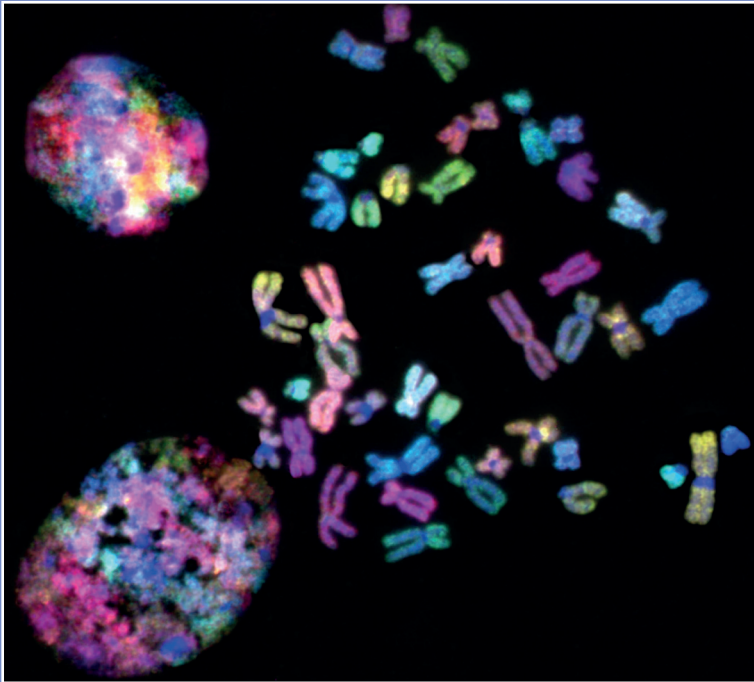


An introduction to the
**Genetics and Epigenetics
of Human Disease**

by Marcus Pembrey



An introduction to the Genetics and Epigenetics of Human Disease

by Marcus Pembrey

An introduction to the Genetics and Epigenetics of Human Disease

by Marcus Pembrey

Published October 2012 by

Progress Educational Trust
140 Gray's Inn Road
London WC1X 8AX
www.progress.org.uk

Editorial board

Professor John Galloway
Dr Tessa Hornfray
Sarah Norcross
Dr Vivienne Raper
Sandy Starr

Front Cover Photograph

The nuclei of three cells in which the DNA of all 46 chromosomes has been made visible, through the laboratory technique fluorescent in situ hybridisation. One 'exploded' cell that is in the middle of replication, shows compacted chromosomes already divided into two chromatids producing the familiar 'X' shape.

Image and photographic credits

Front Cover: Denise Sheer
Professor of Human Genetics
Barts and The London School of Medicine
and Dentistry, London

Figures 1-6: David Smithson, UCL Medical
Illustration Services and Marcus Pembrey,
Emeritus Professor of Paediatric Genetics
Institute of Child Health, University College
London

Figure 7: Neil Sebire
Professor of Paediatric and
Developmental Pathology
Institute of Child Health,
University College London

Contents

| | |
|--|----|
| Executive Summary | 6 |
| Introduction | 6 |
| Your genes and how they work | 8 |
| How you inherit your genes..... | 9 |
| Single gene disorders | 9 |
| Mendelian disorders..... | 13 |
| Dominant inheritance..... | 15 |
| Beyond Mendel's laws | 16 |
| Epigenetic regulation | 17 |
| Genomic imprinting disorders..... | 19 |
| The 'parental tug-of-war' hypothesis..... | 22 |
| Common diseases | 22 |
| Monogenic subsets in 'common disease' | 23 |
| Development and common multifactorial disease | 24 |
| The role of inheritance in common multifactorial disease..... | 25 |
| Case-control studies | 26 |
| Genome-wide association studies..... | 26 |
| Integrating genetics, epigenetics and the environment | 27 |
| Appendix | 30 |
| Glossary | 30 |
| Further reading | 33 |
| Key references..... | 34 |

Executive Summary

The last 20 years have seen great technical advances in the analysis of DNA and the completion of the Human Genome Project in 2003. These advances led to the discovery of which faulty genes cause which of the many simply-inherited (Mendelian) genetic diseases. This knowledge has given rise to greater understanding of the molecular basis of many Mendelian diseases, and why they have the inheritance pattern they do.

Some diseases are caused by a single faulty gene, but do not follow a Mendelian pattern of inheritance. Research into one of these diseases, Fragile X syndrome, revealed that a DNA fault can enlarge over three generations (a dynamic mutation) until a nearby gene is silenced (switched off). Another disease, Angelman syndrome, confirmed that some human genes are normally subject to genomic imprinting, a phenomenon in which a gene is silenced depending on whether it was inherited from father or from mother.

The molecular silencing process (DNA methylation) involved in genomic imprinting is a classic example of epigenetic regulation, in which there is an enduring change in gene activity but without any change in DNA sequence. Epigenetic regulation underpins normal development from the fertilised egg to an embryo with its many different cell types and organs, and may also be involved in response to early life experiences.

Understanding the role of numerous genetic variations and epigenetic regulation in common, multifactorial disorders – such as diabetes, heart disease and cancer – is the focus of much current research, and is proving to be a huge challenge. Adult health depends on a complex interaction between inheritance, nutrition and the physical and social environment throughout prenatal development and childhood.

Introduction

It is the differences that matter. You share about 75% of your genes with your dog but it is the remaining 'doggy' genes – those which make dogs have puppies not babies or kittens – that we see as important. Certainly it is the differences between people that interest us when it comes to health and disease. Why do some live a long and healthy life, whilst others develop life-threatening, chronic diseases in mid-life or indeed are born with a disorder?

A popular response nowadays is to assume most of the difference lies 'in our DNA' – in the 0.1% that we don't have in common with everyone else. This is understandable given the excitement surrounding the Human Genome Project and the ubiquitous use of the DNA double helix icon in all things biomedical. But DNA alone is not destiny.

Human development from conception to adulthood is an inseparable partnership, moulded within our cells, of Nature (the DNA we inherit) and Nurture (the prevailing nutritional, social and physical environment). Our DNA specifies the structure of proteins – the primary chemical building blocks of life – but it is the cell's circumstances that ultimately determine when, where and how much of these proteins are produced.

The response may be a transient adjustment, but sometimes our cells change for life, making different organs as the embryo develops or when a child's life trajectory adjusts to the world in which the child finds itself.

The study of these enduring changes, where life meets the genome, is epigenetics. Epigenetic discoveries will impact upon our understanding of child development, mental health, and how public health and well-being can be maintained in a changing world.

Genetics and epigenetics go hand-in-hand, so this introductory booklet covers both. Much of human genetic variation is a consequence, through natural selection, of life-challenging encounters during our evolutionary history. Much can be learned when this evolved system fails to cope and manifests as disease.

Disease affects us all. Sequencing the first human genome in 2003 was a landmark scientific discovery, promising a revolution in medicine and treatment of disease. Keep these facts in mind, and it is no surprise that thousands of fellow scientists and physicians have devoted their careers to understanding human genetics.

My first encounter with genetic inheritance was through my family's inherited blood disorder, beta thalassaemia, which has a simple pattern of inheritance. Later in my professional life, I encountered two inherited diseases that did not follow this simple inheritance pattern. My work to understand Fragile X syndrome and Angelman syndrome led me to study epigenetic mechanisms. These may have a role in common multifactorial diseases, which are the focus of much current research.

Your genes and how they work

At the heart of every cell in your body is a central control centre, the nucleus. Curled up within the nucleus are 46 chromosomes, each a long molecule of a chemical called DNA. Most of us know that DNA has a double helix structure like a twisted ladder. Each rung of the DNA 'ladder' consists of a pair of two from four chemicals (called nucleotide bases). The chemical cytosine (C) always pairs with guanine (G), and thymine (T) always pairs with adenine (A) (see Figure 1).

This pairing rule means that the sequence of letters on one strand of DNA can be used to predict the sequence on the other, a fact that allows DNA to be replicated when necessary. Scientists commonly write DNA sequences as a single string of letters from one DNA strand, for example AGGGGTCCCC.

A gene is a length of DNA that has several specific features incorporated into its DNA sequence. These are shown in Figure 1. It has a START (three-letter code ATG), a series of three-letter codes (codons) for specific amino acids, and a three-letter STOP codon. In front of the START is a length of DNA where the molecular gene-reading apparatus and associated chemicals (transcription factors) are assembled, called the promoter region (see Figure 2).

