of the cell. The chemical units of DNA by itself are not functional. It was eventually determined that the sequence or order of the DNA letters (A, T, C, and G) is read three letters at a time—each three letters encodes an amino acid, the building unit of proteins.

Can We Actually "See" DNA?

How do scientists actually study DNA—is it something that is visible? Yes and no—DNA is visible under some circumstances, but the actual order of the bases (As, Ts, Cs, and Gs) is not visible. When the concept for a microscope was described in the late 1500s and actually built in the mid-1660s by Antonie van Leeuwenhoek, this was a tool that enabled scientists to eventually see the structure of cells and the organelles (the "organs" of a cell) for the first time and to observe the changes that occurred as cells grew and divided. In 1882, scientists first visualized cellular structures called chromosomes (kromo-zomes), which is DNA condensed in coil-like structures (to be discussed in more detail in Chap. 2). Chromosomes resemble squiggly wormlike structures located in the center of cells. One could actually see that there are multiple chromosomes within cells and later determined that different organisms had different numbers of chromosomes. It was then determined that sperm and eggs had half the number of chromosomes observed in other types of cells (and thus, their union would reconstitute the full number of chromosomes). In 1889, it was proposed that the hereditary material was passed on in these chromosomes, but no experimental evidence had yet proven this theory. Thus, the strands of DNA were not visible, but DNA stored in the chromosomes

was helping scientists learn about DNA replication and movement each time the cell divides.

Furthermore, scientists did not need to "see" DNA to study it or to infer the consequences of genetic changes.

But how does one actually find a mutation? It turns out that the human genome is three billion units long, and contains approximately 20,000 genes. Therefore, searching for just one chemical unit that has been mutated out of 3 billion may take some time. The analogy "searching for a needle in a haystack" aptly describes the unbelievably difficult process of finding a gene linked to a particular disease until this century.

In the first half of the twentieth century, scientists were able to map (or determine the location thereof) a gene believed to be responsible for a specific trait (e.g., eye color) to a specific part of a chromosome. Each chromosome contains many genes, so narrowing the region in which the putative culprit gene was located was a very important first step. This type of mapping is somewhat analogous to determining that a house of interest is located in the city of Baltimore, and possibly narrowing down the part of the city to a certain part of town (e.g., North). But no detailed maps of each chromosome existed to further help navigate (a chromosome is linear), nor knowledge of what the region looked like or how far scientists would have to "walk" to find the culprit gene.

In the 1970s, a method to sequence DNA was worked out for the first time. Scientists could decode the A, T, C, and G's for a given piece of DNA (Fig. 1.5). Thus, once a culprit gene was narrowed down to a chromosome and to a particular "neighborhood" along the chromosome, scientists could "walk" up and down the chromosome by sequencing the regions of DNA to figure out what genes were located in a region and if a mutation resided in one. Several genes for Mendelian disorders were identified using this approach, though it could take years to discover since at that time, scientists were only able to "sequence" small segments of DNA.

In the 1980s, the idea to sequence or decode the entire human genome was introduced—the elucidation of the order of As, Ts, Cs, and Gs of the entire DNA content from a human cell. Considered to be the most ambitious scientific endeavor by many, the feasibility and utility of such an unprecedented effort were uncertain. Much of the genome was considered to be "junk" DNA or DNA of unknown significance. Genes actually accounted for a very small proportion of the genome, but they were the target for understanding the genetic causes of disease.

Fig. 1.5 Section of DNA sequence (source: Adobe Stock)

In the 1990s, technological advances enabled scientists to initiate a massive project to determine the full sequence of the human genome (all 3 billion letters). In particular, it was the development of automated sequencing machines and computational sciences required to analyze and store the DNA data that enabled the project to be completed. As a result, the sequence of the genome was completed, chromosome by chromosome, including the location of known genes, each with a specific coordinate (very similar to longitude and latitude that correspond to a unique location). The complete sequence of the human genome was finalized in 2003, 50 years after the structure of DNA was revealed.

What Have We Learned from Sequencing the Human Genome?

In the 1980's, an idea was being discussed amongst the scientific community about sequencing the human genome. After much debate and securing Congressional funding, the Human Genome Project was launched in 1990 with the singular but enormous goal of sequencing the human genome and creating a reference book of sorts for the scientific community. A draft of the

genome was completed in 2000 and finalized in 2003, 50 years after the structure of DNA was discovered. At the outset of the sequencing project, it was still uncertain how many genes humans had. Some had estimated that the human genome contained up to 100,000 genes. But, to the surprise of many, in the first analysis of the draft of the sequence, it was estimated that the human genome contained only between 30,000 and 40,000 genes. A few years later, an analysis of the nearly complete sequence of the human genome revealed that the number of genes was even fewer, between 20,000 and 25,000 genes. In comparison, the genome of the common gut bacterium *E. coli* is 4.6 million letters long with 3200 genes, the mouse genome is 2.5 billion letters long with 28,000 genes, the dog genome is 2.5 billion letters long with 19,000 genes, the fruit fly is 122 million letters long with 13,600 genes, and the roundworm is 97 million letters long with 19,000 genes.

A second major finding of sequencing the human genome was how genetically similar we all are. Despite our obvious physical and personal differences, the DNA sequence between any two people differs by only 0.1%—that is, we are 99.9% genetically alike. When you multiply that by the total number of DNA units (3 billion), there are 3 million places across the genome that we differ at—which accounts for the diversity of the human population. Thus, each person truly has their own unique genome or genetic fingerprint. Many of these differences are referred to as single-nucleotide polymorphisms (pollymorf-isms) or abbreviated as SNPs (pronounced as "snips")—more than 10 million SNPs have been identified. These SNPs can occur in one person or a handful of people (a rare SNP is considered to occur in less than 1% of the population) or a significant proportion of the population (e.g., 30% of the population). These differences in DNA sequence may or may not be significant with respect to changing the function of a protein encoded by the gene, and thus potentially contributing to the development of a disease.

Why Sequence the Genomes of Other Species?

The human genome was actually not the first genome to be sequenced. In 1995, the genome of the virus that causes flu (*Haemophilus influenza*) was sequenced (a much smaller and feasible first attempt to sequence). More than 40 genomes were actually sequenced before the first draft of the human genome was published. But what benefit is there to sequencing the genomes of other species? To begin, almost all organisms use the exact same genetic code as humans, making it easy to identify DNA regions that have a highly similar sequence. Many species share or have a similar version of genes that are

important to fundamental cellular mechanisms. For example, about 99% of human genes have a counterpart in the mouse genome. It turns out that about 60% of genes are conserved between the fruit fly and human. Conserved genes are thought to be involved in essential physiological functions shared by both species whereas a unique gene (found in only one species) suggests a novel characteristic of that species. Thus, determining the sequence of the genomes of other species, particularly of species commonly studied in the laboratory (referred to as model organisms such as yeast and mice), contributes substantially to our ability to interpret and understand our own genome.

Since we are obviously limited in the types of experiments that can be performed on human research subjects, experiments can be conducted on model organisms such as the mouse or fruit fly that carry a gene in common with humans. These types of studies help us to better understand the function of that gene. In particular, scientists can carefully analyze what happens when that gene is mutated and if a particular treatment will work to prevent or improve the consequences of the mutated gene.

Comparing the genomes of different species also enables scientists to study evolution. For example, having the sequence of the dog genome has helped scientists piece together the family tree of the many different types of dog breeds (e.g., poodle, pug, German shepherd) and place them in the larger evolutionary tree with wolves, coyotes, hyenas, and other similar mammals. In addition, some dogs have particular ailments such as hip dysplasia in German shepherds. Understanding the role of genes behind these diseases will help advance the development of diagnostic tests, treatments, and preventative measures for people. In addition to animals and microbes, the genomes of trees and plants, particularly those used for human consumption such as rice, have been sequenced. Like any other organism, plants are susceptible to disease and thrive only under certain conditions. Understanding the role of genes important to disease and growth may enable scientists to develop more effective pesticides or change growing conditions to maximize yield.

Genes and Disease

If the DNA code is altered or changed, meaning that the letters are changed (often referred to as a mutation or variation—the difference in terms will be discussed later), this may change the triplet code for a specific amino acid and possibly the overall structure or function of the intended protein. One can think of an analogy with the English alphabet—one single typo can change the meaning of a word (e.g., hot vs. hat). In a cell, as a result of a change in

the DNA, the corresponding protein encoded by that gene may not properly function (or it actually may not even be made depending on the type of mutation). Thus, the cell's normal operations are disrupted, potentially giving rise to a disease. Such a change may occur as a result of damage to DNA (perhaps through exposure to ultraviolet light) or a random mistake during cell division when DNA replicates.

Not all genetic changes will result in a disruption of the code, such that the corresponding protein is also impacted and dysfunctional. It really depends on where exactly the change occurred and if that spot is essential to the protein, with respect to whether it is even made or how it functions. For those changes that do disrupt the quantity or quality of a protein, the next question to consider is the specific job of the protein and its significance to the cell's normal activities. Remember, we have two copies of each gene—so if one copy incurs a genetic change resulting in a change to the protein, the other is unaffected and therefore should make a normal-functioning protein. Half of the normal amount of protein may be enough for the cell to survive, though this is not always the case if the change behaves in a "dominant" fashion, as described earlier. Yet, in other cases, an intermediate or moderate level of disease may develop with one affected gene. However, if both copies of the gene carry a genetic change, this will more than likely result in a substantial impact on the cell's health and ultimately cause a symptom or disease (a recessive disease).

Although diseases caused by single genes, either in dominant or in recessive fashion, are relatively rare, they collectively account for thousands of patients. For diseases such as heart disease, cancer, and diabetes, these are referred to as *complex* diseases as they are believed to be caused by multiple genetic changes as well as environmental factors. Teasing apart the interactions between multiple genes and environmental factors is extremely difficult, but newer technologies enable a comprehensive snapshot of a set of molecules at a given time. Think of a picture—if you zoom in on one corner of the picture, you may see a lot more detail of the object(s) in that corner, but you will entirely miss the context and other interactions in the rest of the picture. Likewise, looking at a single gene will only yield a partial understanding of what is happening in a cell. Genome sequencing and technologies enable scientists to assess the total picture with respect to genetic variations, determining which genes are turned on at a given time. Taking it another step further, scientists can determine which proteins and chemicals are present that "run" the cell's operations, potentially with different "teams" of proteins working under different conditions (e.g., with and without a drug). In subsequent chapters, some of these technologies will be described in further detail for specific applications.

Beyond the Sequence

While this book and a great deal of the popular press on genetics focus on changes in the sequence of DNA and its connection to traits and diseases, cells have other mechanisms to control when genes are turned on and off. As noted earlier, the human genome encodes for about 20,000 genes. The "expression" of genes refers to when a gene makes a protein; proteins carry out the specific functions or activities in the cell (proteins carry out the jobs, genes contain the instructions to make the protein). Two analogies might be helpful to consider here to understand gene expression and the importance of controlling expression. In a house, there are many lights and appliances—it would make no sense to turn on every light and appliance in every room in the house and leave them on all day and night. Indeed, it would be a tremendous waste of resources and most will not be necessary to carry out the daily functions in a household (at least not all at the same time!). Instead, as a person enters a room, they can turn on the light(s) and any appliances they need to carry out a specific task. Likewise, it would be a huge waste of resources to "turn on" or express every gene in a given cell and leave it on. In addition, different cells (like different rooms of a house) will have different needs, so only certain genes will be expressed.

While the DNA sequence includes instructions to turn on a gene, the sequence can also be modified, often transiently, to signal that the gene should be expressed or not. The study of these chemical modifications is known as epigenetics. Small chemical groups can attach to the DNA as a signal to the cell regarding a gene's expression, without affecting the exact sequence, sort of like a flag to mark the location of a specific gene. More and more is being learned about epigenetics and the additional layer of complexity it adds to how genes are controlled (or go out of control in certain diseases). Epigenetic modifications are passed on from generation to generation but changes can be acquired.

Conclusion

Understanding the structure and codes hidden in DNA has enabled scientists to enter a whole new dimension of biology, one which has revealed almost something new daily and will continue to do so for years to come. Following the completion of a reference human genome, thousands more human genomes and countless other species have been sequenced worldwide. There are public databases full of genomes! New technologies continue to be

introduced, thereby reducing the costs of sequencing and improving accuracy. It will take some time for scientists to wade through this deep trove of data and understand its significance. Furthermore, we are beginning to learn that the sequence is just one part of a much more complex system that can impact health and disease.

Resources

US National Library of Medicine. Genetics. Available at https://medlineplus.gov/genetics/

Scitable. Gregor Mendel and the Principles of Inheritance. Available at https://www.nature.com/scitable/topicpage/gregor-mendel-and-the-principles-of-inheritance-593/

US National Human Genome Research Institute. A Brief Guide to Genomics. Available at https://www.genome.gov/about-genomics/fact-sheets/A-Brief-Guide-to-Genomics

Genomics England. What is a Genome? Available at https://www.genomicsengland.co.uk/understanding-genomics/what-is-a-genome/

Gregor Mendel. Experiments in Plant Hybridization (1865)

What is Epigenetics? Epigenetics: Fundamentals. Available at https://www.whatisepigenetics.com/fundamentals/

2

My Family's Health History (and Why It Is Important for Me to Know)

Most people have heard of the phrase "family history" or the common saying "it runs in the family," but what exactly does it refer to? Does it refer to any family member, past or present? Does it refer to family members living in a household (maybe just parents and children) or more distant relatives? Does it include step-relatives or half-siblings? Does it refer to a family's origins or ancestry? A family history can actually be quite complex, with respect to not only the size and makeup of one's family but also the different types of information that can be associated with a family history. This chapter addresses many of these questions and hopefully inspires some readers to learn about their own family history.

What Is Family (Health) History?

Almost every visit to a health provider will include questions about family history. Among the numerous forms we are asked to complete at the beginning of a visit to almost every health provider, patients are presented with a list or table of conditions and asked to indicate if any family members are affected with any of those conditions by checking a "yes" or "no" box. Some health providers have transitioned to an electronic format and these questions can be answered in advance of the visit or on tablets in the office. In addition, these forms may include questions about smoking or alcohol use, sleep, lead exposure, ethnicity, and drug use. Some forms may include more detailed questions than others. The significance of these questions may not be clear,

S. B. Haga, *The Book of Genes and Genomes*, https://doi.org/10.1007/978-0-387-70916-1_2

especially if the provider does not review your responses or asks for further information.

The collection and documentation of a family history is actually a very old practice and its purpose has shifted over the centuries (Fig. 2.1). A family history can include lots of different types of information that can serve different purposes. In medieval times, documentation of a family's descendants was important for determining class and societal privileges. In the mid-1500s, the English monarchy required churches to document and store a family's heritage and relationships, and document christenings, marriages, and burials. Land ownership was also included in genealogical records and taxation was determined based on these records. Thus, churches have stored a vast amount of information of important events and families dating back centuries.

In today's medical setting, which is the primary focus of this chapter, a family history typically refers to health-related information of the patient and his or her family members (hence, the phrase "family health history" may be used instead of "family history"). Collecting medical history about family members is a standard practice for almost every type of health provider, although there is no standard set of questions. Increasingly recognized as important factors to our health and disease risk, providers are also collecting information about social factors such as occupation, exercise/activity, education, and home

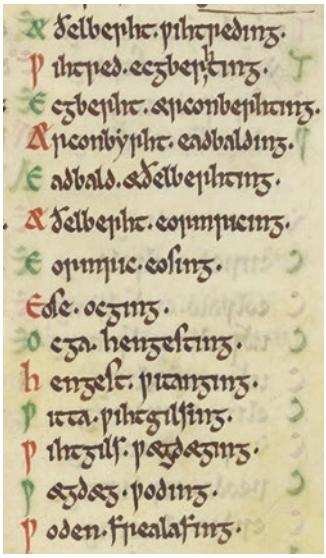

Fig. 2.1 The ancestry of King Æthelberht II of Kent in the Textus Roffensis (twelfth century) (source: https://commons.wikimedia.org/wiki/File:Kentish_tally_(Textus_Roffensis).png)

environment. The information about our environment and family members' health is probably more valuable than many people realize. Both genes and environment can contribute to health risks. Therefore, providing family history information to health providers can help them better estimate future disease risks, recommend appropriate screening to detect disease early, and advise on steps to reduce risks.

So, who exactly counts as "family"? We share our genes with all of our blood relatives. The more distant the relative, the less the genetic similarity (or shared genetic background). For example, we share 50% of our genetic makeup with our parents, children, and siblings. We share 25% of our genetic makeup with grandparents, grandchildren, and aunts and uncles. We share 12.5% of our genetic makeup with cousins, great-great-grandparents, great-great-grandchildren, and great-aunts and -uncles. Recall—our genetic makeup is only shared among biological relatives. So, the woman that married your father's brother does not share any genetic makeup with you or other relatives of her husband.

A household may include both biological and nonbiological family members. The health history of nonbiological members of a household may be of lesser importance with respect to genetics, but there may be some important information about the environment to disclose to your health provider. For example, if you are residing with a nonbiological family member (e.g., boyfriend, girlfriend, roommate, etc.) and they are a heavy smoker, this could impact your health. Shared environment and physical spaces, and shared culture such as lifestyle, physical activity, and food choices, can all impact health.

Collecting a Family History

As mentioned earlier, you have filled out forms at the beginning of a visit to a health provider's office or clinic that includes questions about you and your family members' health. Sometimes it is a long list of conditions that you mark yes/no for yourself and other family members. Alternatively, it could be a series of questions about family members' health. The health provider will review the information you have disclosed on the forms and confirm the details of a family member's condition if checked. If the family health history indicates an increased risk, increased surveillance or referral to a specialist may be recommended.

In some cases, a very detailed family history is required. This is often collected by a specialist, such as a medical oncologist (cancer), medical geneticist, or genetic counselor. If a patient has been referred to a specialist, an inherited

condition may be suspected or present a high risk. A detailed family health history involves review of the health status of three generations of family members, typically the generation before (grandparents) and the generation after (children) for adult patients. Birthdate, current health conditions, age of onset, miscarriages, cause and/or date of death, and ancestry are collected. Depending on the size of the family and health status, this information can take a while to gather and record. More often than not, patients will not know all of the information and may need to follow up with family members after the visit.

Cultural sensitivity is important to keep in mind when collecting and discussing family health information. Different cultures use different terminologies to describe familial relationships or kinships. In some groups, whether someone is biologically related or not may not be distinguished. For example, a person who is a very close friend of the family may be referred to as an "aunt" though she is not related biologically or through marriage. In other cultures, siblings and cousins may not be distinguished. Thus, some care must be taken to insure that common terminologies are not misinterpreted and inaccurate information is collected. Furthermore, discussion of some health conditions (e.g., mental health) among family members may not be accepted in some cultures and therefore may not be shared with health providers.

Showing a Family History: What Is a Pedigree?

In effect, a family history is a record of relationships (regardless of the reason it was collected—medical reasons or for other purposes such as an official record for the church or royal family records). As described above, the degree of relatedness (or genetic similarity) is very important in estimating one's health risks. A diagnosis of a close family member typically carries more importance than that of a distant family member.

Family records were often written in prose or simply memorized or passed down in song. However, for large families, in particular, that may extend for several generations, recollection of a list of names or kinships without some type of documentation was likely difficult and inaccurate. Thus, the information was presented in a format called a "pedigree." The word pedigree is derived from the Latin "ped" (meaning foot) and French word "grue" (crane). The crane's foot represented the connections between the parent and each offspring.

Historically, pedigrees were used to document one's ancestry or line of descent by showing multiple generations of a family. Often referred to as a

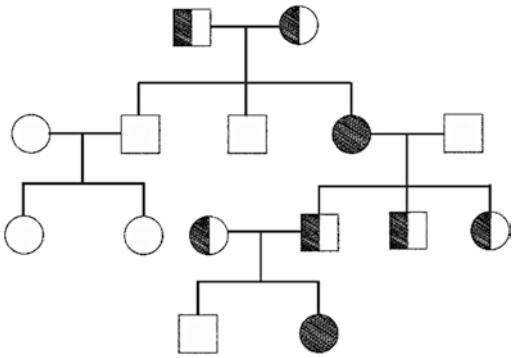

Fig. 2.2 Sample pedigree (circles = female family member; square = male family member; full shading = affected; partial shading = carrier) (source: Biorender)

"family tree," horizontal lines connect family members in the same generation and vertical lines indicate descendants (Fig. 2.2). Thus, one could determine the relationship (or relatedness) between any two family members by following the lines (i.e., uncle-niece). Pedigrees are not limited to human use, and are often used to document the breeding history of livestock, dogs, and other prized animals. In medieval times, members of certain families could wear a coat of arms as proof of ancestry, typically recorded in pedigrees.

When pedigrees began to be used to record health information, additional symbols were created. Shapes were used to depict gender (squares for males and circles for females) (Fig. 2.2). Individuals affected with a disease were typically shaded (e.g., solid circle or square) and individuals who were deceased had a diagonal line drawn through the shape. With the use of a standard set of symbols, less text was needed to include health information enabling patterns of disease inheritance to be rapidly identified. For example, a family with a genetic condition like Huntington's disease, a dominantly inherited disease, would have individuals shaded in multiple generations. For other diseases, only males might be affected and thus only squares would be shaded. The pedigree enables a health provider to skim a "picture" of the family history and immediately note these patterns of inheritance.

Interpreting a Family History

Health providers look for clues in a family history that may suggest that a condition is inherited or passed through a family and therefore that a patient may be at increased risk. As mentioned earlier, not all conditions are caused

by genetic factors (and inherited). Some conditions such as infectious diseases or lung or skin cancer are caused by high exposure to smoking or ultraviolet light, respectively (environmental factors). In reviewing a patient's family health history, health providers will look for multiple members in multiple generations affected with the same condition and that may occur at younger than typical ages of onset. It is possible to have multiple members of a family with the same condition that is caused by nongenetic factors like poor diet or smoking, such as diabetes or respiratory conditions, which is why it is important to collect a wide range of information to better determine causes and patient risk.

Both medical and nonmedical traits, such as intelligence, height, alcoholism, and personality, may occur in multiple generations and family members, thus resembling an inherited trait. However, it is understood that a shared environment can greatly influence the development of a wide range of traits and thus account for the higher than expected number of family members with that trait. For example, a family that prioritizes reading and writing may result in several children of that family growing up to become famous writers. While there may be some genetics involved in excelling at reading and writing, the incorporation of reading and writing activities throughout childhood likely influenced the development of their skills tremendously.

Nevertheless, in years past, some debated whether or not the presence of a family history of a particular trait was due to genetic (an inherited trait) or environmental factors (an acquired trait). While debates over science are not unusual and are actually valuable in interpreting a given set of data, the development of social policies on inaccurate scientific beliefs can lead to devastating consequences. With respect to family history, some scientists believed that nonmedical traits such as intelligence were inherited (not acquired). British scientist Francis Galton was one of the leading proponents of this belief, collecting many family pedigrees with socially "desirable" traits in multiple family members and generations as evidence to support eugenic policies to encourage families with desirable traits to have large families and to dissuade those with undesirable traits not to have children and, thereby, halt the passage of these genes to the next generation.

In the 1930s, Russian scientist Trofim Lysenko also supported the belief that traits could be acquired and were not inherited. His beliefs gained political support in the Communist Party of the Union of Soviet Socialist Republics (USSR), eventually resulting in a ban on the study of genetics in the country. Genetics was considered to be "anti-national" and not a legitimate field of study. Until the rise of Lysenko, Russian researchers were actively conducting genetics research and advancing the field of genetics, particularly agricultural

genetics. The movement led by Lysenko substantially damaged the reputation and credibility of Soviet scientists by extolling ideas not supported by science, usurping other scientists' work, and using science to advance political ideologies. More shockingly, some Soviet geneticists were arrested and killed. After the change in leadership in the 1960s, genetics eventually returned to its place in universities and research commenced.

Digital Family History

Thanks to electronic medical records and many health providers' offices switching to digital intake (where information is collected via a tablet at the beginning of your visit or through a patient portal in advance of your visit), completing, updating, and storing one's family history has become less burdensome for patients. Instead of just checking yes/no for every condition listed on a form, completing a family health history electronically may actually enable collection of a more accurate and complete history. Pop-up questions remind us to think about each family member and what conditions they are affected with or causes of death. And information can be added after the visit. It is particularly important to provide information about family members in multiple generations (e.g., grandparents, great-grandparents) to determine if a condition is "running" in your family. Important details like the age when the person developed the disease, the type of disease if there is more than one kind (like diabetes), smoking status, and other factors may be more easy to recall when prompted or can be added later once you have had a chance to talk with family members.

Thirty years ago, information about family health history would be hand-drawn with a ruler and template for the shapes in the pedigree. Today, several software programs exist that can convert the information collected online into a pedigree (or any other format). Place of birth, date and cause of death, occupation, pictures, and other information can be recorded next to each family member. With some online programs, other documents can also be stored (e.g., birth certificate, military draft card, baptism record). The information is saved so it can be easily updated on subsequent office visits.

Just as important as developing an easy-to-use tool for patients to provide family history, health providers also need tools to quickly analyze the information during the office visit and make recommendations for screening, testing, referral, or lifestyle changes to reduce disease risk. Several electronic family history tools are available to analyze the information provided by the patient and generate recommendations for the health provider based on professional

guidelines to review and discuss with the patient. As recommendations can change occasionally, this helps health providers make recommendations based on the latest clinical guidelines.

What If You Do Not Know Much About Your Family Health History?

More than 100,000 children are adopted each year in the USA and thousands more are conceived via egg or sperm donors. As strange as it may seem, these individuals have something in common—they do not have a complete family health history, if any at all. Yet, all is not lost—even a little knowledge about one's biological relatives can provide real insight into one's health when combined with environmental and lifestyle information from their adopted family.

In some cases, adoptees or individuals conceived with donor sperm or egg may have knowledge about their biological parents or siblings obtained at conception (e.g., from the application of the sperm or egg donor) or at birth (e.g., learned during the adoption process). However, when sharing such information with a healthcare provider, it is important to let them know when that information was obtained, as health information is constantly changing and quickly outdated. This is particularly important, since many people placing children up for adoption are young enough that they likely have not yet developed any diseases.

Some patients with limited knowledge about their family history may consider undergoing genetic testing in order to learn more about their personal health risks (discussed in Chap. 17). Test results, combined with the environmental information from their adopted family, can fill in the knowledge gaps and help some individuals make important choices about their health. However, there are important limitations to consider. There is still much not understood about the genetic causes of disease—so, a negative result does not necessarily mean that an individual does not have an increased risk for a given disease because the test only examines genes that are currently known to be linked to certain diseases. But in the absence of other knowledge, it is one potential source of additional disease risk information.

Why Is Marrying Family Members of Concern?

Based on the "degree of relatedness" discussed earlier in the chapter, we can estimate what percentage of our genes are shared with other biological relatives. Couples that are biologically related will have an increased chance of having a child affected with rare inherited disorders. This occurs because each parent is more likely to carry a rare variant that is shared among family members. A single version of the variant will not likely cause any symptoms for a recessive disease, but having two versions of the variant will result in the development of a disease (recall that we have two copies of each gene). In this case, each parent is considered a "carrier," because they carry an abnormal version that they can pass on to their children.

When marrying someone unrelated to you, the chance that you and your partner will both be carriers of the same abnormal gene is extremely low, so the likelihood that your children will develop the associated disease is very low (even if one of you is a carrier). On the other hand, the closer the relationship between parents (e.g., brother-sister versus second cousins), the higher the chance that both will be carriers of an abnormal gene, inherited from a shared relative. Therefore, their children are at increased risk of inheriting two abnormal versions of the gene (one from each parent) and developing the associated genetic condition.

Marriage between family members is viewed quite differently between different cultures. While some cultures may view marriage between family members as socially unacceptable, it is considered routine or traditional in other cultures including the Amish community in the USA, and in several Middle Eastern and Arab cultures. In some communities, marriages may be arranged between family members. Marriage between two individuals with a shared ancestor, typically between second cousins or closer, is referred to as consanguinity (pronounced con-san-gwin-it-ee). Another term often used to refer to this type of union is "inbreeding."

The Amish community based in Lancaster County, PA, has a high rate of consanguineous marriages, and as a result, a higher-than-expected prevalence of rare genetic disorders in their community. The Amish community are a religious group (Anabaptist Christian denomination), established in 1693 in Switzerland. They migrated to the USA in the 1700s and settled in Pennsylvania and Ohio. The Old Order Amish community in Lancaster number about 35,000, arising from just a few founder (initial) families that first settled in the USA. They do not use any modern technology or electricity, practicing sustainable living through basic tools and manual labor. The Amish only

marry within their community and families are typically much larger than modern American families with 6–7 children.

Discussion of consanguineous relationships can be difficult for patients and health providers. Genetic counselors routinely ask about this during prenatal visits, particularly if there is a history of miscarriage or relatives affected with a rare disease. Some patients may be reluctant to discuss this information due to embarrassment, fear of stigmatization, or acknowledgement of an illicit relationship. About 24 states have laws prohibiting marriage between first cousins. A handful of states do allow marriage between cousins; the state of Maine requires a visit with a genetic counselor prior to marriage so that the couple are aware of the risks.

Taking a Step Back: From Families to Whole Populations

Changes in traits occur over time through multiple generations. If enough detailed data across multiple family generations are collected and recorded, changes may be evident from an analysis of pedigrees or family trees. Since a family is often quite small (even when counting members of multiple generations), some of the gradual change may not be evident over a few generations. These may include gradual physical changes such as the darkening of hair color through the generations or an increased number of family members with heart disease or diabetes. Again, some of this may be due to genetic variation, introduction of new family members (and new genetic variations), and/or environmental factors (e.g., changes in diet).

Scientists can measure and record physical and other changes on a broader scale as well, such as an entire population in a given location (e.g., a bird population on an island). If such differences are noted between generations of a population, scientists will seek to understand the root cause of the change. A physical change in a population of birds, for example in beak size, may be caused by a change in food sources. If food sources shift from plants and soft animals (e.g., worms) to harder seeds or objects that require more physical strength or size to attain, those birds with larger beaks will have a survival advantage and will more likely reproduce and pass the larger beak onto the next generation. Thus, changes in the environment in this example favored a particular trait, the larger beak size (and the genetic variations that contribute to the development of larger beak size). This phenomenon is referred to as natural selection or Darwinian evolution after the scientist Charles Darwin

who first described this theory. The same may also be true of human traits that gradually appear to change over time.

Conclusion

In some ways, family health history is a type of genetic test as the information provides insight into shared genetic risks and can be used to guide preventive care and screening. The value of the information is dependent on the completeness of the information, but nonetheless, incomplete data may alert health providers of potential health risks. While nobody wants to be perceived as the nosy relative and asking detailed questions about family members' health, whatever information can be gathered and shared with a health provider can make a tremendous difference to your own health. With the advent of digital tools, it has become easier to collect, store, and update family health history information.

References

Borinskaya et al. Lysenkoism Against Genetics: The Meeting of the Lenin All-Union Academy of Agricultural Sciences of August 1948, Its Background, Causes, and Aftermath. Available at https://www.genetics.org/content/212/1/1

Genes in Life. How do I collect my family history? Available at http://www.genesinlife.org/genes-your-health/how-do-i-collect-my-family-history

U.S. National Library of Medicine. Pedigree and Family History Taking (Chapter 3). Understanding Genetics: A New England Guide for Patients and Health Professionals. Available at https://www.ncbi.nlm.nih.gov/books/NBK132175/

U.S. Centers for Disease Control and Prevention. My Family Health Portrait. Available at https://phgkb.cdc.gov/FHH/html/index.html

U.S. Centers for Disease Control and Prevention. Knowing is not enough—act on your family health history. Available at https://www.cdc.gov/genomics/famhistory/knowing_not_enough.htm

3

A Savior Sibling

On July 26, 1978, the first "test tube" baby was born in England. Conceived in "laboratory glassware," a 5 lb 12 oz baby girl was born to parents Lesley and John Brown. She was named Louise Brown. The headlines in the *New York Times* read "Woman Gives Birth to Baby Conceived Outside Body." A few days later, another *New York Times* article proclaimed that "… a landmark has now been passed with the birth of a normal baby outside a human body. Probably not since the invention of nuclear weapons has a scientific advance been received with such mixed feelings" (July 28, 1978).

After attempting to conceive for more than 10 years, the event was indeed a blessing for the Brown family. However, not all were ecstatic about this new application of science and its intrusion on, or perhaps, replacement of, a very private and natural process. Known as *in vitro (vee ′trō) fertilization* (IVF), the process was welcomed by those who had long struggled to have a family, but viewed by others as almost surreal and science fiction.

In general, IVF involves the surgical removal of an egg(s), which is then fertilized with a sperm sample in a laboratory dish (rather than a test tube as reported by the press). Either (or both) the egg and sperm can be collected from the natural parents or, if not possible, from donors. Twelve to twenty-three hours later, the eggs are examined to determine if fertilization has occurred. If so, the embryo is further incubated for about 2 days and then implanted in the woman's uterus for gestation and delivery. Three years after the birth of Louise Brown in England, IVF became a reality in the United

S. B. Haga, *The Book of Genes and Genomes*, https://doi.org/10.1007/978-0-387-70916-1_3

States at the Jones Institute at Eastern Virginia Medical School in Norfolk, Virginia. Although data are incomplete, researchers estimated that the 5 millionth baby was born in the fall of 2013 from IVF technology. Today, several different types of IVF services are available and growing social acceptance has contributed to the increasing number of babies born from IVF technologies, accounting for 1–2% of all US births.

So how is IVF relevant to genetics and genomics? As described in further detail below, fertilized eggs generated through IVF can be tested prior to implantation for genetic conditions that may be passed on by one or both of the parents. Following genetic testing of the fertilized egg, only the eggs that test negative for the genetic condition would be implanted in the mother.

Preimplantation Genetic Diagnosis (PGD)

Preimplantation genetic diagnosis (PGD) is a test that is performed after an IVF procedure, typically after a fertilized egg has doubled several times. Specifically, PGD involves the genetic analysis of a *single* cell extracted from a fertilized egg conceived through IVF technologies (Fig. 3.1). The original use of PGD was to identify embryos with inherited genetic conditions. Families with a history of inherited diseases, particularly those that develop in childhood for which little or no treatment was available, were the early users of

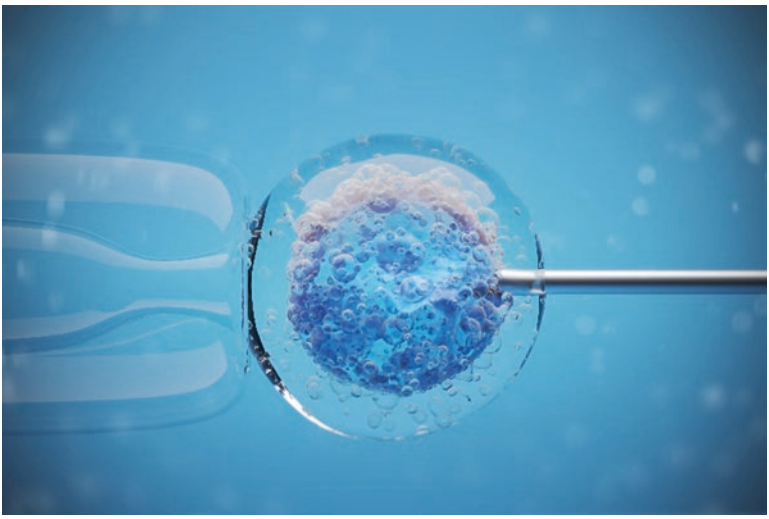

Fig. 3.1 Preimplantation genetic diagnosis (blue background) (source. Adobe Stock Photos)

PGD. Only those embryos that tested negative would be implanted to insure the birth of a child unaffected with that particular condition. For some genetic conditions that affect only one sex, the determination of the sex of the embryo was all that was necessary (embryos of the unaffected sex would then be implanted). Indeed, the first use of PGD in 1990 was to determine the sex of fertilized eggs from a family that was affected with a disorder that predominantly affected males.

Today, more than 200 different genetic conditions (chromosomal and single-gene disorders) can be detected by PGD. In 2012, the Society for Assisted Reproduction Technology estimated that 5% of IVF cycles in the USA undergo PGD (about 8000). Prior to PGD, prenatal diagnosis (testing of fetal cells in the first or second trimester) was the only available option to detect an affected fetus; parents would be faced with the decision of an elective termination or to continue the pregnancy. PGD spares couple from making that decision though they still face the issue of disposal of non-implanted embryos.

Savior Siblings

Bone marrow transplantation is a lifesaving option for leukemias, lymphomas, immune deficiency disorders, and some solid tumor cancers, either curative or driving the disease into remission. In 2018, almost 23,000 bone marrow and cord blood transplants were performed in the USA. Bone marrow transplants involve the transfer of bone marrow, a slurry of cells collected from the center of the pelvic bone. This mixture of cells is essential to the body's immune system and normal functions of the blood. In particular, bone marrow contains cells that are the progenitor (or stem) cells—where all other cells derive from—these give rise to multiple generations of cells that can replenish the stockpile. Stem cells are also present in blood, such as cord blood from newborns, so in some cases, blood can be transplanted instead of bone marrow. Thus, the healthy cells from the donor's bone marrow will grow in the patient's bone marrow, replacing the cells that are missing or damaged and restore health.

In order for a bone marrow transplant (or any other type of transplant) to work, the patient's body must "accept" the donor organ or bone marrow or, in other words, not consider it as "foreign" and reject the tissue. Our body's immune system is constantly on the lookout for invaders—anything that is not recognized as "self"—and will launch an attack if a foreign cell is detected. In some diseases, the immune system mistakenly attacks cells that are "self"—these diseases are called autoimmune diseases. Thus, for any type of organ or

bone marrow transplantation, the organ donor and recipient must be a match with respect to the immune system (the body's defense) genes. If these genes are quite similar (and therefore, the proteins created by these genes are similar) between the donor and the patient, there is a much lower risk of rejection.

In order to match donors and patients, a genetic test is performed. The closest match will be between family members since they already share some of their genetic makeup; about 20% of the bone marrow transplants are from unrelated donors. Specifically, the test looks at a group of genes called the human leukocyte antigen (HLA) genes, which are important to the body's immune defense (this type of testing is also referred to as tissue or HLA typing). The HLA genes make HLA proteins (or molecules) that are located on the surface of cells that our immune system recognizes as its own (self) or foreign (invader)—think of these proteins as sort of an identification tag. If the immune system does not recognize the HLA proteins, it will send a series of signals to other cells to attack the tissue. The more similar the HLA genes, are between donor and recipient the less likely the recipient's body will reject the tissue. As there are multiple HLA genes, many different combinations are possible for each person, which explains why it can be so difficult to find a matched organ donor. Thus, searching for a donor can be a long and agonizing process due to limited supply of donated tissues and waiting for a match for a specific organ/tissue.

In some cases, where no family members or unrelated donors are a match, it is possible for the parents to undergo IVF for the explicit purpose of conceiving a child who would be a match for an affected sibling in need of a bone marrow transplant. While this could happen naturally, undergoing IVF would enable a genetic test to be performed to determine which fertilized egg(s) would be a perfect match and were unaffected by the disease that affected their sibling. Then, those eggs would be selected for implantation. When the child is old enough, he or she can donate bone marrow for their sibling.

In 2001, the first PGD case to identify a matched embryo for an affected sibling was reported. In this case, the family's 6-year-old daughter Molly was born with a disease called Fanconi anemia, a rare inherited condition characterized by bone marrow failure and increased predisposition to leukemia. Bone marrow transplantation is the only treatment for individuals affected with this disease. The parents, Lisa and Jack Nash, underwent several IVF cycles in 2000. A total of 30 embryos were initially tested (out of 33 total) and 24 embryos were found *not* to be affected with Fanconi anemia (the parents had a 25% risk of conceiving another child with the disease).

The 24 unaffected embryos were next tested for the HLA genes. The HLA testing revealed that 5 embryos were matches out of the 24 tested. The embryo

that was implanted from the fourth IVF cycle was the only one to result in pregnancy and the birth of a healthy newborn boy. At birth, the umbilical cord blood of Molly's newborn brother Adam was collected and transplanted to her, resulting in successful bone marrow transplant.

This successful demonstration of the use of IVF and PGD to identify matched donors for their affected siblings gave hope to other families of children with diseases only treatable by a bone marrow transplant. In the USA, no permission is required to use these tests to identify matched embryos, but insurance coverage of the expenses of the procedure and testing may fall upon the families depending on their insurance policy. In contrast, in the UK, prospective parents wishing to undergo any type of PGD must seek a license from the government's oversight body, the Human Fertilization Embryology Authority (HFEA). The majority of licenses granted by the HFEA have been to use PGD to identify embryos affected with inherited diseases such as cystic fibrosis. When the use of IVF/PGD to identify matched embryos was first reported, other families sought permission to use this procedure to have matched donor sibling.

However, not everyone was supportive of the new, potentially lifesaving application of IVF/PGD. One case in particular highlighted the debate about whether this technology should be used to "create" a child for the primary purpose of treating another individual. In the UK, the Hashmi family was initially given permission to use PGD to test embryos in search of a matching donor for their sick 4-year-old son, Zain. Zain Hashmi was affected with a rare blood disorder known as beta thalassemia major and no bone marrow match had been found. He was being kept alive by regular blood transfusions. The Hashmis already had five children, one born after Zain, but none were a match for Zain. After two IVF cycles, two embryos were found to be matches to Zain but one was not implanted and the other was implanted but did not lead to a pregnancy. During this time, the so-called Comment on Reproductive Ethics group in February 2002 challenged the initial approval to allow the use of PGD for this purpose. Specifically, the group was opposed to "designing another child as a therapeutic commodity, as a tissue bank." In December 2002, the High Court ruled that the HFEA did not have the power to license the technique under existing legislation.

The HFEA returned to court to appeal the decision. In April 2003, the UK High Court sided with HFEA, thereby allowing the Hashmis to continue their efforts to use PGD to identify a matched donor sibling. Zain was now 6 years old and mother Shahana Hashmi was older as well. They tried several times unsuccessfully to conceive and eventually announced that they had stopped trying.

Today, applications to perform PGD-HLA typing (referred to as preimplantation tissue typing in the UK) are decided by the HFEA on a case-by-case basis. There is a list of conditions for which PGD-HLA typing can be requested, but approval must still be granted to each family. The family must demonstrate that all other possible treatments have been attempted and exhausted, such that preimplantation tissue typing is the last option. In July 2005, Julie and Joe Fletcher were the first UK couple to have a "savior sibling"—their newborn daughter was a perfect match for their 3-year-old brother who was affected with a rare blood disease known as Diamond-Blackfan anemia. In 2006, the parents of Charlotte Mariethoz from Leicester, England, also affected with Diamond-Blackfan anemia, were given permission by the HFEA to use PGD to create a savior sibling.

Many countries and professional medical organizations support the use of PGD to identify embryos affected with a genetic disorder prior to implantation. However, use of PGD for HLA typing to identify a matched donor for an affected sibling has raised much debate, viewed less of a heroic act (as the term "savior" suggests), but rather more as treating a life as a commodity. The best-selling novel *My Sister's Keeper*, later made into a motion picture, explored a family's experience from the viewpoints of the parents, the affected sibling, and the "savior" sibling (Fig. 3.2). The complex psychological experiences of each family member (the savior sibling was well aware of her role) were explored and highlighted the challenges of families caring for loved ones affected with diseases without proven therapies. In particular, the savior sibling may struggle with knowing that he or she was especially created to help their sick sibling, or else their sibling would likely die. In the same token, that person may struggle to create their own identity, as not only a member of the family but also defining their own identity and life beyond serving as a tissue donor.

Other Controversial Uses of PGD

While the use of originally intended PGD was to screen fertilized eggs for genetic diseases known to run in a particular family, it has not been without some controversy (in addition to the unanticipated use of the technology to create savior siblings). For example, in 2002, a PGD case was reported about the screening of embryos for Alzheimer's disease. A 30-year-old woman was found to have the same mutation that her brother and sister carried for a rare early-onset form of Alzheimer's disease (different from the more common type that develops in the 60s and 70s). Thus, it was almost certain that she

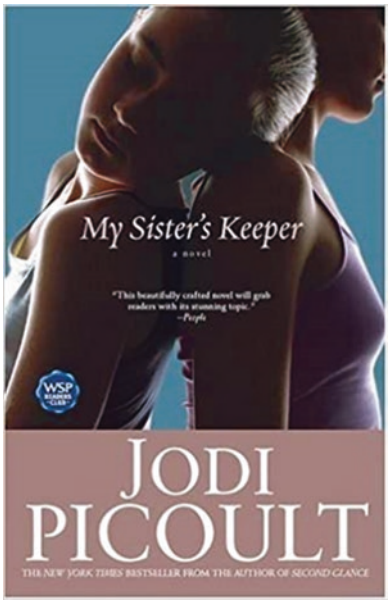

Fig. 3.2 The story of a family confronted with the use of PGD to treat an affected sibling in *My Sister's Keeper* by Jodi Picoult (source: Amazon)

would develop the disease in a few years. Her sister could no longer care for her two young children and was in an assisted living home. The woman and her husband desired to have a child that would not befall the same devastating fate as she eventually would. Therefore, they underwent IVF/PGD to test for the mutation causative of this form of early-onset Alzheimer's disease. Nine of 15 embryos were found to carry the mutation. Four of the unaffected embryos were implanted, resulting in a singleton pregnancy. Prenatal diagnosis confirmed that the fetus was not carrying the Alzheimer's mutation. The couple gave birth to a healthy, unaffected daughter. Following the publication of this case in the medical literature, some raised the strong possibility that while the child may not be at risk for the rare early-onset form of Alzheimer's, there is a good chance that the child could develop the common form of Alzheimer's disease or dementia. In addition, given the timeframe, it may be possible that a treatment would be developed before the child reached the age when symptoms would appear. Lastly, some expressed concern that the mother would likely no longer be able to care for the child once her symptoms began to appear.

In 2007, the same clinic provided PGD services to a family that had been devastated by colon cancer. A particular form of the disease can be inherited (whereas most colon cancer is sporadic) and develop early in life. Four

members of the Kingbury family had died of the disease, and the father of the couple undergoing PGD also carried the flawed colon cancer gene. To avoid passing on this gene to his child, the couple underwent IVF and PGD to screen for embryos and implant only those that were found not to carry that mutation.

Another controversial use of PGD is sex selection, not related to any health reasons, but for purposes of establishing family balance or parental preference. Different cultures have preferences for male children, particularly families of Chinese, Indian, and Middle Eastern descent. This has been an ongoing debate for some time that is now facilitated by this technology to allow preselection of embryos of a certain gender. In some cases, genetic diseases occur predominantly in one sex, and therefore, sex selection in these cases is acceptable. Many view sex selection in the absence of a health risk to be unethical. A 2009 study found that just less than half of parents (~45%) that had PGD to determine gender declined to implant embryos of the non-desired gender, a decision that was associated with parents' heritage.

Ethical Issues

As you might imagine, several ethical objections have been raised regarding the use of PGD, particularly as the number and range of applications using this technology have expanded. To begin with a very general question, does the preselection of embryos for medical conditions or other traits interfere with the natural process of human diversity and life? In deciding which embryos to screen and implant, we are making decisions that have never before been possible, which extends from prior technology of prenatal diagnosis and decisions to continue or terminate a pregnancy based on a positive result from those tests. These technologies and the ability to preselect are reminiscent of the eugenics movement in the USA in the early twentieth century, in which certain traits were deemed inferior or unacceptable and efforts were made to prohibit individuals with those traits from reproducing. Some believe that the power to decide which embryos will be implanted and grow to become living beings belongs to only a Creator or God. From this perspective, it would be considered immoral to select one trait over another for any reason as all human life deserves respect.

The disability community has expressed concerns about the use of reproductive technologies like PGD and prenatal diagnosis and elective termination of fetuses. Many disabled individuals are active citizens of our society and therefore embryos with a perceived "disability" are deserving of the right to

life. Is an affected embryo a less valued life than an unaffected embryo? Scholars Erik Parens and Adrienne Asch state that if decisions are made on a single diagnosis, "a single trait stands in for the whole, the trait obliterates the whole." However, some members with perceived disabilities such as deafness view PGD as a useful technology. Deaf parents who want to have a deaf child can test and preselect embryos with this trait.

A related ethical issue deals with selective implantation or destruction of embryos. Through early diagnosis with PGD, couples can avoid the difficult decision of an elective termination of a pregnancy following an abnormal prenatal diagnosis. However, couples undergoing IVF with or without PGD must still decide on the fate of the non-implanted—both affected and unaffected—embryos. An IVF cycle will most likely produce more embryos than can be implanted and therefore decisions must be made to store the embryos for future use, destroy the embryos after a certain period of time, or donate the embryos for research purposes. If the couple believes that life begins at conception whether in the laboratory or naturally, the weight of the decision to discard an embryo is no less difficult. For those who do not attach any rights of personhood to the embryos, IVF and PGD are considered ethically acceptable.

The application of PGD to create matched donor siblings has raised some debate about the value of another human life for the primary purpose of saving an existing child's life. In 2002, the President's Bioethics Council concluded that any form of selection or manipulation turns the child into a "manufacture." Prior to PGD technology, couples could still conceive with the hope that the next child would be a match for their sick child. If the child turned out not to be a match, would the parents value that child less? When the child grows up and realizes the role they have played in helping their sibling, do they have the right to refuse if further tissue donations are required? Where does a donor's rights begin and responsibilities end? When is the consent of the donor sibling required? What are the psychological implications for the savior sibling?

With respect to screening embryos for adult-onset conditions is the uncertainty that what may be a devastating disease now may well be a treatable disease in the future. For disorders with a strong genetic understanding that develop in adulthood, such as rare forms of inherited breast cancer and the neurological disease Huntington's, support for PGD testing for these conditions has increased. In 2006, the HFEA issued a report stating that PGD testing for conditions that develop later in life is reasonable "because the features of the conditions are not incompatible with them being regarded as serious genetic conditions." But, in a sense, we are placing today's standards

and assumptions on something that will not occur for potentially 30 or 40 years into the future. With the pace of medical research and discovery, it is difficult to predict what the life of someone diagnosed with Alzheimer's or cancer will be like.

