DNA, genes and genomes - Detailed

In our cells

From microscopes to molecules

When researchers first looked at dividing cells through early microscopes, they noticed chromosomes or 'coloured bodies' that underwent intriguing changes when cells divide. We now know that those chromosomes form our genome - well actually, the two copies of our genome that we carry in each of our cells - that contain the entire set of genetic material necessary to make a human.

In most of our cells, people have 23 pairs of chromosomes; one of each pair is inherited from our mothers and the other from our fathers. These chromosomes are made up of long threads of DNA (deoxyribonucleic acid) wrapped in bundles around a protein scaffold.

Although the protein scaffold is important, the DNA in the chromosomes carries the genetic information.

DNA, genes and genomes - Detailed

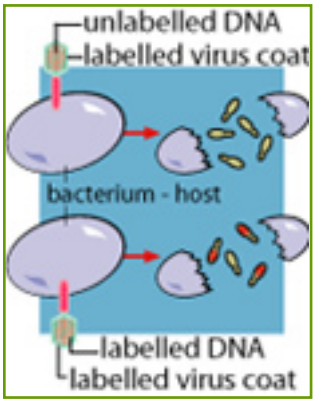
DNA: stuff of life

The chemical

For many years, people who studied genetics thought that DNA wasn't complex enough to contain all of the information needed to make up a genome. However, an elegant experiment carried out by Alfred Hershey and Martha Chase in 1952 began to convince scientists that DNA carried the genetic information, rather than protein.

The Hershey-Chase experiment

In this experiment, Hershey and Chase used a bacterial virus called T2. Although they are only made up of a shred of DNA and a scrap of protein, these viruses can hijack bacterial cells to make more copies of themselves. Scientists knew that either DNA or protein must carry the instructions for making new viruses, but they didn't know which.



When Hershey and Chase added a radioactive label to the DNA of the original virus, they found that the viruses produced were also radioactive. The researchers also labelled the protein of the original virus, but found that the viruses produced then were not radioactive.

Hershey and Chase concluded that the DNA that carried the instructions to make new viruses, and that it was the DNA that was being passed on to subsequent generations.

DNA structure

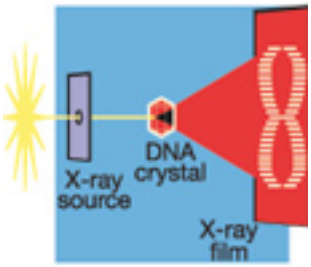
The function of DNA depends to a large extent on its structure. The discovery of the structure of DNA by James Watson and Francis Crick is one of the most famous scientific discoveries of all time. The two scientists used evidence collected by others, particularly Rosalind Franklin and Maurice Wilkins, to deduce the shape of DNA.

DNA, genes and genomes - Detailed

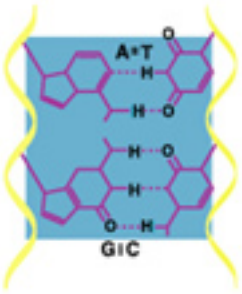
DNA: stuff of life

Franklin's experiments

One of the most important pieces of evidence came from Franklin's experiments of shining X-rays through crystals of the DNA molecule, and using photographic film to record where the scattered X-rays fall. The shadows on the film can be used to work out where the dense molecules lie. This technique is known as X-ray crystallography.



In 1953 Watson and Crick published their idea that DNA must be shaped like a double helix. A double helix resembles a twisted ladder. Each 'upright' pole of the ladder is formed from a backbone of alternating sugar and phosphate groups. Each DNA base (A, C, T or G) is attached to the backbone and the bases form the rungs. There are ten 'rungs' for each complete twist in the DNA helix.



Watson and Crick suggested that each 'rung' of the DNA helix was composed of a pair of bases, joined by hydrogen bonds. Thus A would always form hydrogen bonds with T, and C with G.

Working out the arrangement of bases in the DNA helix could also have been assisted by 'Chargaff's rules'.

Chargaff's rules

Erwin Chargaff was a Czech-American scientist who had noticed that within every DNA molecule, the percentage of A bases was always very similar to the percentage of T bases, and that the percentage of C bases was always very similar to the number of G bases.