A gene acts through the protein it 'encodes'. Proteins are assembled from chains of amino acids that are specified by the codons of the gene. Figure 3 shows that the information in the codons is first transferred to messenger molecules, called messenger RNA, that move from the nucleus to the cell cytoplasm where the amino acid chains and then proteins can be assembled.

Proteins are the chemical building blocks of our bodies. Some proteins (like collagen) contribute to the body's structure, while others (called enzymes) speed up chemical reactions. We have around 20,000 or so genes (making up 2% of our DNA), which are named according to the protein's function or the disease caused by the gene having a disabling mutation.

Most of our DNA is common to everyone, but the sequence varies between people at numerous points along each chromosome. Collectively, these differences can be called 'DNA variants'. The term 'mutation' tends to be reserved for a DNA change that causes a disease.

How you inherit your genes

Our cells have 46 chromosomes or 23 pairs – one set from our mother’s egg and the other set from our father’s sperm. Twenty-two of the chromosome pairs are the same and called autosomes. The remaining pair are the sex chromosomes, XX in females and XY in males. An X-carrying sperm leads to a girl, a Y-carrying sperm leads to a boy.

For all chromosome pairs, it is purely chance which chromosome of any one pair is passed on at conception. This results in the first level of ‘shuffling’ of chromosomes and the genes they carry between generations.

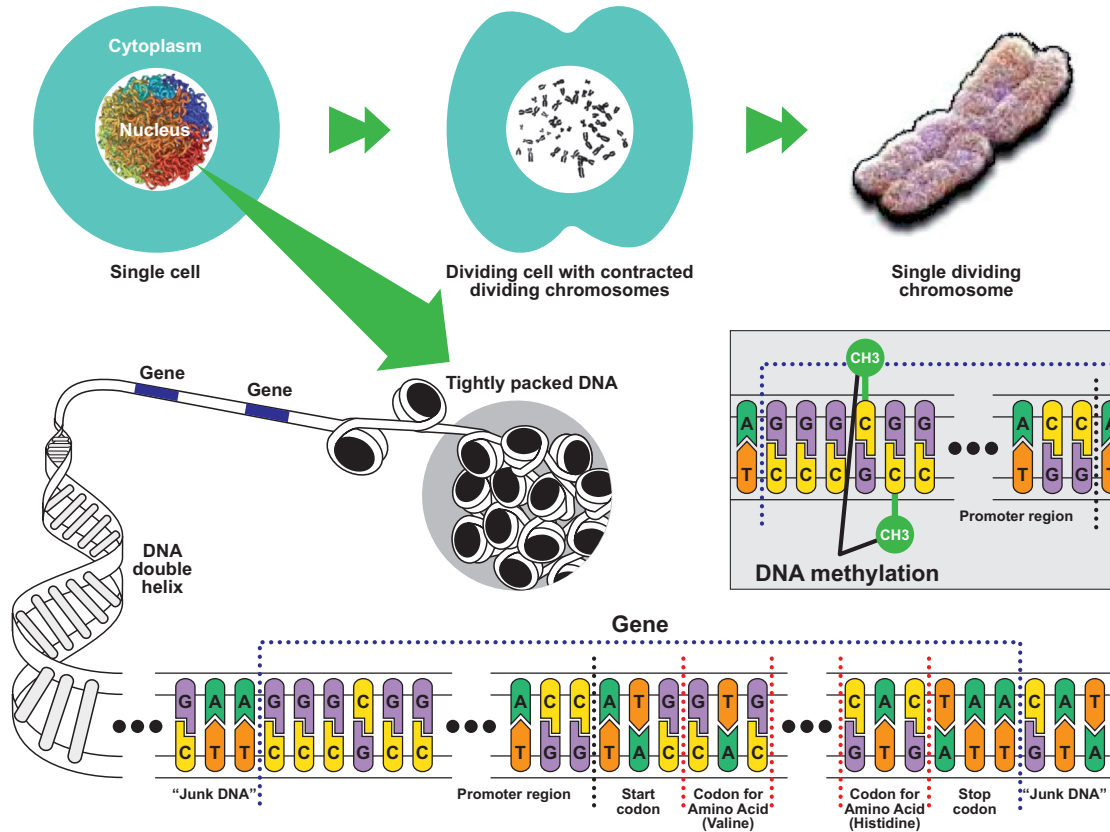
However, there is a further level of shuffling. In the special cell division that leads to egg or sperm formation, the two chromosomes of each pair lie alongside each other. Held in this embrace, sections of DNA are exchanged. The resulting eggs or sperm end up with a single set of 23 chromosomes, but with each chromosome being a mixture of the pair of chromosomes from which it was derived.

All cell divisions run the risk of introducing DNA copying errors, but sperm formation, is especially error prone. Many eggs or sperm with errors make it to the next generation and some errors can result in an inherited disease arising for the first time in a family.

Single gene disorders

As a student nurse, my sister was reprimanded for not taking her iron tablets, until it was discovered that the cause of her mild anaemia was not iron deficiency, but an inherited blood disorder called beta thalassaemia minor. Like millions of people across Asia, one of her two copies of the beta globin gene needed to make haemoglobin – the oxygen carrying chemical in red blood cells – was not working properly.

There was a mutation in the DNA of the beta globin gene on the chromosome 11 that she had inherited from our father. A person has two copies of chromosome 11, each carrying – at the same location (locus) – a beta globin gene. If my sister had also inherited a similar mutation from our mother, she would have suffered from beta thalassaemia major – a life-threatening anaemia requiring lifelong blood transfusions.

**Figure 1****TOP:**

A representation of how 46 chromosomes are packaged into a cell nucleus; a light microscope photo of 46 contracted replicating (X-shaped) chromosomes during cell division; and a scanning electron microscope picture of a single replicating chromosome.

MIDDLE:

A representation of how DNA is packaged in the chromosome; a diagram of 'methyl' groups CH₃ attached to the certain 'C' letters of the DNA molecule.

BOTTOM:

Expanded diagram of the DNA molecule from one chromosome illustrating the promoter and protein-coding regions of a gene. Note the pairing rule: C with G and T with A. Also note that 3 'letters' make up a 'codon' for an amino acid.

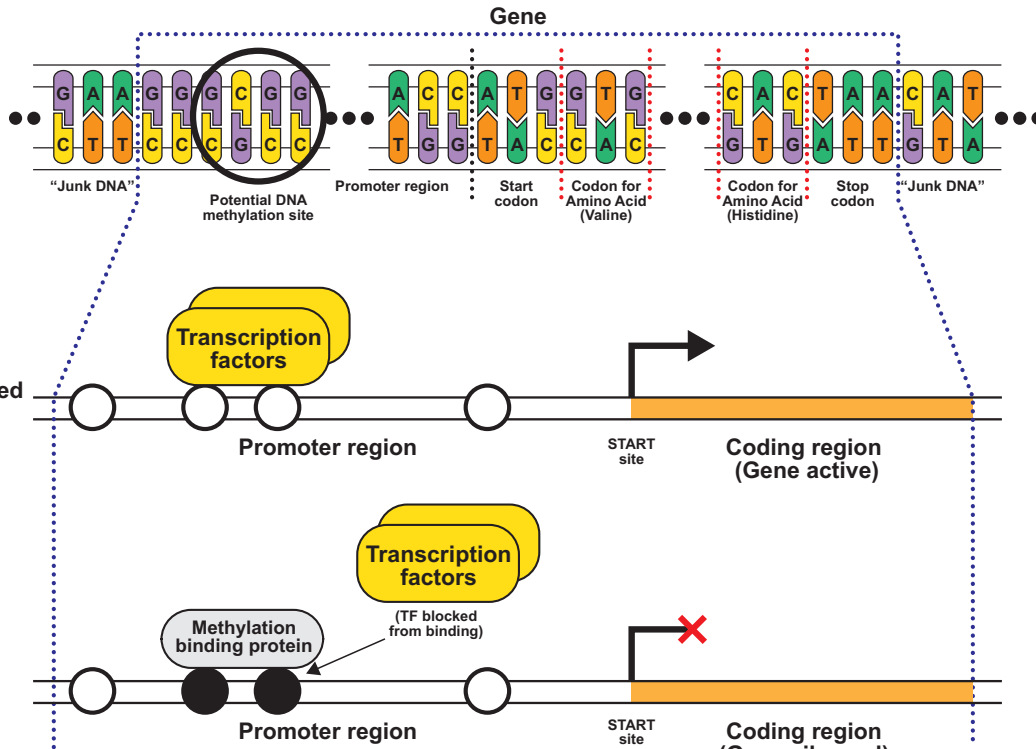


Figure 2

TOP:
Expanded diagram of the DNA molecule from one chromosome indicating the potential site for DNA methylation.