Lastly, while the use of PGD to screen embryos for an untreatable medical condition to reduce the risk of having an affected child is understandable and defendable to many, the selection for or against nonmedical traits such as appearance, personality, or exceptional talents in future generations of children presents more challenges. Although this is not yet a reality, understanding of the genetic basis for more and more nonmedical traits continues to advance as well as the testing technologies that will enable parents to obtain a genetic "portrait" of their unborn child. In contrast to the other uses of PGD where the well-being of the child is the primary issue, the selection of nonmedical traits shifts the focus to parental or societal preferences from the child's health.

Conclusion

As with any new technologies, new applications may be controversial. The use of PGD to pre-screen fertilized eggs to serve as "savior siblings" was not likely envisioned but yet it has been demonstrated to be a feasible option for parents with a child with a devastating condition that could be cured with a bone marrow transplant, if only a matched donor existed. A recurring theme in this book is that new knowledge and technologies can greatly help address medical issues and improve health outcomes, but that other unintended uses arise that force us as individuals, families, communities, and societies to decide what should or should not be allowed. IVF and PGD technologies exemplify how rapid advances in science have impacted family planning and reproductive decision-making, but not without controversy. These debates will continue as technologies and scientific understanding evolve and expand, and what was once the storyline of science fiction novels will be realized in our lifetimes.

Resources

Picoult J. My Sister's Keeper. Washington Square Press. 2004.

US Health Resources & Services Administration. Bone Marrow & Cord Blood Donation & Transplantation. Available at https://bloodstemcell.hrsa.gov/

Belkin L. The Made-to-Order Savior. New York Times Magazine. July 1, 2001. Available at https://www.nytimes.com/2001/07/01/magazine/the-made-to-order-savior.html

Verlinsky et al. Preimplantation Diagnosis for Fanconi Anemia Combined With HLA Matching. JAMA 2001; 285(24):3130-3133. doi:https://doi.org/10.1001/jama.285.24.3130

Ram NR. Britain's new preimplantation tissue typing policy: an ethical defence. J Med Ethics. 2006 May; 32(5): 278–282. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2579414/

Parens, Erik; Asch, Adrienne. Disability rights critique of prenatal genetic testing: reflections and recommendations. Mental Retardation & Developmental Disabilities Research Reviews. 2003, Vol. 9 Issue 1, p40-47. 8p.

4

Too Few, Too Many

Much attention has recently focused on genome sequencing and genetic variation (changes to the individual units of DNA—the As, Ts, Cs, and Gs) associated with disease. Yet the large molecules in which DNA is packaged, known as chromosomes, are also associated with disease. As you will recall, rather than one long string of As, Ts, Cs, and Gs linked together like a "Happy Birthday" sign, DNA is packaged into smaller molecules called chromosomes. In humans, there are 23 pairs of chromosomes. The DNA strand is actually wrapped around a repeating set of eight ball-shaped molecules, allowing for the long strands to be tightly condensed (think of a string wrapped around a tennis ball). One cannot actually see the wrapped DNA around the balls under a standard light microscope, but can see each chromosome when stained with a dye (Fig. 4.1). To the untrained eye, they appear like little squiggly worms of varying sizes (and number if you are looking at different species). But each chromosome is actually unique, based on their size (length) and pattern of "bands" or stripes, sort of like a chromosomal barcode.

A normal human cell contains a total of 46 chromosomes (23 pairs), one of each pair inherited from each parent (Fig. 4.2). Any more or fewer chromosomes could lead to a genetic syndrome or incompatibility with life. A syndrome occurs when several clinical symptoms appear together and are diagnosed as a single disease. It is estimated that up to 60% of known genetic syndromes are due to chromosome abnormalities. A chromosome abnormality may involve (1) a missing or extra copy of an entire chromosome, (2) a missing or extra section or block of DNA in a chromosome, or (3) a rearrangement of DNA between two chromosomes. As will be described in more

S. B. Haga, *The Book of Genes and Genomes*, https://doi.org/10.1007/978-0-387-70916-1_4

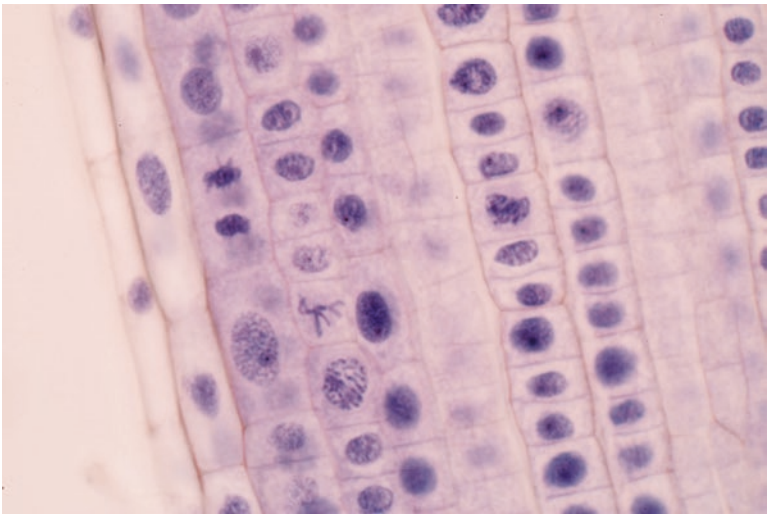

Fig. 4.1 Stained slice of the tip of the root of an onion. Onion cells are lined up in these blocks and it is easy to view the chromosomes (squiggly lines). Each cell looks slightly different as each cell is at a different stage of cell division or growth (source: Adobe Stock Photos)

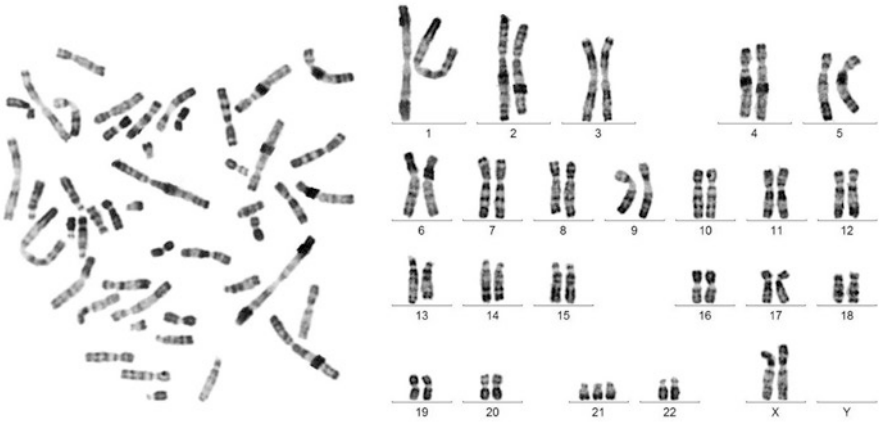

Fig. 4.2 A lab technician has arranged the chromosomes (seen in their natural state on the left side) in order from largest (top left corner numbered 1) to smallest (bottom row, numbered 22) in the picture on the right side called a karyogram (care-ee-o-gram). A pair of chromosomes in the bottom right corner is called the sex chromosomes. Females have two X chromosomes and males have one X and one Y chromosome. Half of the chromosomes (one of each pair) are inherited from our mother and half from our father. The total number of chromosomes is normally 46. In this karyogram, there is an extra chromosome number 21, which causes Down syndrome. Thus, this patient (a girl as indicated by the XX pair of sex chromosomes) has a total of 47 chromosomes

detail in the following sections, the first two types of chromosomal abnormalities may result in missing or extra copies of genes that are on the affected chromosome or section. The third type of change could lead to the creation of a new gene that can cause disease. Normally, every cell in our body contains two copies of each gene (one inherited from our mother, one from our father). If the number of copies of genes is altered, it can affect the amount of protein that is produced overall for that particular gene. In some cases, too much or too little of a certain protein can adversely impact the cell's function. A given chromosome can contain hundreds of genes, so a chromosomal imbalance can disrupt the quantity of many, many proteins and therefore be quite severe for the cell and overall development.

Health Effects of Extra/Missing Chromosomes

The effects of chromosome imbalances can range from being incompatible with life to causing genetic syndromes. For most chromosomes, a newly conceived embryo cannot survive with an extra or missing copy of a chromosome. However, extra copies of chromosomes 13, 14, 16, 18, 21, X, and Y are compatible with life, but result in congenital defects that can range from mild to very severe and the life expectancy for affected children can also vary greatly (with the exception of the Y chromosome). For example, children with three copies of chromosome 18 have Edwards syndrome, and are affected with intellectual disabilities, severe heart malformations, a receding jaw, and enlarged back of the skull. Many children with Edwards syndrome often pass away within the first year of life, though some have survived into their teens. One of the most well-known syndromes caused by an extra chromosome 21 is Down syndrome, characterized by mild-to-moderate intellectual disabilities and heart defects (described in more detail in later section) (Fig. 4.2). In comparison, an extra copy of the Y chromosome in boys (XYY) does not shorten life expectancy and is not associated with disease.

Too few copies of any chromosome other than the X chromosome is also not compatible with life, resulting in a spontaneous abortion (miscarriage). Recall that the X and Y chromosomes are the sex chromosomes—females have two copies of the X chromosome (XX) and males have one copy each of the X and Y chromosome (XY). A baby with a single X chromosome (and a normal number of all other chromosomes) will have a condition called Turner's syndrome. This condition is characterized by a distinct face, incomplete sexual development, short stature, infertility, and a webbed neck.

A third type of chromosomal abnormality occurs when one or two chromosome(s) break apart, and the broken piece(s) reattaches incorrectly. For example, a strand of DNA within the same chromosome can break at two locations; the broken piece of DNA may flip around before reattaching, such that it reattaches upside down (known as an *inversion*). Alternatively, when breaks may occur in two different chromosomes and the pieces switch and reattach to the wrong chromosome, this is called a chromosomal *translocation*. Any type of break in a chromosome can result in a disruption of a gene if there is one present at that location (sometimes the break may occur in a location with no genes). Thus, there may be no protein or a truncated protein produced that is nonfunctional—either scenario may lead to harmful consequences. In addition, if a fusion occurs between two strands of DNA from different chromosomes or within the same chromosome but in the incorrect orientation, this can lead to the creation of a new protein (a hybrid created from part of a gene located on one strand that is fused to part of another gene on the other strand). This newly created protein can be harmful or disruptive to the cell's normal function, giving rise to disease.

One very well-known example of a chromosome break and reattachment onto another chromosome that results in the creation of a new gene is found in many patients affected with chronic myeloid leukemia (CML). In the bone marrow cells of most CML patients, a break has occurred in the bottom part of chromosomes 9 and 22. The broken pieces have reattached, but to the wrong chromosome, resulting in what's called a translocation (Fig. 4.3). The new chromosome that is created, called the Philadelphia chromosome after the city in which it was discovered, thus contains DNA from chromosome 9 fused to the DNA sequence on chromosome 22. As a result, a new gene is formed right where the broken pieces are reattached—part from a gene located on chromosome 9 and the other part from a gene on chromosome 22—that actually causes the cell to become cancerous. A drug has been developed that directly targets the protein created by the new gene from the 9;22 chromosome fusion that blocks its harmful effect, drastically improving the outcome of CML patients.

How Do Chromosomal Abnormalities Arise?

So how exactly do these chromosome abnormalities arise in the first place? If we focus just on congenital (at birth) abnormalities, chromosome abnormalities may develop at preconception or very early after conception occurs, so that the change is present in almost every cell of the person. Biologically

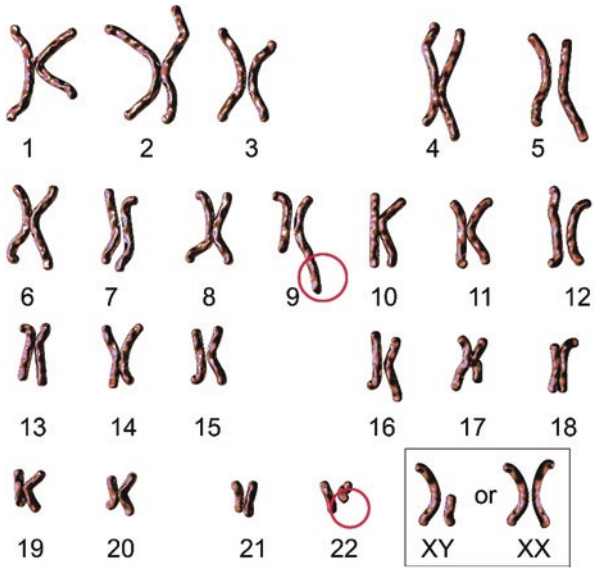

Fig. 4.3 Philadelphia chromosome karyotype (male or female). In the circled tips of chromosomes 9 and 22 is shown the translocation, where a break has occurred and 9 and 22 chromosomes and broken pieces have incorrectly reattached to the wrong chromosome. This 9;22 chromosome is a hallmark of patients with chronic myelogenous leukemia (source: Adobe Stock Photos)

speaking, conception refers to the union of an egg and sperm. Eggs and sperm differ from other cells in our body in that they have half as much DNA. The reason for this is that in order to achieve the normal human complement of DNA (46 chromosomes), the egg and sperm can only contain half as much (23 chromosomes each). Otherwise, the combination of an egg and sperm cell would have double the amount of DNA.

To achieve half as much DNA as other cell types, eggs and sperm undergo a special type of cell division. Typically, most cells in our body go through a process in which the DNA is actually doubled before the cell divides into two daughter cells distributing the DNA equally to each so that the new cells are identical to the parent cell. One cell becomes two cells, two cells become four cells, etc. In contrast, the cell division process for eggs and sperm is slightly different because of the need to reduce the total DNA content in half. Sperm and eggs actually go through a two-phase cell division process. In the first stage, the DNA is replicated and divided into two daughter cells. In the second stage, the two daughter cells undergo a second round of division to yield a total of four daughter cells, but the DNA is not replicated in the second

phase. Therefore, the four daughter cells contain half as much DNA as the parent cell.

For males, sperm production begins during puberty. Sperm production takes about 65 days to complete and up to 200 million sperm can be released per ejaculate. In contrast to the continual generation of sperm, females are born with a finite set of eggs which are stored in the ovaries until menstruation begins. However, the eggs are suspended at the end of the first phase of the cell division process described above. As an egg is released from the ovary (every 28 days or so), the egg continues the cell division process to the beginning of the second phase of cell division. The second phase of the cell division process is only completed if fertilization occurs. It is estimated that 2.5 million premature eggs are present at birth, but only about 400 will mature over a woman's lifetime. While errors can occur at either the first or the second phase of egg division, it is often in this second phase that errors occur in the division of chromosomes between daughter cells. As a result, one cell may end up with no copies (missing) of a particular chromosome and another cell may have an extra copy of that chromosome.

Chromosome abnormalities may arise either in the egg or sperm *or* after fertilization occurs. Every egg or sperm should have one copy of each of the 23 chromosome pairs. But mistakes can occur resulting in an uneven split of the chromosomes between the daughter cells of egg or sperm before conception. If the chromosomes do not divide evenly, the sperm or egg may have extra copies or no copies of a particular chromosome. If one of these eggs or sperm is involved in fertilization, the resulting embryo will also have an imbalance of chromosomes leading to disease or non-viability. Other types of chromosomal abnormalities can arise either in the egg or sperm or very early after fertilization.

Postfertilization, as the embryo grows through successive cell divisions, an uneven division of the duplicated chromosomes can also occur resulting in a daughter cell with too few or too many chromosomes. Depending on what stage this error occurs, the embryo may actually have two types of cells—one with the normal 46 chromosomes and one with an abnormal number of chromosomes. This effect is called *mosaicism* as the embryo will actually have two cell types. The earlier the stage of embryonic development this occurs, the more likely it will be clinically significant.

Down Syndrome

As the medical superintendent at the Earlswood Asylum for Idiots in Surrey, England, in 1858, Dr. John Langdon Down cared for a number of patients with mental disabilities. In those days, the term "idiot" (or idiocy) was a medical term used to refer to patients with severe mental impairments. Dr. Down observed that some patients with mental impairments could be distinguished based on certain ethnic traits. In particular, though of Caucasian descent, some patients showed facial features of Mongolian ancestry; these patients were therefore referred to as Mongoloids. He formally described the combination of the distinct facies and mental retardation in a publication in 1866, entitled "Observations on the Ethnic Classification of Idiots." The term *mongoloid* remained in use for about 100 years until Asian researchers criticized it as derogatory and the disease was hence referred to as Down's or Down syndrome.

The incidence of Down syndrome is estimated at 1 per 700 births. Down children are characterized by several distinct physical features: a single crease across the palm (most people have a double crease across one or both palms), almond-shaped eyes caused by a fold in the eyelid (the outer corners of the eye turn down), a protruding tongue (actually due to a small mouth and enlarged tongue at the back of the throat), and short stature and limbs. Affected children and adults are at higher risk for a number of conditions such as congenital heart defects, gastroesophageal reflux disease, and Alzheimer's disease. It is estimated that only a quarter of Down syndrome pregnancies survive to birth.

The cause of Down syndrome remained a mystery for about a century after its recognition. A number of theories were debated in the early twentieth century as to the cause of Down syndrome—perhaps it was due to damage to the sperm, egg, or fertilized egg; the mother's lifestyle that may have been harmful to the developing fetus; or some unknown inherited factor. In the 1930s, some speculated that Down syndrome may be caused by chromosome abnormalities. But it was not until 1959 that it was finally determined that individuals with Down syndrome had extra genetic material that caused the disease. Indeed, it was found that, these children had an entire extra chromosome 21—one of the smallest of the human chromosomes (it was not until 1956 that the normal number of human chromosomes was finally determined).

For more than a century, it has been observed that older mothers are at higher risk for having children affected with chromosomal abnormalities. Women in their late 20s have a risk of about one in 450. For women between 40 and 45, that risk is increased to one in 38. Why is this? As described above, in contrast to males who continually make new sperm, females are born with

all of their eggs. The eggs are arrested in development and stored in the ovaries until menstruation begins, at which time one egg is released every month and finish development. The older the woman is, the longer the eggs have been stored and the greater the risk of an error in the egg development that will result in a chromosomal imbalance.

Normal But with Missing or Extra DNA?

More recent research has revealed that all of us have imbalances of small and large DNA regions, ranging in size from a single gene (about a thousand DNA letters) to several hundred thousand DNA letters and potentially multiple genes. Known as copy number variation or CNV (as it refers to the difference in number of copies of DNA, where the normal number of copies is 2), it is estimated that about 13% of the human genome is impacted by these types of imbalances (more or less than two copies). Research has shown that CNVs are passed from parents to children. Surprisingly, most CNVs do not appear to impact development or overall health, presumably occurring in nongenic regions (gene deserts). However, certain types of CNVs have been linked to autism, schizophrenia, and increased risk of HIV infection. For children with undiagnosed conditions, testing of the parents can rule out the possibility that a CNV is the cause of the condition if it is found to be present in either parent (and thus presumed not to cause any problems) and the affected child.

Cancer

Since cancer cells have lost their internal controls to prevent cells from growing uncontrollably due to DNA damage, they can accumulate a significant number of changes with respect to chromosome number and structural changes such as rearrangements. With respect to the number of chromosomes, the process of cell division in cancer cells may not occur properly, and the daughter cells may end up with substantially more DNA due to the presence of an increased number of whole or partial chromosomes than the parent cell. Thus, testing of cancer cells can take much longer if there is a very high number of whole chromosomes and or broken chromosomes, compared to the normal number of 46 chromosomes.

As described earlier, a chromosomal rearrangement occurs when breaks occur in two chromosomes and hybrid chromosomes are formed when

those parts reattach to the other chromosome. This type of rearrangement is called a balanced translocation—no DNA has been lost or gained. In other cases, some DNA will be lost or gained (called an unbalanced translocation). The genes (if any) located in the section of DNA where the break occurs or that is lost or duplicated may contribute to the cancer.

Some cancers have very characteristic chromosomal changes that may be helpful in diagnosing or establishing the prognosis of a cancer. Hence, doctors will order testing on the blood or tumor tissue to examine the chromosomes among other tests. As described earlier, the 9;22 Philadelphia chromosome creates a new gene that causes cells to grow uncontrollably in CML. Other cancers such as Burkitt's lymphoma and acute lymphoblastic leukemia (ALL) have a characteristic translocation between chromosomes 8 and 14. In solid tumors, chromosomes 1p and 16q are often missing. In sarcoma, a type of cancer that begins in the bone or in connective tissue such as muscle or fat, a piece of the bottom part of chromosome 12 is often missing (referred to as region 12q13–q14).

Screening Tests for Chromosomal Abnormalities in Pregnancy

During pregnancy, multiple screening tests are now available to determine if a fetus is at increased risk of a genetic condition due to an extra chromosome. A screening test is not diagnostic of a disease, but indicates whether the fetus is at an increased risk of a disease. An abnormal screening result could be caused by a number of reasons. Therefore, if a screening result is abnormal, women are recommended to undergo follow-up testing that can provide a definitive diagnosis.

Prenatal screening tests have changed over the years due to new testing technologies and better understanding of the association between certain biomarkers and disease. For example, a blood test for Edwards syndrome and Down syndrome can be performed early in pregnancy (first trimester). The test analyzes two fetal proteins detectable n the mother's blood in combination with a test called nuchal translucency using ultrasound data. Abnormal levels of these proteins can indicate a risk for these conditions, but follow-up testing of the fetus's chromosomes is needed to confirm the results. Ultrasound or sonography uses sound waves to determine the location of the fetus; a trained technician or doctor analyzes the images on a monitor for specific physical abnormalities, to measure growth and estimate age. A

second-trimester test serum screen, known as multiple marker screen or quad screen, is also available. This screening test evaluates the level of four proteins in the mother's blood. Abnormal levels of these proteins may indicate an increased risk of birth defects like neural tube defects or syndromes caused by chromosome abnormalities. A formula is used to calculate the risk of a chromosome abnormality or birth defect that considers the level of these proteins with the mother's age, weight, race, multiple pregnancy (e.g., twins), and whether the mother is diabetic or not. Again, the result of this screen is not diagnostic and can only indicate an increased risk.

The tests described above involve analysis of proteins, and not DNA. In the past few years, however, newer genomic technologies have been enabling screening for chromosomal abnormalities in the first trimester based on an analysis of cells from the fetus that are present in the mother's blood. The test examines the baby's DNA for extra copies of chromosomes associated with diseases, such as 16, 18, 21, and X, and is called noninvasive prenatal testing or screening (NIPT or NIPS). This noninvasive technology enables the analysis of very small quantities of fetal DNA that are floating in the mother's blood. Before this testing was developed, the only way to analyze the baby's DNA was through an invasive procedure where a long needle is used to obtain some of the tissue surrounding the baby (described in next section). NIPT has been shown to be highly accurate and will likely be further developed to enable DNA testing of single genes, and will not be limited to whole chromosomes.

Diagnosis of Chromosomal Abnormalities in Pregnancy

If a screening result is abnormal, confirmatory testing is recommended. The two most common procedures used to obtain fetal DNA to diagnose chromosomal and other genetic abnormalities are chorionic (kor´ee-onic) villus sampling (CVS) and amniocentesis (am´nee-o-sen-tee-sis). Each procedure carries a risk of miscarriage and therefore it is only recommended for women with a high risk of a chromosome abnormality (e.g., those with an abnormal screening result, family history, advanced maternal age, or history of miscarriage).

CVS can be performed at an earlier stage of pregnancy (9–11 weeks' gestation) than amniocentesis (16 weeks). Tissue surrounding the fetus and comprising part of the placenta from where the baby gets its nutrients are known as the chorionic villi. A sample of the chorionic villi is extracted through a

flexible thin tube inserted through the cervix or a syringe through the abdomen. About 1–2% of women experience complications after CVS including bleeding, cramping, infection, and miscarriage.

Amniocentesis involves removing a sample of amniotic fluid from the fluid cavity surrounding the fetus. The fluid is usually extracted through a syringe inserted through the woman's abdomen. The fluid actually contains cells that are shed from the fetus such as fetal skin, bladder, and gastrointestinal tract lining. Ultrasound is used in both procedures to determine the location of the fetus and placenta.

Once a sample of the chorionic villi or amniotic fluid has been obtained, it is sent to a laboratory where it is placed in a dish and incubated for several days. This allows the cells from the small amount of tissue or fluid to grow for analysis. Laboratory technicians then place a drop of the sample onto a glass slide, add a stain it to better see the chromosomes, and examine the chromosomes under a microscope. Several cells are analyzed to determine the number of chromosomes in each cell and that no rearrangements have occurred the by carefully looking at banding pattern of each chromosome. When viewed under the microscope, the chromosomes appear mixed up inside of the cell. A karyotype (care-e-o-type) is a picture that is created from this analysis where each pair of chromosomes is arranged from number 1 to 21 followed by the pair of sex chromosomes (XX or XY) (See Figure 4.3). A test result of 46 XX indicates that the normal number of chromosomes is present and the baby is a girl. A diagnosis of Down syndrome would be 47 XX (or XY), +21 (indicating an extra copy of chromosome 21). Other more complex tests can be performed to detect CNVs.

One of the most frequent reasons for prenatal screening/diagnosis is advanced maternal age, defined as 35 years and older at delivery. The reason 35 years was chosen as the cutoff for advanced age is that the risk of a chromosome abnormality exceeds the risk of miscarriage associated with the amniocentesis procedure. In recent years, the limited provision of diagnostic tests based on age has been called into question. In 2007, the American College of Obstetricians and Gynecologists recommended that all women be given a choice of having screening tests and/or diagnostic tests, regardless of age.

As with any other test, prenatal diagnosis has its limitations on what it can and cannot predict. A prenatal diagnostic test cannot rule out every possible disease as it can only currently detect major chromosome abnormalities or specific genetic diseases in question. Chromosome analysis is estimated to be greater than 99% accurate, but other genetic tests can have variable accuracy. It is anticipated that more specific and/or comprehensive tests will be

developed to analyze fetal DNA extracted from maternal blood that will generate more detailed results. In a later chapter, we will consider some of the ethical implications of the improved testing technologies that can be performed earlier in pregnancy than ever before and for more and more conditions (or traits) that the baby may carry. While these new technologies may avoid the need for risky, invasive and costly procedures like amniocentesis, it may stir debate about what types of testing should be performed and potential disparities if testing is only accessible to those who can pay out of pocket for it.

Resources

US National Human Genome Research Institute. Chromosome Fact Sheet. Available at https://www.genome.gov/about-genomics/fact-sheets/Chromosomes-Fact-Sheet

US National Human Genome Research Institute. Chromosome Abnormalities Fact Sheet. Available at https://www.genome.gov/about-genomics/fact-sheets/Chromosome-Abnormalities-Fact-Sheet

Stanford Children's Health. Medical Genetics: How Chromosome Abnormalities Happen. Available at https://www.stanfordchildrens.org/en/topic/default?id=medical-genetics-how-chromosome-abnormalities-happen-90-P02126

US National Cancer Institute. Chronic Myelogenous Leukemia Treatment (PDQ®)–Patient Version. Available at https://www.cancer.gov/types/leukemia/patient/cml-treatment-pdq

US Centers for Disease Control and Prevention. Facts about Down Syndrome. Available at https://www.cdc.gov/ncbddd/birthdefects/downsyndrome.html

5

You Have Probably Already Had a Genetic Test (But No One Told You): Newborn Screening

Most people do not realize it, but if you were born in the USA in the mid-1960s onward, you have already had a genetic test of sorts. Of the approximately 4 million children born in the USA each year, more than 95% are screened for a set of inherited disorders. Screening of these diseases during the newborn period enables early identification of affected infants. Interventions can then be administered in the first few days of life that can substantially minimize or prevent symptoms of the disease from even developing. Thus, newborn screening can drastically change the course of the disease, allowing an affected child to have as healthy a life as possible.

Beginning of Newborn Screening: The First Disease Screened

The practice of newborn screening dates back to the 1960s to a man named Robert Guthrie. Born in 1916, Dr. Guthrie was trained as both a scientist and physician. In particular, he was trained as a microbiologist—a type of scientist who studies microorganisms such as bacteria, fungi, and other microbes. Dr. Guthrie had developed a personal interest in diseases afflicting children, as his son was affected with an undiagnosed developmental disorder as well as his niece with a biochemical disorder called phenylketonuria (fee-nil-keeto-nuria) or PKU for short. In 1959, Dr. Guthrie was studying cancer when he was asked by a colleague to develop a simple blood test to monitor levels of a particular chemical called phenylalanine (pronounced fee-nil-al-a-neen). This

S. B. Haga, *The Book of Genes and Genomes*, https://doi.org/10.1007/978-0-387-70916-1_5

chemical is increased in individuals affected with PKU. At this time, other scientists determined that symptoms of PKU could be prevented by avoiding foods containing phenylalanine in the child's diet. However, preventing the symptoms of PKU from developing would require identifying children at a very young age, since many foods contain phenylalanine, including milk.

Drawing a vial of blood was not feasible for infants and young children, but if just a drop of blood could be tested instead, collected from the prick of a finger or heel of the foot, the test would be much easier to conduct. So, working at the State University of New York in Buffalo, Dr. Guthrie set about to determine if a spot of blood collected and dried on a piece of filter paper (an absorbent sheet of paper) would be enough to accurately establish the levels of this particular chemical. After a couple of years of refining this method of screening, not only for PKU but also for other genetic diseases, a small pilot trial was launched.

Following a presentation of his research in the fall of 1961, local hospitals in New York began sending blood spots to Dr. Guthrie to screen, marking the official launch of newborn screening in the USA. He published his work in 1963 in the medical journal *Pediatrics* for hospitals and laboratories around the world to learn about and offer their own testing. Today, blood spots are collected by heel stick with a lancet from infants within the first 48 h of birth. The blood is blotted onto a small piece of filter paper and air-dried for 2–3 h. The blood spots are then sent in a sealed high-quality paper envelope to a state newborn screening laboratory or a private laboratory contracted through the state. Results are generally returned within 7 days.

From the last chapter, recall the distinction between *screening* and *testing*. The same distinction applies here—an abnormal newborn screen requires confirmatory testing before a diagnosis can be definitively made. In the event of an abnormal newborn screening result, all states have established a follow-up procedure for confirmatory testing and referral. Typically, the pediatrician listed on the hospital's record will be contacted about the abnormal screening result and initiate the process for confirmatory testing (Fig. 5.1).

Phenylketonuria (PKU)

PKU is a disease caused by an inborn error of metabolism, or an inability to break down certain chemicals in the body. A buildup of the chemical can harm certain tissues and organs. Children affected with PKU cannot break down a specific amino acid (one of the building blocks or subunits of proteins) known as phenylalanine (feen-ill-al-a-neen). If left untreated, over the

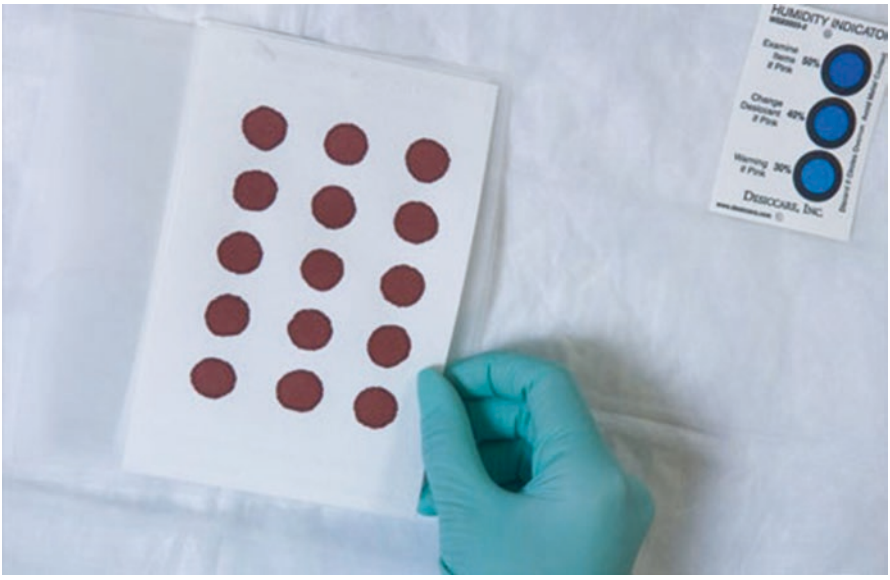

Fig. 5.1 Blood spots (source: US Centers for Disease Control and Prevention; https://www.cdc.gov/nceh/features/newbornscreening-program/index.html)

first few months of life, babies may begin to exhibit symptoms such as vomiting and frequent diarrhea leading to weight loss, light sensitivity, and skin conditions. As the buildup of phenylalanine worsens, children may begin to experience seizures and tremors and exhibit harmful behaviors such as head banging and arm biting. Ultimately, affected children will develop severe mental retardation.

The disease was first described in 1934 by the Norwegian physician Asbjorn Folling. Dr. Folling observed that affected children have a deficiency of a key enzyme important in the breakdown of phenylalanine. It was discovered that children with PKU have high levels of a related chemical in their urine known as phenylpyruvic acid. These high levels could be detected in the urine through a simple chemical reaction—the urine collected in a wet diaper turns green when the chemical, ferric (iron) chloride, is added. However, this test was unsuitable to screen for PKU as the levels of phenylpyruvic acid were too low to be detectable prior to the disease onset.

The treatment for PKU was actually realized before a screening test was developed. In 1953, German and British physicians showed that a diet low in phenylalanine could prevent most if not all of the symptoms of PKU from developing or at least control the symptoms of those already affected. Since phenylalanine is only found in proteins absorbed through the diet, it seemed

logical that limiting those foods would severely reduce the toxic buildup of phenylalanine in the body. However, the low-phenylalanine diet could not reverse symptoms that had already developed, thus raising the importance of early diagnosis or pre-symptomatic diagnosis.

Although the benefits of a low-phenylalanine diet are undeniable, the diet is strict and can be difficult to adhere to as children grow older. Foods high in protein such as milk and dairy products, meat, fish, chicken, eggs, beans, and nuts should be avoided. Low-protein breads and pasta, fruits and vegetables, and cereals make up a substantial portion of low-phenylalanine diet. A special phenylalanine-free baby formula is available to supplement the diet with the necessary proteins, vitamins, and minerals.

Although PKU is a rare disease, some manufacturers of popular food products have chosen to label certain products as unsuitable for PKU patients due to high phenylalanine content. For example, labels on chewing gum or diet soft drinks contain a warning—"*Phenylketonurics: contains phenylalanine*." This is because the artificial sweetener aspartame, used in many diet or sugar-free products, is primarily made up of two amino acids, one of which is phenylalanine.

Although it is recommended that the low-phenylalanine diet be followed for life, often, as children become teenagers and young adults, many gradually drift away from their strict diets. Rising phenylalanine levels have been associated with memory problems and difficulties concentrating and paying attention. For women, increasing phenylalanine levels can be particularly dangerous as they can be harmful to the fetus during pregnancy and cause small head size, growth problems, heart disease, and increased risk of mental retardation. Regular monthly blood tests are recommended to monitor phenylalanine levels.

State by State

Newborn screening programs are administered by the state and therefore programs may differ from state to state. In 1963, Massachusetts was the first state to implement a newborn screening program, screening for just one disorder, PKU. The development of PKU newborn screening programs was slow amidst debate about how effective the treatment was and whose responsibility it would be to collect the blood spots (the hospital, the obstetrician, the pediatrician, etc.). Efforts from several organizations concerned with child welfare such as the March of Dimes Birth Defects Foundation, state health departments, and the Kennedy administration through the Presidential Advisory

Commission on Mental Retardation and the federal Children's Bureau highlighted the importance and benefit of newborn screening programs to reduce unnecessary childhood morbidity and mortality. As a result of these efforts, by the 1970s, more than three-fourths of the states had enacted legislation establishing newborn screening programs.

Since the first program started, the number of diseases screened has increased from 1 (PKU) to 31 in Arizona to 71 in Tennessee (as of early 2021), with much of the growth occurring during the 2000s. New technologies were extremely expensive when first introduced and required trained laboratory technicians to operate. Now, all states use these testing technologies, providing rapid and low-cost screening for multiple diseases. Many of the diseases included in newborn screening programs are inborn errors of metabolism like PKU, where a modified diet can prevent disease or substantially reduce disease severity if started early enough (Fig. 5.2).

The discrepancy between state newborn screening programs has been of particular interest for the past few decades. In 2002, some states were screening for as few as four conditions, whereas others were screening for more than 30 conditions. Families with affected children born in states without a screening program for a particular disease have been extremely outspoken about the discrepancy between states. As a result, several groups have called for a national newborn screening program that would eliminate the discrepancies between the conditions screened in each state. In 2005, a federal committee called for uniform screening of 29 conditions; however, state compliance with the recommendations was voluntary, and differences remained. By 2007, a total of

AL	46
AK	53
AZ	31
AR	32
CA	64
CO	45
CT	66
DE	55
FL	56
GA	33
HI	49
ID	48
IL	65
IN	56

IA	53
KS	34
KY	59
LA	34
ME	52
MD	61
MA	66
MI	58
MN	61
MS	63
MO	76
MT	32
NE	37
NV	57

NH	39
NJ	57
NM	49
NY	60
NC	37
ND	53
OH	38
OK	54
OR	53
PA	38
RI	35
SC	54
SD	50
TN	71

TX	56
UT	53
VT	36
VA	33
WA	37
WV	39
WI	48
WY	52
DC	62

Fig. 5.2 The number of conditions currently screened in each state (as of April 11, 2021) (source: Baby's First test; https://www.babysfirsttest.org/newborn-screening/states)

10 states had implemented screening for the 29 recommended conditions. By this time, the variability in state newborn screening programs had been recognized as a major issue by some members of the Congress. In 2008, Congress passed a law called the Newborn Screening Act Saves Lives, which provides support to states to implement screening for at least 29 of the 31 conditions now included on the Recommended Uniform Screening Panel (RUSP). The Act was renewed in December 2014 to continue federal assistance for state newborn screening programs. The RUSP has been updated and as of 2018 includes 35 conditions.

Some private laboratories offer additional newborn screening services, beyond what most states make available. In most cases, parents can order these tests for their newborns through their pediatricians, but insurance may not cover the costs of additional newborn screening.

How Do States Decide Which Diseases to Screen?

Many states, laws on newborn screening define the for selection criteria of conditions to be included in the screening program, reimbursement, and appointment of a state agency (e.g., Department of Health) or committee to oversee the program and/or determine which conditions are screened. At the beginning of the newborn screening movement, a set of criteria were defined for considering whether a disease should be added to a state's newborn screening panel. More or less, these criteria have held up for almost 40 years:

- The condition being screened for should be an important health problem.
- The natural disease progression should be well understood.
- The disease should be detectable at an early stage.
- Early treatment should be beneficial (if treatment at a later stage is equally beneficial, early treatment is unnecessary).
- An accurate and acceptable test can be developed for early detection.
- The period for initial and repeat testing should be well defined.
- Appropriate healthcare services should be in place.
- The benefits of screening should outweigh the risks.

In general, the incidence of many of the diseases screened is quite rare (Table 5.1). For example, congenital hypothyroidism has an incidence of about 1 in 3000 newborns, the highest of all of the conditions currently screened. The most rare disease screened for is galactosemia, estimated to affect 1 in 53,000 newborns. In total, though, it is estimated that 1 out of 300

newborns in the USA is affected by a condition identified through screening. Some diseases have a higher incidence of individuals of a certain background; for example, sickle cell anemia is twice as common in individuals from families of African, Southeast Asian, or Mediterranean descent (~1 in 3700), compared to those of European descent (~1 in 7400).

One of the diseases that was debated for several years about whether to include it in newborn screening programs was cystic fibrosis (CF). Cystic fibrosis is an inherited disease that is caused by mutations in both copies of the CF gene (recall that we have two copies of every gene, one from each parent). The gene produces a protein that is involved in the production of different bodily fluids such as sweat, digestive juices, and mucus. As a result, CF affects many systems in the body ranging from the lungs to the pancreas to the immune system. Symptoms of CF can begin to show as early as the newborn period and include poor growth and lung infections. The life expectancy of people affected with CF has greatly improved over the past 50 years; many individuals can live well into their 30s and 40s, with early intervention and treatment.

Many states have implemented a two-step test to identify babies affected with CF. First, newborn screening is performed for a protein (called IRT), which is made by the pancreas and is elevated in individuals with CF. If newborn screening shows a high IRT level, a second test is ordered which analyzes the CF gene to determine if the child has mutations in both copies of the CF gene. Colorado was the first state to add CF to their newborn screening program in 1989, followed by Wisconsin and Wyoming. Other states decided to implement a nonmandatory screen, based on the availability of services in each hospital. In 2004, the Centers for Disease Control and Prevention (CDC) recommended that CF be included in newborn screening programs based on "evidence of moderate benefit and low risk of harm." Today, all states include CF in their newborn screening program. CF has the second highest incidence following sickle cell anemia.

Table 5.1 The incidence of some conditions screened for in state newborn screening programs

Genetic condition	Incidence
Congenital adrenal hyperplasia	1 in 19,000
Congenital hypothyroidism	1 in 3000
Galactosemia	1 in 53,000
PKU	1 in 14,000
Sickle cell anemia	1 in 3700/1 in 7400 (European descent)

Another important consideration with expansion of the newborn screening program is the economic feasibility and cost-benefit. The cost of newborn screening is usually nominal, ranging from $30 to $165. Some states do not charge a fee, but for those states that do, government programs and private insurers will generally cover the newborn screening fees. In other states, the cost is included with maternity charges. Only a few studies have been conducted to determine the cost savings achieved through newborn screening programs. For the diseases studied, the reduced costs are mainly attributed to the prevention of onset of disease and thereby the need for expert medical care and services.

One of the bigger and more costly issues is the treatment and care of an affected infant. Given that different healthcare insurers may classify the needed treatment differently, it cannot always be assumed that the expenses to care for an affected child will be covered. For example, low-phenylalanine formula required for PKU babies may be considered as a food by some insurers and will not be covered, whereas other insurers may classify it as a drug and cover it. The discrepancy between state mandatory newborn screening policies and treatment coverage policies is of substantial concern since the goal of screening is to initiate early intervention to prevent or minimize disease symptoms.

Do Parents Have a Choice?

Most newborn screening programs are defined by state law, mandating that every newborn be screened. Except in Wyoming, informed consent (obtaining parents' permission) for the procedure is not required as it is believed that the benefit of identifying an affected child overrides any parental objections to or perceived risks of screening. Most states post information about the screening program on the health departments' websites; however, parents may not be aware of screening nor have been provided information about it during the prenatal or postnatal period. More than 30 states allow parents to decline newborn screening based on religious objections including Alabama, Arkansas, California, and Georgia. Newborn screening can be declined for any reason in 13 states including Alaska, Colorado, Florida, Iowa, and the District of Columbia. Obviously, parents would have to be aware that newborn screening would be performed in order to decline in advance if consent is not obtained.

Moving Beyond Screening for Treatable Diseases

Newborn screening programs have always been characterized as prevention programs since early diagnosis can lead to preventing disease or substantially reducing disease severity. As new technologies are developed, however, the ability to detect different chemicals and proteins in the body associated with other diseases will increase and potentially lead to continuing expansion of newborn screening programs. In the era of genomics, DNA sequencing enables identification of changes in genes that are associated with disease, which might not be detectable by newborn screening. But this does not necessarily mean that early knowledge of disease risk will result in disease prevention or improved health outcomes. Some diseases may not develop until later in life (adulthood), and early knowledge may cause more harm (e.g., parental anxiety, stigmatization) than benefit if there is no intervention that can be administered early.

However, research has found that some parents would like information about disease risks for their child, even for untreatable diseases. Although the information may not improve the health of their newborn if affected, such knowledge may be valuable to future decisions regarding family planning. Other benefits may include psychosocial benefits, access to new though unproven treatments, and access to services such as special education needs, physical therapy, or other support services. Regardless of how the information may be used, many parents believe that they should have the final say about testing of their children.

Conclusion

Newborn screening is one of the most successful public health programs ever developed. With rapid advances in testing technologies, federal and state programs will continue to review newborn screening programs and consider whether new conditions should be added. Technology, however, has a way of running ahead of our policies, system infrastructure, and knowledge. That is not to say that we should not take advantage of new advances early on, but the rule of thumb has generally been to proceed with caution. New technology has and will continue to outpace our ability to understand what it all means and, more importantly, what should be done if an "abnormal" result is detected—two of the criteria that should be met before adding a new disease to a newborn screening panel. Because the technology is now available, should

we change the criteria and screen for diseases in which we do not completely understand what is going on or how to treat? What are the benefits and risks of a positive screening result in the absence of a firm understanding of what to do with that information? Is more knowledge better or can it be harmful? As scientists and society wade through all of the data and new technologies, the issue of whether we *can* is now moot, but rather it is a question of if we *should* (a recurring theme for many applications discussed in this book).

Resources

Baby's First Test. Available at BabysFirstTest.org

US Centers for Disease Control and Prevention. Newborn Screening Portal. Available at https://www.cdc.gov/newbornscreening/index.html

Association of Public Health Laboratories. NewSteps. Available at https://www.newsteps.org/

6

Sweet Blood

Type II diabetes mellitus, or diabetes for short, is actually a very old disease. The earliest known mention of the disease goes back to 1500 BC, when the physician Hesy-Ra observed frequent urination as a symptom. One thousand years later in 150 BC, the physician Aretaeus of Cappadocia (located in what is today Turkey) described the condition as "the melting down of flesh and limbs into urine." Aretaeus used the Greek word *diabetes*, meaning that which passes through, to describe a patient with symptoms of excessive urination.

The second part of the disease's name (mellitus) was added on in the eleventh century. By that time, it had been observed that the urine of people affected with diabetes was sweet tasting. The Latin word for honey—*mellitus*—was added to the name to incorporate this observation. Diabetes has since been described by a number of different names, including the sugar disease, sweet blood, sugar in the blood, sugar sickness, or just sugar.

According to the *National Diabetes Statistics Report for 2020*, about 10% of the US population (34.2 million people) have diabetes. However, about 25% are unaware that they have it and about 7 million people have diabetes but have not yet been diagnosed. About 1.5 million people were diagnosed in 2018. As diabetes is a lifelong chronic disease, the costs of healthcare are staggering—in 2017, the total estimated cost of diagnosed diabetes in the USA was $327 billion. In addition to direct medical expenses, this includes other costs such as disability pay, sick days, and premature death. In addition, a person with diabetes will likely pay $9000 in health-related expenses.

While significant advancements have been made in the last century in the diagnosis and treatment of diabetes, the genetic underpinnings of the disease

S. B. Haga, *The Book of Genes and Genomes*, https://doi.org/10.1007/978-0-387-70916-1_6

remain largely a mystery. While it has been known for some time that a family history of diabetes increases a person's risk, the specific gene(s) behind the disease have proved elusive to researchers. Part of the challenge in identifying the genes for a disease as complex as diabetes is the substantial influence of environmental factors such as diet and exercise. Furthermore, it is likely that more than one gene is involved as diabetes may develop in different ways. However, recent work enabled by new laboratory technologies has led to a number of breakthroughs and even a test to determine one's risk to develop diabetes. Hopefully this work is a harbinger of more discoveries to come that will not only enable doctors to identify who is at increased risk but also lead to new treatments to prevent the onset or development of the many symptoms associated with diabetes.

What Is Diabetes?

Our body's major energy source is glucose, a form of sugar obtained from the food we eat after it is broken down. Glucose is essential for growth and energy, but our body is very sensitive to the levels of glucose in our blood. In order for glucose to move out of the blood and reach the cells in our body that need it, another molecule called insulin must be present. Insulin is not found in foods, but rather is a hormone that is produced by special cells found only in the pancreas, an organ about six inches long located behind the stomach. The pancreas produces as much insulin as needed to keep the body's glucose levels at a safe level.

Diabetics suffer from a problem with insulin—either little or no insulin is produced by the pancreas or if produced, the body does not respond correctly to the insulin. Despite the high levels of glucose in the blood, the body's cells are starved for energy since the insulin is not there to signal cells to take up the glucose (a very controlled, stepwise process). High glucose levels in the blood can cause damage to several parts of the body and lead to numerous conditions including heart attacks, strokes, kidney disease, and blindness (Fig. 6.1).

Despite the long history of diabetes, it was not until 1959 when it was recognized that two major types of diabetes existed: type 1 (insulin-dependent or juvenile diabetes) and type 2 (non-insulin-dependent) diabetes. Type 1 diabetes accounts for about 5–10% of all diabetes cases in the USA. It typically develops in children and young adults though it can develop at any age. Type 1 diabetes develops when the body's immune system mistakenly attacks and destroys the cells in the pancreas that produce insulin. As a result, patients with type 1 diabetes require daily insulin injections to regulate their blood

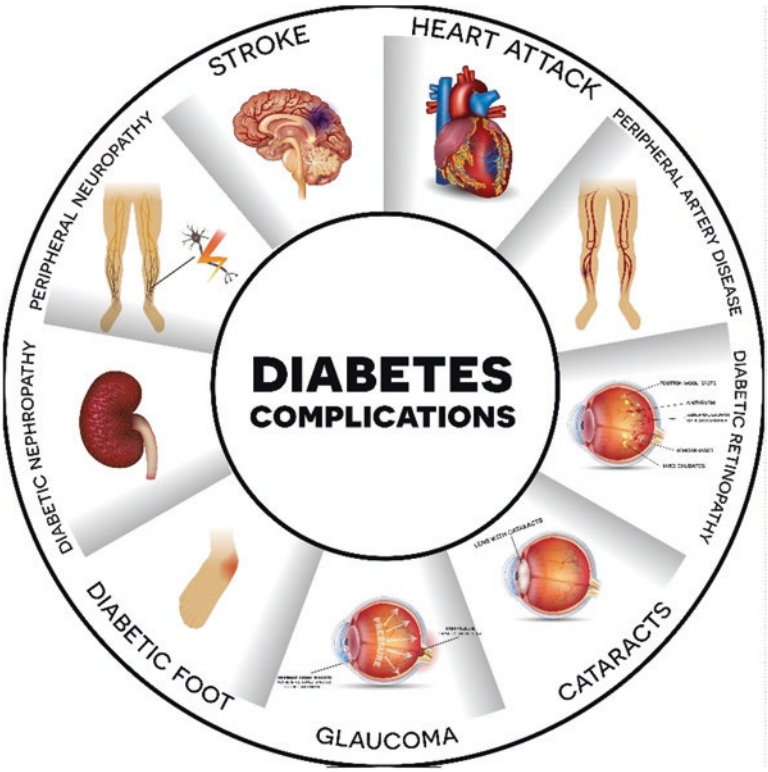

Fig. 6.1 Diabetes complications (source: Adobe Photo Stock)

sugar levels. If left untreated, a person with type 1 diabetes can fall into a life-threatening diabetic coma.