DNA, genes and genomes - Detailed

DNA: stuff of life

From structure to function

The concept that DNA was made of a sequence of paired bases along a chemical backbone allowed Watson and Crick to draw two important conclusions:

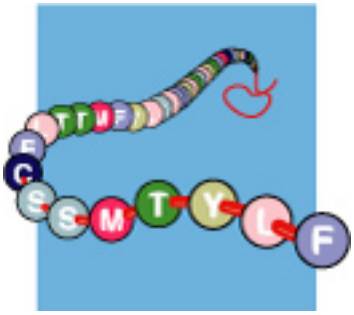
* First, the two sides or strands of DNA provide a mechanism for copying: if both strands are copied then the product is two identical 'daughter' molecules.

* Second, the order - the sequence - of bases is the digital code that carries the instructions for how a cell should behave.

If we can understand the code, we are closer to understanding how cells work.

DNA, genes and genomes - Detailed

Genes: DNA's instructions



Specific sections of DNA carry the instructions to make other molecules, usually proteins. These lengths of DNA are called genes. To interpret the DNA code, the cell first makes a copy of the DNA segment to be read. The copy then travels to another part of the cell, where the code is used to assemble a chain of protein subunits.

These processes are called transcription and translation.

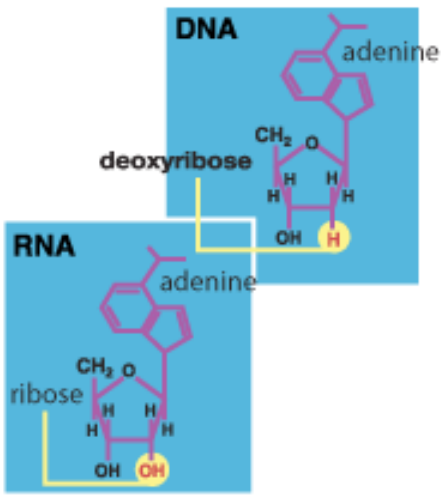
Transcription: copying the code

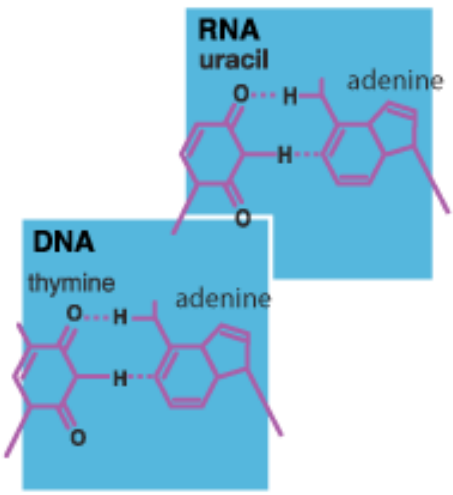
Specific sections of DNA carry the instructions to make other molecules, usually proteins. These lengths of DNA are called genes. To interpret the DNA code, the cell first makes a copy of the DNA segment to be read. The copy then travels to another part of the cell, where the code is used to assemble a chain of protein subunits.

These processes are called transcription and translation.

More about RNA

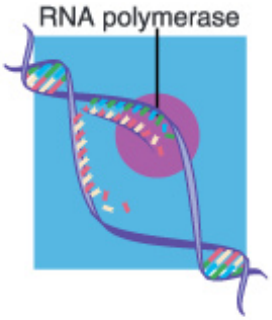
RNA has a different sugar in its sugar-phosphate backbone; it uses ribose, rather than deoxyribose. RNA also uses a different base from the thymine (T) that DNA uses; the RNA base is called uracil (U). Although they are different bases, RNA's 'U' pairs with 'A' just as DNA's 'T' does.