MIDDLE:
A standard way of representing (part of) a gene as 'ribbon', showing the promoter region and the START site for 'reading' (transcribing) the gene. An arrow indicates gene activity. Binding of proteins (called transcription factors) to the gene promoter region facilitates gene activity.

BOTTOM:
Two sites in the promoter region are methylated (solid black circle). This epigenetic change by DNA methylation can then attract a DNA methylation binding protein, that blocks the transcription factors and so silences the gene. A cross indicates gene inactivity.

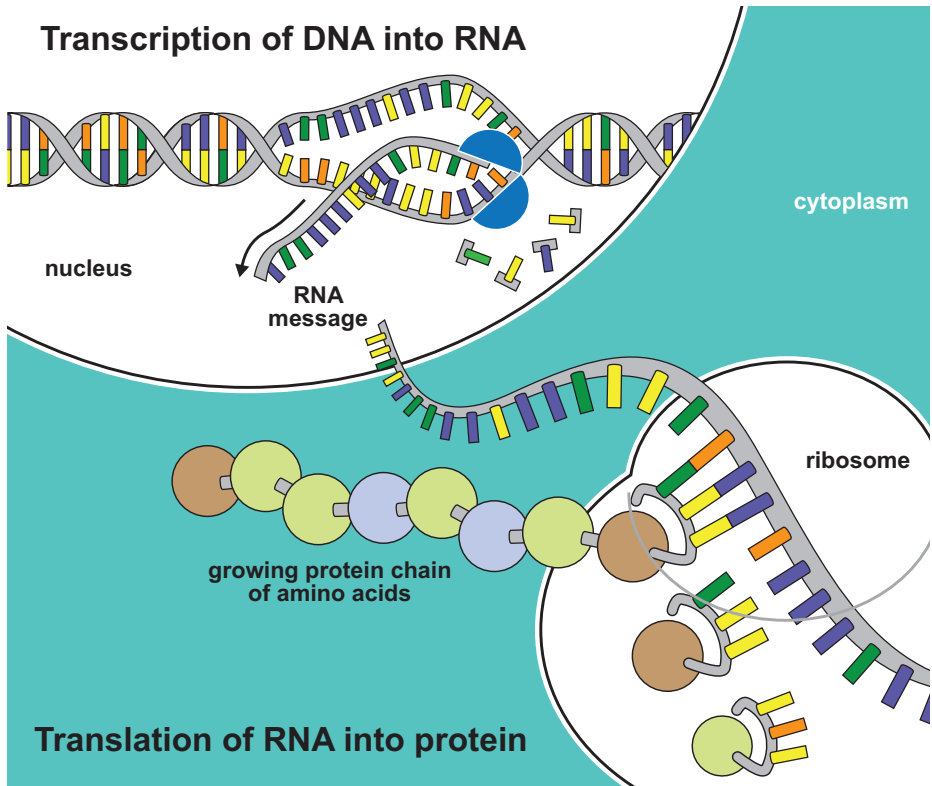


Figure 3

A representation of how the DNA genetic code is read (transcription) and transferred to a newly made messenger RNA molecule that travels from the nucleus to the cell cytoplasm, where ribosomes attach to it. A ribosome acts as an 'assembly plant', building the amino acid chain of a protein (translation) according to the genetic code carried by the RNA.

Our family were pleased that testing for carriers of the mutated beta globin gene was already available in the 1960s, long before DNA analysis was possible. Conveniently, although adults have two different haemoglobins – A and A2 – they can both be measured via a simple blood sample. I had raised A2 and mild anaemia, confirming that I too was carrying a beta globin mutation. My father had nine siblings, and five of them turned out to be carriers.

This one-in-two chance shows that, although everyone has two chromosomes 11, they can only pass one of them on to any one child. Either the chromosome 11 carrying the mutant beta globin gene or the chromosome 11 with the normal gene ends up in the egg or sperm involved in conception.

Mendelian disorders

Long before chromosomes were observed down the microscope, an Augustinian monk called Gregor Mendel deduced how the inheritance of many distinct characteristics must work. Crossing pure lines of different coloured peas, he came up with the odds of the offspring being one colour or the other, comparable to what we see in the inheritance of single gene disorders (see Figure 4). Diseases where one defective gene or section of DNA is involved are named 'Mendelian' disorders after him. Knowing that there is just one place to look for the mutation causing such a condition has helped scientists to discover which gene is linked to which Mendelian disorder.

Understanding these Mendelian inheritance patterns (see Figure 4) coupled with the development of carrier tests has led to effective genetic counselling for families threatened by disorders like beta thalassaemia major. There are many thousands of different Mendelian disorders. Together they affect directly only 1 to 2% of the population but indirectly they worry many more family members.

Mendel also discovered dominant and recessive inheritance. In recessive inheritance, a disease only manifests when a mutant gene is inherited from both parents. The life-threatening anaemia beta thalassaemia major is an example of this type of inheritance. In dominant inheritance a single copy of the mutant gene, inherited from one parent, is sufficient to cause the disease. But what determines whether a Mendelian disorder is inherited in a dominant or recessive fashion? The answer depends on the nature of the mutation.

There are many different mutations that can cause beta thalassaemia major. In fact, there are over 100 different beta thalassaemia mutations. What all these mutations have in common is that they severely reduce the body's ability to produce the beta globin protein that helps to make up haemoglobin.

Most mutations cause the cell to produce less protein. Should a person inherit one faulty copy of the gene, overall protein production drops by 50%. For many cell functions, there is enough slack in the system for the cell to get by on half the usual amount of protein. I have a 'good enough' genome, despite my beta globin mutation.

Protein levels only drop critically, and the disease only manifests, when the mutation is inherited from both parents. This explains why diseases like beta thalassaemia major are inherited in a recessive fashion. A test revealed that my wife did not have the mutation, and so fortunately we did not have to face the risk of having a child with beta thalassaemia major.