Type 2 diabetes is the more common form, accounting for about 90–95% of people with diabetes. The risk for type 2 diabetes increases with age and is often associated with obesity and physical inactivity; hence the rising number of people with obesity will lead to even more people with diabetes. However, although type 2 is typically associated with older adults, the age of diagnosis is getting younger and younger, another significant concern as these people are at increased risk to develop health complications at an earlier age as well.

Another distinction between type 1 and type 2 diabetes is that type 2 diabetes gradually develops over time. This slow onset of symptoms can result in the disease being undiagnosed for some time. People with slightly elevated glucose levels are often diagnosed as "prediabetic." These slightly elevated blood glucose levels can increase a person's risk of developing full-blown diabetes as well as heart disease and stroke. At least 40% of American adults (age 40–74) are believed to be prediabetic, the majority of which will develop

diabetes within 10 years. Symptoms in people with diabetes can be quite mild and go unnoticed, whereas others may experience fatigue, frequent urination, increased thirst and hunger, weight loss, and blurred vision.

Type 2 diabetes is also known to disproportionately affect minorities, which may be due to a combination of genetic and environmental (lifestyle) factors. Family history of diabetes or gestational diabetes (a third type of diabetes that develops during pregnancy) can increase a person's risk of developing type 2 diabetes. About 3–8% of pregnant women develop high blood glucose levels during pregnancy, which can be harmful to the baby if not controlled.

Treatment of Diabetes

For a long time, no treatment existed for diabetes and its diagnosis meant a slow death sentence. The connection between diet and high levels of sugar was observed during the mid-1800s. Interestingly, diabetic patients during the Franco-Prussian war had improved levels of sugar in the urine, believed to be due to their restricted diet because of food rationings. This observation formed the basis for special diets for diabetic patients. In the early 1900s, a number of "fad" diets were recommended to lower glucose levels in diabetics such as the oatmeal diet, milk diet, rice cure, and potato therapy.

Although insulin was known to be produced in the pancreas and was the key to regulating blood glucose levels, it was not until 1921 that insulin was discovered by Canadian physician Dr. Fred Banting. After successfully extracting insulin from the pancreas of dogs, Dr. Banting and his colleagues administered a preparation of the extracted insulin to dogs whose pancreas had been removed. They observed improved glucose levels and recovery from diabetes. Shortly thereafter in 1922, the first patient, a 14-year-old boy, was treated with insulin extracted from cow and similarly showed significant signs of improvement. The worldwide significance of their important discovery did not go unnoticed. In 1923, Dr. Banting and his collaborator, the Scottish physiologist Dr. J.J.R. MacLeod of the University of Toronto, were awarded the Nobel Prize. The mechanism of how insulin works in controlling the entry of glucose into cells is shown in Fig. 6.2.

Following Dr. Banting's successful human testing studies, insulin extracted from other animal pancreases (cow, horse, pig) was used to treat diabetics. However, the often impure insulin extracts would cause side effects in patients. In the 1930s, pharmaceutical companies developed a slow-release form of insulin that would last up to 36 h.

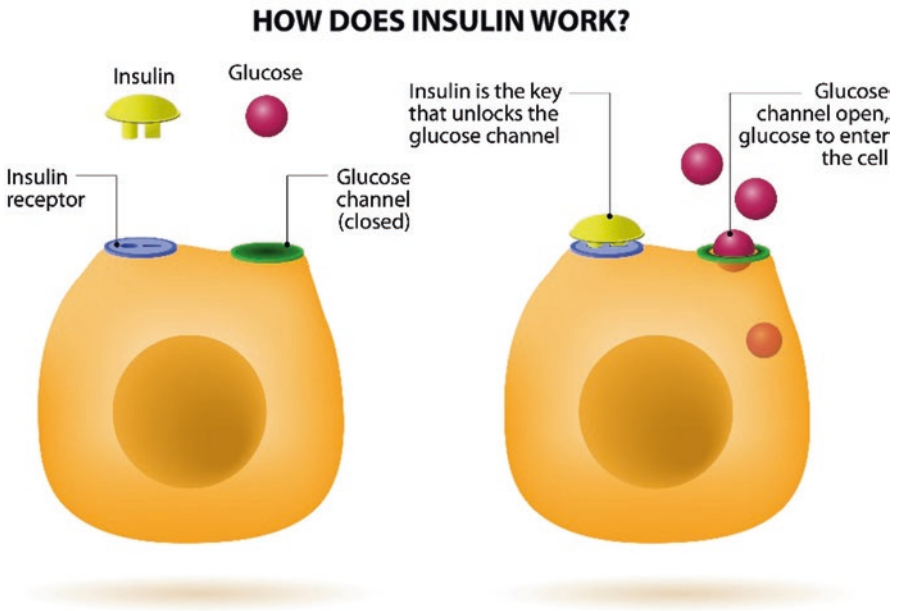

Fig. 6.2 Mechanism of how insulin works to control glucose levels in the body. Our body's cells have very strict controls on the movement of molecules in and out of cells. Many molecules can only move in and out through a specific channel or entryway (like a door); however, the door can only be "unlocked" by another molecule. This is the role of insulin. On the left, glucose cannot be taken up by the cell since insulin is not bound to the insulin receptor, which will send a signal to "open" the receptor (or door) for glucose to enter (source: Adobe Photo Stock)

In 1977, the gene for human insulin was finally identified. At this time, there was much excitement for a new field called biotechnology and the potential for new therapies using DNA. In 1978, scientists from a biotechnology company called Genentech engineered a small circular piece of DNA containing the insulin gene. They then inserted the insulin DNA into a strain of the common bacterial strain *E. coli* in the laboratory. As the *E. coli* rapidly grew, it produced large quantities of insulin that could be easily extracted from the bacterial mixture. In 1980, the genetically engineered form of insulin was first tested in humans and shown to be slightly more effective than the pig-purified insulin in use at the time. Another company called Eli Lilly beat Genentech in developing the first approved form of synthetic human insulin and began marketing the lifesaving treatment in 1982.

People with type 1 diabetes must take insulin to keep their blood glucose level as normal as possible, maintain a healthy diet, and be physically active.

Blood glucose levels can be checked daily using handheld glucose meters that are so sensitive that they require only a finger-pinprick of blood.

Type 2 diabetes is initially caused by the body's failure to respond to the insulin produced by the pancreas. In other words, the insulin levels are normal, but the body appears not to respond as well to the insulin signal. Over time, the cells' "resistance" to insulin may increase. Treatment may initially start with a careful diet and regular exercise, but medication may eventually be needed to control the glucose levels. Thus, as the disease changes, treatment for type 2 diabetes changes.

In women with gestational diabetes, the high blood sugar levels can cross the placenta and trigger increased insulin production in the fetus. The fetal insulin will signal for the glucose to be stored as fat resulting in bigger babies that may need to be delivered by a Cesarean section. In addition, children of mothers with gestational diabetes are at increased risk for breathing problems, obesity, and type 2 diabetes. Therefore, it is critical for expectant mothers with gestational diabetes to maintain glucose levels during pregnancy as close to normal as possible through a careful diet and regular physical activity. Some may need to perform daily blood glucose testing and possibly require insulin injections.

Genetics of Diabetes

For centuries, it has been long known that both type 1 and 2 diabetes run in families, establishing a strong connection between genetics and diabetes. Relatives of diabetics have a higher chance of developing diabetes than those without a family history. While the treatment of diabetes has significantly benefited from the discovery of the insulin gene and biotechnology, our understanding of the genetic causes of diabetes still remains much of a mystery. Both types of diabetes are believed to be caused by a combination of genetic and environmental factors (e.g., infections, diet, physical inactivity). The completion of the sequencing of the human genome and development of new technologies to "scan" DNA for genetic variants have substantially enabled scientists to more rapidly identify genes that contribute to complex diseases like diabetes by analyzing DNA from thousands of patients. However, at this time, there are no clinical genetic tests available to predict a person's risk of diabetes but progress is being made to unravel the genetic causes.

Type 1 Diabetes

At least 20 different genes/gene regions have been associated with susceptibility to type 1 diabetes. One of the strongest genetic links to type 1 diabetes is found in a family of genes that form the basis for our body's immune response. Known as *h*uman *l*eukocyte *a*ntigens (HLA), these genes produce the proteins that sit on the edge of our immune cells exposing foreign pathogens (bacteria, viruses) to the immune defense cells and signaling destruction of these cells. Thus, the HLA proteins are critical in indicating which cells are foreign from our innate (self) cells. Without the ability to distinguish self (our cells) from nonself (bacteria, virus), our body could get confused and mistakenly mount an attack on different cells in our body. Unfortunately, this can happen, resulting in a variety of diseases including type 1 diabetes, multiple sclerosis, and rheumatoid arthritis. These types of diseases are known as autoimmune diseases. In type 1 diabetes, the cells of the pancreas that produce insulin are destroyed or damaged, thereby severely limiting or eliminating insulin production.

Located on chromosome 6, the HLA gene family actually contains hundreds of genes involved in our body's immune response. It is estimated that some of the HLA class II genes in particular, referred to as insulin-dependent diabetes mellitus 1 (*IDDM1*), account for approximately 40–50% of the inherited risk for type 1 diabetes. Despite the strong association between these genes and type 1 diabetes, how exactly these genes contribute to disease development is not entirely clear to date.

Type 2 Diabetes

At least ten genes have been associated with increased risk to type 2 diabetes. However, given the complexity of diabetes and the substantial influence of environment factors, it is estimated that genetic factors may only moderately increase risk, perhaps as high as 20%. As the picture of genetics behind type 2 diabetes is slowly becoming clearer, it appears that a combination of genetic variants will give the best estimate of disease risk than any single gene could. The environment (diet, level of physical activity) could be the major trigger for disease onset.

Three major genes, *KCNJ11*, *PPARG*, and *TCF7L2*, have been associated with an increased risk to type 2 diabetes mellitus in several studies. KCNJ11 was first linked to a type of diabetes that arises in the first 6 months after birth, known as neonatal diabetes. The gene encodes a protein that forms part

of what is called a "channel." The KCNJ11 channel acts like a toll booth regulating the passage of various molecules inside and outside of the cell depending on the cell's needs. In particular, the channel is responsive to signals regarding insulin secretion. If the channel is not functioning properly due to a variation in the DNA sequence of the gene causing some sort of structural defect in the channel, then the correct amount of insulin is not secreted resulting in high glucose blood levels and development of diabetes. Mutations in KCNJ11 can either increase or decrease insulin secretion depending on the type of mutation. Increased insulin secretion will cause a hypoglycemic state (abnormally low blood sugar levels)—the opposite of what occurs in diabetes. The more severe the genetic mutation with respect to the function of the protein, the more severe type of diabetes will develop at an earlier age.

Another validated genetic risk factor for type 2 diabetes is the PPARG gene. PPARG is involved in the creation of fat cells during development and metabolism of fatty acids. Not surprisingly then, genetic variants in PPARG have been associated with increased risk of developing obesity and type 2 diabetes. The PPARG protein is quite sensitive to a number of drugs and therefore, diabetics must be careful in monitoring which drugs they take. Increased PPARG levels can lead to enhanced sensitivity to insulin and reduced blood sugar levels in diabetics. As 75% of the general population carry the genetic variant associated with increased risk (albeit a small risk—about 1.3%), the impact is significant from a public health viewpoint and its impact on diabetes risk. However, some versions of the gene have also been shown to be protective and actually reduce the risk of type 2 diabetes.

Probably the most exciting genetic discovery in type 2 diabetes research is the TCF7L2 gene. The link between this gene and type 2 diabetes was discovered in 2006 by scientists at the Icelandic company deCODE genetics. The Iceland population is believed to be one of the "easier" populations in which to study genetics due to several reasons. First, the population is rather staid—meaning that there has been little movement of people in and out of Iceland, resulting in what geneticists call a homogeneous (ho-mo`jeen-e-us) population. With few immigrants, the same gene versions have been circulating among Iceland's inhabitants for centuries resulting in a rather uniform genetic makeup of the population. Since the genetic differences among a group of people from Iceland are far fewer than a highly mixed population as that found in the USA or Brazil, it becomes much easier for geneticists to identify a gene or gene region linked to a disease. In addition, Icelanders are known for their meticulous record-keeping and have records going back several generations documenting their family's health history, which is extremely helpful to researchers to identify families with a heritable disease.

In a preliminary study of diabetics from Iceland, researchers found that a gene region on chromosome 10 was strongly associated with diabetes. In order to identify which gene in this region of many genes was the culprit gene, researchers looked at several genetic variants from genes located in this region. They found that variants in one particular gene, called TCF7L2, were strongly associated with type 2 diabetes in more than 1100 diabetes patients from Iceland, and almost 600 more patients from the USA and Denmark. Amazingly, individuals carrying two copies of the *TCF7L2* gene variant have an approximately twofold higher risk of developing diabetes, a much higher risk than for any other genetic variant known to date. About one in five people with diabetes (~18%) will carry two copies of this genetic variation. The gene is believed to be involved in regulating insulin secretion.

These findings were subsequently replicated and confirmed in many different populations including Japanese, American, Mexican, West African, and Finnish. While the effect of the *TCF7L2* genetic variant holds up across different populations, its frequency varies between groups with the highest frequencies found in Native Americans (particularly Pima Indians of the Southwest USA). The gene variant is also more common in Hispanics and African-Americans than Whites.

Would You Change Your Behavior If You Knew You Were at Risk for Diabetes?

Although no clinical tests to predict diabetes are available, it is likely that testing will become available in the future. Based on analysis of the TCF7L2 gene and several other genes, a person's risk of diabetes could be estimated. As the contributive risk of each gene is likely to be rather small (the exception being TCF7L2), testing a panel of genes known to be associated with type 2 diabetes would result in a more accurate and reliable test. For example, testing for TCF7L2 in combination with *PPARG* and *KCNJ11* has better predictive power than testing for TCF7L2 alone.

There has been a lot of discussion about whether knowledge of one's genetic risk status would result in improved health outcomes, i.e., the prevention of diabetes onset or early diagnosis and reduced disease severity. Knowing that you are at increased risk for diabetes based on this test may motivate some people to make an honest-to-goodness effort to change their lifestyle to reduce their risk as much as possible. Although you cannot change your genes, other

factors such as diet and exercise are likely to play a much bigger role toward your risk of diabetes that can be improved.

On the other hand, some people who find that they are at increased risk may take a defeatist attitude and think that no matter what they do or change, they will likely develop diabetes anyway because of their genes, disregarding the environmental impact. Or could some people develop a deep anxiety about their increased risk and/or engage in extreme weight loss or obsession to reduce their risk, such that knowledge of their risk is actually harmful? As simply stated by a group of public health scientists from the Netherlands and the USA following the discovery of TCF7L2, "predictive genetic testing is useful when the value it adds to existing interventions outweighs the additional personal and social costs."

So, the 6-million-dollar question is would you change your lifestyle to reduce your risk of diabetes if you were found to be at increased genetic risk. Given the lifelong treatment and the debilitating consequences of untreated diabetes, one would think that this would be a no-brainer. But like other habits that we engage in despite knowing how unhealthy they are (e.g., smoking, not exercising, eating fatty foods), it is not certain that knowing your genetic predisposition to diabetes would really change behaviors all that much. We already know that we should be exercising daily and watching what we eat—would a genetic test that suggests a higher-than-average risk of diabetes motivate us to finally buy that gym membership or start eating healthy? Not to sound too negative, but chances are not for many people. It will require massive public education and changes to our daily lives to facilitate healthier living that extends well beyond a simple genetic test result.

Resources

Diabetes Research Institute. Available at https://www.diabetesresearch.org/diabetes-statistics

National Institute for Diabetes and Digestive and Kidney Diseases. Diabetes. Available at https://www.niddk.nih.gov/health-information/diabetes

American Diabetes Association. Available at https://www.diabetes.org/

7

Will This Drug Work for You?

You might have noticed that people can respond differently to the same drug used to treat the same condition. Talking with your friends, co-workers, and family members, you might learn that a drug worked but caused a side effect in one person, worked somewhat with no side effects in a second person, and did not work at all in a third person. You yourself have likely gone back to your provider and/or the pharmacy to get another drug because the first one prescribed did not work or you experienced a side effect (and possibly needed another drug to fix the side effect from the first drug). The new prescription may be for a different drug altogether or for a different dose of the first drug. This approach to treatment can cost millions of dollars in healthcare costs due to multiple prescriptions, multiple office visits and follow-up care, patient inconveniences and extra sick days, and even hospitalizations caused by side effects.

Providers routinely consider multiple factors when prescribing a drug, but identifying the safest and most effective treatment is still often achieved through a "trial-and-error" approach. Peoples' varying responses to medications may be attributed to a range of factors, such as drug-drug interactions (when taking multiple medications at one time), incorrect prescribed drug or dose, incorrect diagnosis, race, diet, and genetics. Obviously, it is this last factor I would like to discuss here. The study of drug response and genetics is considered one of the most promising genetic applications today. Along with clinical information, genetic testing for drug response may help doctors to more accurately prescribe the safest and most effective drug at the right dose.

S. B. Haga, *The Book of Genes and Genomes*, https://doi.org/10.1007/978-0-387-70916-1_7

In 1959, the combined field of the study of drugs (or pharmacology) and genetics was named *pharmacogenetics*. The term *pharmacogenomics* is often used interchangeably with pharmacogenetics and reflects the more expansive use of genomic technologies. Personalized medicine refers to the tailoring of clinical care and interventions (including drugs) to a patient's unique circumstances instead of relying on population averages or trends ("the average risk of this disease is this"). In the absence of specific information about an individual patient, these population averages are a useful guide to inform preventive care. But consideration of a patient's family health history, health behaviors (e.g., smoking), exercise habits, occupation, age, gender, weight, race, and now genetics will yield a more accurate risk assessment for a given patient. The same also applies to drug response—genetic information may be used, in addition to many other factors, to predict how a patient may respond to a prescribed drug and therefore help healthcare providers select the best drug that will work for a given patient with minimal side effects. As described later, pharmacogenetic tests are considered as one of the early applications arising from the personalized or precision medicine movement. The phrase "personalized medicine" has evolved to "precision medicine" (1) to reflect that current practice already does consider many patient factors in determining the best course of care (it is personalized) and (2) to dispel the misinterpretation that personalized medicine means that each patient will receive a unique intervention or care plan. Thus, the phrase precision medicine aims to convey the use of multiple types of information to more precisely determine a patient's disease risk and drug response to inform and optimize decisions about a patient's treatment or preventive care.

Short History of Genetics and Drug Response

Although there has been intense focus on the genetics of drug response in recent years following the avalanche of genetic research and development of new testing technologies, the study of genetics of drug response has actually been around for quite some time. Going back to the early 1900s, it was observed that, in general, the majority of the population may tolerate a new drug well; however, a small percentage experience side effects (also known as "adverse response)," which may vary in severity and type. The side effects may be due to drug toxicity, caused by inefficient or improper breakdown of the drug.

In the 1950s and 1960s, a few major discoveries were made that demonstrated the connection between genetics and drug response and introduced the concept of drug treatment based on one's genetic makeup. In the summer of 1951, soldiers returning to the USA from the Korean warfront began showing symptoms of malaria. Malaria was a disease that was all too familiar to the US military from their exposure to it in World War II in the South Pacific; about 70 troops out of every thousand stationed in the South Pacific were affected with malaria. In fact, malaria affected most of the major US military campaigns dating back to the Revolutionary War.

The drug quinacrine was the most effective drug available in the 1940s; however, the supply of quinine, the main ingredient of quinacrine, was available in only limited quantity. Quinine, a natural substance found in the bark of the cinchona tree grown in South America and other places, has been used to treat malaria for more than 350 years. During the World War II, supplies of quinine for the USA and Allied forces were cut off by the Japanese. Research immediately commenced to find an alternative drug to treat soldiers before the disease wiped out military units based in the South Pacific and India. A synthetic version of quinacrine called atabrine had been used on a trial basis against malaria in Panama in 1935. When sufficient quantities of atabrine became available in 1945, it was administered to all US military personnel stationed in areas where malaria was endemic. This drug successfully obliterated the devastating effects of malaria on troop readiness, substantially reducing the impact of the disease on troops stationed in these regions.

Despite its effectiveness, however, atabrine had a number of side effects. Soldiers complained of the bitter taste, skin discoloration, headaches, nausea, vomiting, and, although rarely, temporary psychosis. A researcher named Alf Alving at the University of Chicago noticed that African-American soldiers in particular were more likely to experience a bad reaction to some antimalarial drugs. In particular, African-American soldiers would develop a severe form of anemia that would cause severe weakness and fatigue, more often than was experienced by white soldiers. With the drug called pamaquine, a dose of 90 mg was required to effectively treat the malarial strain prevalent in the South Pacific, but a dose of only 30 mg was found to be toxic to African-American soldiers. For another drug called isopentaquine, there was no "safe" dose for African-American soldiers.

To quell the sharp increase of malaria cases being diagnosed in the US military, health officials decided to give antimalarial drugs to infected

veterans as well as to soldiers returning to the USA from the Korean War as a preventative measure. The drug primaquine was administered to returning soldiers, a drug determined to be safer and more effective in halting the infection. However, some soldiers experienced side effects from primaquine and a reduced dose was required for the treatment of African-American soldiers.

In 1956, Dr. Alving discovered the reason behind the severe response to the many antimalarial drugs—a genetic change or variant in a gene located on the X chromosome. As it turned out, the gene, called *glucose-6-phosphate dehydrogenase or G6PD*, often carries a change in its sequence and is known to be one of the most variable genes to date. More than 400 million people carry one of more than 100 known genetic variants that can result in very low levels of the enzyme and potential side effects. The genetic variants are particularly common in individuals of African or Mediterranean descent, explaining the observation of affected African-Americans following treatment with primaquine. In individuals that carry this genetic variant, the red blood cells (a type of cell found in your blood that carries oxygen to tissues throughout your body) are more susceptible to oxidative damage, which eventually kills the cells. The low levels of G6PD protein may go unnoticed until the individual is exposed to a certain food or drug or is affected by illness, which can place stress on the system and trigger red blood cell damage and potentially death. For example, G6PD deficiency can also cause a condition called favism following digestion of fava beans due to certain chemicals found in these types of beans that can lead to liver and kidney problems.

Because G6PD is on the X chromosome (a sex chromosome), the impact of genetic variants primarily affects men, since women have two X chromosomes. The effect of a genetic variant in a gene on the X chromosome is buffered by the second copy on the other X chromosome in women.

Shortly after the discovery of the G6PD variants and its association with adverse side effects with the antimalarial drug primaquine, another gene linked to the adverse response to an anesthetic drug called succinylcholine was identified. Individuals who carry this genetic variant and take this drug will develop a prolonged case of apnea (muscular paralysis).

Understanding the Genetic Mechanisms Behind Drug Response

When a drug is taken, it will move through the body and undergo several stages of "processing" in order for it to be converted to an active form of the drug (some drugs are taken in an inactive form), be transported to the target area (tissue), work as intended, and then exit the body. Many genes that are known to be prone to genetic variation are involved in these stages of drug metabolism, transport, and excretion and thus can impact how efficiently the drug is processed and likelihood of a side effect.

To better understand *how* genes can impact drug response, let us take a closer look at the response to the commonly used pain medication codeine. At one time or another, most of us have probably taken codeine as a pain medication. Codeine is manufactured in combination with other pain medications such as aspirin, ibuprofen (brand name: Advil), or acetaminophen (brand name: Tylenol). Codeine is classified as a controlled substance and generally available only by prescription (some products containing small amounts may be available over the counter).

Codeine is a prodrug, which means that it is inactive until a reaction that occurs to convert it to its active form. So, when a pill containing codeine is swallowed, the body has to convert (or metabolize) codeine to its active form, morphine, before the effects of the drug will be felt (pain reduction). However, approximately 6–10% of Whites, 2% of Asians, and 1% of individuals of Middle Eastern descent carry a genetic variant in the gene encoding the enzyme responsible for converting codeine to morphine, called CYP2D6. Without the correct form of this enzyme, codeine is not converted to its active form and remains in its inactive state in the blood, and no therapeutic benefit is felt. Individuals with a genetic variant in the CYP2D6 gene have poor metabolism due to their inability to normally metabolize or break down certain drugs, including codeine. In contrast, a small number of patients (0.5–2%) carry multiple copies of the CYP2D6 gene; these extra copies cause individuals to quickly metabolize drugs and thus, they rapidly convert codeine to its active form of morphine. Either situation could result in ineffective pain management, and in some cases even death due to drug toxicity.

A second example of how genetics may influence drug response actually involves grapefruit juice. Oftentimes, a warning label is affixed to a prescription bottle to not take the medication with certain foods. For some prescriptions, the pharmacist may advise you not to take certain drugs with grapefruit juice. What is so special about grapefruit juice compared to apple, orange, or

cranberry juice with respect to medications? The effect of grapefruit juice on drug response was discovered by accident when scientists were experimenting using grapefruit juice to mask the bitter taste of a drug and it is believed to be the first food reported to influence drug response. When you drink grapefruit juice while taking a pill, the grapefruit juice acts like a second drug and interferes with the body's metabolism of the actual drug. More specifically, the grapefruit juice inhibits enzymes known as CYP3A (related to CYP2D6) essential to the metabolism of a large number of drugs. Even if drunk hours before taking the drug, grapefruit juice can increase the levels of a drug to toxic levels because the enzymes cannot reach the drug to break it down as they are still preoccupied with breaking down the chemicals in grapefruit juice. This effect has also been observed with other citrus such as tangerines, but not oranges. The CYP3A family of genes is probably the most important group of genes involved in drug response as they metabolize 45–60% of currently prescribed drugs. Thus, patients prescribed drugs that rely on the CYP3A enzymes are advised not to drink this juice specifically during the course of treatment (or a drug and a food).

Likewise, taking two drugs that are both metabolized by CYP3A is not recommended as it can also reduce the availability of the enzyme. As described in the example of CYP2D6 and codeine, genetic variations can also affect the normal function of these important enzymes. For example, genetic variations that reduce CYP3A5 enzyme activity are associated with poor response to therapies such as the immunosuppressive drug tacrolimus used in transplant patients. The prevalence of some genetic variations identified in the CYP3A5 gene is as high as 90% in Whites, but lower in African and Asian populations. Despite the high frequency of genetic variations in this gene, genetics alone does not completely explain the extreme variability seen in responses to drugs that depend on this genetic pathway, suggesting that other factors (genetic and nongenetic) are involved.

Should You Be Tested the Next Time You Take a Drug?

Chances are you have already experienced a side effect to a drug and/or have made multiple trips to the pharmacy in the hopes of getting a drug that will work for you. So what if, with a simple blood or saliva test analyzing those genes known to impact drug response or drug targets, doctors could

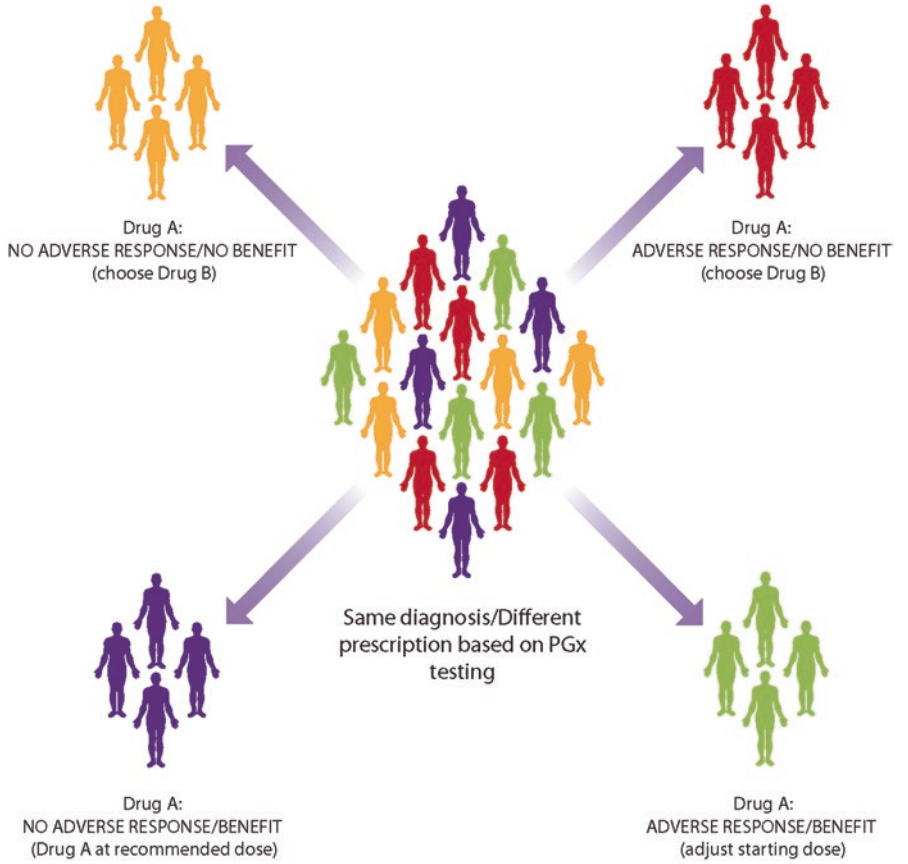

Fig. 7.1 Even with the same diagnosis, we each differ in how we respond to medications, in part due to our genetic makeup. This illustration shows some possible outcomes after pharmacogenetic (PGx) testing of treatment of a group of patients diagnosed with the same condition (source: SB Haga)

substantially improve the chances of prescribing a drug that would be safe *and* effective at treating your condition (Fig. 7.1)? Decisions of which drug will be safest and most effective for *you* and at what dose would then be based on your genetic makeup in addition to other clinical factors like age, weight, race, and other health conditions. The type of testing, pharmacogenetic (PGx) testing is available and being evaluated in many academic medical centers and specialty clinics. If PGx testing was available to you, would this be something you would consider doing? Here are a few things to keep in mind and talk with your doctor about.

First, you have two potential options for when to get testing. Knowing in advance what your chances are of potentially having a side effect or having the drug work could save time and money, and treat the condition as soon as possible. However, this type of testing cannot be quickly performed in the doctor's office yet and thus, it will take some time for the laboratory to complete the testing and return the results back to your doctor. Therefore, if testing is offered at the time a prescription is written, ask about when the test results will be returned. Ordering a test when you need a drug has raised some concerns about the potential delay of treatment while waiting for the testing to be completed. Ask your provider if the recommended dose/drug selection will be safe and effective for you while you await the test results. If available, it may also be possible to wait for a few days until testing is completed before starting treatment or to request a prescription for a lower dose or different drug in the interim until the test results are returned.

Alternatively, some have advocated ordering testing prior to when treatment is needed to avoid any potential delays as you inevitably will need a drug at some point. If you are presented with an opportunity to have testing in advance of needing treatment (such as during an annual physical or checkup), the results would then be stored in your medical record and be consulted when treatment is needed. Since you do not know which drugs you will be prescribed in the future, and therefore which genes should be tested, a pharmacogenetic test ordered in advance (or preemptively) will include analysis of several genes that are known to impact the metabolism of commonly prescribed drugs.

If you have testing, it may be prudent for you to obtain a copy of the test report or a results summary; some labs may provide a wallet-sized card to patients. This will enable you to share the results with other providers or pharmacists, in the event that you receive care from multiple providers and/or switch providers who may not have access to your electronic medical records. In the event of an emergency, having the card with you might be the only way for the emergency physician to access the information. The pharmacist could keep your pharmacogenetic profile on file and check it each time before filling a prescription. No matter who the keeper of this information is, it should be stored in a safe place and only accessible to those who have permission to see it, just like your medical records. Since genetic information is shared with our siblings, children, and parents, you may also consider sharing your results with family members.

The test results will not change over your lifetime, so there is no need to be tested each time a new drug is prescribed unless a new gene is discovered. However, the understanding of the role of genes in drug response will

continue to increase and an updated interpretation of the results may be necessary. There are limited professional guidelines regarding drug or dosage recommendations based on pharmacogenetic data, but as they are released, recommendations based on test results should also be updated. Given the novelty of testing and providers' likely unfamiliarity with the field, it will take some time before the use of pharmacogenetic testing becomes routine.

In summary, understanding the genetic underpinnings of processes related to drug safety and effectiveness can enable more precise treatment decisions, thereby reducing the risk of side effects and improving the likelihood that a prescribed drug will be effective for you. Taking a genetic test before taking a drug may one day become the standard of care in order to determine which drug and at what dose will work best for you. Although genetics will not be able to predict with 100% certainty whether a patient will develop a side effect to a particular drug or whether a drug will be effective, it will soon be considered with the many other factors known to impact drug responses to improve the likelihood of a good health outcome.

Resources

US National Library of Medicine. Medline Plus—Precision Medicine. Available at https://medlineplus.gov/genetics/understanding/precisionmedicine/

US National Library of Medicine. Medline Plus—What is Pharmacogenomics? Available at https://medlineplus.gov/genetics/understanding/genomicresearch/pharmacogenomics/

US National Human Genome Research Institute. Pharmacogenomics. Available at https://www.genome.gov/dna-day/15-ways/pharmacogenomics

US National Institute for General Medical Sciences. Medicines by Design. Available at https://www.nigms.nih.gov/education/Booklets/medicines-by-design/Pages/Home.aspx

US Centers for Disease Control and Prevention. Pharmacogenomics: What does it mean for your health? Available at https://www.cdc.gov/genomics/disease/pharma.htm

8

No Two Cancers the Same

There is probably no one that has not been affected by the disease—either personally or through a friend, neighbor, co-worker, or family member diagnosed with it. Since the War on Cancer was declared in 1971 by President Richard Nixon, we have come a long way in our ability to detect, diagnose, and treat cancer, thereby substantially improving health outcomes and survivorship. In some cases, what was considered a terminal disease (no treatment existed), there are new drugs to significantly improve long-term outcomes in patients with cancer. According to 2019 statistics from the American Cancer Society, the lifetime risk of developing cancer is one in three for women and one in two for men. The risk of dying of cancer is one in five for both men and women, though statistics show that the rate of dying of cancer to be declining over the past decade.

The increasing number of treatments for cancer is due in large part to a better understanding of how cancers develop, grow, and spread to other parts of the body (or metastasize—pronounced "ma-tas-ta-size"). Cancer is inherently a genetic disease; genetic changes in cells accumulate over a lifetime that can be caused by environmental factors (smoking, ultraviolet (UV) light obesity, and certain viruses) as well as inherited genetic variations. Eventually, enough changes in the DNA will cause the cells to reach a tipping point and the normal cells become cancerous or grow uncontrollably. In other words, a series of genetic changes (or "hits") are needed to convert a normal cell into a cancerous cell (Fig. 8.1). Typically, cells have mechanisms to correct DNA damage or can self-destruct if the damage cannot be repaired. However, once the internal mechanisms are damaged, cells can grow unchecked, and more DNA damage accumulates.

S. B. Haga, *The Book of Genes and Genomes*, https://doi.org/10.1007/978-0-387-70916-1_8

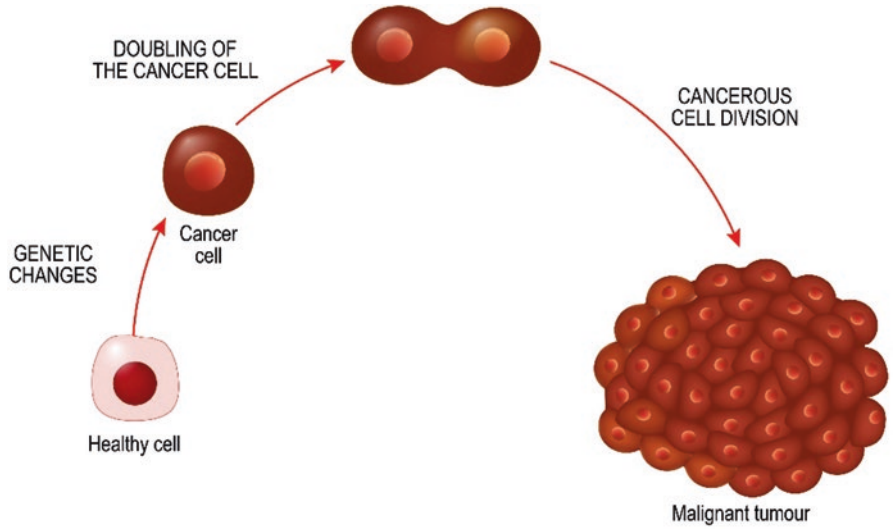

Fig. 8.1 A normal cell can accumulate multiple genetic changes that will cause the cell to grow uncontrollably, leading to a mass (source: Adobe Stock Photo)

In some cases, people are born with a few genetic changes that increase the risk of cancer, or for some rare cancers, a single genetic change almost inevitably will lead to the development of a cancer syndrome. Hereditary cancer syndromes such as Lynch syndrome or hereditary breast and ovarian cancer are adult cancers, but some inherited cancers such as retinoblastoma (cancer of the eye) can affect children.

Given the extent of damage to DNA accumulated over one's lifetime, can it be corrected? For many cancers, interventions include surgery and/or drug treatment (chemotherapy). However, both of these interventions are not always as effective as we would hope for since it may not entirely remove or destroy all of the cancer tissue. For many patients, the cancer evolves and becomes resistant (and nonresponsive) to treatment. With the growing volume of data produced in recent years due to advances in genetic technologies and research, we now know a tremendous amount more about the inner workings of a given cancer type (though still have a tremendous amount to learn). Some of those revelations have led to the development of targeted drugs against cancer cells harboring these distinct genetic aberrations and new clinical tests.

Tools to Classify Cancer

From a genetic standpoint, cancer is a complex disease. It is not caused by a change to a single gene and there is no single path for an individual to

develop cancer. It is like when you look up driving directions on your phone and see 2–3 recommended routes, but imagine if there are 1000 routes to choose from instead. With the exception of some rare inherited forms of cancer, we now know that there are multiple genes that contribute to the development of cancer and thus there are multiple routes a cell can take to become cancerous. While some types of cancers show characteristic genetic changes, other cancers may show a unique combination of genetic changes.

Solid tumors start (originate) in one tissue in the body (e.g., breast), but as the cells evolve and acquire more genetic changes, the cancerous cells can spread (or metastasize) to other tissues. Cancers of the blood (leukemias and lymphomas) also begin within a subset of cells located in the inner core of bones (known as "bone marrow") or the lymph nodes (parts of our immune system found throughout our body). Likewise, leukemias can also metastasize to the brain, spinal cord, and other tissues as the cells accumulate more genetic damage.

The field of genomics has begun to identify which genes are damaged in tumor cells through analysis of the entire DNA of hundreds to thousands of cancer samples. The testing technology has rapidly evolved, jumping from the analysis of one gene at a time to all of the genes from a tumor sample. These genetic "snapshots" of individual cancers allow scientists to identify and better characterize a tumor and to treat the tumor based on its unique characteristics.

Tumor Sequencing

To distinguish which changes are associated with the tumor DNA versus non-tumor tissue (taken from an unaffected part of the tissue or blood), both types of tissue are sequenced. Until now, this type of analysis had not been possible. These data have revealed some common similarities across all cancers as well as unique characteristics between cancer types (e.g., lung cancer vs. colon cancer). To further complicate matters, an affected tissue will actually be comprised of multiple "communities" of cells (sort of like neighborhoods), some cancerous and some noncancerous, and the boundaries are not always distinct (can have normal cells mixed in with cancer cells). Each cancer community has developed a different set of genetic changes and therefore may exhibit different behaviors and response to treatment. Obviously, this can complicate genetic analysis of a given tumor. Some hospitals are beginning to offer tumor sequencing, but testing is not routinely available.

Microarrays

Instead of decoding every single letter of DNA from a tumor cell, another type of technology, known as a microarray, enables rapid analysis of hundreds of thousands of known genetic variants. A small piece of DNA, either normal or containing a different genetic variant, is affixed, or spotted in a specific order, on a slide similar to a microscope slide. When finished, the slide (referred to as a microarray) will have thousands of tiny pinprick-sized spots of DNA. The DNA from a patient's tumor can be applied to the slide and it will bind only to the DNA spot if the sequences match. Thus, the presence or absence of any of the genetic variants on the microarray can be clearly distinguished through a color-based assay (Fig. 8.2). These data are then compared to the pattern obtained from a normal tissue sample. Though not as comprehensive as sequencing, one can rapidly (and more cheaply) obtain a pattern of known genetic changes presented in a given tumor sample.

Microarrays can also be used to determine the pattern of genes that are turned "on" or "off" compared to a normal tissue sample. Distinct gene

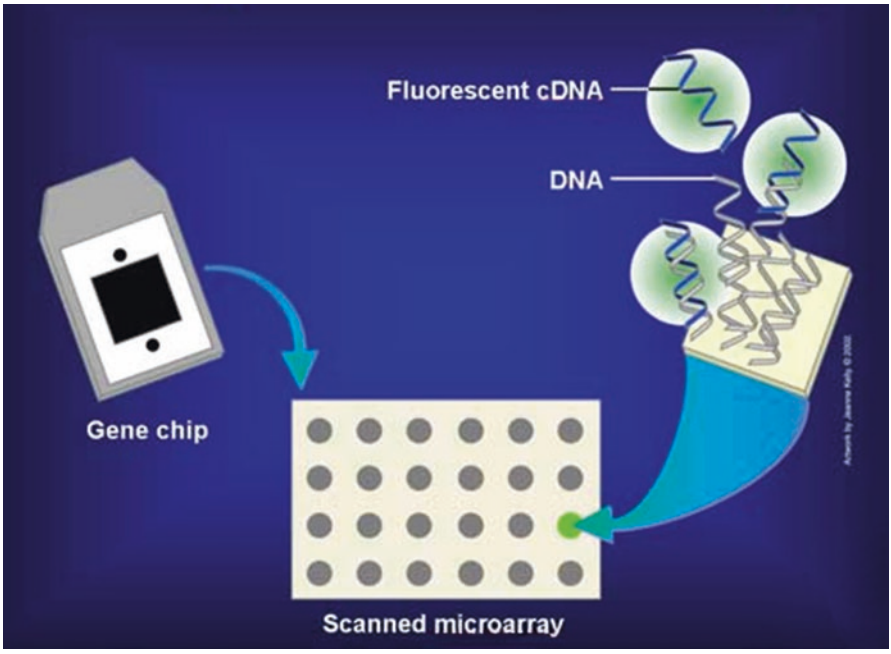

Fig. 8.2 A microarray chip contains complementary DNA to many sequences of interest. The cDNA fluoresces when it binds (or hybridizes) with a matching DNA fragment in the tumor sample (source: US National Cancer Institute; http://www.cancer.gov/cancertopics/understandingcancer/moleculardiagnostics/AllPage)

"expression" patterns have been identified and found to be a good predictor of disease types (such as distinguishing between two different types of leukemias), to distinguish sub-types of the same disease (such as multiple types of breast cancer), or to predict prognosis (how aggressive will this cancer be, or its likelihood to spread). The subclassification of cancer provides a more refined diagnosis, which can help doctors understand the likely long-term outcomes and to select the most appropriate course of treatment.

One of the first studies to use microarrays to analyze gene expression patterns in cancer was published in 1996. Researchers from Stanford University used microarrays to assess the changes in gene expression in melanoma (skin cancer) cells that were genetically altered to become less cancerous. After analyzing more than 1100 genes, they found differences in the expression levels of many genes involved in tumor growth, which would explain the differences in growth rates and other tumor properties. That same year, a group from the Whitehead Institute at the Massachusetts Institute of Technology was the first to demonstrate that microarrays could be used to distinguish between different types of cancer. In their research, they tested two types of leukemia but were blinded as to which sample was which leukemia type. Using only the gene expression data they collected from the microarray experiments, they were able to distinguish the two leukemias and correctly identify samples as either acute myeloid leukemia (AML) or acute lymphoblastic leukemia (ALL).

In 1999, scientists from Princeton University discovered that normal and colon cancer tissue could be distinguished solely by looking at the differences in gene expression using microarrays. And in 2003, researchers from the University of Michigan developed a 158-gene profile associated with pancreatic cancer using microarrays. These genes were distinctly linked to pancreatic cancer compared to normal pancreatic tissues.

Several research groups have used microarrays to define sub-types of a given type of cancer. For example, researchers from several universities and the US National Institutes of Health identified distinct sub-types of disease in patients diagnosed with the most common type of non-Hodgkin's lymphoma known as diffuse large B-cell lymphoma. This disease is quite variable in that about 40% of patients survive for an extended period of time with current therapy, but the remainder do not respond to therapy and die of the disease in a short time. Using microarrays to measure the gene expression levels in 96 normal and B-cell lymphoma samples, the researchers identified two distinct patterns of gene expression levels. The researchers found that the different patterns were due to the type of B-cell the leukemia originated in (recall, a cancer cell was once a normal cell which has gone out of control). Furthermore, the gene expression patterns correlated with the outcomes of the patients. One pattern

was found in patients who had a better survival than those with the second gene expression pattern.

Throughout the past few decades, microarray technology has continued to be used to further characterize cancers and learn about how cancers change. While most of these tests are not available for patient care, they have generated data that has informed drug development and overall understanding of cancer growth.

Inherited Cancers

All cancers are genetic in that they are caused by the accumulation of genetic changes throughout our lifetime leading to uncontrolled cell growth. In some cases, though, genetic changes associated with cancer risk can be passed on from parents to children, meaning that they are inherited. Inherited cancers account for a small fraction of overall cancers, but in some cases are easier to study as they occur in families and display unique characteristics (e.g., early onset).

The first clue that a cancer is inherited is family health history. This information is typically collected at every office visit to identify disease risks that may warrant more screening than is recommended for the general population. If multiple members of one side of your family have been diagnosed with a type of cancer, particularly at younger ages than would be expected, clinical testing for genes associated with those inherited cancers may be ordered. If the patient tests positive for a change in one of the genes tested, it is recommended that close family members should also be tested to determine their risk and inform screening plans.

Probably one of the most well-known inherited forms of cancer is breast cancer. In the late 1990s, variations in two genes, abbreviated as BRCA1 and BRCA2, were found to be associated with an unusually high risk of breast and/or ovarian cancer in some families. Medical tests were quickly developed and clinical testing is currently recommended for women with a strong family history of breast and/or ovarian cancer diagnosed at younger ages. Although this inherited form of breast and ovarian cancer accounts for a small proportion of all cases of breast cancer, it has received a lot of public attention due to the positive diagnosis of celebrities like Angelina Jolie and Christina Applegate, who elected to have surgery to remove breast tissue and/or ovaries to reduce their lifetime risk substantially.

Another inherited form of cancer is a type of colon cancer known as hereditary nonpolyposis colorectal cancer (HNPCC), or Lynch syndrome. Estimated to affect as many as 1 in every 300 people, patients diagnosed with Lynch

syndrome are also at risk for developing other cancers including stomach, breast, and prostate cancers. Multiple genes are linked to Lynch syndrome and testing is available if it is suspected that a patient has Lynch syndrome based on family health history and personal factors such as age of diagnosis.

Cancer Diagnoses, Cancer Sub-types, and Treatment

As you might imagine (or have experienced), patients with cancer undergo a lot of different types of testing. Specifically, patients may undergo imaging, blood tests, and tumor biopsies both before diagnosis and regularly after treatment to monitor response to treatment and recurrence. The complexity of cancer can pose some problems in accurately determining the stage, type, prognosis, and best course of treatment. While sequencing the DNA of tumor cells can identify many genes that have been damaged over an individual's lifetime, the key to predicting the behavior of a tumor or developing targeted drugs is determining which of these genes is essential to the survival of the tumor cell. Drug companies can then focus their efforts on developing treatments targeting these important genes to cripple the growth of cancer cells and minimizing damage to healthy cells. Many of the other genetic variants identified in cancer cells are likely just subsequent (or collateral) damage, and not driving the tumor growth or its behavior.

Because each patient's cancer is genetically distinct, it should not be surprising to learn that patients will respond differently to recommended treatment regimens. To help predict the most effective course of treatment or combination of treatments (surgery, radiation, and/or chemotherapy), genetic and genomic tests are increasingly used to characterize the tumor type. For example, for breast cancer, treatment decisions may be impacted by knowing how likely a patient's cancer is to recur (recurrence risk). In 2007, the first laboratory received approval from the Food and Drug Administration for a gene expression microarray test that predicts the risk of recurrence in breast cancer patients. Known as the MammaPrint test, it analyzes the pattern of 70 genes from a tissue taken from a breast tumor and can predict with 97% accuracy the 10-year survival of a patient. Other companies now offer similar tests to provide a genetic risk assessment of tumor samples that can be used to help providers develop the best course of therapy based on the risk of recurrence. In addition, several laboratories offer a comprehensive analysis of many genes associated with cancer to provide more insight about the genetic characteristics of the tumor that may inform treatment.

Targeted and Tailored Treatments

There have been a number of new targeted drugs developed to treat cancer. A targeted drug is specifically developed to interact with a protein that is important to tumor growth or other related functions. By blocking or inhibiting the action of that protein, which oftentimes is genetically altered due to a change in the sequence, the growth of the tumor can be slowed or halted. Many targeted drugs have been developed for cancer based on observations of unique genetic changes associated with a particular cancer and further characterization of that genetic change, the protein involved, and its role in cancer development and growth. In the late 1990s, two targeted drugs heralded the anticipated personalized medicine revolution: Gleevec for CML and Herceptin for breast cancer.

As described in Chap. 4, patients with CML often have an unusual characteristic chromosome, a fusion between chromosomes 9 and 22. The discovery of the fusion chromosome in 1960 was through careful observation of samples collected from CML patients. But it was not until 1973 that the source of the fusion chromosome was determined to be part of chromosomes 9 and 22.

In the early 1980s, researchers reported that the fusion chromosome led to the creation of a fusion gene at the location where the two chromosomes linked. The placement of the front part of the BCR gene from chromosome 22 to the back half of the ABL gene on chromosome 9 created a new protein, dubbed BCR-ABL (Fig. 8.3). The ABL part of the gene functions as a tyrosine (TY-row-seen) kinase (KY-nays) that disrupts cell growth. Animal studies confirmed that this fusion protein could induce cancer.