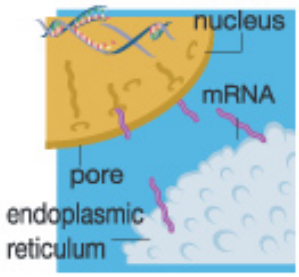


DNA, genes and genomes - Detailed

Genes: DNA's instructions

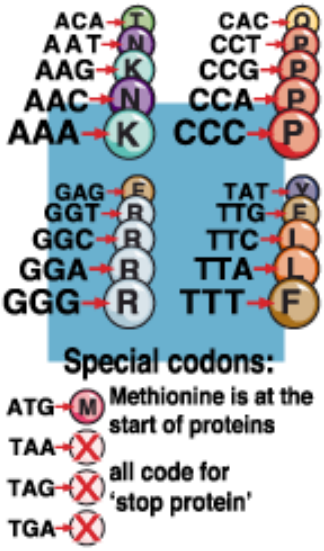


The enzyme RNA polymerase makes the RNA copy, recognizing the 'start here' and 'stop here' signals that appear in the DNA code. It uses available bases, sugars and phosphate molecules from the nucleus to form an RNA molecule that is 'complementary' to the DNA strand. This means that the base A always binds to T (or U in RNA), and C always binds to G. For the codon GTC, the complementary triplet would be CAG.



The RNA molecule that is made - called 'messenger RNA' (mRNA) - then carries its 'message' out of the nucleus to the outer part of the cell (the cytoplasm). The mRNA passes through the pores in the nuclear membrane, and makes its way to cellular components called the rough endoplasmic reticulum (ER), where proteins are made. It is called 'rough' ER because, under the microscope, has a bumpy, blobby appearance. The 'blobby' structures are ribosomes: the factories of the cell.

Translation: reading the genetic code



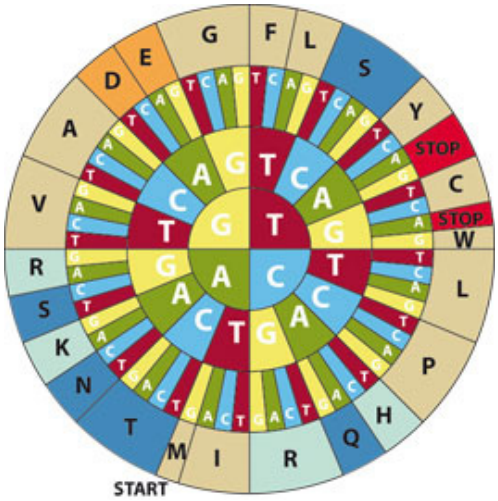
At the ribosomes, the mRNA is used as a template for assembling a protein molecule from its building blocks (amino acids). This process is called translation.

There are 20 different types of amino acids; biologists have given each a code letter. For example, M is methionine, L is leucine, F is phenylalanine (because P is proline). Translation at the ribosomes is very similar to translating from one language to another. In this case, the translation is from the four-letter language of DNA into the 20-letter language of proteins.

DNA, genes and genomes - Detailed

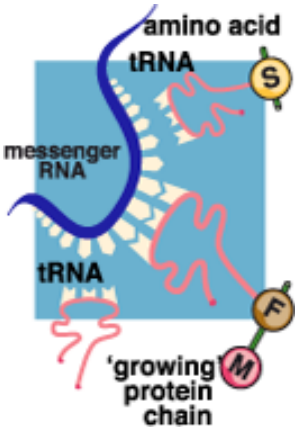
Genes: DNA's instructions

The DNA code is read three letters at a time: these DNA triplets are called codons. Since there are four different RNA letters (A, G, C and U), there are 4 x 4 x 4 = 64 different codon combinations. Most of the codons correspond to a specific amino acid. However, as there are only 20 different types of amino acid, some of the 64 codons code for the same amino acid. Three of the codons are used as 'stop' signals - telling the cell to end the transcript there - and another is the 'start' signal for proteins. (You can get a copy of a codon wheel from our downloads section)



Translation: assembling the protein chain

The genetic code is read by 'adaptor' molecules called transfer RNAs (tRNAs). These deliver amino acids to the ribosomes according to the sequence of the mRNA.



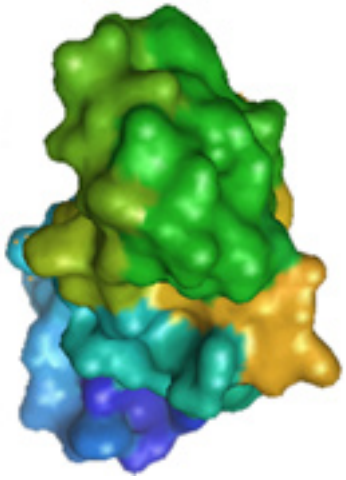
Each tRNA molecule is attached specifically to one of the 20 amino acids and three critical bases recognize the complementary codon in the mRNA. As the tRNAs bind and release, the amino acids on adjacent tRNAs are joined to form a growing amino acid chain.

