Published in *Genes for Life* in 2017 by Future Leaders (www.futureleaders.com.au)

1

Introduction to human genetics

Martin Delatycki and Grant Blashki

No field of medicine has changed in so dramatic a fashion as has been the case with human genetics. It is only 40 years ago this year that the first human gene was isolated. Back then, identifying a single chemical in the human genome took days. Now it is possible to sequence the three billion chemicals in the genome in a similar time. In this chapter, the basic terminology and principles of genetics are outlined. Some of the major leaps forward in gene technology are also explained, and how these have really opened the window to our understanding of human genetics. It may seem a bit confusing for someone who has not learnt about genetics before, and we have outlined some key terms and rules of genetics. You might refer to this chapter when reading other chapters in this book.

There are various metaphors used to explain genetics, and one of the most popular is to think of the human genetic code as a book. In this model, the entire genome is considered as a book and the chromosomes are like the chapters. And within each chromosome there are genes that you might like to think of as individual pages. The genetic code is like the text on each page, and the letters can be thought of as what we call nucleotides, though in genetics there are only 4 letters (A, C, G and T) rather than the usual 26 letters in the alphabet. This makes the genome a somewhat tedious book to read, but if one is patient and digs deep, it reveals amazing information about what makes us who

we are and what the future may bring. To extend the metaphor, the nucleotides form important patterns in groups of three called codons, and it makes sense then to think of these as the words on the page. These codons, like words, do indeed have meaning and instruct the machinery of cells, telling them which proteins to make. So, as you read the following paragraphs keep in mind this 'genetics as a book' approach if you find it helpful.

What are genes?

Genes are sequences of DNA that instruct the body to make proteins. The human genome consists of approximately three billion individual chemicals that are called nucleotides, which are sometimes called bases (like the letters in the book). There are four different nucleotides: adenine (A), cytosine (C), guanine (G) and thymine (T). The genes make up approximately 50 million of the three billion nucleotides in the genome. That is, only around 2% of the genome is made up of genes. The function of the other 98% remains largely unknown. Genes are divided into exons and introns. It is only the exons that instruct the body to make proteins. The introns are removed during the process of making proteins, as is shown in Figure 1.

How do genes result in a protein being made?

Within the exons of the gene, every three nucleotides, called codons, instruct the body to add a particular amino acid to an amino acid sequence that ultimately results in a protein being produced. The amino acids are the building blocks of the proteins. There are 20 amino acids (see Figure 2). Those with a mathematical bent will quickly realise that with four nucleotides there are 64 possible combinations of three nucleotides. Sixty-one of these encode for a particular amino acid. Some amino acids have only one codon while others have

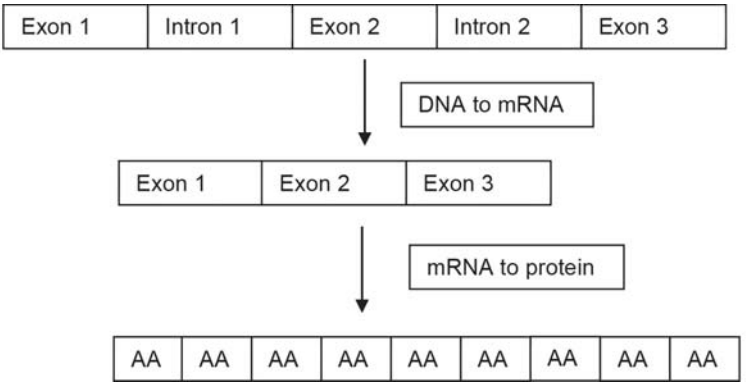

Figure 1: Genes, mRNA and proteins.
Note: AA = amino acid.

up to six codons that encode that amino acid. Three out of the 64 codons instruct the protein building apparatus to cease adding amino acids to the protein. These are called stop codons. The way that proteins are made is that the particular section of the DNA is copied into messenger RNA, and at this point the introns are spliced out and the codons from the messenger RNA are read to produce the protein.