Armed with the knowledge of this fusion protein that behaves to spur cancer growth, drug developers had a new potential target. Scientists at the University of Oregon confirmed that a potential new drug was particularly effective in CML patients with the Philadelphia chromosome, and in effect crippled the leukemia cells by binding to the ABL part of the fusion protein. The drug, imatinib (brand name: Gleevec™), was approved by the FDA in 2001 and has continued to be a first-line treatment for CML patients. Following treatment, when the number of abnormal white blood cells is decreased, the red blood cells, platelets, and other cells in the bone marrow return to normal levels. Patients with CML typically enter a remission phase if the number of abnormal white blood cells is significantly reduced or undetectable. Gleevec is very effective in 60–90% of patients with early-stage CML resulting in almost complete elimination of abnormal white blood cells. A blood test will indicate if the drug is working by measuring the levels of white

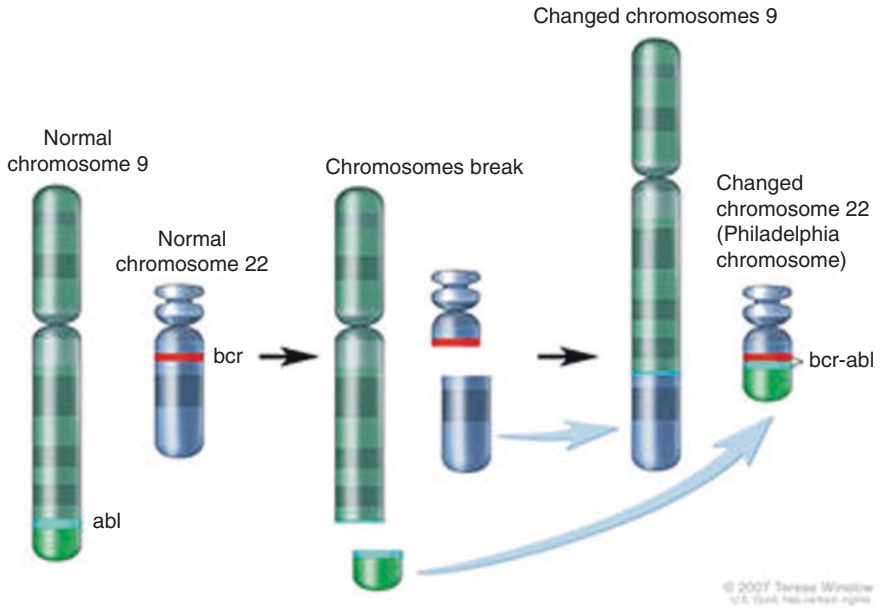

Fig. 8.3 CML is caused by the swapping of pieces between two chromosomes, creating a new gene that combines half of a gene on chromosome 9 (named ABL) with half of another gene on chromosome 22 (named BCR). This new gene, BCR-ABL, is known as the Philadelphia chromosome after its place of discovery and is present in the majority of patients with CML (source: US National Cancer Institute; https://www.cancer.gov/research/progress/discovery/gleevec)

blood cells, red blood cells, and platelets. Periodic bone marrow biopsies (where a long needle is used to obtain a sample of cells from the center of a bone for testing) will indicate how many white blood cells with the chromosomal abnormality are still present in order to track the disease. As leukemia cells evolve and acquire new genetic changes, the drug may become less effective. In this situation, the patient has developed "acquired resistance." Thus, several other targeted drugs have since been developed that patients can be switched to.

Since targeted drugs are specific to a fusion protein or other genetic change and not necessarily the disease, scientists have found that patients with other cancer types such as gastrointestinal stromal tumors that have the same fusion protein respond well to Gleevec. Likewise, other cancer drugs such as axitinib, a drug developed for a type of kidney cancer, have been shown effective after Gleevec is no longer effective due to changes in the fusion gene.

Another major success story is the development of a targeted treatment for breast cancer, and women who have a particular genetic change in their tumor

cells. This change involves a gene that produces a receptor—a type of molecule that is located on the outer edge of the cells (one could think of it as a "receiver" that controls the entry/exit of certain molecules in the cell). In 1986, a gene called human epidermal growth factor receptor 2 (HER2) was first identified as a part of a family of genes involved in cell growth. It turned out that about 15–30% of breast cancer patients have extra copies of the HER2 receptor gene. In 1998, it was reported that the presence of extra HER2 was associated with a form of breast cancer with poorer outcomes.

A drug company developed a new drug that would adhere to the receptor and stop it from working. As with Gleevec™, the drug (trastuzumab) was quickly approved by the FDA (brand name: Herceptin™). All women with breast cancer are tested for this genetic change to determine if they are "HER2 positive" and therefore would respond to treatment with Herceptin (the drug does not work as well in women with the normal number of the HER2 gene). Like Gleevec, the drug has been shown to work in other cancers that have multiple copies of HER2 gene. New HER2-targeted drugs have since been developed as well.

More recent examples of targeted drugs are Tarceva™ (erlotinib), developed for patients with non-small cell lung cancer and genetic changes in the epidermal growth factor receptor (EGFR) gene, and Trikafta™ (elexacaftor/ivacaftor/tezacaftor), a combination drug therapy for cystic fibrosis patients with a specific genetic change in the cystic fibrosis transmembrane receptor. As noted above, in a growing number of cases, drugs developed for one type of cancer with a particular genetic change are demonstrated to be effective in other cancers with that same genetic change. Thus, cancer-specific therapies may be shifting to genetic-specific therapies, placing increasing importance on the genetic characterization of the cancer for treatment decisions.

In parallel with the expanding suite of new technologies with which to diagnose and treat cancer has come a need to define standards to determine which new drugs are effective (how does one determine that the drug works). While the final decision to approve a drug for a certain use rests with the US Food and Drug Administration, researchers, funding agencies, and professional cancer organizations benefit from guidelines and criteria to inform how they should study new drugs and what they should measure and report. By having a standard set of criteria, the comparison of the effectiveness of new drugs is feasible as well as comparison to existing treatments. In the 1990s and 2000s, cancer research groups and medical organizations worked toward developing a set of criteria for assessment of treatment response. However, it became obvious that differences between cancer types posed a challenge to the implementation of a single set of criteria to evaluate the effectiveness of new

therapies. Thus, specialty groups have developed cancer-specific criteria for drug assessment to address unique tumor characteristics and assessment tools. Regular revisions to the guidelines are necessary to reflect changes in scientific understanding, new technologies, and analytical methods.

From Biopsy to Blood-Based Cancer Screening Tests (Risk Assessment)

Everything discussed thus far really pertains to patients already affected with cancer. But what about patients who are not currently affected with cancer—is there a way to estimate what their risk is? While there are some screening processes available for cancer including breast cancer (mammograms), cervical cancer (Pap smears), and colorectal cancer (colonoscopy), many types of cancer do not have any screening tests available, such as ovarian and pancreatic cancer. Ovarian cancer is estimated to affect 1 in 80 women in the USA. Ovarian cancer ranks fifth in cancer deaths causing 14,000 deaths annually—in part because it is often diagnosed at a late stage since the symptoms can be quite mild and go unnoticed for some time. As a result, the 5-year survival rate for ovarian cancer is 48% with surgery and chemotherapy, though it is much higher (92%) if the cancer is localized (based on 2020 estimates from the American Cancer Society). Similarly, there is no screening test for pancreatic cancer and thus the disease is typically discovered in advanced stages and has a low 5-year survival rate of 9%.

Risk assessment of cancer often begins with a patient's family health history. As described in Chap. 2, a strong family history of cancer will typically include multiple biological relatives affected with cancer, often at a younger than typical age of diagnosis. In individuals with a strong family health history of an inherited form of cancer such as breast or ovarian cancer, the doctor will recommend a genetic test to analyze the genes associated with that type of inherited cancer. Blood-based tests are available for colon cancer and breast/ovarian cancer that analyze many genes within a single test for genetic changes associated with that cancer. The test report will look like an alphabet soup of gene abbreviations (TP53, VHL, BRCA1, BRCA2, and others). If a patient receives a negative test result, this does not necessarily mean they will *not* develop the specific cancer running in their family. It could mean that the test is not analyzing the responsible gene for the cancer in your family simply because scientists have not identified all of the genes that could potentially

cause cancer. More frequent monitoring (e.g., mammograms) will also be recommended in order to detect the development of a cancer early.

In addition, particularly for individuals without a family health history of cancer, there is a lot of interest to develop new screening tests for cancers. There are guidelines regarding when people should be screened and how often by mammography, colonoscopy, fecal occult testing, and other methods to detect cancer at early stages. But very few new screening tests for cancer have been developed in recent years. However, with new technologies to analyze very small amounts of cells and DNA, medicine may be on the cusp of a new blood-based test for cancer. Since the 1800s, it has been known that cancer cells, known as circulating tumor cells (CTC), were present in the blood. Presumably, these cells came loose from a tumor and entered the bloodstream. More recently, cell-free tumor DNA (free floating DNA from the tumor) has been detected in blood. While there is only a very small amount of tumor cells or DNA present in the blood, it has been shown to be adequate for testing. Although the technology is still under development as a screening test, this can possibly revolutionize cancer screening, replacing invasive tests (a biopsy of a mass detected on an imaging test) with a simple blood test that can be routinely performed *before* cancer is detected by imaging or the patient is symptomatic. In some cases, a biopsy of the tumor may not even be possible, such as for lung or brain tumors, or in patients where a surgical procedure may be too risky to undergo. Like other genetic tests, these new noninvasive tests (dubbed "liquid biopsies") can provide insight about the characteristics of the tumor without any surgical procedures. One potential limitation at this time is that the detection of tumor cells or DNA may not indicate which tissue the cancer is present in, or where it originated (if the cancer has already spread). This revolution in noninvasive testing is also occurring in the field of obstetrics and prenatal testing as described in Chap. 4.

In 2016, the FDA issued the first approval for a liquid biopsy cancer test to inform the choice of cancer treatments. The test analyzes a blood sample from patients with the most common type of cancer called non-small cell lung cancer (NSCLC), to determine if a genetic variation is present in the tumor that would inform the choice of treatment. Prior to FDA test approval, the test could only be performed on a tumor sample (biopsy). Specifically, the test analyzes the EGFR gene, often altered in non-small cell lung cancer patients.

Conclusion

New technologies have substantially advanced the understanding of the genetic complexities of cancer and helped improve the selection of appropriate and effective treatments, and ultimately disease outcomes. However, despite the wealth of cancer research that has been conducted using microarrays and sequencing, their transition to clinical practice has been slow for a number of reasons. With the generation of huge amounts of data with genomic technologies, it is hoped that new patterns and pathways will be identified that can improve testing, treatment decisions, and drug development.

References

American Cancer Society. Cancer Basics. Available at https://www.cancer.org/cancer/cancer-basics/lifetime-probability-of-developing-or-dying-from-cancer.html

US National Cancer Institute. Available at https://www.cancer.gov/about-cancer

Cancer Net. Available at https://www.cancer.net/

Annual Report to the Nation on the Status of Cancer, part I: National cancer statistics. Available at https://acsjournals.onlinelibrary.wiley.com/doi/10.1002/cncr.32802 (March 12, 2020).

US National Cancer Institute. How Imatinib Transformed Leukemia Treatment and Cancer Research. Available at https://www.cancer.gov/research/progress/discovery/gleevec

9

Correcting Genes

The inheritance of some genetic disorders may have meant extreme morbidity and/or an early death if no intervention was available to counter the effects of the gene (or genes) that had been altered and gave rise to the disease. One cannot "change" their genes in order to correct them to the normal sequence to avert or halt the damage to be caused by the altered gene, or so it was thought. Effort has been ongoing for decades to attempt to do just this—change genes—or alternatively to add a "normal" copy of the gene into a patient's cells with the hope that the normal gene would override the effects of the abnormal protein produced from the altered gene.

Back in the early 1990s, you might have heard news reports of the "bubble boy" who had been treated with something called "gene therapy" and showed remarkable improvement. This case ushered in the excitement and enthusiasm of treating patients with devastating and untreatable diseases, harnessing the new knowledge of genetics. Unfortunately, since that remarkable initial feat, the field of gene therapy has experienced its share of setbacks, including several deaths, serious side effects, and unethical practices of researchers conducting clinical trials. Today, the field of gene therapy has begun to show progress with approved clinical treatments, but much research continues. But there is a lot of hope and excitement for the ability to use newer gene editing technologies to treat inherited diseases and cancer. In this chapter, a brief history and several examples are provided to give you a sense of the challenges, excitement, and ethical issues raised by gene therapy and gene editing.

S. B. Haga, *The Book of Genes and Genomes*, https://doi.org/10.1007/978-0-387-70916-1_9

How Does Gene Therapy Work?

Recall that if someone is affected with an inherited genetic disease, all of the cells in their body will carry the altered version of the gene that is causative of the disease. This is different from diseases such as cancer, which are, in most cases, not inherited from our parents, but rather involve changes in cells in a given tissue (e.g., breast or prostate) that lead to unregulated cell growth.

Genetic diseases that are caused by a change in a single gene were initial targets for gene therapy. In theory, the cells that are most vulnerable to the effects of the altered gene, for example, muscle cells in muscular dystrophy, would be the ultimate target of gene therapy.

But how exactly do scientists insert a normal copy of a gene in a cell carrying an altered copy? One of the major challenges is getting through the protective outer layer of cells. A number of different methods have been used to accomplish this feat, with varying degrees of success. The most common approach is to use a virus as a transporter of the normal copy of the gene. Since viruses are able to easily gain access to a cell (through the same way they infect us) and release its DNA (or RNA), it can be used as a vehicle to transfer copies of normal genes into the patient's cells.

In gene therapy, the virus is referred to as a "vector" or carrier. Although many viruses can cause illness upon infection, scientists only use viruses where the harmful genes have been removed or the virus has been inactivated, rendering the virus as "safe." By removing the harmful genes, even the virus that causes AIDS (human immunodeficiency virus or HIV) has been able to be safely used in gene therapy experiments. There are many different viruses that researchers have studied, with different tissue specificity (e.g., some viruses preferentially infect lung cells or muscle cells) (See Fig. 9.1). There are multiple ways in which a weakened virus carrying the corrected copy of the gene could be administered to the patient. In one approach, a viral solution could be injected into the patient where it would spread throughout the body, perhaps with a preference for certain cells, just as a viral infection typically would. Alternatively, a sample of the patient's cells is removed (e.g., a bone marrow biopsy) and the viral solution is mixed with the cells in a laboratory. After a few days, the infected cells, most of which are hopefully carrying the corrected version of the gene, would be returned to the patient.

Other nonviral approaches involve attaching the normal version of a gene to another molecule, such as a fat molecule, that can be taken up by the cell. Or another method is to directly introduce several copies of the corrected gene to the intended patient cells. Regardless of the method, there are many

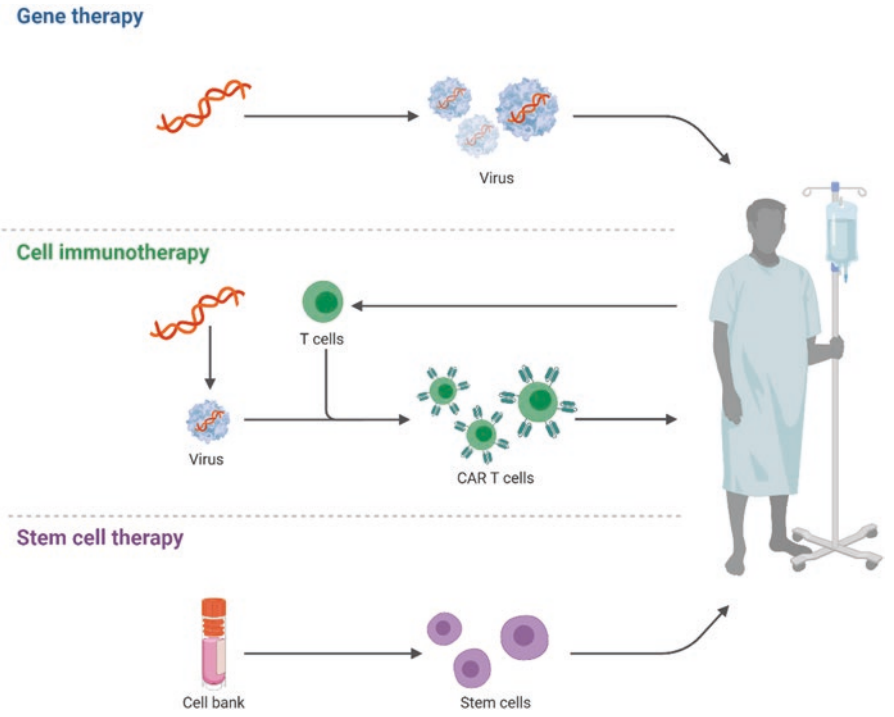

Fig. 9.1 Multiple types of gene therapy. The initial experiments with gene therapy were for rare inherited diseases to correct a mutated gene. More recent work has developed a form of gene therapy to treat cancer called cell immunotherapy. Stem cell therapy can be used to restore healthy populations of cells, similar to what occurs with a bone marrow transplantation (source: Biorender)

hurdles that need to be overcome before gene therapy can be considered to be therapeutic. Even if a normal copy of a gene could be inserted and correct the cell's deficiency, cells have various life spans and some will turn over quicker than others. Therefore, gene therapy may have to be repeatedly administered unless the gene has stably integrated into the cell's genome and is replicated along with the rest of the DNA.

The First Successful Trial of Gene Therapy

Many preclinical animal studies have been conducted to understand the risks of transferring a new gene into a patient's cells (animal studies are required before human trials can be conducted). There are actually two components of the safety concern—the effect of the virus and the gene: whether the process

to insert the normal copy of the gene would be safe and whether a new gene introduced into a cell in the body could function at the level that would provide some relief from the existing, mutated copy.

In 1990, scientists at the National Institutes of Health in Bethesda, Maryland, were approved to conduct the first human gene therapy trials. They first treated a young girl named Ashanti DeSilva who was affected with a disease called severe combined immunodeficiency syndrome (SCID—or "skid"—for short). Children with this inherited disorder have a severely compromised immune syndrome and are extremely susceptible to infection. These children often lead very restricted lives with limited contact outside the home in order to minimize infection. After a research experiment where a young boy had lived for 12 years in a plastic, germ-free "bubble" in a Texas hospital became public, SCID became known as the "bubble boy" disease.

Doctors removed white blood cells from 4-year-old Ashanti's bone marrow and added the normal version of the gene and allowed the cells to grow in the laboratory for a short time before inserting them back into her body. Testing after the cells had been returned showed that Ashanti's immune system was much stronger than before the gene therapy. Over the next 2 years, she underwent about a dozen more gene therapy experiments in which her white blood cells were removed and then transferred back to her body with a normal copy of the gene. As of 2007, Ashanti was in good health and attending college. A second patient, Cynthia Cutshall, received the same gene therapy as Ashanti, and is also doing well today. It is believed that the gene therapy only partially corrected the girls' genetic deficiency, but the fact that the gene therapy was not harmful, particularly at that early time of the field, was a success in itself.

Side Effects of Gene Therapy

Despite almost several decades of research, a lot of uncertainty surrounds gene therapy, which has slowed its progress. One of the major problems with gene therapy is that scientists do not have any control over where the normal gene ends up once it is taken up by the cell. The new gene could integrate or insert itself anywhere in the genome (the DNA in the cell), which is of particular concern if it chooses to insert itself in the middle of another gene, thereby disrupting the function of that gene. If that happens, while the patient's original disease may improve with the presence of a normal working copy of the altered gene, a new condition could develop due to the newly disrupted gene—something technically known as "insertional mutagenesis." Another problem is the effectiveness of the vector or carrier—does it reach the targeted

cell population (without triggering an immune response) and is it able to transport the gene into the cell at a sufficiently high rate that will be effective?

In 1999, a study participant died in an early-phase study conducted at the University of Pennsylvania. Jesse Gelsinger, an 18-year-old male with a rare disease known as OTC (short for *o*rnithine *t*rans*c*arbamylase deficiency), volunteered to participate in a gene therapy study for OTC. Individuals with OTC lack a key enzyme, or protein responsible for a chemical reaction that occurs in the liver. With a very strict diet and daily medication, Jesse could control his disease to the point where he was able to participate in the normal activities of a teenager, although he was hospitalized several times during his childhood.

The goal of the initial trial was to test the safety of a viral vector. The viral carrier was to be directly injected into the artery connected to the liver where the enzyme was missing. Four days after he received the injection, Jesse died due to a massive immune response and systemic organ failure. Jesse's death triggered a series of investigations by university and federal committees. Among the other research violations identified, the investigators failed to inform participants of the deaths of several monkeys in preclinical studies investigating the same viral vector as well as the investigators' financial interest in the success of the trial.

In October 2002, authorities in the USA halted three gene therapy trials when French researchers announced that a young child who had participated in a gene therapy trial had developed leukemia. The 3-year-old boy was being treated for SCID in Paris and, at the time of the diagnosis of his leukemia, researchers were uncertain if it was caused by the gene therapy experiments. Later analysis of the boy's cells found evidence that a part of the viral vector had inserted itself in the boy's chromosomes, specifically into a gene on chromosome 11 linked to cell growth. Until this unfortunate development, 9 of the 11 young boys who underwent gene therapy in this study were able to leave the hospital and lead near-normal lives, effectively cured of the disease.

By January 2003, a second boy involved in the gene therapy study in Paris was diagnosed with leukemia. In response, the US authorities halted 27 more gene therapy trials using a similar type of viral vector, although there had been no reports of severe side effects. Despite the promising results, the puzzling outcomes of the French gene therapy experiments again indicate the need for further research to prevent the occurrence of such serious and life-threatening side effects such as leukemia.

Despite the numerous setbacks and myriad scientific challenges, a handful of gene therapies are available as a treatment today. China was the first country to approve a gene therapy in 2003 for the treatment of head and neck

squamous cell carcinoma (a type of skin cancer). In 2012, the EU approved a gene therapy called Glybera for the treatment of a rare disease called lipoprotein lipase deficiency (also known as familial chylomicronemia syndrome). The manufacturer of Glybera declined to file an application with the FDA and did not renew its market authorization. In the USA, the FDA has approved a handful of gene therapies. In 2017, the FDA approved Luxturna, a virus-delivered gene therapy for a rare, genetic form of blindness called retinal dystrophy. In 2019, another gene therapy, Zolgensma, was approved for the rare disease spinal muscular atrophy. As of 2020, many clinical studies are underway to study the safety and efficacy of gene therapy for a range of genetic diseases and more approved gene therapies are anticipated in the coming years.

Gene Editing

In contrast to gene therapy, gene editing presents a more precise approach to correcting genes with the hopes of restoring natural function and curing disease. Dating back to the mid-1980s, a series of discoveries of special enzymes in bacteria provided the basis for this new field called gene editing. Bacteria have defense systems (like humans' immune system). But unlike humans, bacteria use enzymes to literally cut up invaders and in effect destroy them. The cutting defense complexes do not behave randomly—they bind to specific DNA sequences. As such, these enzymes are actually a large complex—part homing beacon (binds to specific DNA sequence) and part scissors. Thus, when scientists discovered that these bacterial enzymes actually cut at specific DNA sequences, they wondered if they could be engineered to desired DNA targets, like that of a specific gene. Thus was born the era of gene (or genome) editing.

Gene editing actually has multiple tools (or DNA binding/enzyme complexes) at its disposal. Each of these different complexes has its own advantages and disadvantages to engineer an accurate DNA binding/cutting tool that can eventually be reliably used for a commercial or clinical purpose. One of the enzyme complexes that have been well studied by scientists around the world to perform gene editing is called CRISPR-Cas9.

For genetic diseases, different types of "editing" may be required to correct or modify the gene. It is possible to make a single cut (with one complex) in order to insert a piece of DNA or to make two cuts with multiple complexes to remove or drop out a piece of DNA (a cut on either end). The cell has other molecules that assist in reconnecting the cut DNA and sealing up gaps. Scientists' efforts to engineer the enzymes to target specific sequences have

been successful, though often at a low rate of efficiency and not completely accurate. In some cases, the enzyme complex can bind to the incorrect sequences (called off-target) and potentially cause a problem (like cutting a normal gene). In addition to reducing the risk for off-target binding, another challenge is how to get the large gene editing complex into a cell where it can make the gene edits.

One of the first diseases for which CRISPR-Cas9 has been studied is Duchenne muscular dystrophy, a neuromuscular disorder that affects young boys, resulting in gradual loss of movement and early death. In 2014, several research teams began publishing their work showing effective use of CRISPR-Cas9 in mouse models (the intervention must be shown safe and effective in animal models before testing in humans can begin). This disease is actually caused by several different changes to the gene, including some rather large deletions.

Another well-known disease that is a good candidate for gene editing is cystic fibrosis. It is caused by a genetic change in both copies of a gene known as the cystic fibrosis transmembrane receptor (CFTR). The CFTR protein works mostly in the lungs and thus the predominant symptoms are coughing, shortness of breath, lung failure, and increased risk of lung infections. Researchers have been testing whether gene editing can correct the disease-causing genetic variant (it is a much smaller genetic change than occurs in the Duchenne muscular gene) and restore function to cells in the airway where the major disease symptoms arise. Early work has demonstrated that the CFTR can be corrected through gene editing in parent airway cells.

As an aside, several new drugs have been approved in recent years for CF patients. These drugs are known as CFTR modulator therapies that can correct the malformed shape of the CFTR protein and restore partial function, thereby alleviating symptoms. Different types of genetic changes in the CFTR gene can cause cystic fibrosis through slightly different effects on the CFTR protein shape. Thus, there are multiple modulator drugs now available—based on the patient's specific CFTR genetic change, the appropriate drug is selected (though not all CFTR genetic changes have a drug available).

Another disease that has shown promise for the use of gene editing as a cure, though in a slightly different way, is sickle cell anemia (and the related disease of beta thalassemia). Sickle cell anemia occurs in individuals who have a genetic change in both copies of a gene for a molecule called hemoglobin (heem-o-glow-bin) (one inherited from each parent). This genetic change causes the protein to be misshapen as long rods that in turn cause the red blood cell to be misshapen—a sickle-shaped cell instead of round. Often, the sickle-shaped red blood cells will get stuck in the narrow blood vessels,

sometimes when they clump or stick together. The cell clumping and dysfunction of the hemoglobin (not correctly binding oxygen to transport it to other tissues in the body) cause the primary symptoms of extreme pain in the oxygen-starved tissues for patients.

Hemoglobin comes in multiple forms, with each form unique to the age of the individual (embryonic, fetal/infant, and adult)—the adult form is called beta globin. This is a rather unique situation in which two other versions of a gene exist that carry out essentially the same function. Fetal hemoglobin tends to attach more tightly to oxygen than the adult form of hemoglobin. Rather than using gene editing technology to correct the beta globin gene, researchers are testing whether they can turn "on" the infant and fetal version of the hemoglobin gene through gene editing. By having functional versions of the hemoglobin protein produced, although not the adult form, the red blood cells should function properly and essentially cure the disease. To achieve this goal, researchers inactivated a gene that produces a molecule that turns "off" the fetal and infant forms of the hemoglobin gene. One could think of this strategy as removing the safety lock. With the off button no longer functioning for fetal hemoglobin, it will turn on.

In 2020, researchers reported on the first two cases in which they had removed some blood cells from patients, edited the gene to turn off the "off" button, and returned the edited cells to the patient. Analysis of these patients' blood detects the fetal version of hemoglobin. After a year, the first two patients were reported to be doing well and had experienced no symptoms of the disease.

The first clinical trials in humans using CRISPR were launched in 2019 for small groups of patients affected with leukemia, blood disorders, and a form of inherited blindness. Most of the trials involve removing patient cells, editing the cells, and transferring the edited cells back to the patient (as opposed to injecting patients with the gene editing materials to correct the cells within the affected tissue). All initial trials will assess the safety of the CRISPR intervention before determining if or how well it works.

Chimeric Antigen Receptor T-Cell Therapy

In the 2010s, there were some exciting breakthroughs with a slightly different type of gene therapy, a cell-based gene therapy. Known specifically as "*c*himeric *a*ntigen *r*eceptor T-cell" (CAR-T) therapy, each treatment is unique to the individual patient as cells from the patient are removed, genetically modified,

and readministered back to the patient. The cell modification involves the addition of a new gene to a sample of the patient's T cells (a type of immune cell) to produce a chimeric antigen receptor (receptors are proteins located attached to the surface or outside of the cell) (see Fig 9.1). With this new protein, the modified T cells hunt for leukemia cells that have a similar receptor on the surface. One could compare this to giving a bloodhound a sample of clothing of a missing person in order for it to hunt for that specific scent. This type of gene therapy is specifically used for blood cancers at this time. In 2017, the FDA approved the first CAR-T called Kymriah for pediatric and young adults affected with acute lymphoblastic leukemia. Shortly thereafter, the FDA approved a second CAR-T therapy for large B-cell lymphoma. While these treatments have been shown to be very effective in some patients, there is a risk for serious side effects including high fever, flu-like symptoms, and neurological events.

Ethical Considerations of Gene Therapy and Gene Editing

In addition to the scientific challenges of gene therapy and gene editing, the field has been immersed in ethical issues since its inception. In particular, these technologies can be used for nontherapeutic purposes, such as enhancement of traits, and biowarfare. It may be possible that gene editing can be directed to specific cells or tissues (such as muscle cells for a muscle-wasting condition) or administered systemically (any cell could be edited). At the very early stage of life (just a few cells), it may even be possible to perform gene editing, whereby all cells would be edited or corrected. The potential modification of germ cells (egg/sperm) would be passed along to future generations, giving rise to the field dubbed as "intergenerational" research. Some have argued that the inadvertent or intentional modification of germ cells may be ethically acceptable if the end result is elimination of a life-threatening disease (who would want to risk passing the mutated gene onto their children), while others contend it would be unethical to modify germ cells and alter the genetic makeup of future generations. In 2018, a Chinese scientist performed gene editing on early-stage embryos (formed through in vitro fertilization or IVF) to create a deletion (missing DNA) in a gene essential to HIV infection. The gene deletion occurs naturally and these individuals are not able to be infected with HIV. This type of experimentation is prohibited in most countries and was condemned by the scientific community and national governments.

Thus, as the field advances, careful consideration should be given to the safe and appropriate use of the technology.

Conclusion

With continued progress in the field of gene therapy and the infusion of new hope of gene editing, the promise of therapeutic genetic interventions increases. The potential to correct genetic alterations causative of disease that otherwise is not currently treatable has created tremendous hope and excitement. However, much work remains to understand the safety of gene editing and the long-term effects. In parallel with the scientific developments, guidelines and policies should be developed to guide the safe and appropriate use of these new scientific tools.

Resources

US Food and Drug Administration. What is Gene Therapy? Available at https://www.fda.gov/vaccines-blood-biologics/cellular-gene-therapy-products/what-gene-therapy

Gene Therapy Net. Available http://www.genetherapynet.com/clinicaltrialsgov.html

US National Library of Medicine. What are genome editing and CRISPR-Cas9? Available at https://ghr.nlm.nih.gov/primer/genomicresearch/genomeediting

Broad Institute. Questions and Answers about CRISPR. Available at https://www.broadinstitute.org/what-broad/areas-focus/project-spotlight/questions-and-answers-about-crispr

US National Cancer Institute. CAR T Cells: Engineering Patients' Immune Cells to Treat Their Cancers. Available at https://www.cancer.gov/about-cancer/treatment/research/car-t-cells

Renault M. Gene-editing treatment shows promise for sickle cell disease. Washington Post, Dec 5, 2020.

10

Hunting the Invisible Bugs

Microbes or microorganisms are the oldest form of life and are only visible through a microscope. They are also the most common organism on the planet and are found in every continent and ecosystem—in the air, in the water, in soil, and in plants and animals. In fact, we and most other plants and animals could not live without them. Viruses are also considered microbes although there is some debate about whether they are actually a "living" organism as they rely on other organisms to replicate. Microbes play critical roles in all ecosystems and have a range of relationships with other organisms and the host (or organism where it resides and get its nutrients) ranging from a mutually beneficial co-existence to a harmful relationship, either to the host or microbe. But humans could not survive without the microbial communities that exist in the gut, on the skin, and in other parts of the body. They provide huge benefits to us that are critical for our health. At the opposite extreme, however, microbes can cause sickness and even death in the organisms they infect, including humans. In the latter situation, disease-causing microbes are referred to as "pathogens."

While some infectious diseases have been substantially reduced or eradicated due to public health interventions, such as water sanitation and vaccines, infectious diseases remain the major cause of death worldwide, particularly in developing nations. If not fatal, many infectious diseases can cause moderate-to-severe health problems. Symptoms may appear and disappear in a relatively short period (a few days to a few weeks) for some infectious diseases (food poisoning, influenza). In contrast, other infectious diseases become chronic illnesses with potential long-term damage (hepatitis C, HIV),

S. B. Haga, *The Book of Genes and Genomes*, https://doi.org/10.1007/978-0-387-70916-1_10

or remain dormant (no symptoms) and may be triggered or re-awakened by environmental factors (herpes simplex virus, tuberculosis). Lastly, novel or new pathogenic versions may appear and disappear. Examples include the MERS and novel coronavirus-19 virus.

A Quick History of Pathogens

Since ancient times, there have been two major hypotheses regarding the causes of diseases. One school of thought was that most diseases were caused by microorganisms, even before it was understood what microorganisms were. It was believed that these diseases were spread from person to person (or animal to animal) and that microbes were found in bodily secretions though they were not visible by eye. Other scholars in medicine and science held different views, tending to blame the individual as having some inherent weakness and thus succumbing to disease. Diseases were considered to be "spontaneous" and not transmitted between affected persons or animals (the doctrine of spontaneous generation). After all, how can diseases be caused by something that cannot even be seen, only theorized (a pre-microscope argument).

Use of the word "germ" was first documented in the seventeenth century—originating from the Latin word "germen," meaning the sprout or bud of a plant (e.g., wheat germ). The use of the word in reference to "seed of a disease" appeared in the eighteenth century. Microorganisms were first observed by the Dutch scientist Antonie van Leeuwenhoek in the 1600s using his newly constructed microscopes. However, the study of microbes (known as microbiology) did not really take off until the late 1800s. The German physician, Robert Koch, is credited with providing the first evidence to support what is known as the germ theory of disease. His initial studies observed anthrax spores under the microscope in blood samples taken from diseased cattle. He demonstrated that mice infected with the blood from diseased cattle succumbed to disease—in complete opposition of theories of "spontaneous" outbreaks of disease in cattle. In 1882, Koch discovered the bacterium which caused tuberculosis, an extremely common and fatal disease of the time. He was awarded the Nobel Prize for his work on tuberculosis in 1905.

The famous French chemist Louis Pasteur provided further evidence supporting the germ theory and disproving the notion of spontaneous disease. In one of his landmark experiments, he set up a series of flasks containing boiled (sterilized) broth. One of the flasks was connected to a long, open-ended, winding glass tube (resembling a swan's neck); another flask was the

same as the first flask but tilted to allow some of the broth to fill the long tube; and a third flask was completely open (no tube). The second and third flasks eventually became cloudy, suggesting that something from the air was able to reach the sterilized broth. However, in the first flask with limited access to air (the air had to travel down the winding tube), the broth remained sterile. Thus, Pasteur concluded that particles in the air carried germs.

Pasteur also showed that the fermentation (spoiling) of certain liquids like milk was caused by microorganisms. He showed that if the liquid was heated prior to storage, it prevented fermentation since microorganisms were killed by the heat. This process came to be known as "pasteurization," which is still in use today (think "pasteurized" milk).

Microbial Genomics

Microbes, or microorganisms, are single-cell organisms, including bacteria, fungi, yeast, and some primitive forms of algae. In contrast, all members of the animal kingdom are multicellular, or comprised of many cells with different functions (e.g., heart cell, brain cell, muscle cell). Just as with animals, all microbes have a genome. While most microbial genomes are made of DNA, some viruses have a genome made of RNA (a close relative of DNA). For example, the human immunodeficiency virus (HIV) and coronaviruses have an RNA code.

To date, scientists have sequenced the genomes of hundreds, if not thousands, of microbes. Some of the microbes that have been sequenced have not even been discovered or identified—new sequencing technologies can sequence microbes residing in samples from the soil, the ocean floor, and in air samples atop the Empire State Building without first growing them in a laboratory. As discussed later, sequencing is one of the first things performed with the discovery of a new pathogen—it is the quickest way to identify it by comparing the sequence of the new microbe to known ones. As you might imagine, compared to the genomes of plants and animals, microbial genomes are significantly smaller and do not contain the extra noncoding DNA that higher species do. Viruses typically only have 7–9 genes (recall that humans have more than 20,000)!

Tracking an Infectious Disease Outbreak

The cause of a new disease, particularly one that appears quickly and affects many people, could be due to a harmful environmental exposure (e.g., exposure to a chemical in a factory) or a pathogenic microorganism. Rapid collection of data about when the disease occurred or first developed, in whom and where (e.g., patients of a hospital, employees in a factory, an entire town), and the symptoms is critical to predicting the potential cause of the disease—environmental or microbial. If available, patient blood or other biospecimens can be analyzed for infectious agents (through microscopic analysis, culture of the sample under different growth conditions, or direct sequencing) or other signs of an infection (abnormal levels of immune system markers).

If a novel disease is suspected to be caused by a pathogenic microbe, scientists will embark to identify the source of the outbreak. Many diseases begin in animals and eventually the microorganism's genetic code will evolve such that it will become infectious in humans (i.e., it will jump the species barrier). A pathogenic microbe may not cause disease in all animals and humans. Some infectious diseases only occur in animals and are not harmful to humans. In other cases, the microbe may cause different symptoms in animals and humans. If the disease is suspected to be caused by an infectious agent, the next question is what type of pathogen—viral, bacteria, or other type of infectious agent.

Understanding the source will yield clues to the disease's transmission—how does it spread? Is it through consumption of wild animals that are infected (e.g., bats, birds)? Is it through consumption of contaminated meat, vegetables, or other products? Note—acquiring an infectious disease through the consumption of a contaminated product is not quite the same as consuming an infected animal that is not processed. Contamination often occurs in the processing and handling of food products, and is caused by humans. If consumed, the contaminated products will cause illness in humans, often referred to as foodborne outbreaks. Recall when warnings are issued by government agencies for certain goods (e.g., romaine lettuce, ground beef from a specific manufacturer) or restaurants (e.g., Chipotle, Jack in the Box). Once the specific source of contamination has been identified, the product is removed from the market, and the disease outbreak will cease. However, this chapter does not really focus much on contamination as a source of disease outbreak.

The spread of an infectious disease can evolve and have many steps and potential routes. One common method of transmission is through an intermediate, or another organism, that then passes it on to humans. For example, the microorganism that causes malaria is passed to humans via mosquitoes (in

this role, the mosquito is referred to as the vector or carrier). The microorganism responsible for Lyme disease is carried in ticks. Reducing or eliminating these vectors will substantially reduce the number of infections.

Lastly, there is human-to-human transmission, where one infected person can transmit the pathogenic microorganism to a noninfected person. This may occur through contact with certain bodily fluids (e.g., blood, semen) or through the air (referred to as airborne). A microbe may be released into the air by a sneeze or cough and another person nearby may inhale or touch a surface that has been contaminated and then touch their face, nose, eyes, etc. Understanding how a disease is transmitted is key to developing strategies or interventions to minimize exposure to healthy individuals and the spread of disease.

Laboratory Testing

Many conventional tests in microbiology (the field of study of microorganisms) rely on the growth (or culture) of a biospecimen obtained from a suspected infected patient. For example, a cheek swab or vaginal swab collected from a sick patient is wiped across a petri dish or inserted into a test tube that contains a food source for microbes. Typically, this food source resembles jello! The petri dish or test tube is incubated at a warm temperature for a few days and then checked for any growth (easily visible by eye). If growth appears, certain physical characteristics can narrow the identity of the microbe (color, appearance, growth rate, etc.) and further testing can be performed to confirm the specific pathogen. Other types of conventional microbiology tests include microscopy (looking for characteristic microbial shapes and staining patterns under a microscope), antigen detection (looking for pieces of the microbe that triggers an immune response), and serology (looking for signs of an immune response).

Conventional microbial testing is slow and may not be able to identify new microbes. Furthermore, the differences between one disease strain and another may not result in any detectable physical or growth differences, making it challenging to accurately distinguish between closely related strains. Today, the primary way to truly distinguish one microbe from another is by DNA or molecular analysis. Molecular analysis allows a definitive diagnosis and can also enable sequence comparison of samples isolated from other patients anywhere in the world to a related strain of the infectious agent and/or to samples isolated from potential carriers of the infectious agent such as a mosquito or rat. A molecular test can entail the analysis of a small, unique region of DNA

to identify the microbe or for unknown microbes, the sequencing of the entire genome.

Legionnaires' Disease

One interesting example that illustrates the traditional approaches and tools used by epidemiologists (scientists who search for the causes and effects of diseases including infectious diseases) to identify the cause of a new disease is Legionnaires' (lee-jun-airs) disease. It is an infection of the lung by a bacterium known as *Legionella* bacteria and leads to the development of a serious type of pneumonia. Symptoms include cough, shortness of breath, headaches, and fever. Antibiotics are used to treat Legionnaires' disease; about 1 out of 10 people die from the infection.

In the summer of 1976, a mysterious disease outbreak was reported among attendees of the annual convention of the American Legion. Eventually, some two dozen deaths were reported, sparking fear across the city. Scientists were sent to Philadelphia from the US Centers for Disease Control and Prevention (CDC) to interview those affected, retrace their steps and activities, and review their medical and autopsy records to find some commonality between them. Samples were collected and analyzed from various locations and tissues of the sick and deceased. Not only were attendees affected (all men), but so were some wives. However, it did not appear that the disease was contagious, as attendees who had shared a hotel room did not all become sick. Many convention attendees were staying at the Bellevue-Stratford Hotel and socialized and attended events in the hotel's meeting rooms. The media covered the search and the scientists' work in multiple feature stories such as Time magazine's "Tracing the Philly Killer."

Six months after the first cases were reported, the CDC announced that it had identified the responsible microbe, aptly named *Legionella* (and the corresponding disease dubbed Legionnaires' disease) through the analysis of tissue samples from infected individuals. However, this did not yet explain how the microbe was transmitted and its source. Scientists eventually determined that the bacterium grew in water sources and thus deduced that it was transmitted through a water source like the cooling tower (part of a central air-conditioning unit). Unfortunately, by the time the scientists had isolated the bacteria, the hotel's air-conditioning units had been disinfected and no trace was recovered. Ironically, some scientists stayed at the Bellevue-Stratford Hotel, not realizing that the microbe responsible for this pneumonia-like disease was actually in the hotel's cooling unit. So, in contrast to the more

familiar airborne transmission of many infectious diseases, individuals can inhale the infectious microbe dispersed through aerosols emitted from common appliances found in stores, office buildings, or hotels, enabling widespread infectivity in these settings (think of a fine spray emitted from a device such a showerhead or air-conditioning unit). Thus, any water source like hot tubs and cooling towers for air-conditioning units that are not properly maintained and regularly disinfected can be reservoirs for *Legionella* bacteria.

Today, Legionnaires' disease (and many other diseases) remains a worrisome infection, with an increasing rate of infection reported at the turn of the century (increased fourfold between 2000 and 2014). The disease is still being studied with more modern techniques such as genomic analysis to determine why people respond differently to the infection. As is true for many diseases, infected individuals may have a unique experience ranging from mild to severe and experience different combinations of main symptoms (Fig. 10.1).

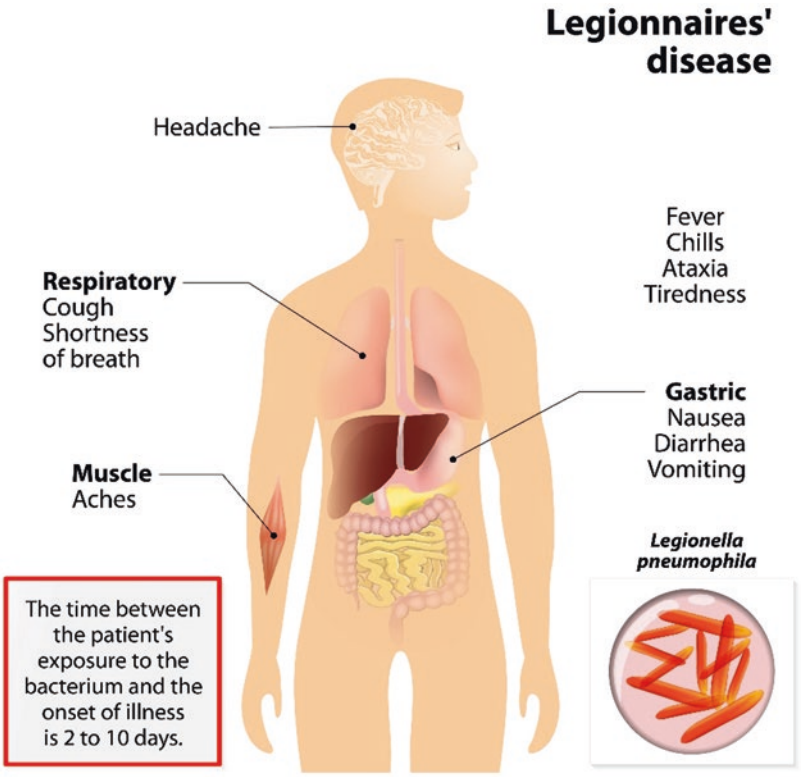

Fig. 10.1 The symptoms associated with Legionnaires' disease (caused by an infection of *Legionella pneumophila*) (source: Adobe Photo Stock)

Recent Infectious Diseases

The following examples illustrate the benefits of molecular analysis to identify, track, and monitor infectious agents from national and international outbreaks that have occurred over the past few decades.

SARS

The international outbreak of severe acute respiratory syndrome (better known as SARS) was officially recognized in February 2003, although it is believed that the outbreak began some time in the fall of 2002. In early February, an adult male in Vietnam with high fever, dry cough, muscle pain, and mild sore throat was admitted to the hospital. His condition worsened over the next 4 days with increasing breathing difficulties and respiratory distress requiring ventilator support. He was transferred to a hospital in China, but died exactly 1 month after he was admitted.

On February 11, the Chinese Ministry of Health reported 305 cases of an acute respiratory syndrome of unknown cause resulting in five deaths in southern China. By March 21, 350 suspected cases had been reported with 10 deaths from 13 countries including Italy, Ireland, Canada, the USA, Thailand, and Singapore. By April 1, that number had risen to a total of 1804 cases and 62 deaths reported from 15 countries. By the end of the month, more than 4300 cases and 250 SARS-related deaths were reported from more than 25 countries. The majority of victims were healthcare workers who had been in close contact with patients with SARS. Numerous travel advisories were issued as it was discovered that several people with SARS had traveled from elsewhere.

The exact cause and mode of transmission were unknown initially and the disease appeared to be rapidly spreading around the world intensifying fears of a potential epidemic. Numerous laboratory tests were conducted on SARS patients in an attempt to figure out what exactly caused SARS in order to determine how best to prevent and treat it. In 2003, genomic technologies were still relatively new and not as accurate as today's technologies. In Canada, DNA was extracted from lung fluid and blood samples from nine of ten patients of the Toronto outbreak. The samples were tested for known viruses by looking for specific DNA sequences characteristic of different viruses such as herpesvirus, parvovirus, influenza virus, Ebola and Marburg virus, and measles virus.

All of the tests for known viruses came back negative. However, a new strain of a virus that belonged to a family of viruses called coronaviruses was

detected in about half of the Toronto SARS patients. Scientists in Hong Kong and at the CDC had also detected this viral type in other SARS patients. The first coronavirus was actually discovered in the 1960s, and by the end of the decade, several strains had been discovered from both animal and human cases. Interestingly, a wide range of diseases resulted from coronaviruses, from stomach to liver disease to respiratory illness. Compared to other coronavirus respiratory illnesses, none have appeared quite as severe as was seen with SARS. SARS was soon shown to be spread by person-to-person contact, most likely through exposure to droplets from an infected person's cough or sneeze.

About 2 months after the SARS virus was identified, two laboratories independently sequenced the genome of the virus. An international team of scientists from the USA and Europe sequenced the SARS virus isolated from a Vietnamese patient, while a Canadian team sequenced the SARS virus from a patient in Toronto. Of the 30,000-base pair-long genome, only about 8 base pairs differed between the two patients. The SARS coronavirus differed in its genetic sequence from other known coronaviruses found in pigs, cattle, and chickens, suggesting that the virus had been evolving for a long period of time before infecting humans.

Further investigation of the SARS sequences showed that virus isolated from the first group of patients infected with the virus resembled a SARS-like virus found in animals. Chinese scientists sequenced the genome of the SARS virus isolated from 61 patients infected in the beginning, middle, and end of the outbreak. In addition, they sequenced the genome of the SARS virus isolated from two palm civets, a wild catlike animal. It was hypothesized then that SARS virus initiated in bats and jumped to palm civets, which were sold in Chinese markets and consumed. Another study detected the SARS virus in the raccoon dog, also a delicacy in China.

By comparing the DNA sequence from different isolates in humans and animals, the scientists were able to develop a genetic timeline based on the number of mutations and confirm the likely animal source from which the SARS virus was initially transmitted. Furthermore, the strain of the SARS virus from patients infected later in the outbreak slightly differed from the strain found in the first group of patients, suggesting a very rapid pace of mutation early in the outbreak. Toward the end of the outbreak, the rate of mutations appeared to level off.

By the summer of 2003, the SARS outbreak appeared contained and was limited to a few cases. It is estimated that this relatively short-lived outbreak of SARS estimated affected more than 8000 people worldwide and was responsible for almost 800 deaths. A handful of cases of SARS were reported in late 2003 and early 2004, but it has otherwise disappeared.

E. coli Outbreak (Spinach)

On September 8, 2006, health department officials from Wisconsin notified the CDC about an outbreak of *E. coli*-caused illnesses. *E. coli* is short for *Escherichia coli*, a very common bacterium that resides in our gut (a good kind), which is used in almost every biomedical research lab. *E. coli* is a common bacterium that actually lives in the gut of all animals, including humans. Most strains of *E. coli* are harmless and actually help the body fight off harmful infections by preventing the growth of harmful bacteria and by synthesizing vitamin K_2, important for blood clotting.