Each codon on the mRNA molecule is read, one at a time. For each codon, the tRNA molecule with the complementary anticodon temporarily binds to the mRNA. The amino acid that is joined to the end of the tRNA molecule is brought in line with the growing polypeptide chain, and the amino acid links to the end of that chain. Once their amino acid is added, the tRNAs disengage from the mRNA molecule, leaving the next codons on the mRNA molecule to be 'read'.

DNA, genes and genomes - Detailed

Genes: DNA's instructions

Folding: Putting proteins to work



Proteins carry out most of active functions of a cell. From the DNA, we can recognize and read off the amino acids that make up the proteins in our bodies. However, we still only understand a little about how the chain of amino acids becomes a working protein. Proteins have three dimensional (3D) structures, which are difficult to predict just from the DNA sequence.

Many proteins are enzymes; these biological catalysts enable or speed up chemical reactions in the cell. They can act as enzymes because of their unique and complicated 3D shapes (like the structure of insulin at right). Each enzyme has a region into which two or more chemicals fit snugly. This region is called the 'active site'. While the appropriate substances are attached to the active site, they react with each other. The reaction means that they are no longer fitted to the active site, so they vacate it, leaving the enzyme free to catalyse another reaction.

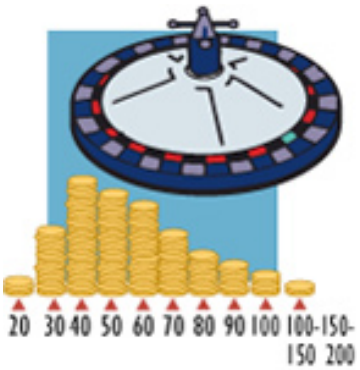
DNA, genes and genomes - Detailed

Genomes: not just genes

Making molecules

Genes were once defined as lengths of DNA that carried the instructions to make a protein. Researchers now know that the instructions in some genes can produce many proteins, and that other genes are transcribed into RNA, but don't ever produce a protein.

No-one is sure exactly how many genes there are in the human genome, but the latest estimate suggests between 20,000-25,000 - barely a third more than a fruit fly! During the Human Genome Project, researchers placed bets on the number of human genes. Their estimates ranged from 27,462 to 200,000 . . .



Although genes make up about a third of our genome, only about 2% of the DNA sequence is transcribed and translated into protein. At the moment, only about 400 non-protein-coding genes have been found, but the number may be far higher - perhaps thousands.

Regulating protein production

Although we carry complete copies of our genome in most of our cells, only certain genes are switched on at any one time. Control sequences in the DNA contain the instructions to switch genes on and off and to produce varying amounts of different proteins at different times.

In humans and other complex organisms, the coding sequences of the genes are separated into small chunks by lengths of noncoding DNA. The coding sequences are called exons, and the noncoding areas between them are called introns. Once the DNA is transcribed into RNA, these noncoding sequences are usually cut or 'spliced out'

DNA, genes and genomes - Detailed

Genomes: not just genes

However, the presence of these noncoding gene sequences is thought to give the cell the ability to generate many proteins from one gene. This way, the protein could be derived from all the exons, use only some of them, or perhaps even use some of the usually noncoding sequences to make different proteins. This is known as 'alternative splicing'.

Junk?

More than half of the DNA in our genome is made up of repeated sequences. The result is as if a printer had made a mistake and scattered lots of copies of one page of a book throughout the story. Some of these repeated areas appear to stabilize the chromosomes; others may have a role in spacing out the coding sequences so that they can be activated independently.

Recent studies have suggested that the amount of noncoding DNA increases according to the complexity of the organism. The 'junk' might conceivably have a role in enabling our genomes to change and evolve. We don't know all the answers yet, but stay tuned for the next update . .

DNA, genes and genomes - Detailed

Many genomes: variation

Am I unique?

Yes. Although, at the level of our DNA, researchers currently think that any two human beings are more than 99% alike, there are differences between us*. These differences can be small changes in a single DNA letter or duplicates and deletions of much larger chunks of DNA.