Why is a skin cell different to a brain cell?

You might be wondering how the complex genetic code is transformed into making all the different types of cells in the body. All cells except mature red blood cells contain the same genome (mature red blood cells do not have a nucleus and therefore do not contain genes). The difference between cells is that in different cells, different genes are active and thus different proteins are produced. Unsurprisingly, since the brain is the most complex organ, more genes are active in brain cells than any other cells and more proteins are produced. So, to go back to the

1st base	2nd base								3rd base
	T		C		A		G		
T	TTT	(Phe/F) Phenylalanine	TCT	(Ser/S) Serine	TAT	(Tyr/Y) Tyrosine	TGT	(Cys/C) Cysteine	T
	TTC		TCC		TAC		TGC		C
	TTA	(Leu/L) Leucine	TCA		TAA[B]	Stop (*Ochre*)	TGA[B]	Stop (*Opal*)	A
	TTG		TCG		TAG[B]	Stop (*Amber*)	TGG	(Trp/W) Tryptophan	G
C	CTT	(Leu/L) Leucine	CCT	(Pro/P) Proline	CAT	(His/H) Histidine	CGT	(Arg/R) Arginine	T
	CTC		CCC		CAC		CGC		C
	CTA		CCA		CAA	(Gln/Q) Glutamine	CGA		A
	CTG		CCG		CAG		CGG		G
A	ATT	(Ile/I) Isoleucine	ACT	(Thr/T) Threonine	AAT	(Asn/N) Asparagine	AGT	(Ser/S) Serine	T
	ATC		ACC		AAC		AGC		C
	ATA		ACA		AAA	(Lys/K) Lysine	AGA	(Arg/R) Arginine	A
	ATG[A]	(Met/M) Methionine	ACG		AAG		AGG		G
G	GTT	(Val/V) Valine	GCT	(Ala/A) Alanine	GAT	(Asp/D) Aspartic acid	GGT	(Gly/G) Glycine	T
	GTC		GCC		GAC		GGC		C
	GTA		GCA		GAA	(Glu/E) Glutamic acid	GGA		A
	GTG		GCG		GAG		GGG		G

Figure 2: Codons and amino acids.

book metaphor, different pages are read for production of different cell types.

Sequencing

Before going into more detail about genetics, it is useful to understand a bit more about how scientists have been able to decode the human genome; in particular, a wonderful technique called sequencing. This major advance in genetics allows the nucleotide sequence in DNA to be ascertained. Once the sequence of nucleotides in a gene is known then the amino acid sequence of a protein produced by the gene can be easily elucidated. It is worth remembering that until the early part of this century, it was only possible to sequence relatively short sections of the genome at a time. Therefore, if a particular gene was suspected to be implicated in a person's medical condition, then short section by short section of the gene would be

sequenced and examined for mutations. This was a relatively slow and labour-intensive process.

A revolution in genetics occurred in the 2000s called next generation sequencing. Here, short sections of the genome are sequenced multiple times and analysed by a computer program that compares each sequence to a reference sequence. The hard work of analysing the sequence is done by computers that can identify changes in the nucleotide sequence compared to the reference sequence. As the technology has evolved, different versions of this sequencing have been developed. Whole genome sequencing refers to next generation sequencing that identifies the sequence of the approximately three billion nucleotides in the genome. Whole exome sequencing refers to next generation sequencing in which the nucleotide sequence of the exons, which encode the proteins, is deciphered.

Mutations

As in books, in genetics, spelling mistakes or even just variations in spelling can occur, and we call these mutations. In some cases, a mutation in a gene can alter the protein structure and therefore alter the protein function and lead to disease. Even a change in one nucleotide out of three billion can be sufficient to cause devastating genetic disease.

But not all variations in spelling cause disease. It is worth being aware that between normal healthy people, the genetic sequence of nucleotides differs quite a lot — on average in 1 in 1,000 positions (0.1%). These benign differences are called polymorphisms and are responsible for the differences between individuals. Differences in height, weight, intelligence, eye colour, hair colour and many other characteristics can be largely explained by these very minor differences.