However, harmful strains of *E. coli* can cause mild-to-severe illness and potentially death. In 2006, Wisconsin health officials believed that the strain of the outbreak was *E. coli* O157:H7, which was quickly confirmed by CDC scientists on September 12 by genetic analysis. The major symptoms are gastrointestinal—diarrhea, severe stomach cramps, and vomiting. There are four strains of *E. coli* that can cause gastroenteritis (gas`tro-en-ter-i-tis) or stomach illness in humans. This particular strain of *E. coli* produces toxins (Shiga toxins) that are damaging to the stomach lining. Although most people recover from *E. coli*-causing gastroenteritis with no treatment, young children and the elderly are susceptible to further complications. Undercooked or raw ground beef, alfalfa sprouts and lettuce, and unpasteurized fruit juices and milk have been the source of *E. coli* O157:H7 in previous outbreaks.

On September 13, Wisconsin and Oregon health officials reported to the CDC that the source of the *E. coli* outbreak was fresh spinach. On the same day, health officials from New Mexico spoke with Wisconsin and Oregon health officials about a suspected *E. coli* outbreak linked to the consumption of fresh spinach. The next day, CDC scientists confirmed that the source of the outbreak was indeed fresh spinach. Interviews with the victims of the outbreak revealed that most had eaten fresh spinach within the past 10 days of becoming sick. Fifty illnesses including one death in eight states believed to be linked to contaminated spinach had been reported. This information was quickly relayed to the FDA and hours later the FDA issued a warning to the public not to eat bagged spinach due to an *E. coli* outbreak.

By September 16, the number of reported cases of illness due to the *E. coli* outbreak more than doubled to 102. Most healthy adults were able to recover within a week, but young children and older individuals were at higher risk for developing kidney failure, which can cause kidney damage and even death. Before it was all over, the spinach/*E. coli* outbreak would sicken 205 people and cause 3 deaths.

This was not the first time (nor will it be the last) that an *E. coli* O157:H7 outbreak involving leafy greens grown in California has occurred. In fact, the FDA had developed the "Lettuce Safety Initiative" to reduce the potential for recurring outbreaks of *E. coli* O157:H7. On September 20, the New Mexico Department of Health announced that it had matched the *E. coli* O157:H7 strain obtained from a contaminated package of spinach to sample from a patient in New Mexico believed to be sickened by the contaminated spinach using DNA analysis. Several other state health departments from Utah to Pennsylvania subsequently confirmed initial reports of *E. coli* O157:H7 in the contaminated bagged spinach. The contaminated spinach was traced back to bagged spinach sold by a single company. State and federal investigators began scouring the processing and packaging plant, where the contaminated products were believed to have been processed. However, no evidence of contaminated machinery was found in the processing plant.

Investigators then turned their attention to where the spinach was grown and harvested. By September 22, the source of the tainted spinach was narrowed down to three counties in California: Monterey, San Benito, and Santa Clara. Genetic profiles of the *E. coli* from the bagged spinach were compared to spinach samples collected from nine farms believed to have supplied spinach to the packaging plant. A match was found between the *E. coli* strain in the bagged spinach to a sample collected from one of the nine farms tested.

After identifying where the contaminated spinach came from, investigators next wanted to determine the source of the *E. coli* contamination. Wild pigs were known to inhabit an area on the edge of the spinach field that tested positive where the fences dividing a cattle pasture and the spinach fields were broken. In addition, the contamination could have originated from irrigation wells located relatively close to the spinach fields as well as waterways that could have carried infected feces from cattle and wild animals. Genetic analysis of samples of cow manure revealed the same strain of *E. coli* as found in the contaminated bagged spinach. However, since the contamination was believed to have arisen before the spinach reached the processing plant, it was not possible to determine the exact source of the *E. coli*. Nevertheless, as a result of the national outbreak, efforts between state and federal officials and industry are underway to prevent contamination and improve the safety of fresh produce.

MRSA

According to the CDC, more than 2.8 million will develop antibiotic-resistant infections in the USA each year, and about 1.25% will die as a result. Antibiotic-resistant infections are just that—bacterial infections that are not treatable by commonly used antibiotic treatments. One example of an antibiotic-resistant infection is methicillin-resistant *Staphylococcus aureus* (MRSA). According to the latest statistics from 2017, about 324,000 patients were infected with MRSA in the US hospitals, slowly dropping from 400,000 cases in 2012. MRSA is not exclusive to hospitals—it can be found throughout community settings (schools, gyms, stores). In mild cases, MRSA results in a skin infection. If the skin is broken, an infection can develop leading to pneumonia (an infection of the lungs) or sepsis (an infection of the bloodstream).

Tracing the source of a hospital outbreak is critical to halting the spread. In the early 2010s, genome sequencing first started to be used to identify the source of MRSA hospital outbreaks and the route of transmission. Prior to that time, conventional microbiology methods of cultures and resistance tests (where samples are grown and exposed to different types of antibiotics) were used to identify and characterize the pathogen.

One early example of the benefits of genome sequencing was published in 2013. An MRSA outbreak in the special care baby unit of a British hospital occurred in 2011. Sequencing of the bacteria obtained from patient specimens and staff was performed. Based on patient records and conventional laboratory testing, three cases occurred at about the same time—and testing indicated that they were likely caused by the same strain of MRSA. Further review of cases from the prior 6 months identified another 13 cases, but 5 were considered to be different, leaving an overall total of 11 near-identical cases. One additional case was identified that had occurred after a deep cleaning of the hospital unit had been completed, which was also determined to be identical to the majority of other cases, bringing the total to 12 cases.

The hospital sequenced all 17 MRSA patient cases to determine if their conclusions based on the conventional microbiology tests were correct. Analysis of the MRSA sequences revealed that two cases were erroneously excluded (and three cases were correctly excluded). Due to the odd case that arose after deep cleaning of the hospital unit, the researchers suspected that a staff member may be carrying the pathogen and have unknowingly transmitted it to the patient after the cleaning. A total of 154 samples were collected and screened by cell culture from 154 staff members. One was positive for

MRSA, which was subsequently confirmed by sequencing to be the same strain. The staff member was treated and no further cases occurred.

Anthrax

Anthrax is a disease caused by the bacterium called *Bacillus anthracis*, which naturally exists in soil. Domestic and wild animals (cattle, sheep, goats, antelope, and deer) may be infected with anthrax through contaminated soil, plants, or water, but it rarely occurs in humans. Infection in humans may occur through inhalation of anthrax spores, eating food or drinking water that is contaminated with spores, or infection through an open skin wound. Symptoms of infection depend partly on the route of infection (i.e., skin, inhalation) and include skin sores, vomiting, and shock (Fig. 10.2). Antibiotics can be given to infected individuals; inhaled anthrax is more difficult to treat.

If an individual has access to spores produced by the bacterium, it is possible for the spores to be dried and used as a form of bioterrorism. Shortly after the attacks of September 11, 2001, letters containing the bacterium anthrax were mailed to news media offices and two US Senators in Washington, DC, Florida, and New York, leading to the death of five Americans and further heightening the nation's fear of more attacks. The largest biological attack in the US history, the investigation of the anthrax letters, dubbed "Amerithrax" by the FBI, relied heavily on genomic technologies.

In 1999, the National Institutes of Health funded a project to sequence the genome of the common strain of anthrax. Little did they know how significant this decision would be in a few short years. Known as the Ames strain, this strain of bacteria was originally isolated from a cow in Texas in 1981, but was named the "Ames strain" as it was mistakenly believed to have originated in Ames, Iowa. The US Army Medical Research Institute of Infectious Diseases in Maryland was the first laboratory to study the Ames strain in its biological weapons program during the 1980s. Since then, it has been distributed to at least 15 laboratories across the USA as well as to a handful of laboratories outside the USA.

The genome sequence of the anthrax strain isolated from the first victim of the 2001 attacks (Mr. Robert Stevens of Florida) was published in early 2002. Comparison of the genome sequence of the Ames strain to the strain from the first victim confirmed earlier reports that the two strains were related. In particular, based on the genomic analysis of other known anthrax strains, the scientists identified the specific Ames lineage that the Florida anthrax

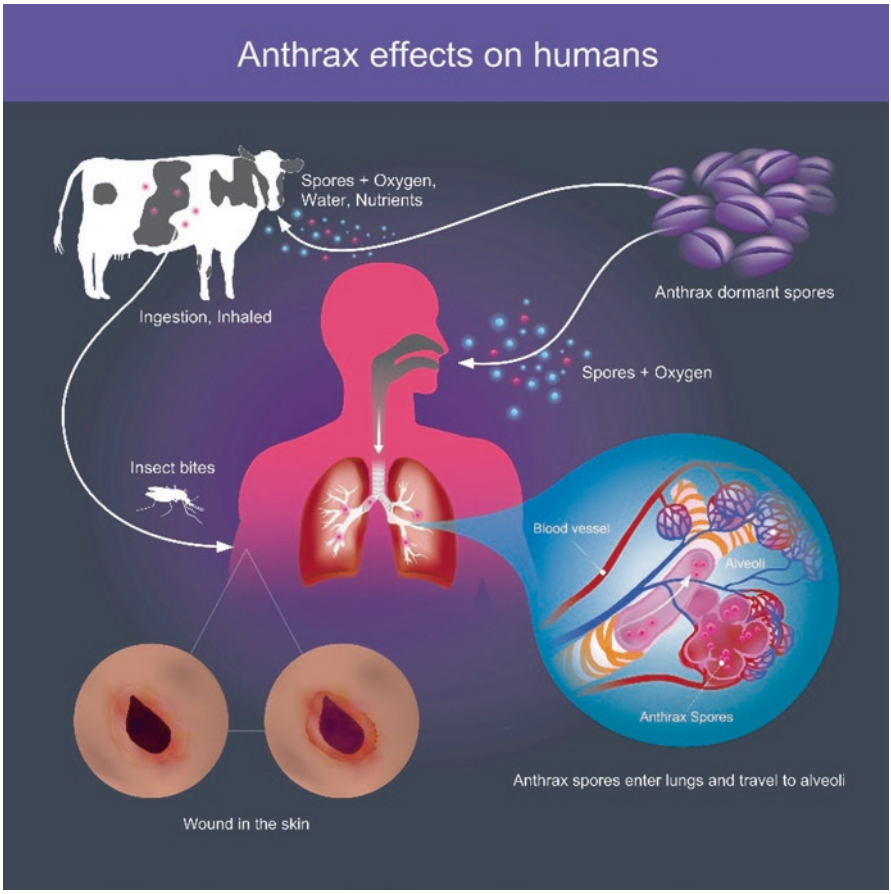

Fig. 10.2 Effects of anthrax infection in humans (source: Adobe Photo Stock)

originated from. However, the genomic analysis could not determine the exact origin of the Florida strain.

There is no doubt that genome analysis substantially aided investigators in identifying the strain of anthrax and its source. A scientist working in a top-security laboratory who had direct access to anthrax and other pathogens has been identified as the person responsible for the attacks. He committed suicide before investigators had the opportunity to learn of his motives.

SARS-CoV-2

In late 2019, cases of a respiratory illness were reported in China. Once patient samples were obtained, the cause was quickly determined to be another

coronavirus through sequencing and genomic analysis. The official name of the virus was SARS-CoV-2 virus, but was also referred to as novel coronavirus-19. The disease caused by this virus is referred to as "Covid-19" (short for coronavirus infectious disease-19). It is a highly contagious disease spread by droplets and close contact with infected persons. Declared a global pandemic in February 2020 by the World Health Organization, it has affected more than 157 million people worldwide and caused almost 3.2 million deaths (as of May 2021). Multiple vaccines have been developed using new and traditional methods.

The SARS-CoV-2 viral genome is about 30,000 bases long and it is made of RNA, instead of DNA though very similar in structure. The SARS-CoV-2 genome is divided into 14 sections of code that can produce 27 different proteins. One of the proteins produced is the characteristic spike-looking molecules present on the outer shell (Fig. 10.3). The origin of the SARS-CoV-2 virus was unclear initially, though the first reported cases were associated with consumers of wild animals sold in local markets in Wuhan, China. As viruses accumulate mutations over time, genetic analysis enabled the tracing of viral spread from region to region. For example, in comparison to the first reported case of SARS-CoV-2 in Wuhan, a patient 650 miles away in another region of China was found to have a single change in the viral genome. Subsequent patients carried a viral strain that had two additional changes. Changes may or may not impact the disease. Changes that make the virus less virulent are good for humans, but bad for the virus and vice versa. Genetic analysis was also key to understanding the spread of the virus in the USA—the East Coast infections appear to have been mostly derived from infected cases coming from Europe based on genetic similarity, and infections on the West Coast were more similar to cases from China and other Asian countries.

Genetic technologies figured prominently into the diagnostic tests and interventions that many companies and academic medical centers raced to develop. Rapid and accurate diagnostic testing is key to monitoring the rate of transmission and identify infected individuals as quickly as possible. Most of the diagnostic testing was molecular based, meaning that it analyzed part of the RNA sequence of virus. Diagnostic testing differs from antibody testing in that antibody testing looks for evidence that a person has been infected with the virus and developed an immune response (production of antibodies).

Furthermore, the race to quickly develop a safe and effective vaccine led to some companies testing new technologies to create genetic-based vaccines instead of traditional vaccines that involve the whole virus that is attenuated or weakened (so as not to make people sick but enough to cause our immune system to create antibodies). In particular, some companies developed a

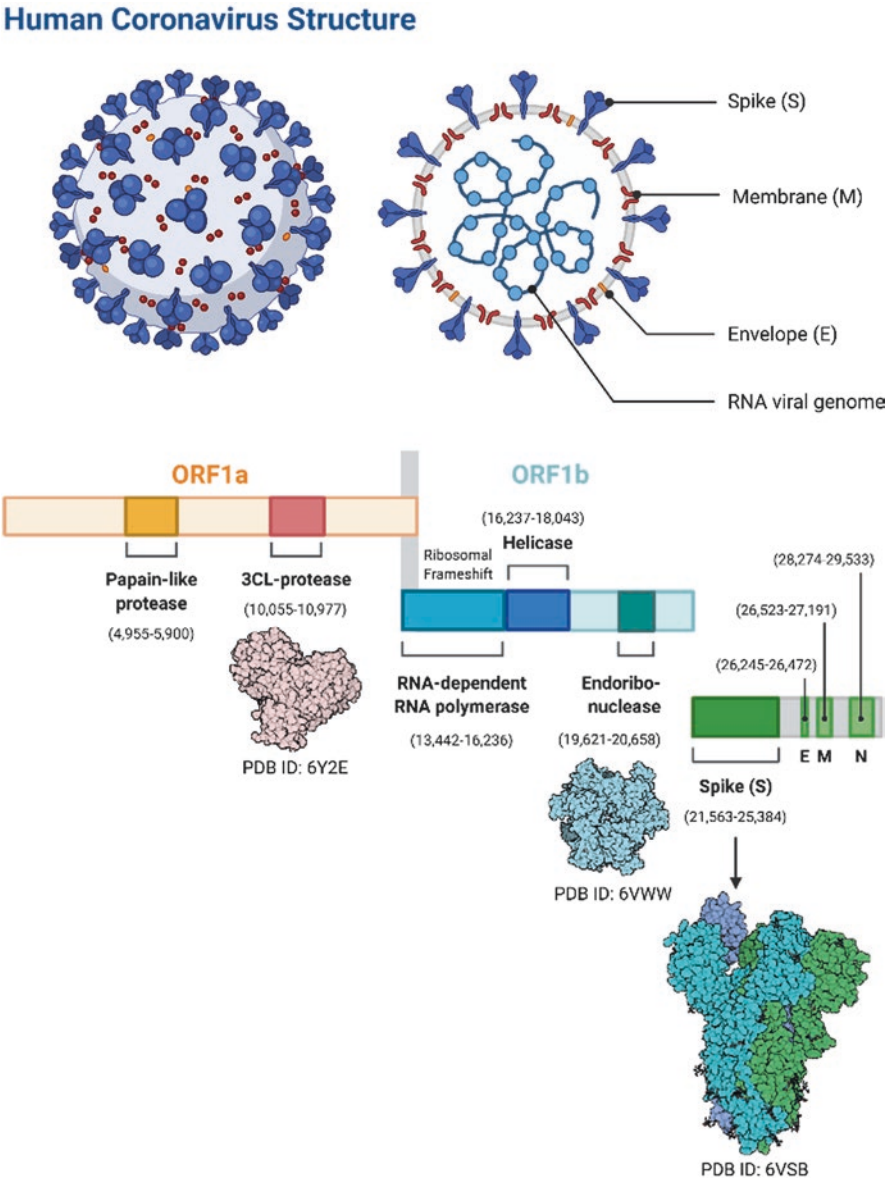

Fig. 10.3 (Top left) An illustration of the outward appearance of the SARS-CoV-2 and (top right) an internal view of the viral genome and key proteins. The virus is almost 30,000 bases long and encodes for 27 proteins; the bottom illustration shows three key proteins (ORF1a, ORF1b, and Spike (S)) critical for its survival (source: Biorender)

vaccine based on RNA- or DNA-based genes from the virus—once injected, the individual's body would produce the protein base immune system.

Microbiome

In recent years, we have learned that an inordinate number of microbes live within and on our body surfaces. These includes bacteria, viruses, and fungi. Although we often associate microbes with bad outcomes (getting an infection), the ones that reside within or on our bodies are actually good and necessary for our health and daily function. One could think of our bodies almost as a hotel for microbes. Scientists have become aware of the wide variety of microbes living on and within us thanks to the advances in genomics technologies. Traditionally, the field of microbiology has used techniques such as collecting samples from the environment or patients, growing them in the laboratory, and examining the growth rate and physical properties in response to different food sources, antibiotics, or temperature. However, many microbes simply do not survive in the laboratory once removed from their natural environment, and thus, new genomics technologies have enabled scientists to directly sequence their genetic code. Since each microbe has a unique genetic code, their identity can be ascertained by comparing it to codes of other microbial species and deducing which group they belong to by their degree of similarity. More often than not, a community of microbes will exist in a given sample, and genomic technologies can simultaneously sequence all of the different genetic codes that correspond to the multiple species in a sample. Furthermore, these technologies can quantify or give us a sense of how much of a particular microbe is present, e.g., perhaps there are two predominant microbial species and the rest exist at very low levels.

As a result, scientists have identified the combination of microbes present at different sites in the human body. In a large study started in 2008, scientists obtained multiple samples from 300 individuals from 15 sites in men and 18 sites in women, including the nose, mouth, skin (crook of our arm/elbow), and lower intestine (stool sample). The genetic sequences from each of the samples from each person were compared to other sites taken from the same person and to other people. Overall, it is estimated that more than 10,000 microbial species exist throughout the human body. The greatest diversity (or combination of different types of microbes) has been reported in tooth and stool samples, with inner cheek and skin samples showing moderate diversity, and vaginal samples showing the lowest diversity. There is overlap in the types of microbes present between multiple sites, but the amount of each species at a specific site may differ. With respect to age, after the first 3 years of life when tremendous growth and development occur, the microbiome at each tissue site stabilizes. In addition, samples between thousands of people around the

world have been analyzed and compared, revealing some geographic differences, likely due to differences in temperature/climate, diet, health status, and other environmental or cultural differences. Globally, the data are interpreted as a continuum rather than discrete differences between regions. Since no two people are exactly the same, it is possible that a unique "microbial signature" exists for each person.

Following many reports describing the range and structure of microbial communities at various sites in humans, attention has shifted to the significance of the microbiome with respect to health and wellness. What does a "healthy" microbiome look like compared to one that increases the risk of disease or is indicative of a disease that has begun to or already developed? As many important functions occur in our gut (stomach/intestine), there has been a great deal of study regarding the composition of the gut microbiome and its relationship to disease. Many reports have identified differences in the gut microbiome between healthy individuals and those affected with obesity, cancer, or other conditions. Gut microbes are essential to breaking down foods, generating important chemicals our bodies need daily, and contributing to our body's immune (defense) system. As changes in the microbiome can occur during the early phases of disease development, it may serve as a warning sign or biomarker of disease before we actually feel ill.

Our microbiome is not constant, meaning that it can be changed based on our environment, diet, medications, and health status. In many cases, it will revert back to its original state if the change in environment is temporary (e.g., a short-term illness, on a diet). While we may not have control over some factors that impact our microbiome, such as climate, we do have some choice over our diet and thus can intentionally make "improvements" to our microbiome through dietary changes. Studies have shown differences in the gut microbiome between vegetarians and nonvegetarians. Many foods are now being marketed as "prebiotic" or "post-biotic," which refers to the effect on the microbiome. Specifically, a prebiotic food would provide sustenance for microbes and stimulate growth; many of these foods are rich in fiber. One could think of it as "feeding our microbiome." In contrast, a post-biotic food contains live bacteria such as yogurt and some juice cocktails (often found near orange juice in the refrigerated section). Unfortunately, limited data exist regarding the health impacts of these foods specifically regarding the impact on the microbiome and ultimately reducing disease risk and infections. Nonetheless, there are plenty of benefits to eating yogurt and fiber!

Medications can also alter the composition of the gut microbiome (and actually the gut microbiome can play a critical role in breaking down drugs into smaller components to be removed from our bodies). The changes in the

microbiome may vary depending on the specific drug, dose, and period of time the drug is being taken. For example, an antibiotic is intended to kill bacteria in general and does not selectively kill off the "bad" bacteria causing an illness and leave the "good" bacteria that live in our gut and perform important functions. Thus, a common side effect for many patients is diarrhea, in part due to the changes in our gut microbiome and its role in digestion. Eventually, the gut microbiome will grow and return to the state it was before the antibiotic was taken. However, the repeated use of antibiotics, often in children, as well as the long-term use for some conditions may contribute to disease risk due to repeated or long-term hits on the microbiome and is an area under active investigation.

Prevention of Infectious Disease

Society has coexisted with infectious disease since the beginning of time. Whereas infectious disease was the top cause of death for centuries, with today's modern medicine, the rate of morbidity and mortality has substantially declined in many countries. Many antibacterial (antibiotics), antivirals, and antifungal medications have been developed in the twentieth century. However, without access to medications and other public health measures that can drastically reduce the transmission of infectious agents, infectious diseases remain a significant cause of death and disability with ripple effects across a population's economy. In addition, there are some diseases that are understudied and/or present more scientific challenges to the development of medications and therefore only a limited number of treatment options are available.

One of the major advances to combat infection and illness related to infection are vaccines. While general awareness has existed of the protections stored in the blood of those who survived an infection (and the ability to transfer the protective components of blood to others) for hundreds of years, the development and use of vaccines really did not begin to make great strides until the late 1800s. In 1879, the famous French scientist Louis Pasteur created the first laboratory-developed vaccine for chicken cholera. In 1914, the first vaccines were approved for typhoid and rabies. Since then, vaccines for several infectious diseases including smallpox, diphtheria, polio, anthrax, cholera, plague, typhoid, and tuberculosis have been developed. Some infectious diseases like HIV have posed some scientific challenges and a vaccine has not yet been developed. Today, new approaches based on genetics and genomics have led

to the development of new types of vaccine that enable quicker production and options to "tweak" the vaccine as the infectious agent evolves.

So, what exactly is a vaccine and how does it work? A vaccine is a way to safely allow our immune system (our body's defense system) to establish an immune response to a specific infectious agent in the event that you are exposed to it sometime in the future. First, a quick overview of the human immune system. Naturally, our body has two types of defenses as part of our immune "response." One line of defense is known as the innate immune system, which is a short-term quick response that triggers the deployment of special cells to seek and destroy infected cells.

The second line of defense is referred to as the adaptive immune system. It involves the establishment of antibodies (ant-ee-bod-eez). Antibodies are developed in response to a specific infectious agent, which will target the agent (tagging it) to be destroyed by other cells. Our bodies will actually retain a memory of past infectious agents for which an antibody response was created, sort of like a catalog, and it will reactivate production of those antibodies should another infection occur. But in the first instance of an infection, that long-term memory does not exist. A vaccine introduces a weakened version of the infectious agent (or parts of it) that does not cause illness to allow our immune system to "view" the infectious agent and to design antibodies. Our body does not realize that the "intruder" (in this case, the vaccine) is not a real infection. As a result, the vaccine enables the immune system to develop a memory for that specific infectious agent that will be triggered should the real version of the infectious agent attempt to infect. For some diseases, a single dose (one shot) will be enough to trigger a strong antibody response. The response may last for many years (as with a measles vaccine). In other cases, a two-shot vaccine is administered, with the second shot referred to as a "booster" to increase the antibody response. An annual vaccine may be needed because the infectious agent has changed enough due to accumulation of changes in its genetic code that make it "look" different to the immune system. Therefore, the antibodies generated to last year's vaccine may not work well enough to prevent infection by the new flu strain. This is the case with the annual flu vaccinations.

Conclusion

Undoubtedly, pathogens, new and existing, will continue to create new health challenges in animals and humans. In recent years, genomic technologies have been used to rapidly identify and monitor pathogens. Such technologies are

increasingly replacing conventional microbiological tests and enabling much more faster and accurate test results. Sequence data from hundreds or thousands of samples can be rapidly shared through scientific databases, used to track mutations and transmission from one region to another, and used to inform development of vaccines and treatments. As new molecular technologies continue to be developed, testing will be able to be performed more quickly, in more places (in a doctor's office, a drive-thru testing kiosk, or even at home), and enable quicker identification of cases and stemming transmission.

Resources

US Centers for Disease Control and Prevention. SARS Basics Fact Sheet. Available at https://www.cdc.gov/sars/about/fs-sars.html

US Food and Drug Administration. Center for Food Safety and Applied Nutrition. What We Do at CFSAN. Available at https://www.fda.gov/about-fda/center-food-safety-and-applied-nutrition-cfsan/what-we-do-cfsan

World Health Organization. Middle East respiratory syndrome coronavirus (MERS-CoV). Available at https://www.who.int/emergencies/mers-cov/en/

US Centers for Disease Control and Prevention. Biggest Threats and Data: 2019 AR Threats Report. Available at https://www.cdc.gov/drugresistance/biggest-threats.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fdrugresistance%2Fbiggest_threats.html

Harris et al. Whole-genome sequencing for analysis of an outbreak of methicillin-resistant *Staphylococcus aureus*: a descriptive study. Lancet Infectious Diseases. February 2013; 13(2): 130–36. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3556525/

Kahn JS, McIntosh K. History and Recent Advances in Coronavirus Discovery. The Pediatric Infectious Disease Journal. 2015; 24(1): S223-S227. Available at https://journals.lww.com/pidj/fulltext/2005/11001/history_and_recent_advances_in_coronavirus.12.aspx

Laupland KB, Valiquette L. The Changing Culture of the Microbiology Laboratory. Can J Infect Dis Med Microbiol. 2013 Autumn; 24(3): 125–128. 2013. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3852448/

Coral J., Zimmer C. How Coronavirus Mutates and Spreads. The New York Times, April 30, 2020. Available at https://www.nytimes.com/interactive/2020/04/30/science/coronavirus-mutations.html

US Centers for Disease Control and Prevention. Key Things to Know about COVID Vaccines. Available at https://www.cdc.gov/coronavirus/2019-ncov/vaccines/keythingstoknow.html?s_cid=10493:cdc%20covid%20vaccine:sem.ga:p:RG:GM:gen:PTN:FY21

The Johns Hopkins University School of Medicine. Coronavirus Resource Center. Available at https://coronavirus.jhu.edu/

Rastogi, M., Pandey, N., Shukla, A. *et al.* SARS coronavirus 2: from genome to infectome. *Respir Res* **21,** 318 (2020). Available at https://respiratory-research.biomedcentral.com/articles/10.1186/s12931-020-01581-z#citeas

11

Can Genes Explain Behavior?

We often notice similarities between relatives. Sometimes, it will be a physical trait like eye color or height, and other times, it will be a disease that runs in a family. Or we may notice similar behaviors between family members—maybe a father and son are both shy, love to play practical jokes, or maybe both are afraid of spiders or heights. As with many diseases, some people think that behaviors are inherited, in that people are biologically (or genetically) predisposed to act or behave in a certain way and thus is immutable or cannot be changed. But a person's environment, upbringing, and exposure to certain experiences may influence their behavior. In this case, a behavior may be acquired; for example, a son grew up watching his father's fear of climbing ladders and as a young adult, he also expressed fear of climbing ladders. Thus, there is an ongoing debate about whether behaviors are inherited (or biologically predetermined) vs. acquired (or learned), also known as nature vs. nurture.

The field of behavioral genetics has existed for a while, but you should not conclude from the name of this field that genetics is presumed to be the predominant factor in predicting behavior. Instead, one should think of it as a field of study that is trying to determine the degree to which genetics impacts behavior and how nongenetic factors may impact the role or contribution of those genes (and potentially change the outcomes). In addition to the complex interaction between genes and environment that may give rise to a certain behavior, there are so many types of behaviors to consider, each with a wide range of symptoms and/or severity; some considered medical and some nonmedical, and some with no standard measure of assessment. Needless to

S. B. Haga, *The Book of Genes and Genomes*, https://doi.org/10.1007/978-0-387-70916-1_11

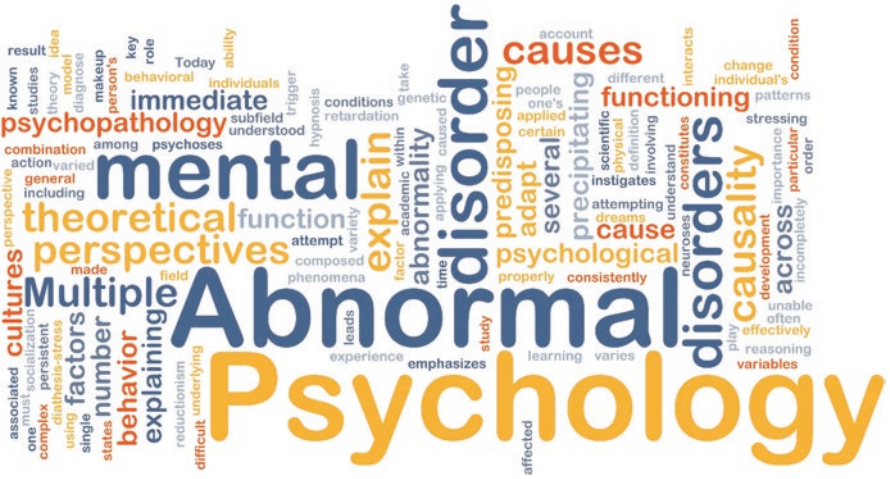

Fig. 11.1 Word cloud illustration of frequently used words including causes, symptoms, and diseases associated with "abnormal psychology" (source: Adobe Photo Stock)

say, it is complicated to study (Fig. 11.1). With new discoveries of the underlying causes of behavior, ethical, legal, and social concerns are raised, some different than those raised by other diseases. In this chapter, many of these issues are described.

What Is a Behavioral Trait?

Ask this question to ten different people, and you are likely to get ten different responses. Behavior is actually quite a large category, spanning a wide range of complex traits, including nonmedical and medical conditions and those that are intentional, conscious acts as well as subconscious or instinctual acts. Some behaviors may be characteristic for certain species (like bird migrations) or unique to each individual.

Let us consider some examples of behaviors. For instance, one's personality is a type of behavior—a person may be shy and quiet or outspoken and outgoing, or could be described as both depending on the setting. Voting preferences, intelligence, and sexual orientation are also considered behaviors. Emotions are considered behaviors—for example, some people may become angry very quickly and return to a calm demeanor just as quickly, while others may remain angry for a long period of time or even exhibit violent behaviors. Furthermore, anger may be expressed in different ways—a person may sulk and become very quiet and withdrawn, or yell and become aggressive.

In addition to personality traits, behaviors also include mental health disorders such as anxiety, attention-deficit disorder, depression, or personality disorders. While some personality traits such as shyness may be considered a medically recognized condition at the extreme (a person is terrified to leave their home out of fear they will encounter another person), they typically are not recognized as medical disorders. A behavior that is defined as a medical condition or disorder is diagnosed by a health professional based on the symptoms exhibited during an examination and testing, if available. Mental health behaviors that are recognized as medical disorders are listed in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), which is updated periodically to reflect new medical evidence and understanding.

Determining If a Behavioral Trait Is Genetic or Not

Unlike Mendelian genetic diseases that display a clear pattern of inheritance and are typically caused by genetic changes in a single gene, the underlying cause of complex diseases is obviously more complex and thus challenging to sort out. For complex traits and behaviors that are suspected to be influenced by genetics, scientists will first study families to calculate what is known as the heritability factor. The value of the heritability factor (abbreviated as "h") falls between 0 and 1—the higher the number, the higher the likelihood that the trait or behavior is influenced by genes and thus passed down from parents. The lower the h value, the more likely that the condition is influenced by environmental factors.

Twins are a natural experiment that have helped determine the value of "h" for many conditions. Identical twins have the exact same genetic makeup, while fraternal twins share 50% of their genetic makeup (like any pair of siblings). Thus, studying twins provides a controlled experiment—most twins are raised in the same household and thus their environment is the same. If genetics strongly drives a trait, we would expect to see the rate of similarity (called concordance) for identical twins to be greater than that for fraternal twins. On the other hand, if environment plays a role, we would expect to see about the same level of concordance for both types of twins.

In addition, studies of adopted children can provide some insight on the role of genetics. Adopted children share the same environment as their adoptive parents and 50% of their genetic makeup with their biological parents and siblings (some of whom may be growing up in different households). The behaviors of adoptive children can be compared to those of their adoptive parents and biological parents; if adoptees behave similarly to their biological parents/siblings, it can be concluded that genetics plays a strong role.

Scientific Challenges of Studying the Genetics of Behavior

As complex traits are impacted by genetic and environmental factors, only the weight of the impact of genes versus environmental factors can be indicated by the heritability factor. Different types of studies are needed to identify the specific genes or environmental factors that influence behavior. It is believed that behaviors are likely due to a combination or group of genes (the effect is called "polygenic"); each gene may account for a relatively small contribution to the overall behavior. Furthermore, it is not yet clear how multiple genes may contribute to the development of a certain behavior. Perhaps the effect of a variant in one key gene involved in a certain brain function is exacerbated by another variant in a key gene, and then a third and fourth variant. This "additive" effect can reach a threshold when the disease will begin to develop. In other cases, one gene may affect another gene for a nonadditive effect. Perhaps the disease will only develop if the person carries one or more genetic variants associated with the behavior and when the person is subject to certain environments (e.g., an abusive childhood)—an example of a gene(s)-environment interaction.

Further complicating matters, there is no straightforward way to measure or quantify many behaviors. Mental disorders and certain behavioral traits that are considered medical conditions are diagnosed based on a clinical psychological examination and some standardized testing, but it is not as definitive as diagnosing high cholesterol, for example, where a blood test can quantify the exact level. Think of a behavioral trait such as aloofness or an easygoing personality—how would you actually determine if a person "has" that trait? One cannot ascertain either trait unambiguously nor generate a score or value (e.g., high, moderate, low) without some ambiguity or subjectivity. This poses a huge problem when trying to study the causes of a certain behavior if scientists cannot even accurately determine who has it or not.

Examples of Genes and Behavior

As alluded to earlier, many people presume that there is a genetic component to behaviors. Why else would certain traits appear to run in families? Science is finally at a point technologically where a snapshot can be taken of the entire genome (meaning the entire genome can be accurately sequenced at an affordable cost and analyzed) of a large number of people. Thus, a study could be

performed that sequences the genomes of 1000 people with severe depression and 1000 people that have never had depression. The genomes, specifically the frequency of genetic variants, are compared between the two groups to determine if there is a statistical difference.

More often than not, however, initial reports of a gene associated with [fill in the blank with any behavior or disease] do not hold up. In other words, another research group will not be able to validate or replicate the initial finding in a different study population. Several reasons may contribute to the inability to replicate or confirm an initial finding.

Nonetheless, the search continues to identify the underlying causes of behavioral traits. Here are some well-studied, and sometimes controversial, traits and the current understanding of the genetic basis.

Schizophrenia

It has long been observed that cases of schizophrenia tend to run in families. Schizophrenia is characterized by confusion, paranoia, and hallucinations, but the onset, severity, and response to treatment can widely vary between patients affected with this disease. In some cases, the symptoms can adversely impact home life, support, and work setting, which may further exacerbate the disease. In other cases, the schizophrenic episodes can be controlled with medication and decline with age.

It has been speculated that some mental disorders and traits tend to pair together, and perhaps involve overlapping neurological pathways, and thus the same genes. This may not be too surprising as psychiatric disorders may involve pathways in the brain that may lead to different diseases depending on when and how they are disrupted in the individual. For example, some individuals affected with schizophrenia also show cognitive impairment. Some researchers have been investigating a potential link between proteins called glutamate receptors and cognitive impairment in patients with schizophrenia. Genetic variants in the genes encoding these receptors have been identified in some patient groups. The receptors play key roles in many brain functions including communication between brain cells, memory formation, and learning. They do so by serving as a docking station for neurotransmitters, or chemicals in the blood that can trigger a cascade of other reactions in the brain. The chemicals are signaling molecules, transmitting messages (and triggering specific actions) in the brain and nervous system. Some examples of neurotransmitters include dopamine, serotonin, and acetylcholine. The receptors could be good drug targets since they sit on the outside of the cells and

are easily accessible. But because these receptors are involved in so many brain functions, it is difficult to limit the effect of a drug to just one specific problem and therefore side effects may occur. Alternatively, other drugs have been developed to try to control the amount of neurotransmitters.

Because there is likely to be more than one path to develop schizophrenia, the combination of genetic variants and the level of exposure to environmental factors remain to be untangled. For schizophrenia, studying pre-disease symptoms that first develop before the full disease spectrum is obviously one approach to reducing the complexity. Known as schizotypal traits, these include moderate psychotic symptoms, social withdrawal, and reduced cognitive capacity. Some of these traits may affect relationships and daily tasks and interactions. These traits are not unique to schizophrenia and could also be indicative of depression, bipolar, and other psychotic disorders. While several genes involved in the glutamate pathways, regulation of dopamine, and brain cell development have been reported to be associated with schizotypal traits, much work remains to confirm their impact on brain functions and development of schizophrenia.

Intelligence

The controversial debate about the heritability of intelligence has fluctuated over the past century or so. Dating back to the late 1800s, British scientist Francis Galton set out to determine if intelligence (which he referred to as "human mental ability") is a heritable trait by studying successful men and their families. He observed that parents of high societal class and reputable occupations (e.g., judge) tended to have children that grew up to have similarly successful careers for themselves. Thus, he concluded that intelligence was passed from successful parents to their children. He also observed that intelligence was a quantitative trait that ranged from very low to high. However, evidence of individuals of high intellect from less than prosperous backgrounds countered Galton's theory of inheritance for intelligence and indicated the importance of access to an education and other environmental factors.

The work of Galton and others gave rise to the eugenics movement in the early twentieth century in Europe and the USA. Efforts were made to promote "selective breeding" (or positive eugenics) between successful parents to produce more children with socially desirable traits. Also, during the early twentieth century, the intelligence quotient (IQ) measure was developed by a German psychologist named Wilhelm Stern. The IQ test aimed to calculate

the ratio of mental ability to age. Such tests were of increasing interest in the USA at the time as a way to assign World War I recruits to roles suited to their aptitude.

With increasing use of the IQ tests in various settings in addition to the military, some evidence suggested that the test was biased and therefore not a fair assessment. There was evidence of cultural biases and that the merging of basic aptitude and learned skills into a single score may be misleading. In addition, broad tests combined different areas such as language (e.g., vocabulary), logic, and mathematics. This led to changes in the scoring of the IQ test and development of new intelligence assessments. Interestingly, scores have increased throughout the past several decades—a rise too rapid to be due to changes in the genetic makeup of populations. Thus, changes in the environment—schools, instructional approaches, and technology—any or all of these things could have contributed to the increasing IQ scores observed over this short period of time.

Twin, adoption, and family studies are suggestive of a genetic underpinning of intelligence. Comparisons of identical and fraternal twins show rates of concordance of about 0.85 and 0.60, respectively, pointing toward genetics (over environment). Oddly enough, heritability values for intelligence appear to increase with age, a phenomenon not seen with other traits. If the genes are stable throughout life, how can heritability of intelligence change over time? The explanation is called "genetic amplification," whereby the effects of genes associated with intelligence increase. The genes themselves do not change over time, but the environment does and likely increases the genetic effect. In other words, the environmental effects are mediated by one's genetic makeup.

Both family-based and large population studies have reported many genes associated with intelligence or a related trait such as memory, though, almost certainly, a good number are spurious findings. A 2018 study examined DNA data that had been collected from more than 269,000 people around the world for different studies, and identified 1016 genes, most of which were known to be involved in brain function. Another line of research has approached the search through an evolutionary lens, specifically looking at whether genes linked to intelligence have been changed over time. With the physical changes in human history with respect to the shape of the skull and brain volume, perhaps some of the genes evolved in parallel that may account for traits such as intelligence, memory, reasoning, and empathy not present in lower order species. Other types of research are investigating the impact of genetic changes on brain cell function and signaling that can provide more insight regarding how genes operate and the effect of variants.

Aggression

Another challenging area of study in genetics research is aggressive behavior. Aggression is broadly defined as a behavior intended to harm another individual, oneself, or the environment. Several types of aggression have been identified, with the two major types known as *proactive* and *reactive*. Proactive aggression refers to an act that is intentional or planned (an offensive act). In contrast, a reactive form of aggression is based on impulse and emotion (unplanned), often in response to a perceived threat (a defensive act). Furthermore, aggression may be expressed in different ways—an individual may be verbally aggressive, physically aggressive, or hostile. Aggressive behaviors also appear as a symptom for many common conditions such as attention-deficit hyperactivity disorder (ADHD) and mental health disorders such as schizophrenia. Thus, it is presumed that these conditions have shared causes for the overlapping symptoms.

As with other behaviors, there is evidence that aggressive behavior may run in some families. For example, there are families with multiple members who have been convicted of violent crimes. Thus, the question arises if this common behavior is due to shared environment or genetics. As with large population-based studies on intelligence, advanced genomic technology and software enable the rapid sequencing and analysis of genetic changes that appear in one group (those with aggressive behavior) and not in a comparison group (those without aggressive behavior), scientists can conduct these studies to identify potential genes involved in regulation of these specific behaviors. In 2018, a European study identified 40 genes associated with aggressive behavior.

Many of the genes associated with aggressive behavior are involved in controlling cell-to-cell signaling (or messages sent between cells to signal a particular action) within the brain. Other genes linked to aggression and violence are involved in the transport of the neurotransmitter dopamine or binding (dopamine receptor) on brain cells.

One of the long-studied genes linked to aggressive behavior is the MAO-A gene, which stands for monoamine oxidase A. It is involved in the breakdown of key neurotransmitters like serotonin, norepinephrine, and dopamine in the brain. These chemicals are involved in regulating several emotions including mood, reward, impulse control, and response to extreme situations (fight or flight response). The genetic variant causes the enzyme encoded by the MAO-A gene to function below normal levels, resulting in higher levels of serotonin, dopamine, and norepinephrine. These individuals may be more

hypersensitive and experience distress or anxiety in more profound ways than those who do not carry the genetic variant. The gene is located on the X-chromosome, so the effect of a genetic change will be more common in males than females.

In 1993, a study of a violent Dutch family identified the MAO-A genetic variant in each of the violent members of the large family. Other studies reported the association between the MAO-A gene variant and antisocial behaviors in school-age boys as well as with aggressive and violent behaviors. However, it was revealed that the effect of the MAO-A genetic change was more likely to appear in children that were abused (verbally or physically). Thus, the effect of the genetic variant may not appear, or not to the extreme, unless triggered by trauma and abuse.

Sort of in Our Genes, But Not Exactly

While this book focuses primarily on changes in the DNA sequence and their association with disease or various inherited traits (in humans and other species), there is another layer of genetic complexity not yet mentioned. Perhaps one could view it as an overlooked sibling, operating in the shadows of her popular and brilliant big sister, but who is stealthily inching toward the limelight. Without changing the DNA sequence, our cells have another mechanism with which to control when genes are turned on and off. This light switch is actually a reversible and mobile one, appearing when the cell sends a signal that is in need of a change in the presence or absence of a gene. The light switch is actually a chemical modification or tag—a small chemical element that attaches to the DNA sequence (which is actually another chemical molecule) but does not alter the code. Known as epigenetic modifications, these chemical tags can block or open a stretch of DNA to the machinery required to turn it on or off.

One common type of chemical modification is called DNA methylation (meth-ill-lay-shun). A methylated sequence of DNA can block other molecules from attaching to the DNA and initiating expression (turning on the gene to make protein). There are specific proteins (enzymes) in the cell that are responsible for adding and for removing the methylation tags to specific DNA sequences.

Epigenetic modifications have been linked to diseases and other traits, including behavior. Epigenetics has been studied in connection with several psychiatric conditions such as schizophrenia and mood disorders as well as neurological diseases such as Alzheimer's. Several genes expressed in brain cells

or involved in the development of the brain plaques characteristic of patients with Alzheimer's have been shown to have different methylation patterns. If such changes in methylation could be reversed through medications, perhaps brain cell function could be restored or, at least, deterioration slowed. A handful of drugs have been approved to treat various cancers by changing the methylation patterns of genes associated with cancer growth. For example, the drug 5-aza-cytidine is used to treat myelodysplastic syndrome and works by removing methyl tags.

Many environmental factors can influence epigenetic modifications such as diet and age. One interesting example is the impact of prenatal stress and disease risk—children born shortly after times of hardship and environment stress, such as during periods of famine during the World War II in Europe, have been found to have increased prevalence of schizophrenia in adulthood, potentially due to alterations in genetic modification acquired before birth that adversely impact brain development. Hormones triggered during stressful events in the mother can impact the DNA of the fetus and cause changes in gene expression through epigenetic modifications.

Risks and Benefits of Defining Genes Associated with Behaviors

In addition to both scientific and medical challenges to identifying the causes of certain behaviors, the field of behavioral genetics has raised some difficult ethical, legal, and social questions. Benefits of increased understanding of brain function and proteins involved in regulating behaviors may lead to new medications and treatments. Furthermore, more tests to improve the diagnosis of medical disorders, potentially before they appear, could lead to substantial savings of healthcare costs through prevention, monitoring, and early intervention.

However, for those behaviors or traits that are nonmedical, such as bullying or intelligence, what is the societal benefit of learning of the genetic factors that contribute to these behaviors? For example, what if we understood the factors that contributed to a high IQ score and a company developed a test for intelligence. Who would use that genetic test and for what purpose? Could school admission offices pre-screen applicants to determine if they were genuinely intelligent (defined by the genetic intelligence test) to distinguish those who are genuinely smart to those who have a photographic memory and can

score well on tests? In other cases, the discovery of a biological cause may excuse or relieve an individual of personal responsibility (culpability) of their actions and therefore they may argue that they cannot be held accountable for their actions (e.g., the "my genes made me do it" defense).

But the benefits of the research should be weighed against the risks, not only to the individual participant, but also to their family members, community, and the group that they identify with at-large. Groups can be defined any number of ways—by religion, gender, geographic locale, race or ethnicity, occupation, etc. For some behaviors like kindness or perfect pitch, perhaps the potential harms are perceived to be low and do not outweigh the benefit of advancing science and unraveling the complex neurological controls of a given behavior and the role of the environment. For other behaviors like gambling, bullying, child abuse, or alcoholism, which carry a heavy social stigma, the risks of genetic research on these traits should be carefully considered.

For example, a study of Maori men (the indigenous Polynesian people of mainland New Zealand) revealed a high prevalence of the aggressive form of the MAO-A gene. After the research findings were published, it was dubbed the "warrior" gene. This was not viewed favorably by the Maori community and within their culture. While some could claim this explained the Maori's strength and resilience to voyage across the Pacific in the fourteenth century and eventually settle in New Zealand, others may highlight this finding as an explanation (or blame) for aggressive and violent behavior of some Maori men. The potential social stigmatization may not have been considered at the outset of the study; careful discussion with community leaders, sociologists, and anthropologists may have helped to avoid the stigma or harm by individuals and communities who participate in research.

Every behavior will likely have unique circumstances surrounding the impact for potential for benefit and harm. In part, the societal views of some behaviors will determine whether the research is even performed (and who pays for it) and how the results are accepted and utilized. Furthermore, different cultural and religious beliefs and values will influence the perceived benefits and risks. Therefore, there is not a one-size-fits-all solution to how this research can be ethically conducted and the results used in a manner to optimize health and well-being while minimizing the likelihood of harm to individuals, families, or groups.

Conclusion

The complexity of human behaviors is being tackled with new technologies to gain insight into the biological mechanisms at play. However, unlike some other diseases and traits, the stigma associated with some behaviors presents new issues that should be considered in parallel (or before) scientific research begins. As human behavior traverses a spectrum from the normal to abnormal, moderate to extreme, to socially acceptable to non-acceptable, additional stigma associated with defining biological causes may outweigh the benefits. The implications of any new research findings cannot always be foreseen or anticipated, but that does not mean careful consideration should not take place before research commences. Behavioral genetics is an exciting area of study, but requires more deliberate discussions at many levels. In this case, expertise of other fields such as law, religion, and sociology may greatly help prepare for the ethical study of behavior and how the results will be used toward the betterment of society.

Resources

Plomin R, et al. Top 10 Replicated Findings from Behavioral Genetics. Perspect Psychol Sci. 2016 Jan; 11(1): 3–23. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4739500/

Plomin R, and von Stumm S. The new genetics of intelligence. Nat Rev Genet. 2018 Mar; 19(3): 148–159. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5985927/

American Association for the Advancement of Science. Behavioral Genetics. 2004. Washington DC. Available at https://www.aaas.org/resources/behavioral-genetics/chapters

Garcia-Arocena, D. The Genetics of Violent Behavior. 2015. Available at https://www.jax.org/news-and-insights/jax-blog/2015/december/the-genetics-of-violent-behavior

Tuvblad, C. and Baker, LA. Human Aggression Across the Lifespan: Genetic Propensities and Environmental Moderators. Adv Genet. 2011; 75: 171–214. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3696520/

Godar SC et al. The role of monoamine oxidase A in aggression: current translational developments and future challenges. Available at Prog Neuropsychopharmacol Biol Psychiatry. 2016 Aug 1; 69: 90–100.