*At the DNA level, only identical twins are expected to be exactly alike, but even then, each individual twin is shaped by their environment and the lifestyle they lead.

Mutations

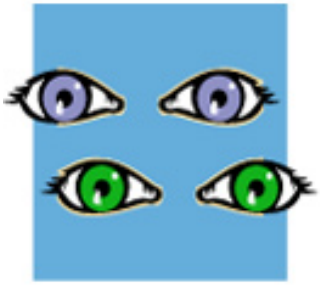
Genomes change - between generations or over a lifetime - these changes are called mutations. Mutations are happening in your cells all the time. Such changes can happen spontaneously and at random. You also inherit mutations from your parents. Environmental factors like smoking and sunlight can increase the rate of DNA mutation in your cells.

If a mutation happens in a part of the DNA that does not control activity or code for a protein, the chances are that the mutation will not even be noticed by the organism. Even mutations in the coding parts of a gene won't necessarily make a difference; for example, if the letter T is swapped for an A in the codon GCT then the protein will still be the same, since both the old and the new codon still code for the amino acid Alanine.

DNA, genes and genomes - Detailed

Many genomes: variation

However, some mutations can be detected because they lead to a change in the structure or the amount of protein produced. These relatively few mutations directly affect the ability of the protein to carry out its job - often resulting in a genetic disorder - or change the way the protein does its job (e.g. changing eye colour from blue to green, or earwax texture from sticky to crumbly).



Using information from the Human Genome Project, researchers are learning more about mutations with small or indirect effects on the way our bodies work. We expect to find that a number of small mutations that switch genes on at the wrong times, or produce too much or too little protein will contribute to common diseases such as diabetes and coronary heart disease.

Mutations: large and small

Mutations come in many different shapes and sizes: a single base may be changed into a different base, a whole segment of DNA sequence may be flipped over and reverse itself, or huge sections of the genome could be duplicated or deleted. Some of these can be seen at the chromosomal level, under a high-powered microscope, but most require techniques that can compare the sequence or activity of specific DNA segments.

These mutations can be passed on from parent to child along with the rest of the gene. This is why diseases can run in families.

Copy number: Small or large sections of DNA can be duplicated or deleted in different individuals. These duplications or deletions can have a dramatic effect on the health of an individual. For example people with Down's syndrome have an additional copy of chromosome 21 causing a condition with significant physical differences and learning difficulties. People with Huntington's Disease have an excessive number of repeated sequences within the huntingtin gene that cause degeneration of their nervous system. However, people who have additional copies of a gene called CCL3L1 have a reduced risk of HIV infection.

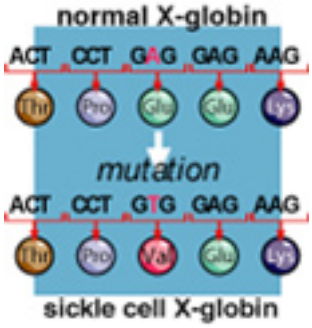
DNA, genes and genomes - Detailed

Many genomes: variation

Translocations/inversions: Small or large sections of DNA can be can reversed or swapped between different chromosomes. Some leukaemias, including chronic myeloid leukaemia (CML), are caused by chromosomes swapping material to make mutated genes. Some forms of haemophilia are caused by short segments of the Factor VIII gene being inverted, or reversed inside the gene.

Mutations: Small changes, big differences

Small changes can make big differences in our bodies. The most common mutation to cause cystic fibrosis - a disorder where a person's internal organs become clogged with thick mucus - is the loss of 3 base pairs in the CFTR gene. A single change to the dystrophin gene sequence can cause one of the muscle-weakening conditions known as muscular dystrophy.



Swapping an A for a T in a gene for haemoglobin - the protein in our blood that carries oxygen around the body - causes a serious disease called sickle cell anaemia. In people with sickle cell anaemia, the haemoglobin includes the amino acid valine where a glutamic acid would usually be. This causes the proteins to clump together and changes the shape and behaviour of the red blood cells. However, this mutation can be beneficial: people who carry this change on only one of their chromosomes are resistant to infection by the parasite that causes malaria.