In 2017, whole genome and whole exome sequencing has now been conducted on hundreds of thousands of human beings. This has helped to identify the benign differences and also assist in differentiating true mutations from benign alterations. Nevertheless, there is still some way to go in our understanding, and it remains the case that it is not always possible to identify whether a particular alteration in the nucleotide sequencing in an individual is a benign alteration or a disease-causing mutation. Such alterations are creatively called Variants of Unknown Significance (VUS).

Next generation sequencing means that while as recently as five years ago where a genetic disorder was suspected that one gene at a time was tested, it is quite remarkable that now it is possible to test for mutations in all known genes. This is especially helpful, as we now know that there are many conditions that can be caused by a mutation in multiple different genes. Whereas in the past it was impractical to test all genes due to financial constraints, it is now possible to identify mutations in genes that are rare causes of conditions, allowing those affected individuals and families to know the cause of their genetic condition. The case study about Max below illustrates how this new technology has been making a difference to people and the decisions that they make about having a family.

Case Study of Max in 2001

> Max is a 30-year-old gentleman who has Charcot-Marie-Tooth disease. Charcot-Marie-Tooth disease is a relatively common genetic condition where the nerves that provide electrical impulses to muscles become progressively damaged. It results in increasing muscle weakness and loss of sensation. At its most severe it can result in an individual requiring the use of a wheelchair for mobility. Max's brother, mother, uncle and grandmother all have or had this condition.

Max is seen in the Genetic Clinic to see if the genetic basis for Charcot-Marie-Tooth disease in his family can be identified. Max is recently married and is keen to take steps to prevent his children having Charcot-Marie-Tooth disease since it has significantly impacted on his quality of life.

In the year 2001 it was known that many different mutations in many different genes can cause Charcot-Marie-Tooth disease. Nevertheless, at that time it was only possible to test a handful of these due to the expense of testing genes and the knowledge that many genes only contribute to a very small percentage of Charcot-Marie-Tooth disease. Max had testing for mutations in PMP22 and MPZ and no mutation was found. He was informed that the genetic basis for his Charcot-Marie-Tooth disease could not be identified and therefore no test was possible to prevent the 1 in 2 chance for his children to have Charcot-Marie-Tooth disease. Max and his partner Jenny therefore decided against having children.

Case Study of Max in 2017

Consider the same scenario in 2017. Max attends the Genetic Clinic. DNA is tested by next generation sequencing. The approximately 40 genes in which mutations are known to cause Charcot-Marie-Tooth are assessed in detail. The causative mutation in LITAF is found. Max and Jenny can therefore utilise either prenatal diagnosis or preimplantation genetic diagnosis (PGD) to prevent their children having Charcot-Marie-Tooth disease.

Max's story is not an isolated scenario, and we are witnessing quite a revolution in the application of next generation sequencing. It is worth appreciating that of the approximately 23,000 pairs of genes, almost 5,000 have so far been linked to human disease. The advent of next generation sequencing

means that the genetic basis for multiple new conditions are coming to light on a weekly basis. Therefore, the number of genes linked to human disease will continue to increase rapidly over the coming years. The information that can be made available to families will help them make better informed decisions.

Of course, as with any new technology it is important to understand that there are some limitations. Next Generation Sequencing cannot identify the genetic basis of disease in every individual. Reasons for this include:

1. Next generation sequencing does not identify the nucleotide sequence of every nucleotide and every gene. Some regions of the genome are difficult to sequence for various technical reasons.

2. Some forms of mutation cannot be easily identified by next generation sequencing. An example of this is triplet repeat disorders. Some conditions, such as Huntington disease, are caused by an increased number of triplet repeats. A triplet repeat is a three nucleotide sequence that repeats over and over. Within the gene that is related to Huntington disease, *HTT*, three chemicals C, A and G repeat over and over — CAG,CAG,CAG and so on. This triplet repeat is normal, but in Huntington disease there are too many. Currently, next generation sequencing technology cannot identify triplet repeats, although this is likely to change as the technology improves. Another example is Fragile X Syndrome, which is discussed in Chapter 6.