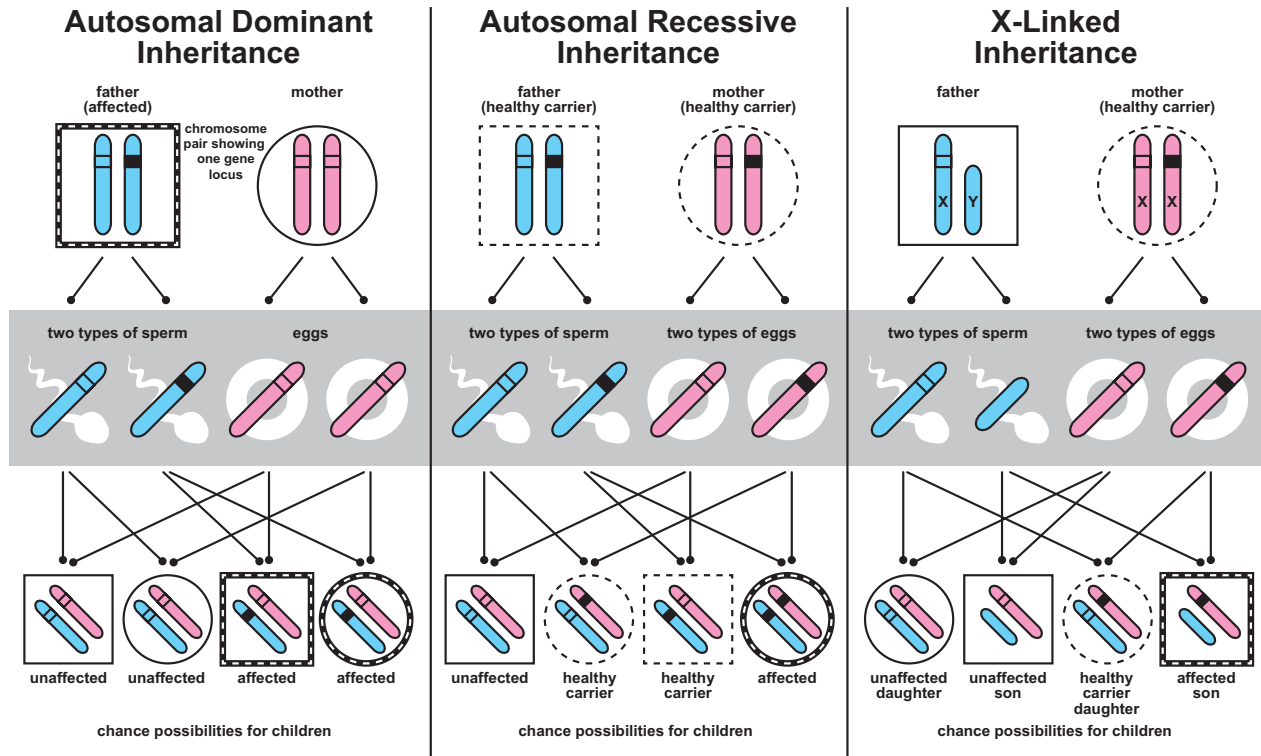


Figure 4

Three Mendelian patterns of inheritance portrayed using one chromosome pair with a single gene locus on it. The mutant copy (allele) is filled in. The four possibilities for a child are illustrated. Any one of these possibilities can be the outcome of a pregnancy. There is a one-in-two chance of being affected in autosomal dominant inheritance, and a one-in-four chance of being affected in autosomal recessive inheritance. Chance has no memory – the risks are just the same for each pregnancy, regardless of what happened before.

Dominant inheritance

Dominant inheritance, by contrast, usually occurs where the mutant gene produces an abnormal protein that overrides or disrupts the functioning of the normal protein. The impact of this abnormal protein depends on what the normal protein does within the cell.

In one scenario, the abnormal protein does not interfere with the function of the normal protein straight away. The cell surveillance system detects the abnormal protein, and tries to neutralise it. For example, chaperone molecules that normally help to fold proteins into the right shape will try to remove misfolded ones. But eventually, the cells defences are overwhelmed, and disease manifests.

Something akin to this happens in the dominant condition, Huntington disease. This is a neurodegenerative condition that typically develops in middle age causing difficulty in muscle co-ordination and eventually cognitive decline. Fragments break off the abnormal protein when important chemicals are being assembled in brain cells. These fragments coalesce, or are rounded up, within brain cells in an attempt to contain the problem. Over the years, the strain on the brain cell maintenance system is too much. There is system failure and cell death.

In other dominantly inherited conditions, the abnormal protein has a profound effect. One example is the dominantly-inherited form of dwarfism, achondroplasia. This is caused by a mutation in the gene encoding a chemical on the outer surface of our cells, which our cells use to receive information. This chemical is called fibroblast growth factor receptor 3 (FGFR3) and is used to receive chemicals called fibroblast growth factor molecules. The mutation makes the receptor think it is constantly receiving fibroblast growth factor molecules. Consequently, the receptor endlessly bombards the cells of the bone growth plates with signals dampening down bone growth.

Mutations like those in FGFR3 are called gain-of-function mutations because they cause something new – and wrong – to happen within cells. Since they create new, abnormal cell processes rather than reducing the efficiency of existing ones, they tend to lead to dominant rather than recessive inheritance.

Mutations arise at some point in many families, and 85% of people with achondroplasia have the condition because of a new mutation in the sperm that led to their conception.

Their parents are normal height. But why are these mutations in sperm rather than eggs, and why is achondroplasia relatively common (about 1 in 25,000 live births)?

The answer seems to be that this gain-of-function mutation gives the sperm-making cell in which it arises a proliferative advantage, leading to a disproportionate number of mature sperm carrying the mutation. This reminds us that we develop from a community of cells, jostling for position, some proliferating and some dying. As we develop, we are actively responding to and dealing with our genetic inheritance through the fate of our individual cells.

Beyond Mendel's laws

In the late 1970s, scientists were making good progress in diagnosing rare single gene disorders and counselling families about their risk based on knowledge of Mendelian inheritance (see Figure 4). Increasingly, carrier tests looking for changes in proteins linked to disease were supplemented by direct DNA studies of possible mutant genes.

But researchers were growing puzzled by certain conditions that clearly ran in families, but did not obey Mendel's laws. One such disorder was Fragile X syndrome, which was becoming widely recognised as among the most common inherited causes of learning difficulties.

Scientists already knew that, when examined under the microscope, the X chromosome from an affected boy would show what looked like damage at one end – a constriction, or 'fragile site' – hence, the name Fragile X syndrome. But Fragile X did not behave like Mendelian disorders linked to the X chromosome, for example haemophilia. In Fragile X families, there were severely affected women with healthy fathers and grandfathers.

There was no scientific basis for estimating the chance of Fragile X skipping a generation, or of a future girl being affected, because Fragile X did not follow a Mendelian pattern. Genetic counselling was difficult. Families find it distressing enough to know their child has a serious genetic problem that might affect future children, without their distress being compounded with vague risk figures.

Initially, researchers explained the Fragile X inheritance pattern by saying that the DNA mutation 'progressed' from one generation to the next. The healthy grandfather would have something called a premutation and, through incremental changes in the DNA, this would progress to a 'full mutation' that affected grandchildren.

Several years later, researchers discovered that the premutation was a small lengthening or expansion of a section of DNA at the beginning of the gene (FMR1) involved in Fragile X. The section of DNA lengthened down the generations to become a full mutation in affected children (see Figure 5). A clue as to why this bit of DNA was unstable came from its sequence of 'letters'. It was composed of CCGCGGCGG repeated up to 40 or so times.

One of our cells' great challenges is faithfully replicating our DNA sequence during multiple cell divisions, including making millions of sperm or eggs. You try repeatedly copying 40 CGG triplets (CGGCGGCGGCGGCGG etc...) as happens when copying the FMR1 gene! It seems that the cell's DNA replication system has the same difficulty, and may write out the CGG triplet too many times.

When the CGG repeat expands to more than 200 triplets, it is mistaken by the cell's machinery for a DNA-invading virus. So-called DNA methylating enzymes swing into action to switch off what they think is 'dangerous' DNA inserted by a virus trying to hijack the cell. Although the CGG repeat is located before the coding region of the FMR1 gene, the enzymes deactivate or silence the whole FMR1 gene. The ensuing loss of FMR1 protein disrupts many processes within brain cells.

Fragile X was the first of several conditions to be discovered that worsen down the generations due to these unstable or dynamic mutations. Another is Huntington disease.

Epigenetic regulation

Our genes can be likened to the keys of a piano. When they are 'switched on' and producing proteins, that's like the keys being pressed. So what plays the music? What generates the different 'melodies' – those intricate, co-ordinated patterns of gene activity that distinguish brain cells, liver cells or other cell types? The answer, in short, is that the process of living and developing 'plays the music' through epigenetic regulation of gene activity.

Epigenetics is the study of how genes are switched on and off in an enduring fashion – which happens, for example, in response to conditions within the embryo. The cell can switch genes on and off using DNA methylation, tagging DNA with the methyl chemical group, CH₃ (see Figures 1 and 3). Once added during development, these epigenetic marks can remain, even if the cells and DNA replicate. The epigenetic marks do not change the DNA sequence, just the level of gene activity.