Richardson, S. A Violence in the Blood. *Discover Magazine*. September 30, 1993. Available at https://www.discovermagazine.com/the-sciences/a-violence-in-the-blood

What is Epigenetics? Available at https://www.whatisepigenetics.com/fundamentals/

12

I Will Have the Genetically Modified Foods, Please

When shopping for vegetables at a produce stand or a grocery store, have you ever stopped to wonder how they came to be? This is not being philosophical or talking about where or how it was grown—locale, hothouse, organic, or otherwise. Even before all of these things are decided, the farmer must first decide which seed to buy. The choice of which strain of a particular fruit or vegetable to grow will be influenced by several practical factors including customer demand, labor, time, cost, need for pesticides, soil requirements, etc. Farmers also have the choice of buying a genetically modified (GM) seed or not. What exactly are GM foods?

What Is GM?

As described in the first chapter, all organisms have genes including fruits, grains and vegetables. Just as Mendel studied the inheritance of certain traits in peas, each vegetable has certain characteristics that are genetically controlled and are passed on from generation to generation. Some of these traits are of particular interest to farmers (e.g., ability to grow under certain conditions, reduced susceptibility to disease, hardiness) and consumers (e.g., taste, appearance, nutritional content). Thus, crops may be genetically modified for many different purposes to improve yield, taste, quantity, or appearance. In addition, some crops may be modified to produce vitamins or drugs. Livestock may undergo genetic modification as well to improve meat quality, nutritional value, and quantity, or as a new source of organs for transplantation.

S. B. Haga, *The Book of Genes and Genomes*, https://doi.org/10.1007/978-0-387-70916-1_12

Traditional or conventional breeding techniques depend on cross-fertilization to create a plant with a desired trait (Fig. 12.1). Although only one trait may be sought after in a particular breeding experiment, it is not possible to target the gene or genes responsible for that trait and to enhance it (e.g., make strawberries bigger) through traditional breeding techniques. So, all of the genes between two plants with the desired trait are naturally combined to hopefully yield a new generation with that desired trait, but the process is entirely random. As a result, this process may take many plant generations (years) to achieve the desired trait, and unwanted traits may arise along the way.

The genomes of many plant and livestock species have been sequenced, enabling more detailed analysis of individual genes and their functions. Scientists can move or "transfer" a gene or genes from one species to another to achieve the desired trait. For example, if one desires to create a plant (or animal) with a

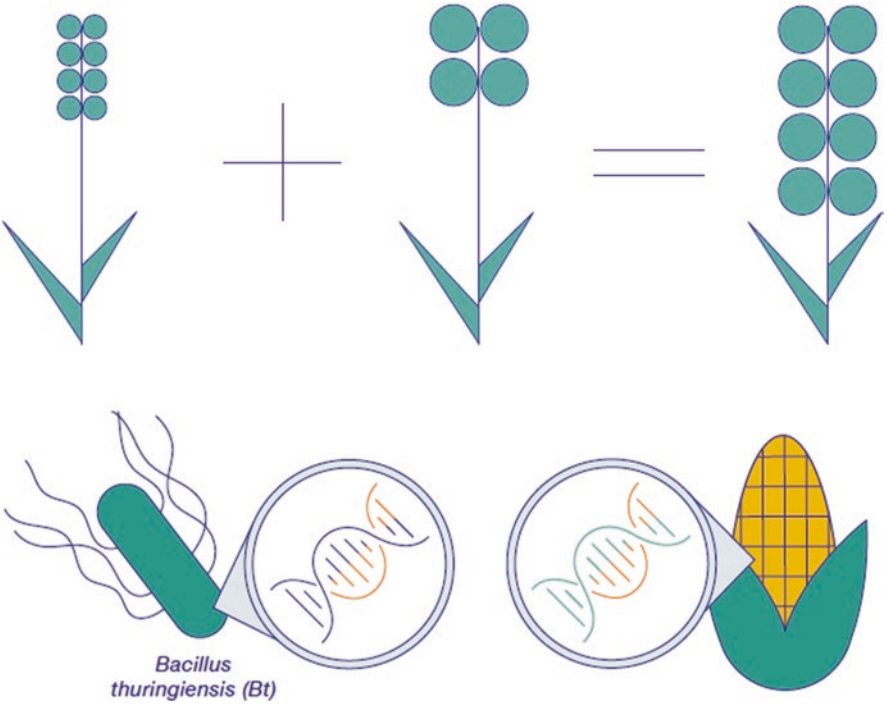

Fig. 12.1 The top illustration depicts a traditional cross-fertilization between two plants with different traits to yield a hybrid. The bottom illustration shows the process of genetic engineering to insert a new gene into the DNA of corn (source: US Food and Drug Administration; https://www.fda.gov/food/agricultural-biotechnology/types-genetic-modification-methods-crops)

trait it does not naturally have, a gene from another organism can be transferred and inserted into the genome of that species. Yet another approach to genetic modification is to alter (either increase or reduce) the expression a particular gene that controls a particular trait. For example, if strawberry size is the desired trait and scientists know that gene A within the strawberry genome is responsible for growth, they can modify that gene to increase the amount of that gene produced. In this hypothetical example, the more of gene A that is produced, the larger the strawberry will grow. In other situations, it may be beneficial to actually reduce or "knock out" the expression of a certain gene altogether.

Alternatively, new techniques have been developed called gene or genome editing as described in Chap. 9, enabling scientists to make very precise changes (gene modification) to alter specific genes linked to particular traits of interest. Thus, the combination of more genetic knowledge about plants and livestock and new techniques for modifying genes enables the development of new seeds or offspring with a specific desired trait very rapidly.

The main difference between traditional breeding and genetic modification is the ability to alter discrete gene(s) and the degree of alteration. However, similar to traditional breeding processes, changing the genetic makeup of a plant or livestock, even of a single gene, could result in unexpected consequences since the balance of genes (and their corresponding proteins) is being changed. In the other example where a new gene is introduced, the species' genome may not have the ability to control when the new gene product is turned on and off, and thus, the growth or characteristics of that plant or livestock may not be what was intended.

Genetically modified (GM), genetically modified organisms (GMO), and genetically engineered (GE) are terms generally used synonymously to refer to the use of genetic technologies to alter the natural genetic makeup of plants and livestock. In addition to foods, these same technologies can be applied for other purposes such as environmental sciences and pest control. The field has generated some controversy as will be discussed at the end of this chapter. In particular, genetic technology to create new or enhanced traits in foods may cause unforeseen risks to humans and/or the environment.

How Many GM Crops Are There?

In 1990, the first GM food product, chymosin (ki-mo-sin), was approved by the FDA for the production of cheese. Chymosin is the active ingredient of rennet, an enzyme complex used to curdle milk. Prior to GM chymosin, the enzyme was obtained from calf stomach. GM chymosin was created in order

to increase the availability needed for cheese production—although produced in bacteria, it is no different than natural chymosin. Today, 80–90% of cheese produced in the USA is curdled with GM chymosin.

The first GM crop was approved and planted in 1996. As of 2018, almost 500 million acres (or almost 200 million hectares) of GM crops are grown in 26 countries, with the USA, Brazil, and Argentina growing the most GM crops. Most of today's GM crops that are commercially grown are pesticide resistant or insect resistant and much is not for human consumption (used for animal feed, for example). The four most commonly grown GM crops are soybean, maize, cotton, and canola. According to the USDA's Economic Research Service, in the USA in 2000, 25% of corn was GM; in 2019, 92% of corn was GM. Similarly, most (98%) of the cotton grown in the USA is GM.

Examples of GM Crops and Livestock

There are a wide range of GM crops and livestock that have been developed for a multitude of reasons. While the public may associate the use of GM technology to enhance foods for desirable features (e.g., taste, size, appearance) or to reduce the use of pesticide use, a wide range of applications are under development. For example, food availability and accessibility are of concern with the growing global population. Researchers are studying the use of GM technologies to increase yield, plant hardiness, fruit size, etc. to meet anticipated food shortages. Many GM plants are not produced for direct human consumption, but instead, indirectly through products like cornstarch, corn syrup, or cooking oil. Not all have reached the market, but these examples highlight the breadth of possibilities.

Enhanced Taste, Appearance, and Size

Flavr Savr Tomato

One of the first GM products created but no longer in circulation was the Flavr Savr tomato. Traditionally, tomatoes are harvested when still green and firm and ripen off the vine following chemical treatment to reduce crushing and rotting during shipment. The Flavr Savr tomato was developed to alter the natural ripening process, so that it ripens on the vine but remains firm after picking and shipment to markets. Instead of adding a new gene into the tomato genome,

the California biotechnology company Calgene modified a tomato strain to produce substantially less or none of a specific gene involved in the natural ripening process. Without the gene product, the ripening process is halted.

In 1994, the Flavr Savr tomato was approved by the FDA. The GM tomatoes were initially sold under the name brand MacGregor's tomatoes and labeled as "grown from Flavr Savr seeds," but the label did not explicitly state that they were genetically modified. The consumer response was initially positive but declined rapidly. In less than a year, MacGregor halted sale of the Flavr Savr tomato. A combination of factors likely contributed to its disappearance from the US market including the fact that it was more expensive than non-GM tomatoes and they were not found to be all that "flavorful," and increasing public awareness about the uncertain health consequences of GM foods.

In 1996, Safeway and Sainsbury stores agreed to sell Flavr Savr-based tomato puree in the UK. In contrast to the US label, the UK product clearly stated that the tomato puree was derived from "genetically modified" tomatoes and was priced cheaper than the non-GM competitor. Similar to the US experience, sales were initially encouraging but gradually declined and the product was withdrawn from store shelves by 1999.

AquAdvantage Salmon

In 2015, the FDA approved an application submitted by AquAdvantage Salmon. This GM Atlantic salmon contains a gene that makes a growth hormone (from the related Chinook salmon), but scientists inserted another piece of DNA derived from a different fish (the ocean pout) to control when the gene is turned on. With increased growth hormone, the AquAdvantage salmon can grow to full size faster. The GM salmon is only allowed to be grown at fish farms with multiple forms of physical and biological environmental containment measures in place, and therefore will not be introduced into the wild or pose a threat to wild populations.

GalSafe Pigs

In 2020, the FDA approved an application for a genetically modified pig that does not produce a molecule called alpha-gal sugar. This molecule is attached to the outside of pig cells and individuals with alpha-gal syndrome or other conditions cannot consume pork and other meats due to an allergic response

to alpha-gal sugar. In addition to providing a safe food source, GalSafe pigs could also be used as a source of medicines such as the blood-thinning drug heparin. By removing the alpha-gal allergen, these medical products would be safe to use in patients with alpha-gal syndrome. Any products that contain porcine components (or other animal-based components), such as some types of makeup, can also cause allergic reactions due to alpha-gal, and thus, the GalSafe pig can help reduce those adverse responses.

Insect Resistance

According to the US Environmental Protection Agency, almost $9 billion in the USA was spent on pesticides (includes herbicides, insecticides, fungicides, and other pesticides) in 2012. Farmers may spend 4–5% of their total costs on pesticides. Reducing pesticide use can benefit many groups—farmers that have to spend extra money to save their crops, the supply chain and consumers to whom costs will be passed along, and the environment. For some crops, multiple pesticide applications are required.

Bt Corn

The European and Southwestern corn borer is a particular nuisance to farmers, causing more than a billion dollars in losses due to damage and control costs each year. Scientists discovered that a bacterium found in soil, *Bacillus thuringiensis* (Bt), produces a family of proteins that are toxic to the larva (caterpillar stage) of the European and Southwestern corn borer. The larva, not adult corn borer, actually causes the most destruction to crops. For the past several decades, sprays and powders of Bt have been used successfully against European corn borer and other insect pests on a variety of crops since the 1960s. Multiple applications may be necessary as the toxins in the insecticide break down rapidly in sunlight or are washed away by rain.

Upon ingestion of this endotoxin, digestive enzymes activate the toxic form of the protein. The toxin causes holes in the gut lining, causing massive fatal infection in the bloodstream or starvation. It poses no risk to animals or humans since the proteins required to convert the endotoxin to its active (harmful) form are not present. In addition, it has not been shown to impact other insects such as beetles, flies, bees, and wasps. However, farmers may be reluctant to use insecticides due to the high costs and environmental concerns.

The gene that encodes for the Bt endotoxin protein was eventually discovered. Thus, instead of pesticides, scientists created GM strains of corn that contained the Bt endotoxin gene. In 1996, Bt corn was introduced. In comparison to insecticides, GM Bt corn can control up to 96% of the corn borer larvae whereas an insecticide controls approximately 67–80%. Today, in addition to Bt corn, Bt potatoes and Bt sweet corn are also grown. Different strains of Bt corn are available that target other pests such as rootworm.

Herbicide Tolerance

Herbicides are commonly used by farmers to control weeds, which can substantially affect crop yield and quality through competition for water, soil nutrients, and sunlight. Herbicides are costly and some are harmful to the environment. Of the $9 billion spent on pesticides (including herbicides) in the USA in 2012, 58% ($5 billion) was specifically spent on herbicides. The chemical glyphosate is a commonly used herbicide and shown to be less environmentally harmful. It works by deactivating proteins in plants to kill them. As most plants contain these proteins, it does not discriminate between "good" plants like crops and "bad" plants (weeds). Thus, the type of herbicide is known as a "broad-spectrum" killer. One popular brand is called Roundup™, marketed by the company Monsanto and used by farmers and homeowners alike. With the advent of genetic modification, Monsanto developed several GM crops that would be resistant to their Roundup™ herbicide. Specifically, a bacterial gene known to confer resistance to glyphosate was introduced into soybean allowing the herbicide to be safely applied to soybean fields.

The use of GM herbicide-tolerant crops reduces the use of costly and environmentally harmful pesticides as well as reduced tilling to disrupt weed growth. Today, there are several "Roundup Ready" crops including soybeans, corn, canola, alfalfa, cotton, and sorghum.

Enriched and Drug-Producing Plants

Pharmaceutical crops are GM crops engineered to produce drugs or vaccines for human use. As a number of biotechnology drugs such as insulin are produced in bacteria and other organisms, the idea of producing drugs in crops is not that far-fetched. As with other GM crops, "pharma crops" or plant-made pharmaceuticals are mired in debate and the FDA has not approved any drugs or other substances produced in pharma crops intended for use as

pharmaceuticals. The idea behind the concept is that drugs can be produced more cheaply and stored long-term in the seed or grains until needed. Pharma crops can be grown in the same regions where their food counterparts are grown, often by the same farmers. Rice, corn, barley, tobacco, and safflower are just some of the crops that have been genetically modified to produce drugs such as vaccines, human antibodies, and human blood proteins. Thus, one concern that has been raised is the potential contamination of food crops and threatening of the food supply.

In addition, several livestock are susceptible to disease, raising the possibility of modifying animal genomes to reduce infection. For example, African swine fever is a viral disease that has caused major problems to pigs in some African countries. Some pigs are naturally resistant to infection due to a single gene. Thus, scientists have taken this gene from resistant pigs and introduced it into swine livestock to create a resistant breed.

Golden Rice

Rice is a staple crop for half of the world's population and rice production has been steadily increasing. In 2003, 395 million tons of milled rice was produced, making rice the second largest produced grain in the world. In 2018, almost 500 million tons of milled rice was produced. Western and Eastern Asia produce the majority of the world's rice, with China and India alone accounting for more than 50% of total rice production (China and India also consume the highest amounts).

As rice is the major source of calories for many people, particularly those residing in lower income countries, the nutritional value of rice is an important consideration to the health of many populations. Rice is high in complex carbohydrates and a good source of many vitamins and minerals; however, much of the nutritional value is lost through the milling process. As a result, the US law requires all white rice to be enriched (or fortified) with vitamins B1 and B3 and iron.

The absence of vitamin A in rice has contributed to vitamin A deficiency in many parts of the world, leading to millions of cases of severe eye problems including permanent blindness and one to two million deaths annually. In the early 1990s, European scientists began work to genetically engineer a common rice strain to produce a provitamin A nutrient known as beta-carotene. To do this, they inserted three genes—two from the Daffodil genome and one from a bacterial genome. The addition of these genes resulted in the production of lycopene (a vitamin rich in tomatoes) which the rice's natural enzymes convert

to beta-carotene (which gives the rice a golden hue). In the late 1990s, "Golden Rice" was announced to the world and hailed as a major breakthrough for biotechnology. Since then, scientists have added a fourth gene, this time from the maize genome to the Golden Rice strain to further increase the production of beta-carotene levels closer to the Recommended Daily Allowance.

Golden Rice represents an example of bio-fortification compared to current practices of fortification of plant and dairy products. Other varieties are under development including Golden wheat, Golden cotton, and Golden potato.

Since 2017, several countries have been approving the production of Golden Rice including the Philippines, Australia, Canada, New Zealand, and the USA. This represents a major step forward and may finally lead to farmers' planting Golden seeds.

Reduced Nicotine Tobacco

It is well known that the chemical nicotine within tobacco products promotes addiction and dependency. But what if there was a way in which nicotine levels could be lowered to the point at which they were not bad for humans? Genetic modification of the genes that are involved in nicotine production could yield tobacco with reduced levels of nicotine.

Scientists at a company called Vector Tobacco in North Carolina did just that by developing a GM tobacco strain that produced substantially decreased levels of nicotine. Following the same approach to make the Flavr Savr tomato, scientists were able to reduce the levels of a gene important to the production of nicotine and related chemicals. Known as Vector 21–41, this GM tobacco strain was tested in several states including Hawaii, Illinois, Mississippi, and Louisiana in the late 1990s. The USDA determined that it met the criteria to be reclassified as a "nonregulated" GM crop in 2002. In 2003, Vector Tobacco manufactured a low nicotine and nicotine-free cigarette brand called Quest. Available in three nicotine levels, Quest products were intended to help smokers gradually reach nicotine-free smoking. This product was discontinued in 2010, but research continues to explore GM tobacco.

Vaccines

Infectious disease continues to be a major cause of morbidity in many low-income countries. Vaccines could substantially reduce the morbidity and mortality associated with these diseases. However, costs of development, production, and delivery have presented major challenges to getting these drugs

to the people who need them most. The idea of an "edible vaccine" was conceived in order to easily produce and distribute vaccines to affected populations through staple food crops. For example, bananas have been genetically modified to contain a vaccine against *E. coli* and *Vibrio cholera*-causing diarrhea. GM tomatoes have also been developed to produce an edible vaccine against hepatitis B virus. Animal studies have found that GM potatoes and alfalfa are effective in creating an immune response against diarrhea and cholera.

Genetically Modified Enzymes for Food Production

Food manufacturing requires many ingredients and steps to generate the final product. Many food products are made from an extensive number of ingredients (just read the label of virtually any packaged good!), including natural ingredients and others which are purchased or can be made through a series of steps. For some things, such as cheese, bread, and beer and wine, enzymes are required to change some of the natural ingredients. For example, we purchase baker's yeast to bake bread or cinnamon rolls at home from flour, milk, eggs, and other natural ingredients. Likewise, but on a much larger scale, food manufacturers also need enzymes to produce certain products, though they require much larger quantities than we do in our kitchens. Thus, industrial enzyme manufacturers mass-produce these enzymes and sell them to food manufacturers as well as to paper, textile, and pharmaceutical industries.

Like natural ingredients, there may be different versions of a given enzyme. While each version performs the same chemical reaction, they may operate just a bit differently. For example, an enzyme called glucose isomerase (i-som-er-ace) converts or changes the common form of sugar (glucose) into another form (fructose) with a slightly sweeter taste that is the key ingredient of high-fructose corn syrup. High-fructose corn syrup is a sweetener used in many products such as soda, candy, canned fruits, and juices. Bacteria naturally have a gene for the glucose isomerase enzyme in order to break down cellulose as a food source. But different bacterial strains have slightly different versions of the glucose isomerase gene. Some bacteria called "thermophiles" reside in very warm environments and can survive at higher temperatures (which means their enzymes can function at higher temperatures, a necessary condition in some food manufacturing processes). Thus, scientists have studied the characteristics of glucose isomerase genes from thermophiles and other bacterial strains to find more efficient and stable versions of the enzyme to use in different conditions for food manufacturing. To evaluate a given version,

scientists can study how the enzyme operates under different conditions. Once a good strain is identified, the enzyme can be mass-produced from its native bacterial strain or the gene can be inserted into another type of bacterium (such as *E. coli*) that can be grown easily. The enzyme is then extracted from the bacteria, purified, and sold to food manufacturers.

In some cases, genes encoding glucose isomerase or other enzymes can be modified in order to improve its efficiency, stability, or other features. DNA sequences can be added or modified to change the characteristics of the enzyme to perform better under certain food processing conditions. So, although a food manufacturer may use all "natural" ingredients, they may use GM enzymes to create the final product.

Regulatory Oversight of GM Foods

In the USA, three federal agencies oversee the approval and use of GM plants, animals, and products: the Food and Drug Administration (FDA), the US Department of Agriculture (USDA), and the Environmental Protection Agency (EPA). Each agency has a specific function: the FDA oversees safety and labeling of all GM plant-derived foods and feeds; the USDA's Animal and Plant Health Inspection Service monitors the field trials of new GM crops; and the EPA oversees crops that produce pesticides. GMOs are under stricter regulation in the European Union than in the USA. The FDA's review and approval process of GMOs is risk based. The GM plant or animal must be demonstrated safe for humans and animals to consume, and the genetic modification must be safe for the animal and should pose no or limited risk to the environment. In regulatory terms, the new product must be considered "substantially equivalent" or very similar to the natural food or comparator. Safety assessments consider the physical and/or biological containment measures developed to minimize the environmental impact. Physical containment measures may include restricted places where GM plants or animals can be grown (e.g., fish farm). Biological containment refers to changes made to the organism that would inhibit reproduction and spread of the GM plant or animal into the wild population such as plant seeds or animals are sterile.

The regulations continue to be revised as new evidence emerges and technologies are developed. In 2018, the USDA announced that it would no longer regulate plants with small genetic modifications to its native genes (the plant's DNA) or introduction of genes from related plant species.

Product Labeling

In the USA, initially special labeling of GM foods was not generally required since GM foods are not considered to be materially different from their traditional counterparts. Over the years, however, with growing public awareness, labeling policies have changed. In 2016, the National Bioengineered Food Disclosure Standard legislation was passed, requiring that human food containing GE be labeled to indicate that it is bioengineered. The USDA was charged with implementing a mandatory standard for disclosing whether a food is "bioengineered." In 2018, the USDA issued a regulation implementing that law. In the European Union, all products containing more than 0.9% of an approved GMO must also be labeled. Some manufacturers have chosen to label their products as "GMO-free" (Fig. 12.2).

Public Opinion and Debate over GM Foods

As noted, GM technologies have been used for lots of different purposes in addition to the focus of this chapter on plants and livestock. In Chap. 6, the use of GM technologies in the early 1980s enabled production of insulin, the much-needed treatment for diabetics (recall that prior to the production of insulin through biotechnology, it was extracted from animals for medical use). While there was some early opposition to the use of biotechnologies in the 1970s when first introduced (prior to successful development of human insulin), many of the concerns appeared to dissipate with the absence of reported harms. Similarly, when the first GMOs were approved for sale, some sectors of the public were vehemently opposed. While acceptance appears to be increasing, perhaps due to the limited evidence of harm, some groups though remain opposed today.

Fig. 12.2 Examples of some labels used on products to indicate that they do not include GMO ingredients (source: Adobe Photo Stock)

The portrayal of GMOs as "Frankenfoods" in the early 1990s presented GMOs to the public as very negative, unnatural, scary, and the creation of reckless scientists. Before the public had any real understanding about the science, potential benefits, and evidence of harm of GMOs, the imagery and language regarding GMOs triggered public fears and led to a backlash in some countries. Organizations opposed to GMOs and anti-GMO activists have organized protests, burned and destroyed fields of GM crops, and lobbied governments to block approval or importation of GM crops. Opposition has been stronger in parts of Africa, Asia, and Western Europe compared to the USA.

There are several reasons behind the debates surrounding GM foods. In general, one could categorize them into three main areas: (1) environmental concerns, (2) health concerns, and (3) ethical issues related to genetic modification and engineering of crops and livestock.

Environmental Concerns

Several environmental concerns have been raised with the use of GM crops and livestock. Obviously, the major impact would likely be on the ecosystem in proximity to the GM crops in particular, but also potentially cause downstream effects. It is not feasible to evaluate the effect of GM crops on every organism in an ecosystem and thus adverse effects could potentially appear after the GM crop or livestock is approved. Nevertheless, lots of research is ongoing to assess the environmental impact on certain plants and species. In the early days of GM technology, a small study was conducted at Cornell University that found that monarch butterflies under laboratory conditions might be harmed by eating pollen from Bt corn plants. However, in 2001, scientists from the USDA published a series of studies demonstrating no effect to monarch butterflies under natural field conditions, quelling some of the public uproar over the initial findings. These types of seemingly conflicting findings often occur in science as more researchers begin to investigate a specific question and develop new ways of assessing harm and repeating and expanding prior studies.

Other potential consequences may arise from unanticipated crossbreeding. For example, though unlikely, a weed may cross with a Bt corn generating a new weed strain that would now be resistant to the glyphosate herbicide. Similarly, insecticide resistance could also develop in Bt corn or other strains that would negate all of the benefit of these crops. As with any plant species,

changes in DNA can occur in GMOs that can affect disease susceptibility, growth, or other characteristics, rendering them ineffective.

With hundreds of GM crops approved, there remains limited evidence of environmental harm. However, one consequence that has been reported is the development of pest or weed resistance to GM crops. For example, some cases have been reported of insects like the corn borer developing resistance to Bt crops. This can occur as an insect adapts to its new environment through genetic changes that enable it to consume Bt crops without dying. This advantageous genetic change can then perpetuate in a population and a new resistant strain is created.

Health Concerns

Among some of the main health concerns to humans is the possibility of GM-containing foods causing new allergies or some type of adverse effect due to toxicity or an immune response. When genes are "switched" from one species to another, individuals with an allergy to the source of the gene (for example, peanuts) may develop an allergic response to the GM food that has a single gene from that source. The source of the new gene in the GM product may not be disclosed in the labeling. In other cases, some individuals may develop some type of immune response if the body does not recognize the new protein or break it down.

Another potential concern that has been raised with consumption of GMOs is the development of antibiotic resistance. This is a slightly different concern than use of antibiotics in livestock. With respect to GMOs, scientists will often use antibiotic resistance genes when they are engineering the DNA. In order to determine if the experiment worked and the new DNA is present in the organism's cells, scientists will include an additional piece of DNA that encodes for something they can measure, such as a fluorescent marker or antibiotic resistance. For a fluorescent marker, the engineered cells would produce a fluorescent light (think of the light emitted from a firefly). For antibiotic resistance, the scientists can analyze the engineered cells to determine if they will grow in the presence or absence of an antibiotic. If the antibiotic resistance gene is present, the cells will be able to grow in the presence of the drug; if it is not there, the cells will not grow.

It is possible that these genes can be "transferred" to humans or organisms in the environment that consume the GMO plant or animal. Called "horizontal gene transfer," this phenomenon has been shown to occur between other species such as bacteria, but not yet demonstrated to occur following

consumption of a GMO food. If antibiotic-resistant genes in GMOs could be transferred to bacteria, they could potentially acquire antibiotic resistance. Antibiotic-resistant bacterial strains (unrelated to GMOs) have presented an increasing threat to human health as current antibiotic treatments may not work on infected patients. For plants or other animals in the environment, horizontal gene transfer is believed to be exceedingly rare.

Ethical Concerns

There is a recurring question regarding human interference in the natural world. Are scientists going too far by engineering crops and moving genes from one species to another in order to meet human preferences or reduce environmental damage? But aren't they, in effect, just speeding up an otherwise natural process for the selection of particularly desirable traits through modification within a species' DNA such as with the Flavr Savr tomato or nicotine-free cigarettes? Is it acceptable to swap animal genes for animal genes and crop genes for crop genes but not to mix genes between animals and crops? Where do we draw the line to prevent someone from going too far? Or has GM technology already overstepped the boundaries that we value as a society? Simply put, just because we can do something, should we?

There are lots of questions and, depending on who you ask, a wide range of answers but no clear solutions.

Conclusion

Back in the supermarket, does the decision boil down to personal preference? Are people guided more by environmental concerns, nutritional value, taste, quality, appearance, or fear of the unknown or potential harms? If given the choice, how would you make a decision of whether to purchase a GM product or not?

The practical/pragmatic circumstances that have spurred the development of GM foods such as the desire to reduce pesticide use or to increase production volumes to feed more people are understandable to many and laudable. As the world's population continues to grow and outpace food production, it seems logical to take advantage of new technologies to ameliorate hunger, malnutrition, and related diseases. As we continue to push the limits of land use, combat deteriorating soil conditions from overuse or runoff, and reduce the harmful effects of fertilizers to surrounding ecosystems, again it seems not

only logical but also prudent to take advantage of new technologies to sustain and improve current growth and production. On the other hand, the benefits that may be achieved through GM technology must be weighed against not only the ethical concerns but also the desire to rush through careful scientific experimentation and testing including both environmental and human studies and the need to educate the public about these products. Some of the backlash against GM foods may well be because these data have not been collected or communicated openly to the public.

Lastly, whether we choose to buy these products for their enhanced taste, appearance, nutritional value, lower costs, or simply convenience, we each make these decisions based on our own preferences and values. For some, the fact that a food has been genetically modified makes absolutely no difference and there is no need for a special label if the food is deemed to be equivalent to its natural product. For others, the fact that the genes of the food have been manipulated may cause ethical, religious, or just general uneasiness despite its demonstrated safety. The reluctance to eat GM foods due to the unknown origin of all genes in the modified food is easy to understand in persons with certain food allergies (e.g., peanut allergies), for religious reasons (e.g., Muslims and pork), or personal preferences (vegetarians). Transparency about GM foods and public resources can help consumers make informed decisions.

Resources

U.S. Department of Agriculture. Biotechnology FAQs. Available at https://www.usda.gov/topics/biotechnology/biotechnology-frequently-asked-questions-faqs

U.S. Department of Agriculture. Economic Research Service. Adoption of Genetically Engineered Crops in the U.S. Available at https://www.ers.usda.gov/data-products/adoption-of-genetically-engineered-crops-in-the-us.aspx

U.S. Department of Agriculture. List of Bioengineered Foods. Available at https://www.ams.usda.gov/rules-regulations/be/bioengineered-foods-list

U.S. Food and Drug Administration. Understanding new plant varieties. Available at https://www.fda.gov/food/food-new-plant-varieties/understanding-new-plant-varieties

U.S. Food and Drug Administration. Agricultural Biotechnology. Available at https://www.fda.gov/food/consumers/agricultural-biotechnology

U.S. Environmental Protection Agency. Pesticides Industry Sales and Usage 2008–2012 Market Estimates. Available at https://www.epa.gov/pesticides/pesticides-industry-sales-and-usage-2008-2012-market-estimates

Ricepedia. Available at http://ricepedia.org/rice-as-food/the-global-staple-rice-consumers

Golden Rice Project. Available at http://www.goldenrice.org/

International Service for the Acquisition of Agri-biotech Applications. 2018. Global Status of Commercialized Biotech/GM Crops: 2018. ISAAA *Brief* No. 54. ISAAA: Ithaca, NY. Available at https://www.isaaa.org/resources/publications/pocketk/16/

US Department of Agriculture USDA (2018) Secretary Perdue Issues USDA Statement on Plant Breeding Innovation https://content.govdelivery.com/accounts/USDAAPHIS/bulletins/1e599ff

13

Cleaning Up the Environment

Environmental pollution is a global problem that continues to increase with urbanization, development, and reduction of natural lands (forest clearing). Pollutants include both heavy metals and organic compounds and can contaminate soil, air, and water. Plants, animals, and microbes that reside in or near contaminated areas may experience harmful effects from exposure or consumption of polluted materials and the effects may reach the top of the food chain—humans.

While several approaches are available to monitor and decontaminate polluted areas, the scale and range of the problem have outgrown current solutions. As genetic and genomic technologies can be used to improve the health of humans, the same technologies can be used to improve the environment and address problems due to natural causes or human-made. And, as with medical applications, safety is the first issue to be addressed when developing a genetic-based intervention to address an environmental problem. But unlike an individual patient, the introduction of a genetic-based intervention into a habitat or ecosystem may need to have more assurances (or controls) in place to limit unintentional harms on other species in that habitat and other habitats if the intervention can possibly be carried or transmitted to other distant habitats, including humans. Thus, while genetic and genomic technologies hold great potential to address environmental concerns, they must be demonstrated to be safe in an open setting and have appropriate controls to limit potential harms.

S. B. Haga, *The Book of Genes and Genomes*, https://doi.org/10.1007/978-0-387-70916-1_13

Applications

Around the world, soil and water pollution has caused great economic, agricultural, and environmental problems. Environmental problems can be either natural (animal or plant - based such as an infectious weed or large mosquito population) or due to human activities (e.g., nuclear waste, oil spills). Heavy metals such as zinc, lead, mercury, and cadmium are by-products of many manufacturing processes in petrochemical, agrochemical, coal, and mining industries. Known as inorganic pollutants, many of these metals exist at low levels, but become toxic at higher levels to ecosystems and humans when deposited in the soil or waterways. In addition, nitrogen and phosphorous from fertilizers and animal waste contribute to pollution of waterways. Another group of pollutants is referred to as organic pollutants, often abbreviated due to their long chemical names such as polychlorinated biphenyls or PCBs. Organic pollutants do not dissolve in water and tend to persist in the environment for very long periods of time.

In many cases of environmental contamination, the polluted area can be cleaned with traditional methods including removal of contaminated soil (excavation to a landfill) or ex situ remediation, soil incineration, soil washing with chemicals, and groundwater treatment pumps. Other devices can be installed to reduce the amount of pollution such as air or water filters, and recycling programs. However, traditional methods may not be very effective, are costly, and/or are not feasible for some locations or areas (some polluted areas may be too large). In addition, methods to remove contamination and waste products may, in turn, generate additional pollutants or cause other harms to soil or water.

In environmental sciences, much research has focused on the development of new methods and technologies to detect, monitor, and decontaminate pollutants, which can supplement or replace traditional methods. In particular, biological or genetic-based interventions have been explored to control populations or clean up pollutants to reduce risks to human health and/or restore the habitat or region to its pre-affected state. An example of a current biological - based method to facilitate decomposition is a compost bioreactor, or a more sophisticated process of composting. These methods vary by costs, equipment or facilities, time, and feasibility. The use of genetic-based technologies may help overcome barriers faced by current approaches to decontamination, though it may introduce new issues. A sample of genetic-based technologies under investigation are described in the next section.

Pest Control

A major area of environmental research is pest control, particularly insects that can act as carriers (known as vectors) and transmit disease to humans (referred to as vector-borne diseases). Mosquitos are a common carrier for many devastating diseases such as malaria, Zika, and dengue fever. Ticks are a carrier for the bacteria that cause Lyme disease and Rocky Mountain spotted fever. As some species currently occupy an important role in a given ecosystem, the impact of reducing or eliminating a species must be carefully evaluated on the broader level (or potentially another issue will arise). Traditional interventions involve the use of chemical applications (pesticides) to reduce larvae (egg) growth, capture/baiting, and physical barriers (e.g., netting) to reduce population size and/or interactions with humans.

One area of genetic research to control population size of pests involves a technology called a gene drive (Fig. 13.1). A gene drive is a technique that inserts a particular version of a gene in the next generation (or offspring). Whereas the natural process (the passing of gene version from parents to offspring) is totally random, a gene drive will enable a specific version of the gene to be passed onto future generations. For example, if a researcher wanted to grow extra-large strawberries, a gene drive would be created for the particular gene version for large size or growth that would be passed onto each new crop

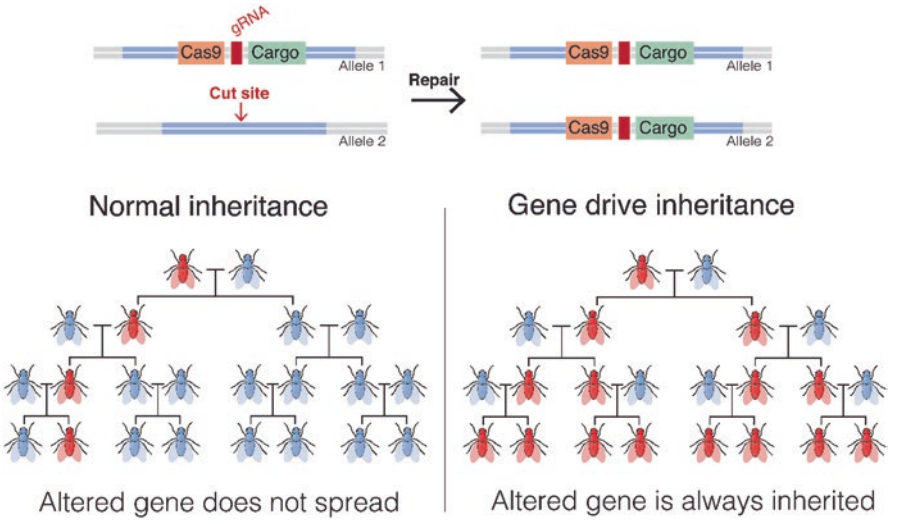

Fig. 13.1 A diagram of how a gene drive is made using the CRISPR gene editing technology and how it works (source: Marius Walter, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons)

of strawberries. A gene drive utilizes the recently discovered gene editing technology (called CRISPR—discussed in Chap. 9) that is attached to the gene to be passed into the next generation. Dubbed a "copy-and-paste function" for genes, a gene drive will cut and insert the desired gene version.

Gene drives have been studied with respect to mosquito control. As certain environmental conditions may favor large spurts in population size (e.g., high rainfall), thereby posing risks of increased vector-borne diseases to humans, a gene drive could be used to reduce fertility rates of female mosquitos, which would drastically slow the growth of the population. Another approach to control population growth would be to increase the number of male species in a population, something that is also under genetic control.

Genetic modification is another approach to controlling pests. As discussed in Chap. 12, it has been extensively used in the agricultural sector to change the physical state of crops and animals to achieve certain societal food preferences (larger strawberries) or needs (larger salmon). The same genetic modification techniques can be used toward the goal of modifying a characteristic of a species that may be harmful to humans. For example, genetic modification of mosquitos to disable their ability to carry and transmit an infectious pathogen (like the malaria parasite) may have huge benefits for human health without disruption of the natural ecosystem. The health and long-term survival of the genetically modified species would need to be carefully studied in pilot studies. It is possible that removing certain features through genetic modification would alter the long-term survival of the altered species in a wild population. For example, wild species may recognize or sense differences and not select the genetically modified species for mating.

Bioremediation and Biodegradation

In larger areas of environmental contamination, traditional methods of cleanup may be poorly suited, creating a niche for biological solutions. Two biological-based approaches, bioremediation and biodegradation, are actively being investigated as they may cause less structural damage and environmental harm to a site, though, in some cases, it may require a longer period of time for cleanup. Bioremediation refers to adding nutrients and other factors to speed the growth of natural organisms that can help remove or detoxify polluted areas. Related, *biostimulation* and *bioaugmentation* are parts of overall bioremediation efforts. Biodegradation refers to the process of using organisms to actually degrade components of polluted areas.

Both plants and microbes such as bacteria have been investigated as solutions for bioremediation and biodegradation, with plants serving as possible long-term and sustainable interventions, and bacteria as a rapidly deployed solution. The degradation of some pollutants requires a multistep process to break it down into a nontoxic form, and thus, depending on the extent of the pollution, a sizable amount of plants or microbes will be needed. Genetic engineering of plants and animals may enhance more rapid growth and efficiency of the degradation process. If more than one gene is required for a multistep chemical reaction to break down a pollutant, engineers can improve the overall efficiency of the chain and "strengthen" the rate and efficiency of a reaction by replacing less efficient components of the reaction (think of an assembly line, or more apropos, a "disassembly" line). A chemical reaction begins with a starting material (the pollutant), which I will refer to as "A." "A" is converted to another form called "B," which is then converted to "C," and each step yields a potentially less toxic form of the pollutant. After the final step of the chemical pathway, is reached the pollutant is degraded.

Bacteria

A wide array of bacterial species naturally exist that have evolved and adapted to survive in unusual environments, e.g., in extreme temperatures, acidity, and with scarce "food" sources. New genetic technologies have enabled scientists to identify genes unique to these bacteria that equip them with these unique advantages to survive. As it turns out, some of these genes may be beneficial to human needs for distinct applications such as biodegradation or bioremediation.

Two major environmental issues for which genetically modified bacteria have been studied are oil spills and the growing threat of plastic pollution. The devastating immediate and long-term consequences of oil spills have been extensively documented. Oil spills of all sizes have occurred around the world though the number of accidents has substantially declined since the mid-1970s according to the International Tanker Owners Pollution Federation. In recent memory, the two major oil spills in the US waters were the Exxon Valdez oil spill in Alaska in 1989 and the BP Deepwater Horizon rig explosion off the coast of Louisiana in 2010.

Containment, removal, and degradation are the major strategies to managing an oil spill. Containment of the spill involves erecting physical or mechanical barriers, such as a boom, that can confine the oil to a limited area and if near the shoreline, prevent it from reaching more densely inhabited

ecosystems such as swamps. A boom resembles a long flotation device but may also include a "skirt" that can hang below the water surface to entrap the oil in a confined area. The trapped oil can be pumped, skimmed, or absorbed using sorbents (a type of material that soaks up oil).

In addition to mechanical and physical strategies, chemical and biological based options are available. Chemical dispersants are substances that can break up oil into smaller droplets. Though debate exists about their effectiveness (typically work better in warmer waters and immediately after a spill) and environmental toxicity, dispersants were used to clean up the Exxon Valdez spill and BP oil rig explosion. The study of dispersants used in the BP Deepwater Horizon spill showed some interesting findings with respect to not the direct effectiveness of the dispersant, but to how it actually stimulated the growth of naturally existing oil-eating bacteria. There are several types of bacteria that live in water that can digest or break down components of petroleum (oil). Molecules called hydrocarbons (hi-dro-car-bons) are the main component of oil, and hydrocarbon-degrading bacteria, or more technically, hydrocarbonoclastic bacteria, have been studied as a solution to oil spills. Since these bacteria are not naturally widespread or present in sufficient quantity to make an immediate impact on a rapidly spreading oil slick, bioremediation and use of bacteria-friendly dispersants could help stimulate growth.

Alternatively, genetically modified bacteria could be developed to rapidly digest oil, a potentially less toxic solution to wildlife and habitats compared to dispersants. The first genetically modified bacteria were announced in 1971 by scientists at General Electric. Many questions continue to be studied including how quickly genetically modified bacteria can integrate into the natural environment, how quickly they multiply (to determine how much should be applied to a given area), and the environmental impact once the oil is significantly reduced or dispersed. To limit harmful environmental impacts, it may be possible to engineer a halt in the growth of genetically modified bacteria (aka "suicide" gene) once the "food" source (e.g., oil) is gone. In addition, new mechanisms for rapid application or dispersion of oil-degrading bacteria are under investigation. For example, freeze-dried oil-eating bacteria may enable rapid and widespread application, potentially by plane or boat.

Bacteria are also being investigated to address another growing problem: the massive quantities of plastics polluting the oceans. For example, two enzymes (proteins) have been identified in bacteria that can degrade polyethylene terephthalate (PET), a commonly used plastic in food packaging.

A type of bioremediation under investigation involves the engineering of what are called "bacterial microcompartments." These microcompartments exist naturally in bacteria and function as a sort of a storage unit or place

where specific reactions occur. Thus, the contents of the microcompartments are separated from the rest of the bacterial cell, perhaps to protect it from toxic materials within the microcompartment. Engineering bacterial microcompartments could fulfill a very specific function—to enable bacteria to take in, break down, and store certain pollutants. This would potentially allow the bacteria to withstand otherwise toxic levels of the pollutant. In addition, artificial microcompartments can be created for plants where they can store pollutants absorbed through the roots to reduce toxic effects.

The engineering of bacterial microcompartments does not involve the artificial creation of the actual structural components, but instead modification of the bacterial genes that are responsible for the creation of the compartments. Furthermore, scientists can potentially modify the enzymes (proteins involved in chemical reactions) contained within specific microcompartments in order to break down specific pollutants.

Plants

Phytoremediation is a process in which plants can remove pollutants, or "detoxify" the soil. Some native plant species may be able to absorb the pollutants through their roots and in effect remove them from the soil. Molecules called "transporters" can help move the pollutants to other areas within the plant where it can be degraded and/or stored. Genes that encode for molecules that assist in the absorption, transport, degradation, and storage of chemicals can be genetically engineered or enhanced in plants to carry out specific jobs with respect to bioremediation. For example, scientists can engineer plants to develop an "absorption" process that does not naturally exist, or perhaps boost a naturally existing one. The genetically modified plants can be sowed, harvested, and disposed in the contaminated area. For water pollutants, algae are being studied as a potential phytoremediator for polluted lakes and waterways.

Plants can also be utilized in the bioremediation of plastics. For example, scientists are working to genetically engineer algae to produce two key enzymes with the goal of degrading PET in waterways and oceans.

Biosensors

With respect to the environment, a biosensor is a biological-based method to sense or detect chemicals such as pollutants that are not easily detected by visual detection or smell. In addition, biosensors can also be used to monitor plant health, as many common agricultural crops are susceptible to disease, costing millions of dollars in crop losses. These are distinct from biosensors that have been developed for medical purposes in humans to detect molecules in the body predictive of disease. The fields of biology and nanotechnology have combined to produce nano-inspired biosensors. Nanotechnology refers to the development and use of very small devices (measured in nanometers; 1 cm is equal to 10,000,000 nm). Nano-inspired devices are being evaluated for the detection of plant viral or fungal DNA and molecules that are released in the plant under stressful conditions (pollutants, drought, etc.). Not only have plant biosensors been developed for the detection of environmental pollutants and plant disease, but they have also been particularly useful in elucidating natural processes such as plant growth. Another application of a biosensor is to monitor pollutants in drinking water. Access to safe drinking water is a major challenge in many parts of the world.

One type of biosensor involves something called microbial fuel cells. Microbial fuel cells are synonymous to a traditional fuel cell except that they typically use bacteria (or yeast) and oxygen as the fuel source to generate an electric current instead of fuel and oxygen. In a microbial fuel cell, microbes assist in the conversion of chemical energy to electrical energy. If a pollutant was the "fuel" source, the higher the level of pollutant, the greater the amount of electricity would be produced. Thus, a microbial fuel cell could easily function as a biosensor based on the linear relationship between the amount of pollutant and electricity generated. Although various types of microbial fuel cells have been around since the 1900s, it has not been until the 2000s that they have been used for removal of wastewater contamination and detection of heavy metals and organic compounds (microbial fuel cells can operate in water or soil environments). In 2018, scientists reported the first paper version of a microbial fuel cell. This paper-based sensor can be easily used anywhere without electricity or special reagents, allowing for the detection of bioactive compounds in water samples.

Genetically - encoded biosensors can detect certain chemicals through a biological reaction caused by the production of a protein such as a fluorescent protein upon connecting with a specific molecule in the system. For example, if a biosensor detects and binds to calcium, it will emit a signal that can easily

be detected. In some cases, the signal produced will indicate the level of the specific molecule in a plant, soil, water, or other setting.

Concerns

As discussed earlier, the development of food-related GMOs has raised a range of concerns regarding the safely of altered plant and animal species to other plants and animals in a given habitat as well as safe consumption by humans. The same concerns would also apply to GMOs developed for environmental applications (bioremediation and biosensors). Unlike some agricultural settings that are contained or enclosed (e.g., a greenhouse or fish farm), GMOs for remediation and monitoring would be applied in natural (non-enclosed) settings. Thus, the safety concerns for the natural habitat and ecosystems are higher, and warrant development of some internal controls that will enable the GMOs to disappear from the environment such as a self-destruction mechanism or withholding of essential food sources for survival. Extensive testing (field tests) is required to understand how well the intervention works and the environmental impacts, including potential harms to humans.

In addition to environmental harms, the uncontrolled spread of a genetic-based intervention can cause distress as it may cross local or national boundaries. Nearby distant communities who do not support or wish to use them may have no way to control or block the spread. The mixture of genetically engineered interventions with natural species could have ramifications for economic trade (as it pertains to agriculture), violate ethical or religious beliefs, or raise other local concerns. Thus, reaching a consensus across a large region regarding the use of genetically - based interventions may be extremely challenging and impact their overall use.

Conclusion

With lots of different genetic-based technologies under evaluation, we may expect to see more pilot testing or introduction in various regions to help address the myriad environmental issues in need of attention. Yet, careful assessments and study are required to maximize safety to habitats and ecosystems as well as humans. Open debate, transparency, and involvement of communities and stakeholders can contribute to the development of acceptable guidelines and protocols as well as enhance local public awareness and support.

Resources

Rylott EL, Bruce NC. How synthetic biology can help bioremediation. Current Opinion in Chemical Biology 2020, 58:86–95. https://doi.org/10.1016/j.cbpa.2020.07.004

Burén S, López-Torrejón G, Rubio LM. Extreme bioengineering to meet the nitrogen challenge. Proc Natl Acad Sci U S A. 2018 Sep 4;115(36):8849–8851. doi: https://doi.org/10.1073/pnas.1812247115.

Chouler J, Cruz-Izquierdo Á, Rengaraj S, Scott JL, Di Lorenzo M: A screen-printed paper microbial fuel cell biosensor for detection of toxic compounds in water. Biosens Bioelectron 2018, 102:49–56.

Lee et al. Biotechnological Advances in Bacterial Microcompartment Technology. Trends Biotechnol. 2019 Mar;37(3):325–336. doi: https://doi.org/10.1016/j.tibtech.2018.08.006.