3. Where the mutation is only present in some organs of the body. It is becoming increasingly apparent that some genetic disorders are not due to a mutation in every cell in the body but rather to a mutation in some cells in the body. The most common tissue to be tested for mutations is DNA from the

white blood cells. If the mutation that has resulted in disease is present in another tissue — for example, the brain — then testing DNA from white blood cells will not identify the genetic basis of the condition.

Modes of inheritance

Just when you thought you were getting a handle of some the complexities of genetics, there are some further rules that are critical to understand about the way in which genes are passed on to the next generations.

There are three main modes of inheritance which are critical to understand — autosomal dominant, autosomal recessive and X-linked recessive. As you read about these, keep in mind that our genes exist in pairs, half from our father's side and half from our mother's side.

In autosomal dominant inheritance, if there is a fault in only one of the pair of genes, the person will experience the condition. Because we get half our genes from each parent, there is a 50-50 chance for the child of a person with a dominant condition to inherit the genetic mutation, and a 50-50 chance they will inherit the copy of the gene without the mutation and therefore will not have the condition. Examples of autosomal dominant conditions include Huntington disease, neurofibromatosis and Marfan syndrome.

In autosomal recessive inheritance, a fault in each of the pair of a particular gene is required to cause the condition. Where a couple are both carriers of the same recessive condition, there is a 1-in-4 chance for each pregnancy to be affected by the condition and a 3-in-4 chance that the pregnancy will be unaffected. Examples of autosomal recessive conditions include cystic fibrosis, haemochromatosis and thalassaemia.

In X-linked inheritance the genetic mutation resides in a gene on the X chromosome. Because females have two X chromo-

somes, they are generally more mildly affected or unaffected compared to males. If a woman is a carrier of a faulty gene on the X chromosome, one in two of her sons will be affected by the condition and one in two of her daughters will be a carrier as she is. When a man with an X-linked condition has a child, all his daughters will be carriers and none of his sons will be affected. Examples of X-linked conditions include haemophilia, fragile X syndrome and Duchenne muscular dystrophy.

To add a further caveat, even though most genes' DNA reside in the what is called the nucleus (like a genetic storage centre in the middle of the cell) some genes are actually based in the mitochondria, a cellular organelle often described as the little battery packs of the cell. So, the vast majority of genetic disease is due to mutations in nuclear DNA, which consists of 23,000 genes. However, some rare diseases arise in the mitochondrial DNA, which encodes for only 37 genes. A key point is that mitochondrial genes are only inherited from the mother and so conditions due to mitochondrial DNA mutations can only be inherited from the mother.

Chromosomes

The DNA in the cells is organised into chromosomes which are long strings of nucleotides and as described earlier are like the chapters in the genetic code book. There are 23 pairs of chromosomes numbered 1–22, with the 23rd being either an X or a Y chromosome. Males have an X and a Y chromosome (see Figure 3), whereas females have two X chromosomes.

There are many chromosomal disorders that result from excess or missing whole or sections of chromosomes. The most common chromosomal disorder is trisomy 21, which results in Down syndrome, where there are three rather than two copies of chromosome 21. Microarray technology, also called array