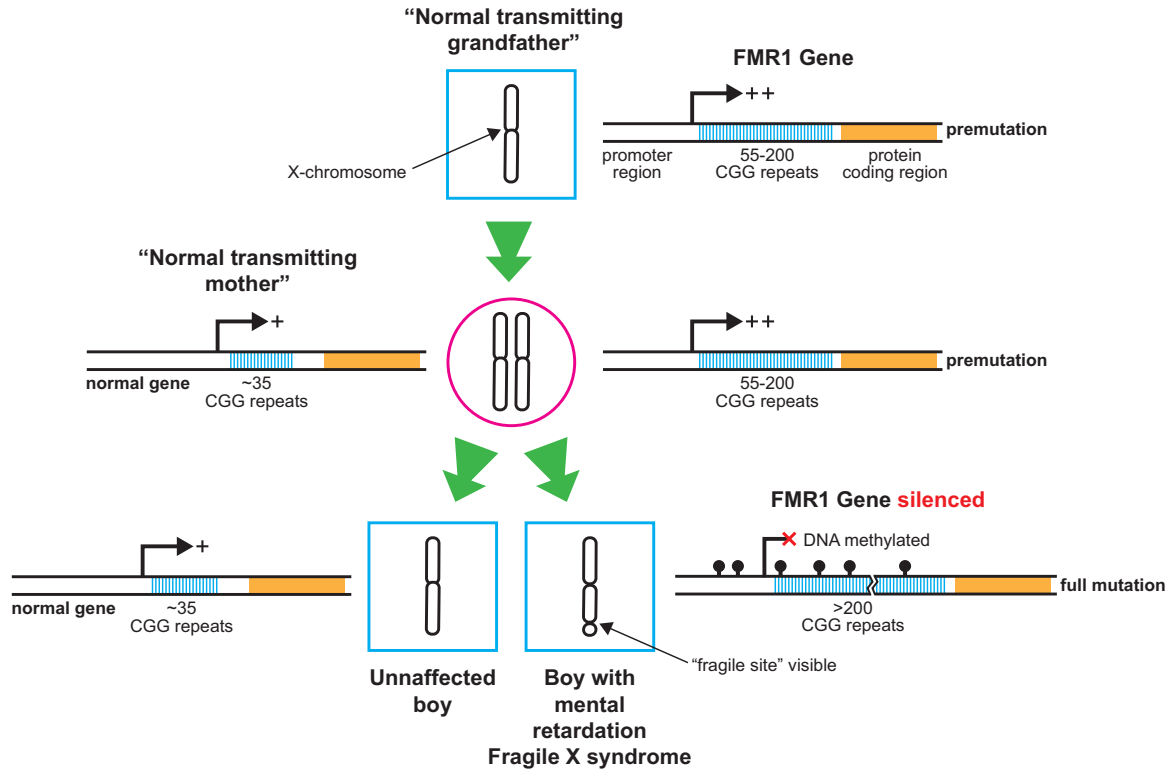


Figure 5

A representation of three generations of a family where a boy has Fragile X syndrome. We see him and his unaffected brother (bottom), his mother (middle) and his maternal grandfather (top), and the appearance of the X chromosome in each of them. Alongside each person, the state of their FMR1 (Fragile X mental retardation 1) gene is portrayed. The grandfather and mother carry a premutation with 55-200 repeats of CGG in the gene promoter region, but with the gene still active (black arrow). The affected boy has a full mutation with more than 200 repeats. This expansion results in methylation of the DNA (black circles) and silencing of gene activity (red cross).

An extreme example of epigenetic regulation is a group of genes called imprinted genes. Here, only one copy of the gene is active, depending on whether it came from mother or (in other instances) from father. The other copy is silenced by DNA methylation during sperm or egg formation.

Genomic imprinting disorders

Another genetic disease that does not obey Mendel's rules is Angelman syndrome. People with this disease have marked learning difficulties, no speech, and often epilepsy. They have an unsteady, jerky gait, and often laugh. In the late 1980s, scientists looked at the chromosomes of people with Angelman syndrome and found that some had a bit missing (a deletion) on chromosome 15. The deletion was small, close to the limit of resolution of microscopes, which explained why it had not been spotted before.

But scientists soon faced a problem when it became clear that this deletion was identical to the one known to cause Prader-Willi syndrome, a different disorder from Angelman. Furthermore, not all people with Angelman syndrome had this deletion. The next breakthrough came when a US study investigating Prader-Willi reported a patient with two (intact) copies of chromosome 15 from mother, but none from father (see Figure 6).

Researchers realised that Angelman syndrome might involve genomic imprinting, a normal phenomenon where a gene is switched on or off, depending on which parent it came from. Genomic imprinting had been recently revealed by experimental studies in mice. The earliest of these experiments manipulated fertilised eggs to produce embryos with either both sets of chromosomes from mother, or both from father.

This replacement of one set with another intact set of 23 chromosomes, but from the same parent, proved disastrous for the mouse embryo. With both sets from father, only disorganised placental tissue develops; with both from mother, the growth-retarded embryo fails to progress. Clearly, chromosomes from mother and father do not work the same. For a baby mouse to develop, it needs to inherit chromosomes from both parents. What is true of mice is true of humans (see Figure 7).

When a *single* pair of chromosomes comes from the same parent, the impact on development is less severe because fewer imprinted genes are affected. Researchers realised that, if Angelman syndrome involved an imprinted gene, then any child born with

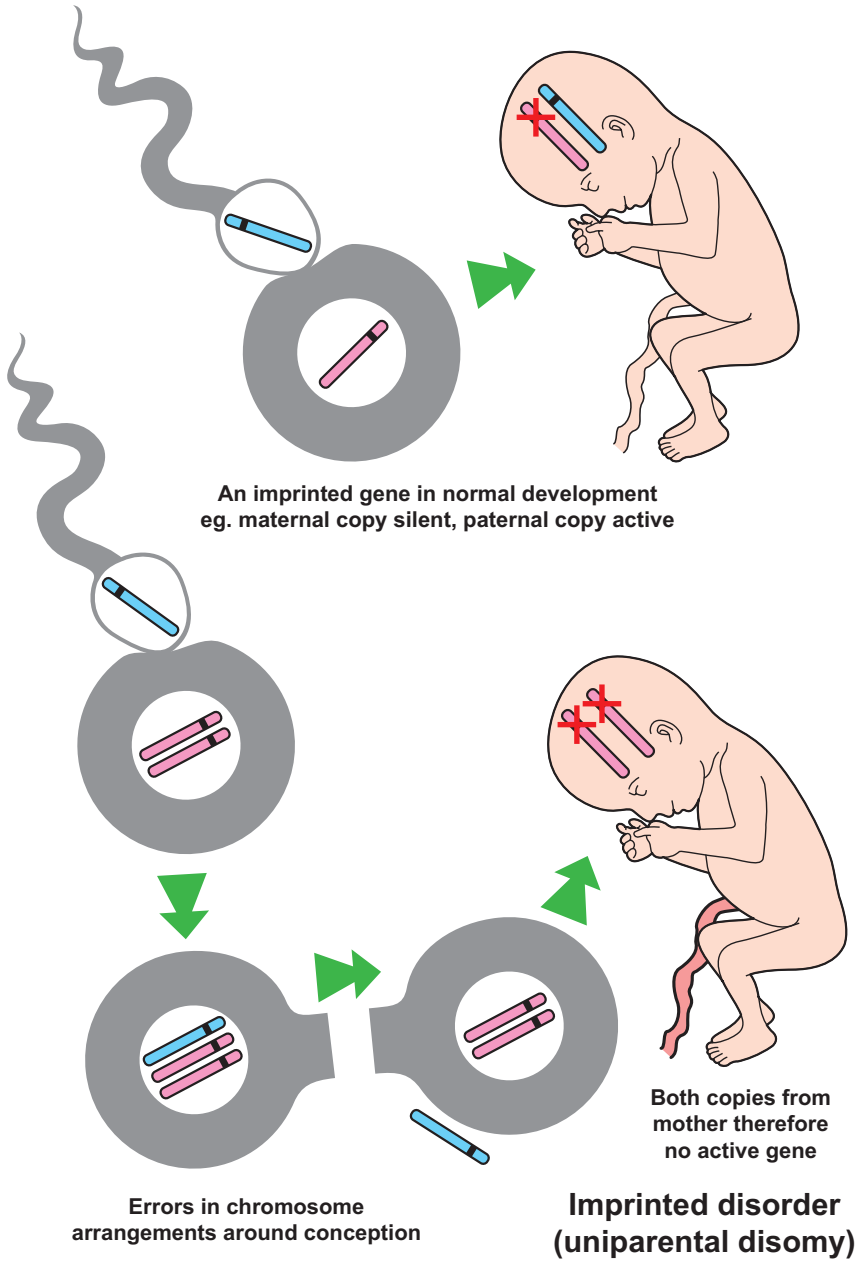


Figure 6

An illustration of one chromosome pair containing a single normally imprinted gene locus. In the example at the top, the maternal copy (allele) is silenced (red cross) with the baby relying entirely of the activity of the paternal allele. The bottom diagram shows how chromosomal errors around conception can lead to a baby with two maternal alleles and no active gene, despite having two intact chromosomes. Such a uniparental disomy of chromosome 15 is one cause of Prader-Willi syndrome.

two intact chromosomes 15 from their father and no chromosome 15 from their mother would have the syndrome.

Studying the parental origins of DNA from children with Angelman syndrome (who did not have a deletion) revealed that two children had, indeed, inherited both of their chromosomes 15 from their father. Thus the key Angelman gene was imprinted. When there is a mutation in an imprinted gene, it produces a non-Mendelian pattern of inheritance controlled by the sex of the parent passing on the mutation, not the sex of the child.