Kahn J. The Gene Drive Dilemma: We Can Alter Entire Species, but Should We? The New York Times Magazine. Available at https://www.nytimes.com/2020/01/08/magazine/gene-drive-mosquitoes.html?action=click&module=Editors%20Picks&pgtype=Homepage

Our World in Data. Oil Spills (2013). Available at https://ourworldindata.org/oil-spills

US National Oceanic and Atmospheric Administration. Who Thinks Crude Oil Is Delicious? These Ocean Microbes Do. Available at https://response.restoration.noaa.gov/about/media/who-thinks-crude-oil-delicious-these-ocean-microbes-do.html

Nina Dombrowski, John A. Donaho, Tony Gutierrez, Kiley W. Seitz, Andreas P. Teske, Brett J. Baker. Reconstructing metabolic pathways of hydrocarbon-degrading bacteria from the Deepwater Horizon oil spill. *Nature Microbiology*, 2016; 16057 DOI: https://doi.org/10.1038/NMICROBIOL.2016.57

McFarlin et al. Biodegradation of dispersed oil in Arctic seawater at −1 °C. PLoS One. 2014; 9(1): e84297. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3885550/

14

Genetics and the Crime Scene: Just Like TV?

Isn't it ironic how time and time again we can start out doing one thing but end up in a completely different place from hence we began? Sometimes this turns out to be a good thing, and sometimes not. This scenario often occurs in science because we do not always have a clear understanding of where research will lead us. In science, learning the answer to one question only raises many more questions. Hence, this is the story of DNA fingerprinting and its now routine use in criminal investigations.

The omnipresence of DNA-based identification, from the real-life, sensational O.J. Simpson trial to celebrity paternity suits to numerous fictional television crime shows, makes it hard to recall a time when we did not rely on DNA-based technology. Despite its pervasiveness today, the use of DNA identification in forensics is relatively young, dating back to the 1980s. As with all new technologies, DNA-based identification technologies have evolved substantially, and alongside, many societal and ethical concerns have arisen.

Identification Testing Before DNA

Dating back to the 1920s, the most common type of identity testing was by blood type. There are four blood type groups: A, AB, B, and O. These blood types are actually due to the inheritance of different variations of a single gene. For example, a person with type AB blood has inherited an "A" version (or allele) from one parent and a "B" allele from the other parent. Given that

S. B. Haga, *The Book of Genes and Genomes*, https://doi.org/10.1007/978-0-387-70916-1_14

there are only four blood types, blood type testing does not have a high rate of exclusion—in other words, knowledge of blood type can eliminate only about 30% of the population. Thus, even if a positive match is found between a suspect and a sample from a crime scene, the test cannot definitively identify a person as the perpetrator given the commonality of blood types.

In the 1930s, serum testing was introduced, which involved testing for several more biomarkers found in blood, thereby increasing the accuracy of identification. Many of these markers are clinically tested for prior to blood transfusions and organ transplantation to determine compatibility between the donor and recipient. As with blood type, differences between these markers are common and are more useful to exclude or rule out individuals.

In the 1970s, additional biomarkers related to a person's immune defense system, which were even more highly individualistic, replaced the previous biomarkers. Known as "HLA" markers, these are routinely used in matching donors and recipients for organ transplantation. The closer the match, the more likely the recipient will not reject the donor organ. This is why the best donor often is a close relative as they will have the highest likelihood of sharing or having the same HLA markers. Again, depending on how common the HLA marker is in the general population, it may or may not be possible to definitively confirm a match between a suspect and a specimen left at a crime scene.

Rise of DNA Testing

So when did DNA testing start to be performed by criminal investigators and what does it entail? DNA fingerprinting, as it has come to be known, actually arose in an unexpected manner. In the late 1970s, a British scientist by the name of Alec Jeffreys was working on understanding the evolutionary process of a group of genes through the study of small genetic changes or variations. These variable DNA sequences were referred to as "markers" or signposts on the genomic "map" to help determine the order or arrangement of genes. In some ways, one could think of these as genetic "landmarks." At the time, scientists were working blindly (so to speak) trying to figure out where genes were located in the vast genome.

In the 1980s, continuing with his work on evolution of gene families, Jeffries and his colleagues were studying a gene that had been isolated from skeletal muscle tissue of a seal called myoglobin. A region of the gene had a curious sequence of short sequence repeats, or a string of letters that repeated like "TA-TA-TA-TA-TA." As it turned out, DNA repeat regions were not

unique to the myoglobin gene, but are present throughout the genome in many different species including humans. These repeats ranged in size from as short as two base pairs or letters (in the example above) to greater than 100 base pairs. Furthermore, these repeat regions were found to be highly variable between individuals, where one person may have 20 repeats in a given region and another person may have only five repeats.

By determining the number of repeats in these variable DNA regions, a genetic profile could be established for a given individual based on the number of repeats in multiple variable regions (Fig. 14.1). For example, variable region number one may range from 4 to 36 repeats, variable region number two could range from 40 to 90, and variable region number three could range

Person 1: (11 repeats)	ATGG ATGG ATGG ATGG ATGG ATGG ATGG ATGG ATGG ATGG ATGG
Person 2: (3 repeats)	ATGG ATGG ATGG
Person 3: (7 repeats)	ATGG ATGG ATGG ATGG ATGG ATGG ATGG

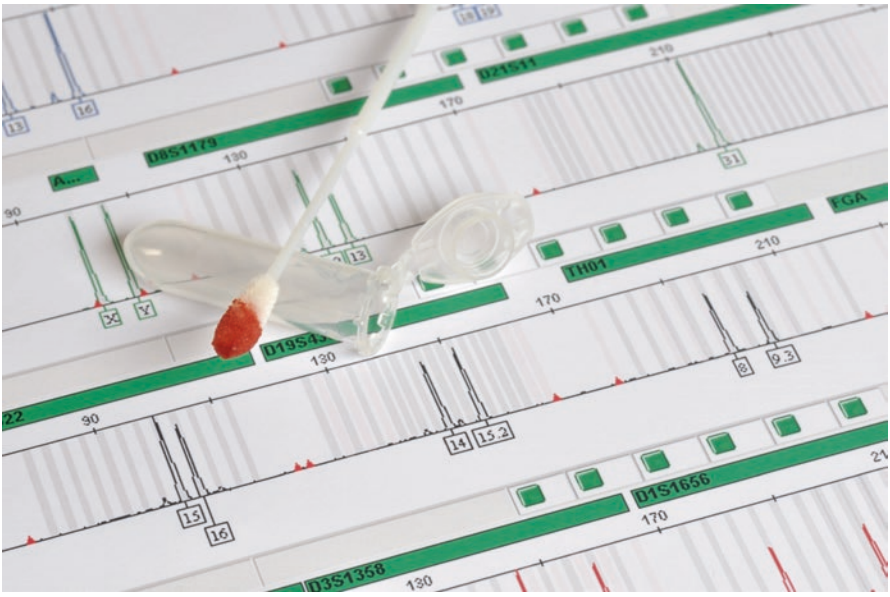

Fig. 14.1 (Top) An example of different number of a repeated DNA segment (ATGG) in three different people. (Bottom) The photograph of a printout of a DNA profile shows the different lengths for multiple repeats across the genome. Recall that each individual will have two versions at each location (inherited from each parent), so two numbers are shown for each specific location (source: Adobe Photo Stock)

from 14 to 56. Thus, the more regions analyzed, the more unique the genetic profile could be established. The probability of a match is calculated based on the frequencies for each variant at a given site. Think of a lottery drawing of six winning numbers selected from 1 to 100—what are the odds of getting all six numbers correct?.

Analysis of these variable repeat regions in a person's DNA could be applied not only for identification purposes but also to establish or confirm familial relationships since family members would have overlapping (or shared) genetic profiles. In his acceptance speech in 1992 for an award honoring his achievements, Jeffreys acknowledged that based on his early research, "the implications for individual identification and kinship analysis were obvious."

In the early days of the DNA forensics era, sequencing technology was not available. Instead, scientists generated genetic profiles based on the size of DNA fragments. Called restriction fragment length polymorphisms (RFLP), the technique relied on the use of enzymatic proteins which would cut the DNA at certain regions with a distinct sequence, producing different size fragments depending on the number of repeats in a given DNA region (Fig. 14.1). Thus, a person with 4 repeats in a certain DNA region would yield a smaller fragment than a person with 20 repeats at the same region. Although the actual sequence of the region could not be determined using this method, the differences in sequence could be inferred from the size of DNA fragments and produce a sort of DNA barcode.

DNA can be isolated from virtually any tissue. In humans, this includes saliva, skin, hair, semen, and of course blood. Around the same time that the understanding of DNA variation and its use for personal identification purposes were realized, a new technology known as polymerase chain reaction or PCR was developed. PCR is a method that can amplify or copy any given region of DNA, a very valuable tool that enables the analysis of DNA obtained from small amounts of specimens, such as blood droplet or single hair (and thus, very small amounts of DNA). Later, in the 1990s, when DNA sequencing machines were developed and could rapidly and accurately sequence DNA, the preferred method of DNA analysis shifted to the combined PCR and sequencing method, which is still in use today.

The First DNA-Based Criminal Investigation

In 1986, the first criminal case believed to have used DNA forensics involved the rape and murder of two schoolgirls in England that occurred more than 2 years apart. A young man who had confessed to the murders was in custody, but DNA testing revealed that he was not a match to the semen samples

retrieved from the victims. As such, the first use of DNA forensics resulted in the exoneration of a suspect. But it was clear from the DNA test that one man was the perpetrator of both crimes. After releasing the innocent man, the police asked for the men between the ages of 13 and 24 in the two towns where the crimes were committed to voluntarily donate a sample of DNA. More than 5500 men came forward to donate a blood sample; only one man refused to submit to a blood test, but the police had already confirmed his alibi. Already knowing that the perpetrator had type A blood, the blood samples collected from the dragnet were screened for blood type. Only those samples which were type A were then submitted for a DNA test.

As you might imagine, the actual perpetrator of the crime was a bit reluctant to voluntarily donate his DNA sample. So, he convinced a co-worker to pretend to be him and submit a blood sample, so the police would not get suspicious. As often happens, the co-worker bragged about the blood swap one night at the local bar. A woman who overheard the conversation reported it to the police who apprehended the suspect who proceeded to confess to both murders.

The dragnet approach has raised a number of issues regarding civil liberties. An individual may decline to submit a blood sample for any number of reasons unrelated to the crime, but the appearance of guilt linked to the refusal to submit a blood sample is reason enough for many to consent. One concern linked to a dragnet is what happens to the blood sample after the case has been solved and the perpetrator apprehended. Some police departments promise to destroy the samples after the case is closed, while others may store the samples and/or profiles indefinitely. The violation of personal privacy has been at the center of a debate surrounding the idea of developing a national DNA databank. In addition, dragnets have not always resulted in the capture of the perpetrator and are very expensive to conduct.

The Growth of DNA-Based Forensics

In addition to the use of DNA analysis for criminal investigations, other applications became apparent, namely the use of DNA analysis to confirm familial relationships for paternity and immigration cases. The first reported use of DNA forensics for familial relationships occurred in 1985. Dr. Jeffreys was contacted by a lawyer who wanted to know if DNA testing could be used to confirm a familial relationship. A young boy was about to be deported back to his home country unless evidence could be submitted documenting that the boy was a member of the family. The blood type evidence was not strong

enough to convince the court that the boy was a member of the family. The DNA testing, however, did confirm that the boy was a member of the family and the court dropped the deportation suit. Shortly thereafter, in 1986, the first paternity case using DNA testing as proof of paternity was heard by the UK courts.

Using the same technologies as with DNA identification, paternity testing has become more accurate and rapid over time. With access to tissue (usually blood or cheek cells) from the mother and child, the probability of a person suspected of being the father can easily be determined. Unlike DNA identification testing where a 100% match is needed, a positive paternity test requires a 50% match of the DNA regions between the alleged father and child. Recall that a child inherits 50% of his or her DNA from the father and 50% from the mother. In situations where the mother is not available, paternity testing can still be performed and the results are equally conclusive.

In the USA, as the use of DNA forensic evidence came to be accepted by the courts, it soon became evident that a standardized genetic profile needed to be established. Without a standardized profile, police labs would not be able to compare samples tested from different labs and databases of genetic profiles would be next to useless if every lab was testing a different set of DNA regions to generate their profiles. Therefore, the US Federal Bureau of Investigation (FBI) chose 13 distinct DNA sites to be analyzed for forensic purposes. Each of these 13 regions were chosen due to their highly variable nature, which would produce a highly unique profile for any individual. As a result, it is highly unlikely for any two people to have the same DNA profile. These 13 regions are not universally accepted as other countries have developed different standards.

Forensic DNA Databases

As many crimes go unsolved even with the collection of evidence and generation of DNA profiles, criminal laboratories began to store these DNA files to enable comparison to evidence collected from other crime scenes and new suspects. Since the DNA profiles could be stored electronically, these databases were first developed at the local level, later expanding to the national level. In 1988, Colorado was the first state to enact laws requiring genetic profiles of sex offenders to be stored in a criminal database. In 1990, Virginia became the first state to enact laws requiring DNA from all major crime offenders. In 1994, the US Congress passed the DNA Identification Act,

officially establishing the *Co*mbined *D*NA *I*ndex *S*ystem (CODIS), a network of databases at the local, state, and national levels of DNA profiles from convicted offenders, unsolved crime scene evidence, and missing persons. In 1998, the FBI launched the CODIS database. All DNA profiles are generated at the local level, and then are added to the state and national databases.

By 1996, most states had created sex offender DNA databases and soon thereafter began expanding DNA databases to include other violent crimes and burglary. In 2005, 43 states had all-crimes databases. Today, every state has policies regarding DNA collection and analysis from arrestees and convicted individuals of certain crimes, creating a web of policies across the country. The move to expand DNA databases to include arrestees was (and remains) a controversial policy since the DNA of many innocent people will be collected and stored in these databases. As of 2018, 31 states had laws that allow for the collection of DNA from arrestees (individuals arrested or charged but not convicted). State forensic laboratories are behind (or backlogged) with samples awaiting to be analyzed. According to a 2014 survey by the US Bureau of Justice Statistics, approximately half a million samples had not been processed by more than 400 publicly funded crime scene laboratories in the USA.

Thus, DNA profiles generated from evidence collected at a crime scene can be compared to DNA profiles stored in local, state, or national criminal DNA databases for a potential match or hit. If no matches are returned after searching the local database, criminal labs can expand their searches to the state or national database. According to the FBI, in July 2020, almost 14 million offender profiles were stored in databases across the country. In addition, the database has yielded more than 500,000 hits or matches.

One issue that has garnered attention regarding database searches is the potential for partial matches. As biological family members will share some percentage of their DNA depending on the relationship (closer relatives will have a higher percentage than more distant relatives), the possibility exists that a partial match will be found if a relative's DNA profile is in the database. A partial match means that some of the 13 parts of the DNA profile will be shared. States have different policies regarding partial matches, with Maryland and the District of Columbia banning them, and a handful of other states opting to use this search strategy only if no other methods have returned a match.

DNA Genealogical Databases and Criminal Investigations

As described throughout this book, there have been many different applications of advances in DNA analysis. One of the popular nonmedical uses of DNA analysis has been to identify one's ancestry or heritage. This is possible due to the understanding that certain regions around the world show different DNA patterns, not related to genes or certain traits. As a result, these lingering patterns in people's genomes have served as a general connection to different geographic locations. While not typically precise enough to pinpoint exact locations or uncover new distant relatives from hundreds of years ago, it may reveal shared lineage (you and Christopher Columbus share a common DNA pattern!), providing a general picture of your heritage and perhaps revealing unexpected connections to different parts of the world or cultures.

These DNA-based genealogical services are available through a number of online companies, where you can order a test kit to collect a sample of saliva or a cheek scraping, and access your results through an online account.

To facilitate broader analysis and comparison between individuals, the ancestry test results may be stored in a database because you wish to identify distant relatives or of the testing company. Since some databases only require that you set up an account, they are essentially accessible to anyone. This fact did not escape some clever criminal investigators working on a case that had been unsolved for decades. The case involved a series of robberies, rapes, and murders that terrorized California in the 1970s and 1980s (dubbed the "Original Night Stalker" and "Golden State Killer"). The investigators believed that they had a sample of the DNA of the perpetrator from a rape kit collected from one of the crime victims. However, they could not find a matching DNA profile in any of the state or national criminal databases. They then submitted the sample to an ancestry testing company under the disguise as a regular consumer. The company compared the results to their database; however, no strong matches were identified. The investigators then uploaded the DNA profile of the unknown perpetrator to another genealogy database, which returned a partial match, suggesting a potential relative. Through further investigation of the match and the person's family members and physical descriptions of the perpetrator through eyewitness accounts, the investigators narrowed the search and identified a single suspect—a one-time police officer, living not far from where some of the crimes were committed. Upon obtaining a discarded coffee cup from the suspect, a DNA analysis revealed a perfect match between the suspect and the DNA from the rape kit. Joseph James

DeAngelo Jr. was arrested in 2018, at his home in Citrus Heights in Sacramento County, the same county where he began his string of crimes. This was the first demonstration of the use of a DNA genealogy database and the successful identification and apprehension of a suspect and other cases have subsequently been solved using this approach. However, this approach has raised concerns about DNA privacy and sharing of DNA results in public databases.

Humanitarian Applications

DNA forensics has been applied to identification of human remains retrieved from natural and man-made atrocities. Typically reserved for identification purposes when physical identification methods (e.g., direct facial recognition and identification based on unique physical characteristics such as scars or tattoos, dental records, or fingerprint analysis) are not feasible due to extreme degradation of remains, it has been extremely successful and provided closure to thousands of families, aided the prosecution of war criminals. DNA can be isolated from bones, teeth or hair depending on the level of degradation of the remains. In the human body, dental enamel is the hardest substance produced, providing excellent protection of DNA which is found in the center of the tooth, known as the pulp. The DNA profile from the unidentified remains would be compared to a DNA profile from a known sample of the victim (e.g., from a toothbrush, hairbrush, or leftover blood sample).

The DNA used for forensic analysis is found in the cell's nucleus (think of the nucleus as the central command center of a cell)—those 23 pairs of chromosomes described in Chap. 1. For old and severely damaged remains (e.g., fire), the DNA in the nucleus may be degraded and therefore cannot be analyzed. In these circumstances, another type of DNA found in a different location of the cell, known as the mitochondria (my-toe-kan-dree-a), may still be intact and can be used for DNA analysis. In the cell, the mitochondria are a separate entity from the nucleus. In addition, there are multiple copies of mitochondria within a cell (compared to one nucleus), each containing mitochondrial DNA. Because multiple copies of mitochondria are present in a cell, mitochondrial DNA may survive under extreme conditions compared to nuclear DNA. The mitochondrial genome is also much smaller (only 16,000 bases—smaller than any single chromosome) than the nuclear genome. But mitochondrial DNA is only inherited from the mother whereas nuclear DNA is inherited from both parents. Therefore, a mitochondrial DNA profile can be compared to a DNA profile generated from either a sample of the missing person if available or a *maternal* relative.

DNA identification of human remains has been used around the world from the mass graves of Bosnia-Herzegovina to the remains of members of the Davidian cult who perished in a fire in Waco, Texas, to the victims of the September 11 attacks in New York, Washington, DC, and Pennsylvania. One of the first historical analyses of human remains was that of the Russian Romanov family. Executed by gunfire in 1918, the assassins attempted to cremate the bodies but then moved and buried them in another location. Some of the bodies were discovered in 1979, but the remains were not exhumed until 1991. Not surprisingly, the remains were degraded and broken to the point that identification based on skeletal analysis was impossible. DNA analysis was performed by US and Russian scientists, comparing the profiles to living (though quite distant) members of the royal families in parts of Europe, and confirming the remains of the Russian family.

Another example of the use of DNA analysis for identification occurred following the overthrow of a military dictatorship in 1976 in Argentina, where an estimated 30,000 Argentinians "disappeared," including infants and children, some who were born to women pregnant at the time of their kidnapping. In 1977, the grandmothers of the kidnapped children formed a group to find their missing grandchildren. Using local resources to investigate the disappearance of their grandchildren, particularly what appeared to be suspicious adoptions in surrounding communities, the grandmothers gathered a wealth of circumstantial evidence regarding the location of many children. However, the identities of the children still could not be conclusively determined. Eight years following their disappearance, a new Argentinian Government established a commission to investigate the whereabouts and fates of the children. Under the compassionate leadership of Dr. Mary-Claire King, a geneticist at the University of Washington, mitochondrial DNA testing was performed on young men and women suspected of being abducted as infants and compared to the mitochondrial genetic profiles of the grandmothers. Several positive identities have been made, reuniting the grandchildren with their biological grandparents and other family members.

For more than 10 years, the International Commission on Missing Persons (ICMP) has been assembling data to determine which remains collected from mass grave sites belong to the more than 40,000 boys and men that disappeared following the collapse of the former Yugoslavia. The genetic profiles of thousands of family members in search of the remains of lost relatives are stored in a central database established by the ICMP. By 2006, the remains of 10,000 people had been positively identified through DNA analysis. In addition to the shear number of victims, DNA identification efforts faced a major challenge given that remains were often moved from the original burial site to

prevent discovery of the crimes. The displacement of the bodies resulted in the fragmentation of skeletal remains among multiple grave sites and the mixture of remains from different massacres.

Other Applications of DNA-Based Forensics

While most of us probably think of criminal court cases as the prominent use of DNA forensics, its application is far from limited to humans. DNA can be isolated from tissues of other species as well, which has resulted in some interesting and unforeseen uses of DNA forensics. An industry that may not immediately come to mind as taking advantage of genomics is the wine industry. But this industry, more than others, may particularly benefit from genomics given the many varieties of grapes and their susceptibility to both disease and counterfeiters.

In 2007, the first grapevine was sequenced through a joint effort by French and Italian scientists. The grape variety, *Vitis vinifera*, is derived from Pinot Noir, and grown for fruit and beverage (think table grapes, grape juice, wine, and raisins). The first fruit plant to have its genome sequenced, grapevine is the most commonly grown fruit, accounting for more than 7 million hectares and grown on every continent except Antarctica. Italy, Spain, and France are the largest producers with over 1 million hectares each.

Some interesting findings have come out of grape genomics research. The grapevine genome contains 480 million bases arranged on 19 chromosomes; it is estimated to have more than 30,000 genes. There are two main grape varieties—white and red—due to the presence or absence of a single protein known as anthocyanin in grape skin. It turns out that white grapes are actually a "freak of nature" due to mutations in two genes that are involved in the production of anthocyanin. These mutations block the production of anthocyanin resulting in the absence of the red berry color, the parent grapevine. Not surprisingly, analysis of the grapevine genome revealed a large number of genes associated with wine flavor. There are more than 100 genes and former genes (known as pseudo-genes) that are involved in the production of terpenoids and tannins, the chemicals that contribute to a wine's aroma and flavor.

The sequencing of the grape genome may benefit vintners for two major reasons. First, grapes are susceptible to several fungal diseases such as black rot and downy mildew that can lead to increased production costs to remove the unwanted chemicals caused by disease, decreased shelf life, reduced wine quality, secondary infections, and ultimately crop loss. Second, grapes are extremely sensitive to hot or dry climate conditions that can affect harvest

times, sugar production, and ripening process. Therefore, the creation of heat-tolerant or drought- or disease-resistant varieties may be extremely appealing to growers. In addition, vintners may be interested in understanding the genetic mechanisms behind the aromas and flavors of certain grapes that may eventually lead to the development of new varieties. However, high consumer familiarity with centuries-old wine varieties creates substantial market challenges to the introduction of new genetically modified varieties, regardless of enhanced flavor or reduced cost, making these engineered strains unlikely anytime soon.

A second major benefit to understanding the DNA of grapes is its use to detect counterfeiters. The practice of DNA grape profiling now exists to identify unknown species, confirm or certify the identity of a grapevine, and determine the identity of grapes sold to wineries. The University of California (UC) Davis created and maintains the largest database of grape DNA profiles in the world. The database enables scientists to compare the DNA profile of a given grape to more than 600 profiles of major wine grape, table grape, raisin, and rootstock varieties grown in California and France. Six to eight DNA markers (recall that 13 markers are tested in humans) are tested to create a grape DNA profile, which is then compared to the database.

In 1999, scientists from UC Davis shocked the wine industry when they discovered that 16 French wines, including the highly reputable Chardonnay and Pinot, are all related. But most surprisingly, their newfound heritage was not exactly a source of family pride in the wine community. Through DNA profiling, it was found that these grapes were the products of the highly respectable Pinot variety and the mediocre gouais blanc variety (no longer grown in France). Similarly, it was also discovered that Sauvignon Blanc (white) and Cabernet Franc (red) are the parents of Cabernet Sauvignon (red). Thus, as with people, it is now possible to map the lineage of grape varieties.

A long-standing mystery about North American grapes was also solved through DNA profiling. Two red wine varieties, known as Norton and Cynthiana, are believed to be the oldest native North American varieties commercially grown today. In the 1800s, the Norton variety was a staple grape of the Virginia wine industry. Over the next century, the crop migrated westward to Missouri and Arkansas where it picked up the name Cynthiana, or was nicknamed as the "Cabernet of the Ozarks." Wines labeled as Norton are typically dark and aromatic wines with hints of raspberry, coffee, and chocolate. In contrast, wines labeled as Cynthiana are lighter and fresher. These differences are likely due to environmental factors such as soil and temperature. However, based on comparisons of the DNA profiles of the two putative

strains, we now know that Norton and Cynthiana are actually one and the same, although their origin is still up for debate.

Conclusion

Undoubtedly, the use of DNA in criminal investigations and other industries to definitively identify the source of a DNA sample has been revolutionary. The combination of new technologies and discovery of unique features in the human genetic code as well as other species have yielded a highly accurate method of identification, with a range of applications. As technologies continue to advance that enable faster analysis on minute samples and scientific databases continue to expand with new genetic codes, we can anticipate more widespread use. As with other genomic applications, policies regarding the collection, storage, and removal of genetic profiles from local and federal databases should consider individuals' privacy and public safety goals.

Resources

Bureau of Justice. Publicly Funded Forensic Crime Laboratories: Resources and Services, 2014. (November 2016). Available at https://www.bjs.gov/index.cfm?ty=pbdetail&iid=5827

Lee, HC and F. Tirady. Blood Evidence. Perseus Publishing. Cambridge, MA: 2003.

Jeffreys AJ. 1992 William Allan Award Address. Am J Hum Genet 53: 51–55, 1993.

National Conference of State Legislatures. Forensic Science Laws Databases. Available at https://www.ncsl.org/research/civil-and-criminal-justice/dna-laws-database.aspx

National Institute of Justice. What Is STR Analysis? 2011. Available at https://nij.ojp.gov/topics/articles/what-str-analysis

S. Panneerchelvam and M.N. Norazmi. Forensic DNA Profiling and Database. Malays J Med Sci. 2003 Jul; 10(2): 20–26. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3561883/

Arnaud CH. Thirty years of DNA forensics: How DNA has revolutionized criminal investigation. Clinical and Engineering News 2017; 95(37). Available at https://cen.acs.org/articles/95/i37/Thirty-years-DNA-forensics-DNA.html

US Federal Bureau of Investigation. CODIS—NDIS Statistics (July 2020). Available at https://www.fbi.gov/services/laboratory/biometric-analysis/codis/ndis-statistics

Katz B. DNA Analysis Confirms Authenticity of Romanovs' Remains. Available at https://www.smithsonianmag.com/smart-news/dna-analysis-confirms-authenticity-remains-attributed-romanovs-180969674/

15

Are Humans Related to Cavemen?

We have all heard the jokes and seen the cartoons about cavemen. In general, based on the caveman stereotype, they were not too bright, carried around a club, dressed in fur loins, and were mostly men. While there are obviously no written records or photographs of these early humans, a lot has been learned between archaeologists and geneticists to advance our understanding of the lives, movement, and history of our ancient ancestors. Regarding early humans that lived several thousand years ago, we are still learning about their lives, ailments, and lifestyle from examination of the remains left behind. While much has been learned through archaeological studies, genetic and genomic analysis can provide further insights regarding relationships, movement, and health. As a result, maybe our stereotypes will change to reflect what we have learned in recent years.

Some Ancient History

As to be expected, over time, particularly thousands of years, the biological materials in the human body degrade, leaving primarily a fossilized record of their existence. Based on these fossils, it is estimated that modern humans have roamed the Earth for about 500,000 years. The human body has greatly changed over this period (as has our environment and lifestyle). In particular, the shape and size of the brain/skull, height, and facial features have evolved.

S. B. Haga, *The Book of Genes and Genomes*, https://doi.org/10.1007/978-0-387-70916-1_15

It is believed that there were several species of primitive or archaic humans before modern humans (*Homo sapiens*) arose. The earliest fossil evidence of modern humans was discovered in Africa and dates back to about 300,000 years ago. Earlier human species were *Homo habilis* and *Homo erectus*, when they began to move out of Africa, and evidence shows use of fire. There are other groups of humans that are less well understood that coexisted with the main groups.

Much of the understanding of early humans has been derived from fossilized remains and dwellings, which are measured, dated, and mapped. With current genetic analysis or sequencing technologies, analysis of DNA extracted from fossils is now possible and can provide more insight regarding relationships and movement across continents.

How Long Can DNA Last?

One of the obvious questions to consider is how long can DNA last? Probably tens of thousands of years, maybe 100,000 years, under ideal conditions. But fossilized remains are generally not found in ideal conditions, and in the majority of cases, the DNA is severely degraded or not present at detectable quantities. In addition, one of the major problems in extracting DNA from old tissues, particularly those that have been exposed to the natural elements for long periods of time, is contamination by other DNA sources, particularly other humans (including the scientists) and bacteria. Bacterial contamination is easier to detect given the unique characteristics of bacterial genes and the differences between bacterial DNA and other species.

Mitochondrial DNA, due to its smaller size and multiple copies, may be present in higher quantities in old remains. The trade-off is that the data that can be gained from mitochondrial DNA analysis regarding evolutionary history or health of the species is more limited. But methods to amplify or make many copies of DNA in limited quantity are now possible.

Genetic Analysis of Ancient DNA

The field "paleogenetics" was thus born with the partial success of genetic analysis of fossilized remains and can complement more traditional methods of analysis of fossilized remains. For instance, carbon dating is the most common technique used to date fossilized remains—bone, teeth, bark, etc. This technique measures the amount of the chemical element carbon found in all

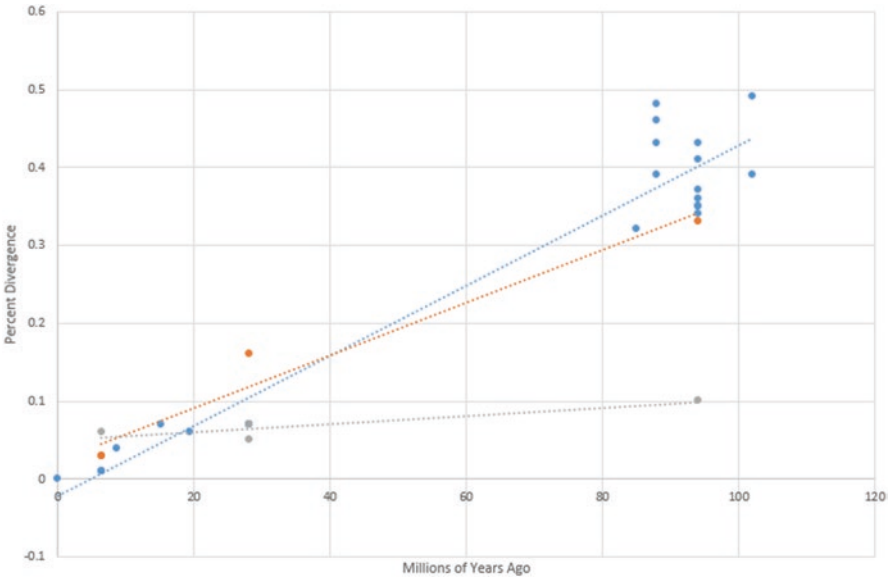

Fig. 15.1 This graph shows a comparison of the degree of changes (divergence) of three genes over time (cytochrome C (grey), fibrinogen (orange), and PRR30 (blue)). The gene of the grey line at the bottom shows the least amount of change over time (source: Wikimedia)

organic materials. As scientists know the rate at which carbon decays, they can estimate the age of a specimen based on the current amount of carbon present. Scientists can now date some remains using a method known as DNA dating. Similar to the carbon dating method, this technique is based on the assumption that mutations occur at a constant rate. The more mutations detected in a specimen, the older the sample is predicted to be. In addition, as the genomes of many species have been sequenced and are publicly available, scientists can compare the DNA of different samples of ancient species to today's species (Fig. 15.1). For humans, DNA changes in medically related genes can provide some insight about disease, health risks, or other traits.

Neanderthals

The Neanderthals (*Homo neanderthalensis*) have probably garnered the most amount of attention of the human species other than ourselves (*Homo sapiens*). Neanderthals are the closest known human relative to modern humans. These two groups coexisted throughout Europe and Asia, but it is not clear when or where the split occurred between modern humans and the Neanderthals.

Fossilized remains have been found throughout Europe and parts of Western Asia dating back 400,000 years. They were physically distinct from modern humans in height (shorter), with a broad chest, and stocky physique. Evidence has contradicted the stereotype of an unintelligent species with discoveries regarding the use of fire and building of hearths, clothing and blankets, boats, and use of plants for medicinal purposes. It is estimated that the Neanderthal population disappeared about 30,000 years ago with the last populations of Neanderthals found in Spain and Portugal. Modern humans were thought to have arrived on the scene about 45,000 years ago. It is presumed that modern humans to some degree drove the Neanderthal population to extinction due to competition, while others suggest climate change. While genetic analysis is not likely to reveal the exact reasons for extinction of the Neanderthals, it has begun to reveal the degree of interaction between the Neanderthal population and modern humans.

The first remains of the Neanderthal was found in 1856 in the Neander Valley in Germany, hence its name (some have added an "h" to the end to replicate its phonetic pronunciation, while others use "Neandertal"). Attempts to extract DNA from 24 Neanderthal remains resulted in only four samples with actual Neanderthal DNA. Instead, modern human DNA was found in most of the Neanderthal samples as well as DNA from cave bears from the same caves, highlighting the significant problem of contamination of ancient remains. One of the leading researchers on the genetic analysis of Neanderthals is Dr. Svante Paabo from the Max Planck Institute for Evolutionary Anthropology in Germany. In 1997, Dr. Paabo and colleagues reported on their DNA analysis of the first Neanderthal specimen. They had extracted mitochondrial DNA from the humerus (leg) bone and were able to sequence a little. After analyzing the mitochondrial DNA sequence, they estimated that the split between early humans and Neanderthals occurred around 550,000–690,000 years ago, matching the estimation based on archaeological data. However, this conclusion data was based on the DNA analysis of only a single Neanderthal specimen, and the accuracy of the DNA dating would need to be confirmed through analysis of additional specimens as they were discovered.

In 2000, the initial finding was confirmed when analysis of DNA extracted from a 29,000-year-old Neanderthal infant rib bone found in the northern Caucasus (southern Russia) showed little similarity to modern humans. Furthermore, comparison of the German Neanderthal sample to the northern Caucasus sample showed that they both came from the same population. At least ten more unique Neanderthal specimens have been analyzed and all show that they are more similar to one another than to modern humans.

Of particular interest is whether or not Neanderthals interbred with early humans. Archaeological evidence suggests that interactions between Neanderthals and modern humans occurred (e.g., stone tools associated with modern humans have been found with Neanderthal remains), although analysis of the structure of bones taken from this area suggests that no interbreeding took place. In 2003, a study was published by an Italian team that analyzed mitochondrial DNA from Neanderthal, Cro-Magnon (early modern human), and modern humans. They found similarities in the mitochondrial DNA sequence between Cro-Magnon and modern humans, but found very little similarity between Neanderthal and modern humans. As a result, they concluded that Neanderthals made no contribution to the gene pool of modern-day humans—in other words, no interbreeding between Neanderthals and early humans took place.

In 2006, Dr. Paabo and his team isolated DNA from a 38,000-year-old Neanderthal fossil from Croatia. Previously, only the small mitochondrial DNA could be extracted from Neanderthal remains. Although the stretches of DNA were much smaller than normal due to degradation, researchers were able to sequence about a million base pairs in total. Contamination is always a concern, but it was estimated that 94% of the DNA extracted from the bone was of Neanderthal origin. This is a tremendous feat in itself given the age of the specimen and will provide unprecedented insight into understanding the evolutionary history of modern humans.

Current DNA evidence suggests that the common ancestor of Neanderthals and modern humans extends back to as far as 500,000 years, much earlier than was initially estimated. The new dating is based on the analysis of a single gene known as FOXP2, a gene that is associated with human language and speech. Chimpanzees also have a copy of the FOXP2 gene, but with a slightly different sequence in a key portion of the gene, which is hypothesized to be important to speech. Analysis of the FOXP2 gene sequence in two Neanderthal samples extracted from remains found in a cave in Northern Spain showed that the same sequence was present in modern humans, suggesting that the gene sequence began to change prior to the split from the common ancestor of Neanderthals and humans.

Also in 2006, a team of US and German scientists (led by Dr. Paabo) announced their intention to sequence as much of the Neanderthal genome as possible. A draft sequence of the Neanderthal genome was published in 2010. Because of the degradation to the DNA, DNA was extracted from three Neanderthal remains and the genome sequence was re-created by aligning the short sequence fragments together to form as complete a sequence as possible. In their publication, the team concluded that, based on the comparison with

modern human DNA, Neanderthals likely interbred with modern humans. Specifically, the paper reports that 1–4% of Neanderthal DNA can be found in modern humans. However, this interpretation of the data was controversial, with other research groups disputing the conclusion that any interbreeding took place between Neanderthals and modern humans as the two groups are about 99.5% genetically identical. Since 2010, other publications have provided further evidence of interbreeding, perhaps putting to rest this controversy. Specifically, analysis of Neanderthal DNA shows traces of modern human DNA, and analysis of modern human DNA shows traces of Neanderthals. These exchanges occurred across Europe and Asia.

A pair of publications in 2014 supported the conclusion of interbreeding with two interesting claims: (1) the Neanderthal DNA may have benefitted modern humans, providing them with traits to survive in colder northern regions, and (2) the children of Neanderthal and modern humans may have been infertile, effectively halting the expansion of this mixed human population.

Caveman's Health

Many interesting stories abound that help us understand the lives (and health issues!) faced by our distant ancestors in different parts of the world. For example, the bacterium that is responsible for acid reflux (or GIRD), *Helicobacter pylori* (*H. pylori*), which affects many people today, was also a health issue several thousand years ago. Today, it is estimated that more than half of all people are infected with *H. pylori*, though only about 10% of people will actually develop disease symptoms. Several prehistoric remains have detected *H. pylori* sequences. In 2016, an international team of researchers published the sequence of *H. pylori* based on DNA sequencing of stomach content from a 5300-year-old male from Southern Europe, close to the Eastern Italian Alps (dubbed the European Copper Age mummy) [interesting fact: the food he ate just before his death could be ascertained through imaging analysis of his stomach].

Because this bacterium has been around so long and DNA sequences often change over time (evolution), scientists have been able to trace the movement or migrations of people long ago based on the DNA sequences of this bacterium. Thus, a map and timeline of human migrations can be generated. For example, the Iceland sequence of *H. pylori* is predicted to be of Asian origin (based on similarity to a sequence obtained from human remains in Asia) that had presumably migrated to Europe, as this particular sequence has also been

found across Europe. *H. pylori* strains from Africa had not yet moved into Europe, but evidence from more recent fossilized remains shows intermixing between African strains of *H. pylori* and Asian strains in Europe, giving rise to a hybrid European strain.

Fast-forwarding to more recent times and the coronavirus pandemic, researchers reported that a particular group of genes on chromosome 3 (six genes to be exact) was associated with a higher risk of severe Covid-19 (the disease caused by the coronavirus). In November 2020, a study reported that this stretch of DNA on chromosome 3 was derived from Neanderthal. Interestingly, it turns out that there are wide differences in the number of individuals around the world with this particular version of genes. While differences do typically exist between gene versions across populations, the range of people carrying this Neanderthal version is quite large. In Bangladesh, it is estimated that 63% of individuals carry one copy of the Neanderthal version, whereas in Europe, only 16% are estimated to carry it. In another study published in June 2020, a comparison of Italians and Spanish patients that became seriously ill and those who did not found that the Neanderthal genes were more common in those who became seriously ill. The significance of the genes is not clear at this time but researchers are continuing to look for clues to explain the huge range of outcomes upon infection, ranging from asymptomatic to death.

Egyptian Mummies

The oldest known Egyptian mummy dates back to 3300 BC, and was nicknamed "Ginger" because of the color of his hair. While not as famous as some Egyptian mummies (namely, the pharaohs), Ginger had a modest burial site surrounded by pottery. The word "mummy" is derived from the medieval Latin word *mumia*, which was derived itself from the Arabic word *mūmiyyah* meaning bitumen—a black, sticky tar-like compound used in the embalming process. It was believed that mummification would provide safe passage to the afterlife.

The practice of mummification results in the preservation of soft tissue such as skin and muscle. Sometimes the preservation is so good that the facial features are distinguishable thousands of years after their death. Mummification was not a quick process, lasting up to 70 days from the time of death with much of the time spent on drying out the body. This could occur naturally (dried out by extreme temperatures) or through a chemical process known as embalming. This ritual of mummification was not performed for everyone. It

was typically reserved for high-ranking officials, priests, and other nobles in the royal court, but anyone who could afford it could be mummified. It is estimated that 70 million Egyptians were mummified over a period of 3000 years. The practice of mummification declined as Egypt embraced Christianity and no longer believed that it was necessary to ensure passage to the afterlife.

Despite the huge number of mummies believed to exist, most of them have been destroyed by vandals and treasure hunters. Fortunately, the most important individuals had the largest and more elaborate burial sites, protecting them until today. Archaeologists can tell a great deal about the individual and the culture of the time from examining the place of burial and artifacts within the tomb as well as analyzing bone structure to determine the cause of death, lifestyle, and age and sex of the individual.

Probably the most famous of the Egyptian rulers was King Tutankhamun or King Tut for short. A boy king who died at the age of 19 after ruling for 6 or 7 years, his is the only pharaoh's tomb discovered intact in 1922, in the Egyptian Valley of Kings. It is presumed that he died of tuberculosis, although a bone fracture in his left thighbone was revealed by a CT scan of the wrapped body. Although King Tut is believed to be of royal heritage, his exact lineage is unclear as he may be either the son of King Amenhotep III or Amenhotep's son Akhenaten. If permitted, DNA analysis could reveal the true lineage of King Tut. The Egyptian Government has granted and rescinded permission to remove a sample of tissue from King Tut for DNA analysis. However, given the decayed state of the mummy, some doubt that any viable DNA could be extracted if permission were granted again.

In 1985, Dr. Paabo published the first report on the extraction of DNA from an Egyptian mummy. Twenty-three mummies dating as far back at the Sixth Dynasty to Roman times (~2370 to 2160 BC) were sampled. Most of the mummy specimens contained no DNA, except for that of a 2400-year-old 1-year-old boy. This work confirmed reports by others that DNA is more likely to be found intact in superficial tissues rather than tissues from inside the body cavity, likely due to the harsh drying-out process of the body.

In 1993, Dr. Woodward of Brigham Young University was asked if he could analyze DNA from six Egyptian mummies from the Fourth and Fifth Dynasties (2570–2290 BC). The museum caretakers were particularly interested in confirming the sex of and establishing the relationships between these six individuals. Based on the visual inspection, it appeared to be a three-generation family—two grandparents, two parents, and two children. The face masks and names written on five of the sarcophagi indicated whether the occupants were male or female. DNA analysis revealed that these six

individuals were indeed a family. However, DNA analysis of the sex of the two parents showed that the bodies had been switched (the male was found in the female-labeled sarcophagus). X-ray analysis showed that the family had been executed as each had a broken neck.

Further work by Dr. Woodward and his team established relationships between several pharaohs discovered in the late 1800s and early 1900s. King Ramesses III, Ramesses II, Seti I, Amenhotep I, Seknet-ra, and others are on display in the Cairo Museum. Dr. Woodward was given unprecedented access to each of the mummies as they were being moved to a new exhibit hall. Of 11 samples taken, DNA was extracted from 7 mummies. DNA analysis confirmed that Ahmose I married his full sister, Seknet-re, as they shared mitochondrial DNA sequences which would have been directly passed on from their mother. In addition, it was found that Amenhotep's mitochondrial DNA was different from Ahmose's, which was expected as his mother was not a direct descendent.

Molecular analysis of Egyptian remains has also revealed something about the types of diseases that were rampant during that time. As elsewhere, infectious diseases seemed to be quite common in ancient Egypt. Infectious diseases were described in ancient Egyptian medical papyri and genetic analysis provided an opportunity to detect the presence or absence of microbial DNA. In 2003, a survey of 85 Egyptian mummies discovered in Thebes West, Upper Egypt, an area known to be used for burials of the upper social classes was conducted to determine if the parasite *Mycobacterium tuberculosis* was present in the ancient remains. *M. tuberculosis* was positively identified in 25 samples, supporting earlier reports that ancient populations were afflicted by tuberculosis.

In 2006, a similar DNA analysis revealed that some ancient Egyptians were affected with a disease that is still present today, leishmaniasis. Leishmaniasis is an infectious skin disorder that can be particularly painful and fatal in some cases. Almost half a million people worldwide will die from this disease each year. Dubbed the "black fever" in India, the disease is believed by some to have originated in what is now Sudan. Prevalent in northeastern Africa, the Middle East, and Central and South America, researchers were curious as to whether ancient Egyptians were afflicted with this disease. A team of researchers led by Professor Albert Zink of Ludwig Maximilian University in Munich, Germany, analyzed DNA extracted from bone samples of 91 Egyptian mummies and 70 from old Nubia—modern Sudan. The team discovered mitochondrial DNA of the leishmaniasis parasite in nine of the Nubian mummies and four of the Egyptian mummies. The Nubian mummies date back to A.D. 550 and the Egyptian mummies are much older, dating back to 2050 to

1650 BC. Interestingly, no traces of disease were found in Egyptian mummies dated before or after this period.

Yet another common infectious disease still affecting thousands of Africans was present in ancient Egypt. Schistosomiasis, or commonly known as snail disease, is caused by a flatworm that often infects humans through the skin after exposure to contaminated rivers. The parasite was detected in samples of the liver and intestine of two mummies through genetic analysis and reported in 2014.

Implications for Evolution

Obviously, one cannot study ancient humans and other organisms without reconsidering the evolutionary paths to the present day. In simplest terms, some species became extinct, some diverged, and some evolved into the modern species. For humans, the last common ancestor of modern humans and Neanderthals is still uncertain. Across Asia, another species coexisted called the Denisovans. The relations between modern humans, the Neanderthals, and the Denisovans were believed to be relatively limited, though genetic data confirms that modern humans have traces of Neanderthal DNA as described earlier, indicating that some interactions took place. The birthplace of modern humans is Africa and the migration out of Africa led to the eventual displacement and extinction of Neanderthals (or they coincidentally disappeared at that time).

Genetic analysis has and will continue to revise and refine the evolutionary relationships and the timelines heretofore not possible based on archaeological data and dating of fossils. DNA sequences continuously change over time—some will persist and some will disappear. Those changes that do persist and increase in frequency with each generation contribute to the evolutionary process. Microevolution is a term that refers to changes in the frequency of genetic variations in a given species. While most singular genetic changes do not result in obvious physical or behavioral features, over time, the accumulation of these changes leads to the gradual transition of different measurable forms. In contrast, macroevolution refers to more global changes that affect multiple species or populations. As science continues to collect and document changes in life that preceded present-day humans, a more challenging question is to understand what factors caused these changes.

For those who believe in other theories of human evolution and origination, these data will not likely change personal beliefs and may be viewed with skepticism that modern humans, in particular, evolved from lower species.

The connection between modern humans and the stereotypically portrayed "lesser" human species of Neanderthals may be viewed as an insult, though archaeological data suggest that the Neanderthals were intelligent and skilled hunters with their own system of communication. Given the sheer scale of data and the increasing contribution of genetics and environmental impacts, some believe that the complexity can only be the work of divine intervention.

Conclusion

Genetic analysis of fossilized remains could revise or even rewrite current understanding of how humans came to be. And it is possible that there is more than one story, as the populations of different continents may have derived from different ancestors. For example, there is debate about some remains in China dubbed Peking Man and how this group fits into the history of human migrations. In addition to understanding human origins and relatedness between groups and to modern humans, genetic analysis of ancient remains has begun to yield some insight about the health of these populations. But genetic data provide just another piece of the story, considering that scientists are working with very degraded specimens and do not have a "full" picture. Nonetheless, these technologies are providing new insight that was not possible 50 years ago. So, yes, traces of Neanderthal DNA exist in us, and that is probably a good thing.

References

Paabo, S. *Neanderthal Man: In Search of Lost Genomes.* Basic Books: New York, 2014.

Qiu, J. How China Is Rewriting the Book on Human Origins. Scientific American. July 2016. Available at https://www.scientificamerican.com/article/how-china-is-rewriting-the-book-on-human-origins/

Callaway E. Evidence mounts for interbreeding bonanza in ancient human species. Nature. February 17, 2016. Available at https://www.nature.com/news/evidence-mounts-for-interbreeding-bonanza-in-ancient-human-species-1.19394

Roebroeks W, Soressi M. Neandertals Revised. Proc Natl Acad Sci USA 2016; 113 (23): 6372–9.

Akst J. Neanderthal DNA in Modern Human Genomes Is Not Silent. The Scientist. Sept 1, 2019. Available at https://www.the-scientist.com/features/neanderthal-dna-in-modern-human-genomes-is-not-silent-66299

Medline Plus. What does it mean to have Neanderthal or Denisovan DNA? Available at https://medlineplus.gov/genetics/understanding/dtcgenetictesting/neanderthaldna/

16

(Re-)Creating New Life

Recall the blockbuster film *Jurassic Park* based on a novel with the same title written by Michael Crichton. In the movie, dinosaur DNA was extracted from a mosquito that had been feeding on a dinosaur. After its dinosaur meal, the mosquito gets caught in some tree resin and unfortunately died shortly thereafter. Within the hardened tree sap, the insect and its bodily fluids, including the dinosaur blood, were protected and conveniently sealed from the environment for millions of years. The discovery of the fossilized tree resin, also known as amber, with the mosquito perfectly intact provided a source of relatively intact dinosaur DNA, whereas prior, the only remaining artifacts from which DNA could be extracted from, although severely degraded, were dinosaur fossils. In the novel, scientists used the well-preserved dinosaur DNA extracted from the mosquitoes trapped in amber to reconstruct the dinosaur genome and ultimately re-create dinosaurs. Gaps from the deteriorated parts of the dinosaur genome were replaced by frog DNA, a presumably close relative. But from there, the experiment goes haywire as the dinosaurs roam wild beyond human control.