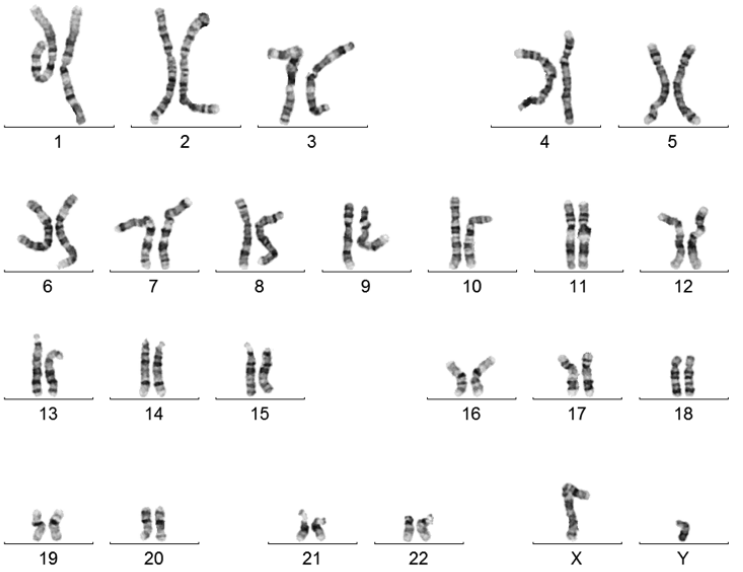

Figure 3: A normal male chromosome complement.

comparative genomic hybridisation (array CGH) or molecular karyotyping, has increased the identification of chromosomal abnormalities. This technology allows the identification of tiny slivers of extra or missing genetic material, which are called copy number variants (CNVs), that were not possible to identify by the chromosome assessment technologies used in the past.

Most chromosomal problems arise as a one-off, but it is important to be aware that some occur due to an inherited predisposition. This can be from a parent with the chromosomal copy number variant or because the parent has a so-called balanced chromosomal rearrangement that becomes unbalanced when passed on to the child. The main example of this is a balanced *translocation* where sections of two chromosomes swap place (see Figure 4). This is no problem for the parent because the correct amount of genetic material is present, albeit in a different order to the usual. When a person with a balanced

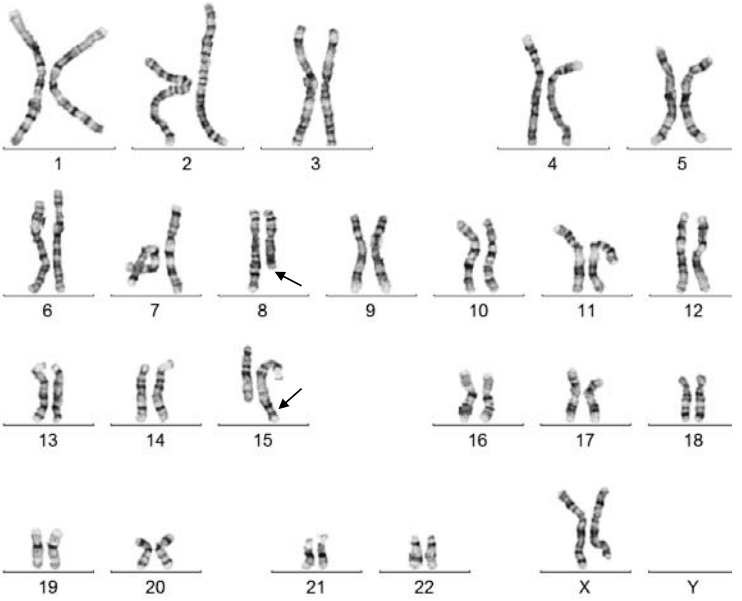

Figure 4: A chromosome translocation between chromosomes 8 and 15. The arrows point to the missing genetic material from chromosome 8 and where this is attached to chromosome 15.

translocation makes an egg or sperm, it can contain too much and/or too little genetic material, which can result in a miscarriage, or the birth of a child with significant problems such as intellectual disability and problems with organ structure.

Conclusion

We hope that this chapter has provided a basic outline of the principles of human genetics. With rapidly evolving technology, the level of detail and dropping costs of tests are at levels unimaginable even 10 years ago. We have provided an overview of the various ways in which conditions can be caused or passed on by genetics. At the same time, we have tried to paint a realistic picture of the uncertainties and limitations of

genetic technologies. We believe it important for the community to have a sound understanding of the basics of genetics in a world in which genetic technologies are going to play an increasingly important role, and hope that this chapter will contribute to this discussion.