In Angelman syndrome, the child is only affected when the mutant gene on chromosome 15 is passed down by the mother. A mutation can be transmitted by a child's father unnoticed for generations, because the gene is normally switched off by the imprinting process, so the presence of the mutation is irrelevant.

To date, scientists have shown over 50 human genes to be imprinted when working normally. With some it is the mother's copy, and in others the father's copy, that is silenced. But what 'marks' a chromosome as having come from mother or father, and what triggers the silencing?

The answer: the DNA methylating enzymes that, with the Fragile X gene, jump on the runaway CGG repeat and silence it. With imprinted genes, a non-mutant gene is being silenced, depending on which parent it came from. Work on Angelman syndrome and other imprinting disorders has shown that DNA methylation and silencing of some genes is, unlike in Fragile X, a normal part of sperm and egg formation.

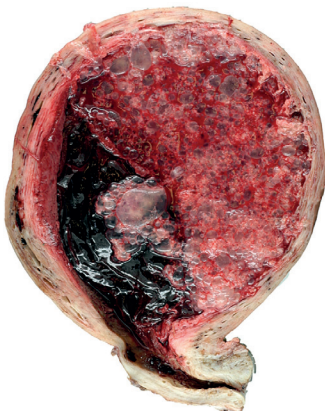


Figure 7

Longitudinal-section of the uterus of a woman who had an emergency hysterectomy was found to have a hydatidiform mole. When an embryo ends up with two sets of 23 chromosomes from father and none from mother, a tumour (or 'hydatidiform mole') of placental tissue grows in the womb.

The 'parental tug-of-war' hypothesis

Reconstructing the selective advantages that led genomic imprinting to evolve in mammals is difficult. But genomic imprinting seems to be linked to the parental conflicts of interest generated when fetal development switched from eggs to inside mother.

According to the 'parental tug-of-war' hypothesis, fathers want to maximise survival of their offspring after birth, so genes where the father's copy is active should lead to *increased* fetal growth. Imprinted genes with an active mother's copy should counter these effects and *suppress* fetal growth, to conserve mother's energy for childrearing and her own survival for future pregnancies.

A classic example supporting this idea involves two oppositely imprinted genes. Only the father's copy of the insulin growth factor 2 (IGF2) gene is active, and this increases fetal growth. A nearby gene, called H19, is only active from the mother's chromosome and suppresses growth.

Errors naturally occur during imprinting. In humans, imprinting errors causing IGF2 genes from both parents to be active cause Beckwith Wiedemann syndrome, a disorder where babies are born with large organs, a large tongue and a risk of cancer in early childhood. Imprinting errors that increase production of H19, a regulatory RNA molecule, cause Silver-Russell syndrome, where the embryo's growth is severely restricted in the womb.

The genomic imprinting story reminds us that a regulatory system has evolved that can lead to enduring changes in gene activity, even though there are no changes in the DNA sequence. A gene can be silenced for a whole lifetime, provided that the silenced state is passed between cells when they replicate during development. This system is called epigenetic regulation.

Common diseases

The genetic diseases mentioned so far in this booklet are caused by a single mutation in the genome of the affected family. There is a strong relationship between the mutation's location and the symptoms of the disease.

Common, chronic diseases of Western societies, such as eczema, asthma, diabetes, coronary heart disease, depression and cancers, are very different. They are nearly all multifactorial, so defining *the* cause is impossible. There are many contributing factors, including the person's genetic makeup. Imagine driving along a narrow, winding country lane in the rain, having just had a row with your partner. The windscreen is fogged up, and your car has rather bald tyres. You round the corner to find the lane blocked by a tractor backing into a field, and crash into the tractor. What is the cause of the crash?

One cannot be certain, and so it is with multifactorial diseases. Building up the picture with more and more risk factors, including DNA variants, provides some sort of explanation. Family studies show that many people with a common disease have rather more ancestors and relatives with the same disease than one would expect by chance, which suggests that inheritance has something to do with it. Inheritance is clearly important in a small minority of families whom we will consider next.

Monogenic subsets in 'common disease'

A small proportion of people with a strong family history of a common disease carry a genetic mutation that confers such a high risk for that disease, that it can reasonably be called 'monogenic'. Showing dominant inheritance, these families need special clinical management.

A well-known example is mutations in the BRCA1 or BRCA2 gene, which raise the risk of developing breast cancer. About 5% of women who develop breast cancer have one of these mutations. The percentage rises to about 10% if the woman has already had a cancer at a young age and has a family history, and drops to about 2% if there is no such history. If one carries a BRCA1 or 2 mutation, the risk of getting breast cancer by the age of 80 ranges from about 50% to 90%.

About 4% of bowel (colorectal) cancers are due to dominantly inherited mutations causing familial adenomatous polyposis or hereditary non-polyposis colorectal cancer. These mutations confer a lifetime risk approaching 100%.

About 1% of people presenting with diabetes in early life have the condition due to a dominant mutation. Making this genetic diagnosis can sometimes allow the person to stop injecting insulin, and control their diabetes with tablets.

Development and common multifactorial disease

Outside of examples like BRCA1 and 2, families still have tendencies to develop certain common multifactorial diseases more often than expected by chance. But this tendency for a disease to run in the family is a complex matter. It is not all due to the DNA variations inherited at conception, and even that influence is often indirect.

The DNA mutation causing a monogenic disease is like a car having a puncture or a flat battery – if it happens, you are in trouble. The DNA variations contributing to common disease risk or resistance are more akin to the differences between a car and a tractor. How each fares depends on the challenges ahead. Both can travel on a smooth road, but the car is faster, more fuel efficient and can go further. However, the car is more likely to fail if forced to leave the road.

It is helpful to think of common adult-onset disease risk emerging as a result of a person's developmental journey. Studies suggest that factors operating before birth and during childhood determine much of our likelihood of developing diseases as adults. For example, there is an association between low birth weight and greater risk of developing coronary heart disease, high blood pressure or type2 diabetes in later life.

What might explain such associations? First, it may be that the same DNA variants cause *both* the slower growth of the fetus *and* the increased risk of type 2 diabetes or coronary heart disease in later life. An extra twist needs to be added to this idea. It may be the presence of the DNA risk variant *in the mother* rather than in the fetus that is influencing fetal growth. The mother's genes influence her metabolism and the chemical composition of her blood, which in turn can influence what goes across to her baby. So in a sense the mother's genetics is part of the fetal environment.

Another possible explanation of the link between fetal growth and adult disease could be the family's bad habits. If a mother smokes, her children are more likely to be smokers. If a mother smokes in pregnancy, her babies will be born small. If those babies grow up to be smokers, they will be at greater risk of coronary heart disease. And as always with multifactorial disease, there is a yet another twist – there is increasing evidence of genetic influences on behaviour, including all aspects of addiction.

A final possibility is epigenetic regulation. Epigenetic studies suggest that people's early life experiences can alter their gene activity for decades or even for life, without affecting

their DNA sequence. Early results from animal experiments suggest that this is a promising avenue for research into the origins of adult health.

Going from the single fertilised egg to 100 trillion cells during development means that the cells, and the genome within, have to divide many times and replication is not perfect. DNA mutations occur, and the effects on the cell can range from inconsequential to lethal. Some of these, so-called somatic mutations, are transmitted to descendant cells. Some induce dangerous behaviour in the cell such as unrestrained cell division, which might be the first step on the multi-mutation road to cancer.

To summarise, development and health depend on five broad factors: inheritance at conception; DNA mutations in some cells after conception; how some cells reproduce and proliferate faster than other cells; early developmental experience; and the prevailing social and physical environment.

The role of inheritance in common multifactorial disease

Inheritance at conception consists principally of the chromosomes (with their DNA and variations) from mother and father. However, we also inherit a tiny amount of DNA from our mother's mitochondria in the egg cytoplasm, tiny energy-producing structures vital to cell function. Mitochondria carry their own genome, only 0.0005% the length of the human chromosomal genome, and carrying just 37 genes. Other molecules capable of influencing development, such as RNA, are also transmitted with the sperm and egg, but knowledge of their effect on offspring is still rudimentary.

Scientists have got some measure of this biological inheritance and its role in common disease from twin studies. Identical (monozygotic) twins come from the same fertilised egg and share the same genetic variations, whereas non-identical (dizygotic) twins come from two fertilized eggs and have, on average, half their genetic variations in common, just like siblings more generally.