While the movie is based on some facts, a dinosaur has never been re-created in real life. Until recently, the idea to actually re-create or synthesize an entire organism from scratch had only been possible in the imaginations of science fiction novelists. Fiction has now become reality as the ability to re-create (or create anew) the DNA code from scratch is possible today.

S. Haga, *The Book of Genes and Genomes*, https://doi.org/10.1007/978-0-387-70916-1_16

It All Starts with DNA

While the idea of a dinosaur park may not be believable, the premise of the movie is not as far-fetched as you might think. In the 1970s, scientists had figured out how to manipulate genetic sequences by cutting and reattaching them and switching sequences from one species to another. Although these experiments were done with very small segments of DNA and primarily in bacteria or other simple organisms, the power of such technology was scary even back then. Scientists were quick to respond to public fears if the field proceeded unchecked and set up an oversight system to prevent, as much as was possible, potential harmful consequences of the field without halting its progress.

As you might imagine, recombinant DNA technologies, as it was initially called, have advanced substantially in the past 40 years. Advances in technologies to manipulate gene sequences combined with information gained from sequencing hundreds of genomes have led to the creation of a new field called synthetic biology or synthetic genomics. Simply put, the new field centers on the synthesis or construction of genes, pathways, or whole genomes. Scientists can modify different parts of genes, combine one or more genes together from potentially different species, and knock genes out in the DNA code or genome—analogous to the major functions of any writing software program today: cut, paste, copy, insert, and delete. But compared to recombinant DNA technologies of the 1970s, today's tools are much more precise and scientists are working with much larger sequences than ever before. While most of these technologies are primarily being tested and used on organisms that can be easily manipulated in the laboratory such as bacteria, viruses, and yeast, they can also be used on more complex organisms such as mice and even human cells.

Today, it is possible to create a DNA sequence from scratch (instead of taking an existing piece of DNA and moving it or cutting it). So what do we mean "from scratch?" As in cooking, a scientist would start with the main ingredients of a genome—DNA. As we already know, DNA is comprised of four chemical bases (A, T, C, and G) arranged in a very precise order or sequence. Different parts of the genome (and different sequences) correspond to different functions—the best known are the actual genes that encode for proteins. But other sequences control when these genes are turned on (activated), what signals can activate them, how much gene product (protein) is made, etc. In some cases, scientists know about each of these parts, but in many instances, they do not. Potentially parts can be switched for other parts to attain a slightly different function. One can think of it like those

make-your-own salad eateries—you would choose an item from each of the following categories to make your complete salad: greens, protein, additional vegetables, and dressing. Each time you visit, you may vary your selections based on your mood, available items, specials, etc. Sometimes, the combination of ingredients selected did not taste as good as you imagined. The same could be true for a gene—the combination of parts selected may not yield a fully functional gene.

Synthesis

If one wants to re-create an exact copy of an existing gene present in any species, scientists can utilize lab protocols to do that quickly and accurately. But if one wanted to make changes to an existing gene, the new "parts" may need to be synthesized and then linked together, depending on how substantial the change is. To do this, scientists would order a specific DNA sequence to be made—companies that synthesize DNA would construct the sequence in the exact order of A, T, C, and G as requested. For small changes, however, it may be possible to do this within the laboratory using "gene editing" technologies (imagine "editing" the letters of DNA like you would when you edit a document) right within the genome.

If one wanted to re-create a genome from a species that was no longer living, they would have to rebuild it one small piece (sequence) at a time. The recipe for the genome or the exact DNA code may have been determined from analysis of fossilized remains. Therefore, one already knows the exact order of the DNA letters—the challenge is trying to reconstruct it from scratch.

As of the time of this writing, it has only been possible to synthesize an entire genome for small organisms (bacteria and viruses). Scientists start with small snippets of DNA fragments (between 50 and 100 bases long) and attach them together in the correct order to create larger pieces, and then these larger pieces are linked to make even larger pieces. Eventually, an entire intact and complete genome can be built in the laboratory (see Fig. 16.1). This can be a very long and laborious process as every time another piece is attached, scientists must confirm the sequence of the newly lengthened piece of DNA before attaching the next DNA fragment.

These technologies for synthetic genomics extend from what was described in Chap. 15 with genetically modified organisms (GMO). While GMOs typically have used less complex laboratory processes, it is certainly possible for future GMOs to take advantage of more sophisticated tools. Newer synthetic

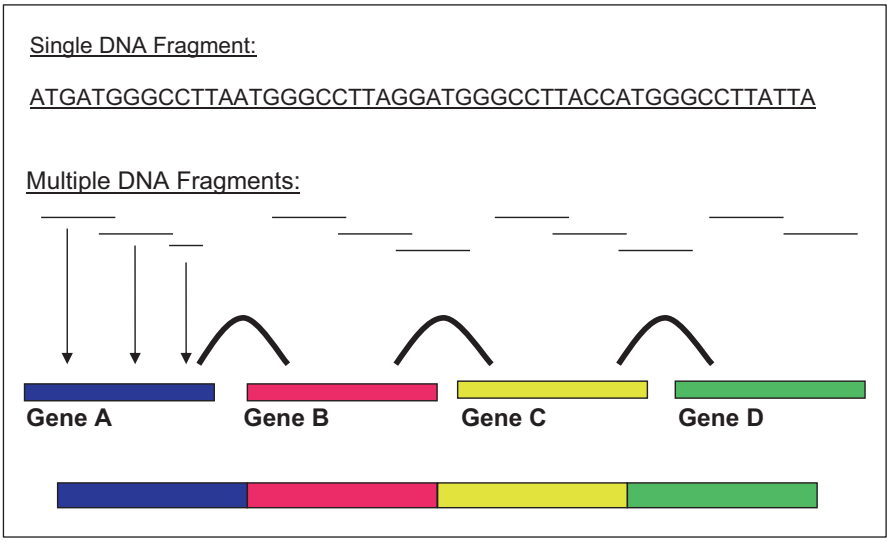

Fig. 16.1 Synthesizing a genome by taking smaller pieces of DNA fragments and "stitching" them into longer pieces until a single long piece is achieved that is the genome of an organism

genomics projects are now incorporating gene editing technologies as will be described later in this chapter.

At this point, you might be beginning to wonder how exactly this technology can be used to improve our health and/or environment or advance our understanding of basic biological processes. Some examples are described in the next section. Alternatively, you might be thinking whether this is the dream of a mad scientist (or Hollywood) to use the technologies to build the DNA backbone for some creature that only exists in one's imagination? If fear or alarmist thoughts are rising in your mind, you are not the only one. More often than not, significant concerns have dominated the discussion regarding the potential misuse of these technologies. While good can come of it, so can harm. The potential synthesis of a deadly pathogen may much more likely pose greater problems for society than the re-creation of an out-of-control *Tyrannosaurus rex*.

Re-creation of Deadly Microorganisms

The history of the field of synthetic genomics is relatively short, but as we know from past experience, it can accelerate at a very rapid pace. In 2002, Dr. Eckard Wimmer and his colleagues from the State University of New York

were the first group to assemble a live, infectious poliovirus from customized mail-order DNA fragments. Once a destructive and ultimately fatal disease, polio has been eradicated in much of the world due to global vaccination campaigns. While the sequence of the polio genome is publicly available, samples of polio are locked away in a few secure facilities across the country. It took his group approximately 3 years to assemble the 7500-base-long polio genome. To test that what they had assembled was indeed infectious and "alive," they infected cells grown in the laboratory as well as mice to demonstrate that their synthetic version of the virus had similar qualities to the natural strains.

In 2003, Nobel Prize-winner Hamilton Smith and colleagues from the Venter Institute in Rockville, Maryland, came up with a much faster method of genome assembly when they constructed a bacteriophage (a type of virus that infects bacteria) in only 2 weeks (a little smaller than the polio genome at 5389 bases). Thus, in just a year, the dramatic reduction in time to assemble a genome was astounding but not surprising.

But perhaps what has attracted the greatest amount of public attention to this field was the re-creation of one of the most deadly flu viruses of all time to date. In the spring of 1918, a relatively mild but highly contagious flu virus spread from town to town across Europe. The flu was dubbed the "Spanish" flu as one of the first towns to be struck was San Sebastian, Spain. After a brief remission, it returned with a vengeance and struck populations around the world with much deadlier symptoms in the fall of 1918. For the majority of those infected, its initial symptoms seemed like the typical flu—body ache, fever, chills, and headache. But for a subset of flu victims, in only a matter of days, they succumbed to a buildup of bloody fluid in their lungs, literally drowning to death.

In contrast to other flu epidemics, this strain particularly affected young and healthy individuals.

When all was said and done, it affected more than 25% of the US population, culminating in a worldwide death toll estimated to range from 20 to 50 million. It devastated not only families, but also entire villages and a substantial proportion of the populations of small countries as well as the several military service members battling in the World War I, including the USA.

In the 1990s, scientists from the US Armed Forces Institute of Pathology (AFIP) searched their archives to find tissue specimens that had been preserved from US servicemen that had purportedly died of the flu in 1918. As most of the those infected with the flu actually died of a second infection (bacterial pneumonia), viral particles were not present in the majority of specimens that had been stored. Of 78 samples examined by the research team,

two were found to contain some remnants of the virus. In addition, the scientists traveled to Alaska where they exhumed four bodies believed to have died from the Spanish flu from a mass grave. Lung tissue from one of the victims, an Inuit woman, was found positive for the flu virus, preserved for decades due to the frozen permafrost conditions.

Between 1997 and 2005, through painstaking analysis of very small sections of the viral genome, the virus's eight genes were eventually sequenced one by one, albeit with some gaps due to degraded DNA, totaling about 13,000 DNA bases. Then, in 2005, a team of scientists from the AFIP, the CDC, Mt. Sinai School of Medicine, and the USDA announced that they had reconstructed the Spanish flu virus using synthetic genomics techniques. Since a complete viral genome could not be determined from victims, scientists used the backbone of a closely related flu virus to fill in the gaps (sound familiar?).

After the virus was re-created, scientists tested its virulence in mice. In other words, could what they re-created in the lab actually be brought back to life? Experiments with mice infected with the man-made Spanish flu resulted in death in 3–5 days, showing severe lung inflammation similar to what was reported in human victims.

The re-creation of the Spanish flu virus generated a range of public response as evidenced by the major newspaper headlines around the world. While some papers chose to herald the scientific achievement (the New York Times proclaimed "Experts Unlock Clues to Spread of 1918 Flu Virus"), others focused on security issues (the London paper, The Guardian, announced "Security Fears as Flu Virus that Killed 50 Million is Recreated"). The San Antonio Express-News asked its readers frankly "Do We Really Want to Fool with 1918 Virus that Killed 50 Million?" Unlike the other earlier synthetic genomic experiments, the Spanish flu virus did not currently exist anywhere in its natural state. The paper was therefore a recipe for how to essentially re-create a lethal virus from a sketchy map of its genome.

To date, the largest genome to be re-created is from the bacterium *Mycoplasma mycoides*, with just over 1000 protein-coding genes (recall that humans have an estimated 20,000 genes) and a genome size of 1,077,947 bases. It is not harmful to humans, but can infect livestock, mostly cattle and goats. A team of scientists aimed to reconstruct this large genome piece by piece and actually bring it to life. Indeed, they were successful in doing so, laying the groundwork for the next team to build something even bigger.

Minimal Genome

Unlike bacteria and viruses, higher organisms have a lot of DNA that does not encode for genes. For much of these nongenic (or noncoding) sections of DNA, scientists are not exactly sure if it serves a purpose. In other words, are these long stretches of DNA that do not have any genes necessary for survival? In humans, only about 3% of the genome encodes for genes that make proteins—so that is a lot of noncoding DNA that we are carrying around. For a while, the noncoding DNA was referred to as "junk" DNA. But recent study has suggested that maybe it is not junk; it turns out that some of these DNA gene deserts play important roles in regulating gene expression (serving as an on/off switch for genes).

Other higher level species also face a similar situation—is all that extra DNA necessary for survival? Imagine how much more streamlined and efficient the process of cell growth and division would be if the entire genome did not need to be replicated each time a cell divides. Scientists curious about this have wondered what the minimum number of genes actually is for survival (for an organism to reproduce and perform basic functions).

Referred to as the "minimal genome," scientists have relied on computer modeling and knowledge of basic biochemical pathways to estimate the minimum number of genes. But only lab experiments can determine if the computer models are correct. Two approaches can be used to create an organism with a minimal genome. The first approach begins with an existing organism and actually knocking out (deleting) those genes that are deemed unnecessary for survival—a top-down approach. After removing those genes deemed nonessential, a series of tests can be performed to determine if the cells are growing and behaving normally. Experiments in which genes are randomly knocked out of the bacterium *Mycoplasma genitalium* have demonstrated that only a subset of the genes (between 265 and 350) are actually required for the bacterium to survive under laboratory conditions. In contrast, an organism's genome can be created from scratch by building a DNA sequence with only the genes hypothesized to be necessary for life—a bottom-up approach. As no genome greater than 13,000 bases (the Spanish flu virus) has been synthesized, the type of experiment has not been performed yet although it is probably possible to do.

Creating a "leaner" cell, so to speak (or trimming the fat!), could serve a number of industrial purposes, such as food processing (think beer, cheese, etc.), environmental applications (ridding the soil of chemical contaminants), or medicine. Potentially, a shell of a cell might be constructed and then specific genetic functions added to achieve the desired application or use. At this time, these are still futuristic scenarios but have been contemplated.

Re-creation of Extinct Species: Step 1—Obtaining DNA Sequences

One of the possibilities of re-creating genomes is the potential to bring back extinct species. De-extinction, as it is known, would require a genome to be synthesized, which would subsequently be transferred into an egg to develop (Fig. 16.2). With the exception of species that became recently extinct (let us say within the last 150 years), only partial genomes would likely be available from many extinct species (extracted from fossilized remains). In those cases, it may be possible to use the genome of a related species to fill in the gaps of the extinct species' genome (recall Jurassic Park). However, this would likely not yield an exact re-creation of the extinct species but rather a hybrid of two species, the extinct species and the other species whose DNA was used to fill in the gaps. An oversimplified description of the steps to recreate an extinct species are described below is with many details left out (or even unknown at this time).

Ancient DNA: Dinosaurs

In 1962, Dr. George Poinar found a piece of amber on a beach in Denmark with a fly embedded in it. An insect pathologist, this marked the beginning of Dr. Poinar's extensive collection of amber. In 1982, 20 years after his initial finding, he and his wife, Dr. Roberta Poinar, observed a completely intact mosquito imbedded in one of their amber specimens. More than just the skeleton of the mosquito, actual cells were visible in the 40-million-year-old specimen, a shocking discovery to say the least. The exquisite preservation of the insect gave way to the idea that if cells were still intact, perhaps the DNA inside the cells was also still intact. Furthermore, maybe the cells (and DNA) from the animals the mosquito had been feeding off of were preserved in the mosquito. The idea behind the premise of *Jurassic Park* was born.

Amber is actually fossilized tree resin. The formation of amber remains a mystery, but this million-year process produces a range of colors (off-white to black) and transparencies, with the golden color being one of the most popular. Although amber is found throughout the world, one of the largest amber deposits is found along the Baltic and North Seas. Amber is a very valuable commodity, dating back to prehistoric times, particularly to Greek and Roman cultures, with known amber trade routes between the North Sea and Rome and Greece. Amber is used to make ornaments, jewelry, and small sculptures, but was also believed to have medicinal properties and could ward off evil forces.

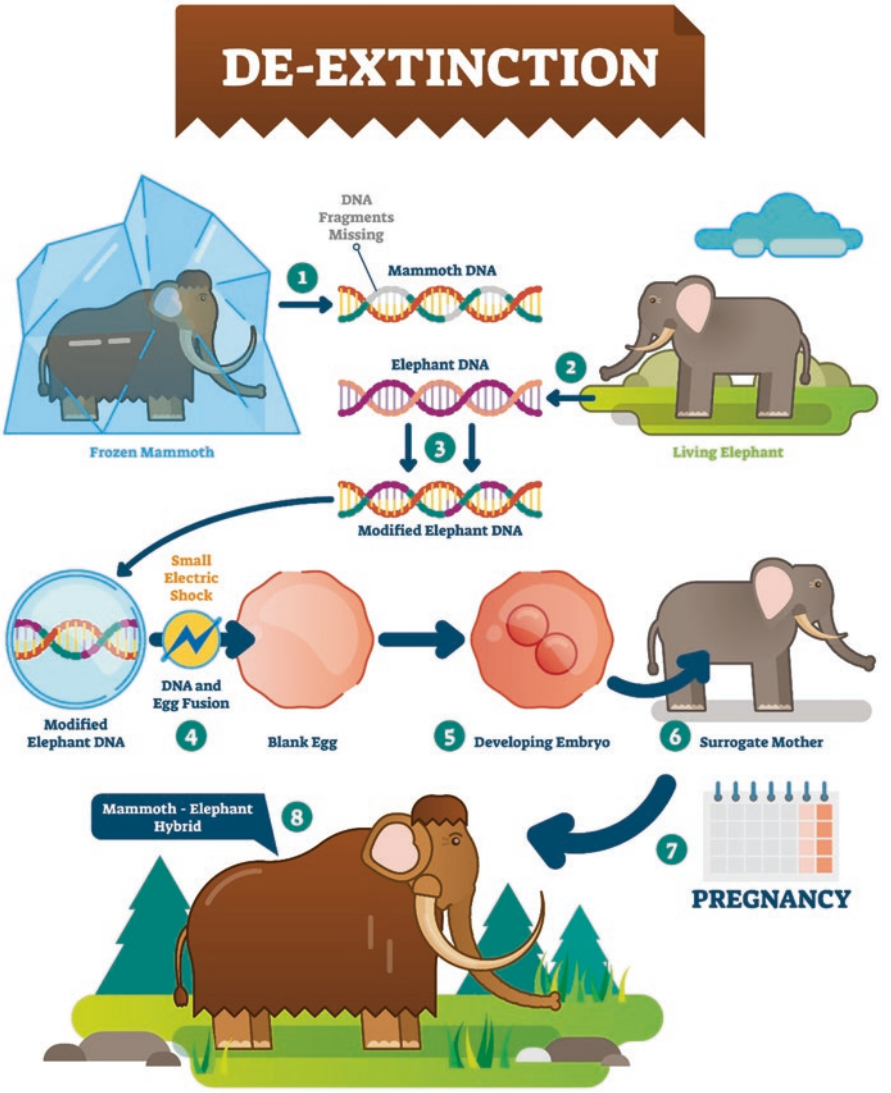

Fig. 16.2 Illustration of the steps for de-extinction of a woolly mammoth (source: Adobe Photo Stock)

While the amber served as an excellent protective barrier, scientists wondered if any viable DNA could be extracted and actually sequenced to compare to sequences of modern species. The opportunity to test the powerful technologies of genetics on fossilized materials and potentially change our understanding of evolutionary history had finally arrived.

The answer it turned out was yes. But it was not from the original discovery of the 40-million-year-old amber-trapped mosquito. Remarkably, small DNA fragments were extracted from a weevil trapped in amber estimated to be 120–135 million years old by the Poinar team (George, Roberta, and son Hendrik). In addition, DNA has also been extracted from a 20-million-year-old fossilized magnolia leaf, a bee, termites, and wood gnats preserved in amber.

In 1994, researchers from Brigham Young University reported extracting DNA from fossils found in a coal mine in eastern Utah. The fossils were embedded in rock found in an area where other dinosaur fossils had been discovered; however, their small size precluded definitively identifying them as dinosaurian. Estimated to be 80 million years old, only very small segments could be sequenced due to severe degradation. Comparison of the DNA sequence showed no similarities with any modern animal species, including birds which are presumed to be the descendants of dinosaurs. Following publication of the report, substantial debate ensued about whether the fossils were really dinosaurian.

In 1997, an astonishing report was published announcing that red blood cells were observed in bone from a *Tyrannosaurus rex*. The discovery was mind-boggling as soft tissues such as blood vessels and muscle decay over time, whereas hard tissues like bone become mineralized, lasting millions of years. Found in the Hell Creek Formation in Montana by renowned paleontologist Jack Horner, the fossil was estimated to be 68 million years old. By all accounts, traces of biological material in bones this old should have been completely degraded and undetectable. After soaking the bone fragments in acid, which will dissolve the bone but not the soft tissue, a slice of the bone was made and examined under a high-powered microscope. Furthermore, proteins such as hemoglobin (the molecule that carries oxygen in red blood cells) and collagen were detected and sequenced, similar to how a strand of DNA is sequenced. Comparison of the protein sequence to a variety of modern species showed that the *T. rex* protein sequence was most similar to that of the chicken. These data confirmed the long-standing hypothesis that birds are descendants of dinosaurs.

A major problem with any DNA extracted from remains, even relatively recent samples, is the issue of contamination. Depending on the length of time of decay, the type of tissue, and the environmental conditions or location, contamination is more than likely from other species (e.g., scavengers), present-day humans (or humans from when the fossils were originally found, discarded, and then rediscovered), and bacteria or other microbes. While DNA analysis can distinguish between some species, in some cases there may not be a comparison available, or the DNA is too degraded for analysis.

Woolly Mammoths and Other Recently Extinct Species

More recently extinct species may be easier to bring back for several scientific reasons, though they probably will not raise the same level of excitement as the *T. rex*. First, a recently extinct species may have well-preserved tissue stored in a museum or freezer, enabling scientists to extract and sequence its entire intact genome. Thus, an exact replica of the extinct species is possible.

Quagga

In 1984, DNA was extracted from an extinct species for the very first time. A team from the University of California led by Dr. Allan Wilson extracted and cloned DNA from an extinct animal known as the quagga. A quagga looks like a cross between a zebra and horse, but the exact lineage of the species was not clear. Native to Southern Africa and driven to unsustainable numbers by farmers trying to protect their crops, the last living quagga died in captivity in 1883 at the Amsterdam Zoo. Tissue was preserved and stored in a frozen repository at the San Diego Zoo.

The team extracted DNA from muscle and connective tissue cut from the skin section. Although the DNA was significantly degraded, molecular analysis was possible on short DNA fragments. After comparing the sequence of the quagga DNA to that of modern zebras and horses, the team determined that the quagga was more closely related to the zebra than horse. The molecular analysis was repeated in 1988 with new, more accurate techniques such as PCR. Fortunately, the findings of the second experiment exactly confirmed the initial findings.

Marsupial Wolf

In 1989, DNA from another extinct species, the marsupial wolf of Australia, was extracted. DNA analysis of this doglike animal found it to be related to other Australian marsupials but not to South American marsupials. Analysis of quagga and marsupial wolf DNA demonstrated that DNA extraction and analysis were possible from recently extinct species preserved in museums or biorepositories. The next obvious question was whether DNA could be extracted from fossilized remains.

Woolly Mammoths

Woolly mammoths were long-haired, elephant-like creatures with long, powerful tusks with the last remaining ones dying out about 4000 years ago. The majority of woolly mammoths were believed to have died out about 12,000 years ago at the end of the Pleistocene era. Several carcasses have been found across Siberia and Northern Europe. Although many were substantially degraded, a handful have been found fairly intact due to the frozen tundra conditions enabling preservation and minimizing degradation. But even under these conditions, only very small portions of the woolly mammoth's genome are able to be sequenced making it highly unlikely that the species could be resurrected from its genetic code.

In the early 1990s, DNA was extracted from a 40,000-year-old woolly mammoth known as "Dima." The fossil was discovered in 1977 in the frozen tundra in Siberia. Initial attempts to clone the DNA in the early 1980s failed due to technical difficulties. In the early 1990s, using the new technique of PCR, the experiment was repeated and short fragments of mammoth DNA about 350 base pairs long were amplified and sequenced. Comparison of the DNA extracted from the woolly mammoth to living elephant species proved that the woolly mammoth was related to both Indian and African elephants.

Preserved hair specimens from woolly mammoths suggest a range of dark- and light-haired animals. In humans, the gene MC1R has been linked to red hair and red or yellow hair in other mammals such as dogs, mice, and horses. Some of the hair color may be inaccurate due to pigment degradation or bleaching. In 2006, scientists sequenced this gene in DNA extracted from several woolly mammoth specimens. Based on the sequence and the different colors associated with different gene versions, the researchers confirmed that the coat color of woolly mammoths was variable, most likely brown or black. So not only can DNA analysis teach us something about the evolutionary history of extinct species, but it can also provide a more accurate picture of what these creatures looked like.

Re-creation of Extinct Species: Step 2—Genome Synthesis

Second, in the quest to bring back an extinct species, a synthesized genome must be re-created. No synthesis of a genome this large has been accomplished at this time, either for an extinct or living species with or without an

complete genome blueprint. There have been experiments performed where the DNA/genome from a cell from one animal has been moved and inserted into an egg of the same species (a "donor" egg) and implanted in that same species, giving rise to an exact replica of the source of the DNA. This is not an easy process, by any means, though certainly a lot easier than what would be required to synthesize a full genome and bring back an extinct species.

Even if DNA could be obtained, as noted earlier, there will be lots of gaps or holes in the DNA sequence. If the extinct species has a living relative, it is possible to use the DNA sequence of the living relative to fill in the gaps of the extinct species' genome (Fig. 16.2). Alternatively, the full genome of the living species could be edited to replace sections with sequence from the extinct species. With either approach, the end result will be a hybrid of two related species and the resulting product (the animal) is not certain. For example, while elephants and woolly mammoths are related, there are features that distinguish the two.

Re-creation of Extinct Species: Step 3—Development

After a synthesized genome is re-created, it must then be inserted into an egg to initiate development, and then implanted into a surrogate. Identifying a surrogate or closely related species for egg donation and gestation may be challenging for some species. Fortunately, for the woolly mammoth, an elephant egg could be used. However, elephant populations are dwindling, and many eggs would be needed for experimentation, which would require subjecting several female elephants to multiple rounds of egg collection (only 3–4 eggs could likely be collected from one elephant annually, and hundreds may be needed). Even if an egg with the synthesized genome started to grow and divide and was ready to be implanted, the elephant presents a number of physical challenges to transfer the embryo to the uterus that potentially would result in death. For other species such as the extinct wolf or dolphin even however, this process could be much more favorable for implantation and surrogacy. Alternatively, it might be possible to create an artificial egg (or environment) in which to inject the synthetic genome, but much still remains unknown in reproductive biology that would make this a viable option anytime soon.

Vaccine Development

Synthetic genomic technologies can be used for a range of medical applications. Vaccine development is one area that is being explored in this field and have been successful in the development of COVID vaccines. New methods of vaccine development are needed as the current approach does not provide the type of precision and flexibility to quickly develop new vaccines in response to a viral or bacterial threat. Current approaches typically use an inactivated (or attenuated) virus or bacteria, which can pose some risks. Furthermore if the sequence of the virus or bacteria changes, the vaccine may not be effective and a new vaccine would have to be developed. Scientists have been experimenting with using genes that produce specific viral or bacterial proteins, enclosing the DNA (or related molecule RNA) and some proteins in a bubble, and injecting it into an organism to determine if it would trigger an immune response. In 2020, two of the approved Covid-19 vaccines developed by the companies Pfizer and Moderna used a new technology called mRNA vaccine technology. Unlike traditional vaccines that use a weakened version of the virus to trigger an immune reaction in the body to fight an infection if exposed, these vaccines work through a piece of RNA (sister molecule of DNA) that encodes for part of a viral protein so that the body will then mount an immune response. So, the end result is the same—your body is primed to fight an infection if exposed to SARS-CoV-2 (the virus that causes Covid-19); however, these new mRNA can be made extremely rapidly (unlike a traditional vaccine) and can be modified as needed if the virus acquires genetic changes. With this approach, there would be no need to use the entire virus or bacterium and therefore no risk of infection.

Ethical Issues of Synthetic Genomics

To recap, synthetic genomics enables a lot more creative experimentation and new applications than genetic engineering or modification technologies, which have been typically limited to single genes. Whether attempting to replicate the DNA code of a known organism or experiment with the unknown through the synthesis of a new code, the power of synthetic genomics will take us to another level of scientific experimentation that will raise many ethical issues, reminiscent of the public concerns raised in the 1970s with the introduction of recombinant DNA technology.

In general, the concerns can be divided into two major groups: (1) practical concerns regarding safety and prevention of harmful misuses and (2) appropriate use of the technology in accordance with societal values and needs. Intertwined with the second concern are philosophical, moral, and religious beliefs that pose substantial challenges to reach a consensus about the best path forward.

To begin with, the safety of synthetically created organisms to the environment will likely be an ongoing concern. As with GMOs, genetically synthesized organisms may be indistinguishable from their natural counterparts and novel surveillance techniques may be required to promptly detect "man-made" organisms if they pose a threat in the natural environment or to human health. Despite careful testing in the laboratory, adverse impacts on ecosystems/environment cannot always be anticipated. To reduce the likelihood of escape of synthetic organisms into the natural environment, as with GMOs, two types of containment measures can be developed: (1) physical containment and (2) biological containment.

In addition to direct harm to the environment or ecosystem posed by the synthetically created organisms, there are other harms or threats to consider. Any technology or product can be used for multiple purposes—good and evil. This is referred to as "dual-use" technologies. A weapon can be made from common household supplies. A cellular phone can be used to detonate a bomb. Similarly, once the essential genes or the function of certain genes for an organism have been defined, a biological weapon can be constructed by combining the most destructive genetic components known or, worse yet, created anew. Imagine a genetic recipe for a biological weapon—one part Anthrax genes, two parts Spanish flu, one part botulism, etc. Genes can be mixed and matched, and enhanced and altered at random as long as the viability of the organism is not compromised. A genetic engineer will become just as dangerous as a nuclear or chemical explosives engineer.

Following the Spanish flu pandemic in the last century, there were a number of theories that the Spanish flu virus was actually a weapon of mass destruction—that the Germans had tainted aspirin sold to the USA by the pharmaceutical company Bayer with it or that it was brought by a German ship that docked in Boston Harbor and dispersed through the air into the city (recall that this was the time of the World War I). These scenarios sound eerily familiar to our society today, 100 years later. In 2003, before the re-creation of the Spanish flu was announced, the US Central Intelligence Agency convened a meeting of experts to ascertain the application of new scientific techniques in genetics and genomics to create biological weapons of mass destruction. In its report, they concluded that "the same science that may cure

some of our worst diseases could be used to create the world's most frightening weapons." The announcement of the re-creation of the Spanish flu and other genomes in effect provides the instructions for how to synthesize a genome, regardless of the specific genes or species.

This brings us to the second concern about the appropriate use of these technologies and who oversees or makes these choices. For example, should private laboratories be allowed to work on bringing back a dinosaur? Are there any rules that would prohibit a company from doing this? They could make a huge profit either by offering to make any animal that someone was willing to pay for and/or they could charge a fee to allow the public to view the animals they re-created. Aside from the legal issues about doing this type of activity, there are also environmental concerns (e.g., how would a large animal be contained to limit harm to the environment), and concerns about the health of the animal itself (nothing or very little may be known about the habitat, diet, or behavior of an extinct animal and uncertainty if the animal could even live in today's environment) and the safety of its caretakers.

Is this an endeavor in which federal tax dollars should be used to support? Who would decide which species to re-create? What could be the benefits? Potentially, scientists could learn a great deal about extinct species, promote public awareness about conservation and the environment, or benefit the current environment. Do we even have the right to bring extinct species back, much less deciding which ones should be brought back from extinction? Is that a decision that should be reserved for higher powers?

Perhaps lurking in the back of our minds is the question of whether a human could be genetically synthesized. Again through the imaginations of science fiction writers, one could foresee the technologies being used to create an elite fighting human machine (think of the Ork armies created to defeat man in J.R. Tolkien's The Hobbit). Genetic traits responsible for physical and mental prowess and strength could be combined or enhanced to create an indomitable army. Although the imagination is limitless, our scientific and technological capabilities are still a long way off from understanding the genetic controls of complex traits such as athleticism or behavior.

Can life as we know it actually be reduced to a minimum set of genes? This reductionist view purports that being "alive" is due solely to the physiological properties of an organism, as defined by the genetic code. It does not take into consideration nonphysiological experiences such as spirituality and an interconnection between members of a group that cannot be physically defined or influenced. Will the applications produced by the field of synthetic genomics place our society at risk of too narrowly defining what life is and in essence force us to reconsider what it means to be human?

Conclusion

The future is exciting with the advent of synthetic biology and the advances in science and potentially wide range of applications that this field may yield. But, as with many technologies, there is also a downside to consider. Developing guidelines and more public engagement will help define what is and is not acceptable and enable technologies to move forward with the proper oversight and restrictions. No doubt, the tools will continue to advance, and subsequently the potential uses of those tools, for good and for bad. Society needs to keep apace to maximize the greatest good and limit potential harms.

References

Worrall, S. We could resurrect the Woolly Mammoth. Here's How. National Geographic. Available at https://www.nationalgeographic.com/news/2017/07/woolly-mammoths-extinction-cloning-genetics/

Pickrell J. Ticks That Fed on Dinosaurs Found Trapped in Amber. National Geographic. Available at https://www.nationalgeographic.com/news/2017/12/tick-dinosaur-feather-found-in-amber-blood-parastites-science/.

Wray B. Rise of the Necrofauna: The Science, Ethics, and Risks of De-Extinction. Greystone Books: Vancouver, 2017.

Crichton M. Jurassic Park. Knopf: New York, 1990.

O'Connor MR. Resurrection Science: Conservation, De-Extinction and the Precarious Future of Wild Things. St. Martin's Press: New York, 2015.

Novak BJ. De-extinction. Genes (Basel). 2018 Nov; 9(11): 548. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6265789/

Roberts WS. The booming call of de-extinction. The Scientist. Oct 19, 2020. Available at https://www.the-scientist.com/news-opinion/the-booming-call-of-de-extinction-68057

17

Home-based Genetic Testing: A Brave New World

In medicine, patients do not typically order their own tests. The typical process involves multiple steps, most of which would be difficult for patients to complete on their own. To begin, before a patient can get a clinical test, it must first be authorized by a health provider. In most cases, only a health provider would have the expertise to determine which test(s) is indicated (or needed) for a given patient. There is typically a form, called a test requisition form (or order form), where the health provider indicates which tests the laboratory should perform for a patient and a signature is required. If you have ever seen one of these forms, they are often full of abbreviations of test names that are not typically understood by the average patient. Once a test requisition form has been signed off by a health provider with the test(s) to be performed checked off, most clinical tests require a biospecimen (e.g., blood) that cannot always be obtained by the patient directly. While a urine or fecal sample can be self-collected, blood samples are obtained by nurse or certified phlebotomist. Lastly, the proper storage and shipment are typically done by the health provider's office or at the testing laboratory.

The results of the test are then returned to the ordering health provider, who in turn will review the results with the patient. Sometimes results are communicated by phone to the patient, while other providers will require an office visit to review the results and discuss next steps. Many health systems and testing laboratories now have online patient portals where test reports and other medical records can be accessed directly by the patient. Interpreting the test results and deciding if any further actions are required is a complex step.

S. B. Haga, *The Book of Genes and Genomes*, https://doi.org/10.1007/978-0-387-70916-1_17

There are some exceptions to this general practice of ordering clinical tests. For example, pharmacists can order some tests related to medications and disease monitoring. Other tests can be purchased directly by individuals, such as over-the-counter tests like a pregnancy test. Some devices or machines can be purchased (some may require a health provider's authorization) to be used at home to monitor blood levels for sugar or other chemicals. And now, some genetic tests are available to purchase online with results shared directly with the consumer instead of the health provider. Known as direct-to-consumer (or DTC) testing, delivery of genetic tests without a provider has been both welcome and of concern.

Clinical Genetic Testing

There are lots of different types of clinical tests and lots of testing technologies. Standard blood tests may analyze levels of potassium, iron, and other chemicals that are important to your body's normal function. Other tests may examine levels of glucose (sugars), fats (cholesterol), or a hormone (testosterone). And yet other tests may look for the presence of a virus or bacterium.

A genetic test is another type of clinical test. While many genetic tests sequence DNA of one or more genes, other genetic tests analyze the function of a protein (is it functioning at the normal level?) or measure the level of a chemical in a pathway that is believed to be dysfunctional due to a genetic variant. For example, if chemical "A" is to be converted to chemical "B" but the protein responsible for the chemical reaction is not functioning, there could be high levels of chemical "A" and/or low levels of chemical "B" (sometimes there are backup pathways that may be triggered to address chemical imbalances).

Like other clinical tests, genetic tests can only be ordered by a health provider—in many cases, genetic testing is ordered by a clinical geneticist, or a provider who is specially trained in genetics. These tests are typically diagnostic, meaning the tests were ordered for patients that already had symptoms of the disease. In other cases, genetic tests are available to predict disease risk or susceptibility. These tests are typically ordered for unaffected family members that have one or more relatives diagnosed with cancer or another inherited disease who want to learn of their own risk. These tests are not available for all types of cancer, but only those that are highly heritable forms for which the gene(s) has been identified, such as breast cancer or colon cancer. For some diseases, a patient may wish to learn if they *will* develop a heritable disease,

such as the neurological disorder Huntington's disease or cystic fibrosis, if they have a family member that has been diagnosed with that disease.

Many of the examples described above are single-gene genetic tests. With greater understanding that diseases can be caused by one or more genes and with new testing technologies that allow for multigene analysis, single-gene tests are transitioning over to tests of multiple genes or even the entire genome. "Gene panel" tests enable quicker results instead of ordering testing for one gene at a time until a positive result is returned. For some patients, the disease or potential causative gene is unknown, and thus, a test that sequences the entire genome would need to be performed.

New Genetic Discoveries

With the sequencing of the human genome, many research teams raced to discover the genetic variants associated with a wide range of traits, disease risks, causes, recurrence, and response to medications. Thousands of discoveries have been published. However, not all of the discoveries held up to further scrutiny and were dismissed. The discoveries that were confirmed by other research teams were subsequently developed into clinical tests by commercial and academic laboratories. These tests could then be ordered by healthcare providers for their patients as needed. This new wave of tests fell into the category of "risk assessment"—the presence of a specific genetic variant increased one's likelihood or risk for that disease, but it was not diagnostic (for certain).

Many of the traits and diseases of interest to scientists, health providers, and the public are called "complex" diseases caused by multiple factors, both genetic and environmental (in contrast to single-gene diseases such as cystic fibrosis). And they also affect a large proportion of the population—diseases such as heart disease, diabetes, asthma, and cancer. Traits such as height, athletic ability, and intelligence are also complex—influenced by multiple genetic and environmental factors. New genomic technologies and declining costs enabled a more comprehensive analysis of individuals' genetic makeup to hunt for these genetic variants associated with disease. For many complex traits and diseases, the genetic and environmental contributions are still unknown—including the specific combination of genes and environmental factors, time of exposure to the environmental factor(s), degree of exposure, age of exposure, etc.—and the resulting level of risk. Even for well-understood risk factors, such as obesity, it is not totally clear what other factors can mitigate risk, length of exposure (e.g., how long someone has been obese to cause an increased risk), or degree of obesity (overweight or morbidly obese).

Advances in Testing Technologies

One of the major obstacles to sequencing the first human genome was that the sequencing technology did not initially exist to complete such a large project. When the idea was first considered, nothing remotely close to the size of a genome (not even teeny tiny microbial genome) had been sequenced. In total, the Human Genome Project cost about $3 billion and took 10 years to complete to sequence a single human genome. At that cost and pace, it would be highly unlikely that technology could be widely used by researchers, let alone as a clinical testing platform. Shortly after the Human Genome Project was completed, the US National Human Genome Research Institute set a goal of a $1000 genome, a comparable price to other medical technologies. In the not-so-distant future, it was envisioned that most people will have their genome sequenced (at least the parts relevant to health), which may become parts of the medical record and/or stored on a portable storage device such as credit card-like magnetic strip or memory stick.

Private initiatives were also announced to stimulate innovation. The X Prize Foundation, an educational nonprofit institute, created a $10 million Genomics prize. To win the Archon X PRIZE for Genomics, teams must successfully sequence 100 human genomes within 10 days for less than $10,000 per genome. Unfortunately, the prize was canceled in 2013, citing that the premise of the competition was outpaced by innovation.

The first genome of a known person to be published was that of Nobel laureate James Watson, the co-discoverer of the structure of DNA. Costing approximately $1 million (substantially less than the Human Genome Project), the sequencing was performed as a collaborative effort between the sequencing technology company 454 Life Sciences, the Rothberg Institute, and Baylor College of Medicine's Human Genome Sequencing Center. Since then, millions of genomes have been sequenced, mostly in research labs for the purposes of discovering causes of disease and other health conditions such as response to medications.

In the late 1990s, toward the end of the work on the Human Genome Project, other research teams were developing another type of testing technology called microarrays. As described in Chap. 8, a microarray resembles a microscope slide spotted with thousands of short sequences of DNA. Some of the sequences represent a copy of the normal sequence and others contain a genetic variant. An individual's DNA sample can be prepared and affixed to the microscope slide; if the sequence in the sample matches any of the individual DNA spots, it will adhere, resulting in a color change that can be measured and recorded. Therefore, instead of actually sequencing all of the As, Ts,

Cs, and Gs, a microarray can quickly identify genetic variants for a substantially lower cost. This is the main type of testing platform that DTC companies currently use. However, once the price of sequencing comes down further, companies may switch.

Bringing Genetic Testing Directly to the Public: No Doctor Required

The new knowledge of the influence of genetic variants on traits and diseases combined with the availability of more rapid, accurate, and cheaper testing technologies created a golden opportunity for the establishment of a new industry to provide consumers direct access to genetic tests online, without authorization from a healthcare provider.

In early 2000s, companies were established to offer genetic testing services directly to the public. Tests could be purchased online without a physician and just a credit card. These tests were not intended to provide information to inform medical decisions, like a typical clinical test would that is ordered by a health provider. However, the DTC tests included a wide range of traits and diseases. Once a test was ordered online, the company mailed a DNA collection kit directly to the consumer. Technologies had advanced such that a blood sample is no longer required for testing. Enough DNA can be obtained from a sample of cells scraped from the cheek lining or small amount of saliva that could easily be collected by the consumer without assistance from a health provider. The consumer then sends the DNA sample to the lab to be analyzed. The lab report is sent directly back to the consumer, typically accessible via the online account that was created when the test was ordered. It is up to the consumer whether they choose to share the report with their regular health provider.

This new approach to the delivery of genetic tests caused an upheaval in the testing and genetics communities, primarily due to concerns that the public did not have a good understanding of what they were purchasing and potential harms that could arise based on their test results. In response, the FDA and other government agencies began to take a closer look at these companies. In response to federal requests of clinical evidence in support of their tests and FDA review, many companies closed (described in more detail below). A few companies remain in business today, such as 23andMe, and some other companies allow consumers to request testing but provider authorization is still required. As the cost of testing continues to decline and more evidence

emerges about the genetic basis of diseases, we may expect more companies to return and make their services available again directly to consumers.

Recreational and Ancestry Testing

Aside from health-related DTC testing, many people are interested in tracing their family's origin based on DNA analysis. Genealogy research has traditionally been based on archived records such as birth and death certificates, census data, and other official documents. However, DNA analysis now presents another tool to ascertain one's family roots. Through analysis of DNA samples collected from people around the world, scientists have found certain genetic variants or signatures that are more common in one population than another. This is likely due to the age of population, migration (influx/efflux), hardships that cause high number of deaths, isolation (e.g., island populations), arranged marriages, culture, and other factors. Thus, genealogical DNA tests will look for the presence of some of these genetic variants that differ between populations to predict ancestry. In addition to learning about distant ancestors, it is possible to discover living relatives by the amount of shared genetic variants present between two samples (more than a coincidence).

It has been estimated that millions of people have purchased ancestry testing kits through companies like 23andMe and Ancestry DNA. Like the health-related kits, consumers can send in a brush of cheek cells or a sample of spit for analysis.

Other DTC companies offer what has been referred to as "recreational" testing to predict (or confirm) physical traits like earlobe shape or red hair color, diet preferences, behaviors, athletic build or ability, and other non-disease-related traits. Given the complexity of traits and behaviors, however, it is highly unlikely that most of these traits could be accurately predicted based on today's scientific understanding.

Government Oversight of DTC Companies

Manufacturers of medical equipment, devices, and medications (either prescribed or over the counter) must obtain approval from the FDA before they are allowed to sell their products in the USA. The approval is based on a thorough review of years of research to demonstrate that the drug or clinical device is safe and effective for the specific clinical indication for which it is intended

to be used. In contrast, government oversight of clinical testing is complex and sometimes unclear. Clinical testing falls into two general categories: (1) test kits manufactured by a company (all of the components are manufactured, packaged, and sold as a single use to testing laboratories) and (2) a test developed by a single testing laboratory. Manufacturers of test kits must obtain approval from the FDA before they can be marketed; however, clinical tests developed within a laboratory are not necessarily required to obtain FDA review, but subject to the agency's jurisdiction and position on the potential harm to patients (or consumers).

All clinical laboratories, regardless of if they are purchasing test kits or are using their own tests developed in-house, are subject to inspections of the laboratory environment, personnel, quality control and assurance, and test protocols. These inspections may be done through the state public health department or accredited professional laboratory organizations. Inspection reports are publicly available and labs must respond to violations cited in the inspection or risk losing their laboratory certification. Some have criticized that the inspections do not examine each test in-depth, particularly with respect to the clinical validity and utility of the test (e.g., is there evidence to demonstrate that the test is useful for clinical care?).

Beginning in 2010, the FDA began to take action on companies offering DTC testing services. The federal agency sent warning letters to five DTC companies regarding their testing services and the lack of FDA approval. In November 2013, the FDA sent warning letters to several DTC companies again, informing them that their testing services were not approved and that they should meet with FDA officials. In 2015, the FDA sent yet another round of warning letters to testing laboratories about their DTC offerings (Pathway Genomics, DNA4Life, DNA-CardioCheck, Inc., and Interleukin Genetics, Inc.) and the lack of approval for these tests. As a result, several companies decided to cease offering DTC testing or to close.

In addition to the FDA, the Federal Trade Commission has expressed concerns about the potential for harm of DTC tests to consumers. In particular, the major concerns are regarding the strength or validity of the claims—do we really understand how much a given genetic variant influences or impacts a specific trait or disease? In other words, what data are DTC companies using to support their test reports and are that data of high quality?

Currently, there is only one company offering approved health-related DTC testing, 23andMe. In April 2017, the FDA announced that it had approved a small test panel to be marketed by the company. Called the

23andMe Personal Genome Service Genetic Health Risk (GHR), the test is for ten diseases or conditions. In 2018, the US FDA approved marketing of two more DTC tests from 23andMe: a test to predict medication response and risk of adverse events called "Personal Genome Service Pharmacogenetic Reports" test (see Chap. 7) and a test to analyze genetic variants common in women of Ashkenazi (Eastern European) Jewish heritage in two genes associated with inherited breast and ovarian cancer.

Risks and Benefits

The DTC genetic testing market has raised lots of excitement and lots of concern. No other clinical tests are as easily available as DTC tests. While the industry has been drastically curtailed, it remains to be seen if new companies will open (or existing laboratories will offer a DTC service) if they can gather the necessary data to submit and obtain FDA approval first.

For health-related data, a positive result could cause anxiety or stress, or drive consumers to change their behaviors that may cause adverse outcomes. For example, consumers may urgently schedule a visit with a health provider and potentially undergo further testing or other interventions that are not medically indicated, despite their positive result, wasting time, healthcare dollars, and resources. A negative result can also be harmful if a consumer thinks that they are not at risk for developing a particular disease based on their test report and participates in risky health behaviors or disregards recommended preventive health screenings.

A second concern is whether consumers really understand what they are purchasing given the complexities of genetics and genome sciences. With no person to speak with and explain the benefits and limitations of testing, consumers are asked to review the information presented by the company and agree to undergo testing before they purchase the test. But how much of that information is really understood by consumers? In a clinical setting, for a patient considering genetic testing, a session with a genetic counselor is often scheduled to review information about the test and how it may impact one's health. This session can easily take 30 or 45 min or more to complete.

However, DTC tests offer the general public an opportunity to learn about themselves in ways not possible before. Patients and consumers alike may enjoy the freedom to gain insight about their genetic makeup regarding their family's roots, some fun things about themselves (e.g., such as a genetic explanation for why they do not like spicy foods), and some health-related things.

And despite the concern raised by government agencies, health providers, and ethicists, to date, no substantial harms have been reported. Perhaps consumers do not take the test reports as seriously as feared and understand that they should consult a health provider about their disease risks and prevention. Although still cost-prohibitive for many people with a price of around $99, the cost of DTC tests is likely to continue to decline with the development of cheaper testing technologies, increasing affordability.

Conclusion

With the new wave of genetic information and new testing technologies, the advent of DTC testing capitalized on these factors and public interest in DNA. The market has vastly changed since the early 2000s, but we may see a second wave emerge as the understanding of genetic and environmental factors related to disease continues to increase and testing costs decline. Development of good educational resources for the public by institutions and the government may help increase the public's understanding about the benefits and limitations of DTC testing and promote more informed decision-making.

References

U.S. Food and Drug Administration. FDA allows marketing of first direct-to-consumer tests that provide genetic risk information for certain conditions. Available at https://www.fda.gov/news-events/press-announcements/fda-allows-marketing-first-direct-consumer-tests-provide-genetic-risk-information-certain-conditions (April 2018)

U.S. Food and Drug Administration. Direct to Consumer Tests. Available at https://www.fda.gov/medical-devices/vitro-diagnostics/direct-consumer-tests (Dec 20, 2019a).

U.S. Food and Drug Administration. Lists of Direct-To-Consumer Tests with Marketing Authorization. Available at https://www.fda.gov/medical-devices/vitro-diagnostics/direct-consumer-tests#list (Dec 20, 2019b).

Regalado A. More than 26 million people have taken an at-home ancestry test. MIT Technology Review 2019. Available at https://www.technologyreview.com/2019/02/11/103446/more-than-26-million-people-have-taken-an-at-home-ancestry-test/

23andMe. Available at 23andme.com

Ancestry DNA. Available at ancestrydna.com

Index

S. B. Haga, *The Book of Genes and Genomes*, https://doi.org/10.1007/978-0-387-70916-1

Q