If a monozygotic twin has a disease influenced by inheritance, their co-twin is more likely to develop it than if they were dizygotic twins. This higher likelihood indicates that some disease risk was inherited at conception. Twin studies show that inheritance makes a major contribution to most common diseases, although they cannot explain which genetic variations have an influence. This information has to come from other types of research.

Case-control studies

Another approach is to compare DNA samples from people with a disease (cases) with DNA samples from people without that disease (controls), and look for genetic differences between them. This is called a 'case-control' study.

Early case-control studies had one big drawback. Analysing many whole human genomes was not possible, so researchers looked at a few candidate genes that they suspected might be involved in a disease. Choosing these candidate genes required the researchers to know already what the genes did, and how the disease developed. Unfortunately, only a few of these candidates turned out to have a major role in common diseases.

One example is filaggrin (nature's skin moisturiser) and its role in eczema, which affects some 20% of children. Mutations in the filaggrin gene are remarkably common, at least in Europeans. A careful population study (the Avon Longitudinal Study of Parents and Children) found that 6% of babies carried one of the two most common mutations. These mutations account for a significant percentage of the risk of eczema and allergic asthma in the population – 15%, according to the ALSPAC study.

Genome-wide association studies

After the Human Genome Project was completed in 2003, it became technologically possible to look at thousands of genes. Big case-control studies of common diseases were launched, to ask which of the 500,000 known common DNA variations – if any – are associated with a particular disease. The genome-wide association study (GWAS) was born.

The pioneering Wellcome Trust Case Control Consortium (WTCCC) led one of the first GWAS studies. It looked at DNA samples from adults with bipolar disorder, coronary artery disease, Crohn's disease (an inflammatory bowel condition), high blood pressure, rheumatoid arthritis and type 1 and type 2 diabetes.

The WTCCC published its first findings in 2007, after analysing the DNA of 14,000 patients (2,000 for each disease) and 3,000 controls. It discovered that 24 regions of the human genome were associated with one or other of these diseases. But these common DNA variants (or nearby ones) each made only a tiny contribution to the relevant diseases.

Many a GWAS since the WTCCC has painted the same picture. Many DNA variants associated with disease have now been discovered – for example, more than 20 for type

1 diabetes, more than 35 for type 2 diabetes, and more than 70 for Crohn's disease. But, even taken together, they often explain only a small percentage of someone's risk of common disease. However, for Parkinson's disease – which affects 1-2% of the elderly – 11 genetic variants seem to explain 60% of the population risk.

With the latest DNA sequencing technology, we will be able to study rare DNA variants that, although carried by fewer people, will probably explain more of their disease risk. Some DNA variants may one day help hospitals and clinics to diagnose disease, but this is as yet uncertain.

Integrating genetics, epigenetics and the environment

In common disorders, genetic and environmental influences go hand-in-hand, with epigenetic regulation underpinning many of the long-lasting effects of early experience. Some studies have forced us to revise the way we think about the effect of genetic variants. Over the last decade five DNA variants were found to be associated in some way with antisocial behaviour and poor self-control in young people, particularly following an adverse childhood with poor parenting.

However, when researchers looked at good as well as bad outcomes, it turns out that these DNA variants are not associated with poor self-control *per se*. They are associated with *how responsive* children are to parenting. The same set of variants associated with poor self-control following bad parenting are also linked to very good self-control after good parenting. So here, the genetic variation seems to be influencing the *degree of response* to the social environment. Some children are deeply affected for better or worse, whereas others much less so.

Gene-environment interactions during development are usually very difficult to study because so many factors are involved. But there is one area where the environmental exposure is well defined, and that is treatment with a prescription drug. With many drugs, a proportion of patients do not get the beneficial response expected, and a few get seriously detrimental side-effects.

How a person's genetics affect how they respond to drugs is studied in pharmacogenetics, which – it is hoped – will lead to 'personalised medicine'. After several false dawns over the

last 20 years, pharmacogenetics is finally being introduced into medical practice. Clinical trials are establishing whether genetic tests can help determine the appropriate dose of the widely used, blood-thinning drug Warfarin, so as to reduce dangerous side-effects such as internal bleeding.

One area that promises to move much quicker than pharmacogenetics is the use of new DNA sequencing techniques to discover which somatic DNA mutations have occurred in cancer tissue. This will help doctors to select drugs that are 'tailor-made' to a patient's tumour.

We are also beginning to understand how DNA methylation affects common-disease risk. We know there are numerous places in the human genome where the pattern of DNA methylation is highly variable, even among monozygotic twins who have the same DNA sequence.

There is some evidence that the pattern of DNA methylation is influenced by the diet of one's mother around the time of conception. Research has shown that the ongoing health and DNA methylation of people in rural Gambia is affected by the season in which they were conceived. People whose mothers went hungry during the rainy season at the time they were conceived have a poorer outcome for future health.

A Dutch study found that children whose mothers experienced famine around the time they were conceived kept the same altered patterns of DNA methylation in their blood cells for 60 years after their birth. Blood cells do not last that long, which suggests that methylation patterns were carried in the children's self-renewing stem cells. Alternatively, the famine could have changed the children's hormone response levels for life.

We should not be too surprised if we discover that the early environment can cause big shifts in a child's development, by permanently changing methylation patterns. After all, the worker bee and the queen bee have identical DNA but are very different in structure and behaviour. This is due to DNA methylation changes triggered by social circumstances, and by a diet of royal jelly!

Other links between DNA methylation and adult disease are emerging, particularly in cancer, but researchers still have many questions. For example, do the changes in

DNA methylation affect whether genes are switched on or off? And is the altered DNA methylation a cause or a consequence of disease?

The relationship between DNA methylation and cancer is difficult to study, but scientists have found that some people with hereditary non-polyposis colorectal cancer – a bowel cancer – have abnormal DNA methylation instead of the expected DNA mutation. This occurs in all the cells of the body, but is not usually passed on to the next generation.

An aspect of these methylation changes that is of particular interest to researchers is that they are more reversible than changes to our DNA sequence. Studies are already underway to test whether drugs that alter DNA methylation might help to treat cancer, but we do not yet know whether this will work.

Understanding more about the genetics and epigenetics of human disease will depend on many things, not least people's willingness to actively participate in research. Participation by families with rare monogenic diseases has resulted in valuable insights into the workings of development and inheritance – insights that are relevant to the common, multifactorial diseases that affect us all.

Further progress will require big commitments from researchers, funders, and large numbers of willing participants – and, being genetics, a commitment across the generations.

Appendix

Glossary

Amino acid The chemical building blocks that make up proteins.

Autosomal dominant inheritance A pattern of inheritance in which a gene change has an effect even if only one copy is inherited – for example, the mutated gene that causes Huntington disease (see Figure 4).

Autosomal recessive inheritance A pattern of inheritance in which a gene change has an effect only if two copies are inherited – for example, the mutated gene that causes cystic fibrosis (see Figure 4).

Autosome A chromosome that is not one of the sex chromosomes (X and Y).

Carrier An individual who has a disease-causing gene mutation on one chromosome of a pair, and a normal version of the same gene on the other chromosome of that pair. Usually refers to unaffected carriers of recessive conditions.

Chaperone molecule A protein that aids in the folding of a second protein. The chaperone prevents proteins from taking conformations that would be inactive.

Chromatid During cell division chromosomes consist of two sister chromatids. Each contains a complete double helix and the two are exact copies of each other.

Chromosome Tightly packaged bundles of DNA, the chemical that encodes genetic information. Nearly all human body cells have a set of 46 chromosomes. Egg cells and sperm cells have 23 chromosomes.

Codon A unit of genetic code consisting of a sequence of three nucleotide bases (see Figure 1).

Deletion A missing portion of genetic material in a gene, a section of DNA, or part of a chromosome.

DNA (deoxyribonucleic acid) The chemical that encodes genetic information. It contains four different nucleotide bases, known as adenine (A), cytosine (C), guanine (G) and thymine (T).

Epigenetics The science of enduring, but ultimately reversible, changes in the pattern of gene activity, during embryo development and beyond, that do not involve alteration of the DNA sequence. These changes occur in response to conditions within the embryo and more generally.

Gene The basic biological unit of inheritance. Genes are made out of DNA, and most of them are coded instructions for making proteins. Other genes make molecules that control gene activity.

Genetic condition/disorder A situation where problems are caused by one or more genes affecting the way the body looks, develops or works.

Genetic counselling Information and advice given to patients affected by, or at risk of, a genetic condition. An explanation of risks and options may include the findings of specific gene tests.

Genetic predisposition/susceptibility Increased probability (compared to the general population) of developing a disease, due to the presence of one or more genetic mutations.

Gene silencing The switching off of a gene by a mechanism other than genetic modification. That is, a gene which would be expressed (switched on) under normal circumstances is silenced (switched off) by machinery in the cell (see Figure 5).

Genetic test A test that gives genetic information, for example about disease status or susceptibility.

Genetic variant A relatively common (say, carried by 2% or more of the population) alternative version of the DNA sequence in a gene, that by itself does not lead to a genetic disorder.

Genome The total genetic information of a living thing, a complete copy of which is found in most body cells. The human genome is made up of around 2.9 billion chemical letters (base-pairs) of DNA.

Human Genome Project An international, publicly funded effort to read and decode the entire genetic information of a human being, the results of which were published in 2003.

Imprinting The process by which certain genes are silenced (switched off) according to whether they were inherited from father or mother.

Junk DNA DNA sequences that do not encode for protein sequences, also called non-coding DNA.

mRNA (messenger RNA) An intermediate stage between a gene and the protein it codes for. A cell uses an mRNA template, rather than reading the DNA code directly, when making a protein.

Mutation A change in the genetic information of an individual, which may have a harmful, neutral or beneficial affect.

Nucleotide The building blocks that make up DNA and RNA molecules. A single nucleotide consists of a nitrogenous base (adenosine, cytosine, guanine or thymine in DNA), a phosphate group and a sugar. In the molecule of DNA, the base adenine (A) always pairs with thymine (T), and cytosine (C) always pairs with guanine (G).

Premutation An expansion that is not long enough to lead to a disease caused by expanded nucleotide repeats, but is long enough to destabilise the repeat, so that later

generations are affected.

Protein A large biological molecule made up of a string of amino acids. There are thousands of different proteins in the human body, each with a particular job. For example, haemoglobin carries oxygen around the blood.

Ribosome The component of a cell that creates proteins from amino acids and RNA.

RNA (ribonucleic acid) A chemical similar to DNA. Like DNA, RNA contains the chemical bases adenine (A), cytosine (C) and guanine (G), but instead of thymine (T), RNA has uracil (U).

Sex chromosome The two chromosomes (X and Y) that determine a person's sex.

Single gene disorder A disorder caused by a mutation in a single gene. Cystic fibrosis, for example, is caused by a mutation in the CFTR gene.

Somatic Pertaining to all body cells apart from sperm, eggs, and the cells that give rise to sperm and eggs. Because a somatic mutation does not affect any cells that are directly involved in sexual reproduction, it cannot be passed on from one generation to the next.

Transcription factor A protein that binds to specific DNA sequences, thereby controlling the flow (transcription) of genetic information from DNA to mRNA.

Uniparental disomy A situation where a person receives two copies of a chromosome, or of part of a chromosome, from one parent and no copies from the other parent (see Figure 6).

X-linked inheritance A pattern of inheritance caused by a gene mutation located on the X chromosome. This is usually recessive, in which case it affects only boys (XY) while females (XX) can be carriers (see Figure 4). But in some cases it can be dominant, in which case it affects boys and girls.

Further reading

Books

Jess Buxton and Jon Turney. *The Rough Guide to Genes and Cloning* (2007)

This clearly written book provides an excellent introduction to genetics, considering everything from Dolly the sheep to designer babies.

Professor Matti Häyry. *Rationality and the Genetic Challenge: Making People Better?* (2010)

This book addresses the moral issues and ethical debates surrounding recent advances in genetic science.

Mark Henderson. *50 Genetics Ideas You Really Need to Know* (2009)

This accessible book distils the central concepts of genetics in a series of 50 lucid and engaging short essays.

April Lynch and Vickie Venne. *The Genome Book: A Must-Have Guide to Your DNA for Maximum Health* (2009)

A concise book which explains lots of the jargon. It covers information specific to issues the reader may face: pregnancy, nutrition, cancer, the heart and mental health.

Websites

BioNews

www.bionews.org.uk

A free web- and email-based source of news, information and comment in human genetics, embryo research and assisted conception, produced by UK charity the Progress Educational Trust. It covers both scientific and ethical developments in these areas.

GeneTests

www.ncbi.nlm.nih.gov/sites/GeneTests/

A US publicly-funded medical genetics information resource developed for physicians, other healthcare providers and researchers. It contains a laboratory directory, a clinical directory and educational materials.

Genetic Alliance UK

www.geneticalliance.org.uk

A national charity of patient organisations with a membership of over 140 other charities, supporting all those affected by genetic disorders.

Unique

www.rarechromo.org

A source of information and support for families and individuals affected by any rare chromosome disorder, and the professionals who work with them.

Key references

Belsky, J and Beaver, KM (2011) 'Cumulative-genetic plasticity, parenting and adolescent self-regulation', *Journal of Child Psychology and Psychiatry* 52, 619-26

Bodmer, W and Bonilla, C (2008) 'Common and rare variants in multifactorial susceptibility to common diseases', *Nature Genetics* 40, 695-701

Buiting, K (2010) 'Prader-Willi syndrome and Angelman syndrome', *American Journal Medical Genetics C Seminars in Medical Genetics* 154C, 365-76

Chittka, A and Chittka, L (2010) 'Epigenetics of Royalty', *PLoS Biology* 8,e1000532

Chonchaiya, W et al. (2009) 'Fragile X: A Family of Disorders', *Advances in Pediatrics* 56, 165-86

Eichler, E et al. (2010) 'Missing heritability and strategies for finding the underlying causes of complex disease', *Nature Reviews Genetics* 11, 446-450

Feinberg, AP (2010) 'Epigenomics reveals a functional genome anatomy and a new approach to common disease', *Nature Biotechnology* 28, 1049-52

Feinberg, AP et al. (2010) 'Personalized epigenomic signatures that are stable over time and covary with body mass index', *Science Translational Medicine* 2, 49ra67.

Gluckman, PD et al. (2009) 'Towards a new developmental synthesis: adaptive developmental plasticity and human disease', *The Lancet* 373, 1654-57.

Goriely, A and Wilkie, AOM (2010) 'Missing heritability: paternal age effect mutations and selfish spermatogonia', *Nature Reviews Genetics* 11, 589

Henderson, J et al. (2008) 'The burden of disease associated with filaggrin mutations: a population-based, longitudinal birth cohort study' *Journal of Allergy and Clinical Immunology* 121, 872-77

Hesson, L et al. (2010) 'Epimutations and cancer predisposition: importance and mechanisms', *Current Opinion in Genetics & Development* 20, 290-298

Hirasawa, R and Feil, R (2010) 'Genomic imprinting and human disease', *Essays in Biochemistry* 48, 187-200

International Parkinson Disease Genomics Consortium et al. (2011) 'Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies', *Lancet* 377, 641-9

Mavaddat, N et al. (2010) 'Genetic susceptibility to breast cancer', *Molecular Oncology* 4, 174-91

Ross, CA and Tabrizi, SJ (2011) 'Huntington's disease: from molecular pathogenesis to clinical treatment', *Lancet Neurology* 10, 83-98

Tobi EW, et al (2009) 'DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific', *Human Molecular Genetics* 18, 4046-53

Ubeda, F (2008) 'Evolution of genomic imprinting with biparental care: implications for Prader-Willi and Angelman syndromes', *PLoS Biology* 6, e208

Waterland, RA et al. (2010) 'Season of conception in rural gambia affects DNA methylation at putative human metastable epialleles', *PLoS Genetics* 23, 6:e1001252.

Wellcome Trust Case Control Consortium (2007) 'Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls', *Nature* 447, 661-78

An introduction to the Genetics and Epigenetics of Human Disease

Why do some people live a long and healthy life, whilst others develop life-threatening, chronic diseases in mid-life, or indeed are born with a disorder? Disease affects us all. Sequencing the first human genome in 2003 was a landmark scientific discovery, promising a revolution in medicine and treatment of disease.

This booklet aims to provide a basic introduction to the role of genetics and epigenetics in diseases ranging from the well known (breast cancer and eczema) to the more rare (Fragile X and Angelman syndromes). It begins by examining conditions where there is a single faulty gene to blame, and moves on to diseases where there are likely to be many contributing factors, including one's early developmental experience and the prevailing social and physical environment.

© 2012 Progress Educational Trust



Progress Educational Trust

informing debate on assisted conception and genetics