

**What is epigenetics?**

There are over 50 trillion cells in the human body, and 200 different types. Almost all these cells share exactly the same DNA sequences in their nucleus, so how do they become so different? One of the biggest factors in understanding this is epigenetics. “Epigenetics” refers to the study of marks that are found on or near DNA but **do not** directly affect the DNA code itself. Instead, epigenetic marks change how the cell “reads” the DNA.

We can think of our genome as a book containing instructions on all the different proteins needed to make a human body. However, as each cell type has specific functions, they might only need a small number of these proteins. Epigenetics act like bookmarks and highlighter within that book. They let the cell know what part of that book it needs to pay attention to and what parts to ignore for the cell type they want to be.

We can also think of this as acting like a traffic light system, where epigenetic marks can tell the cell what genes should be “on” or “off”. Cells only generating RNA, and therefore protein, from genes which are “on”.

**Types of epigenetic mark**

Epigenetic marks we study are typically either 1) chemical marks added directly onto the DNA or 2) chemical marks added onto the proteins that DNA is wrapped around.

1. The main chemical mark added directly onto DNA is a methyl (CH3) group added onto cytosine “C” DNA bases to generate a methylated cytosine “mC”. We often call this **DNA methylation**. We often visualise methylation status of DNA bases using a “lollypop” symbol on top of the DNA, where the head of the lollypop is coloured in to represent a methylated DNA base and left blank to represent an unmethylated DNA base.
2. If it were completely unwound, your DNA would stretch to around 2 metres! All that must be compacted into the nucleus of each cell. To achieve this, the DNA is tightly wound around proteins called nucleosomes, which are made up of an octamer (8 subunits held together) of proteins called histones. Each histone protein has a long tail structure which can contain many epigenetic marks. These marks can include methyl groups and acetyl groups (CH3CO), and their complicated sounding names are just descriptions of their position. For example, H3K27ac is an acetyl mark found on the 27th lysine (K) on the tail of histone number 3. These marks are typically referred to as **histone modifications** and are similarly visualised by various shapes/colours on sticks on top of DNA to show specific histone marks and should be explained in the legend of any figures they appear in.

**How epigenetic marks affect gene expression**

Epigenetic marks affect the expression of genes through several different mechanisms. Some epigenetic marks can increase the compaction of DNA, which prevents the expression machinery from accessing the DNA and turns the gene “off”. Others help decompress the DNA to increase access, turning the DNA “on”. Other marks can act as “landing signals” to help actively guide expression machinery to the right areas of the DNA. Some parts of DNA can have both ON and OFF epigenetic marks, ready to quickly turn ON or OFF. Through these mechanisms epigenetic marks can directly impact the **central dogma** (see definitions), controlling what DNA regions are copied into RNA, and therefore the number of proteins made.

Epigenetics plays a major role in controlling the cell type. Stem cells (like the cells that all embryos begin life as) can turn into almost any other cell type. The DNA of stem cells is held in a specific “stem” epigenetic state, and when stem cells differentiate their epigenetic landscape changes to establish the specific gene expression required for the desired cell type, turning ON required genes and turning OFF genes which are not needed. Many of these epigenetic changes to the DNA then prevent cells from suddenly changing into another cell type. This is one of the reasons why our skin cells don’t spontaneously begin to change into taste buds, thankfully! Cancer, however, sometimes disrupts the epigenetic landscape and as a result some cancerous tumours (called teratomas) can contain many different cell types such as hair, muscle and teeth.

**Heritable epigenetics (imprinting)**

Normally during healthy fertilisation and embryo development most epigenetic marks are removed and reset to allow cells to return to a “stem” state. However, some epigenetic marks remain and are therefore passed on from parent to child. The process of maintaining these heritable epigenetic marks is called **imprinting**. At imprinting regions one of the parental copies of the DNA is epigenetically silenced and therefore the child will only express the other parent’s copy of the gene. Currently there are only 9 chromosomes which are known to have imprinted regions. This process is extremely important for inactivation of the second X chromosome in females (called **X-inactivation**) to prevent them from having twice as much gene expression than is needed for healthy development.

**Factors affecting epigenetics**

Many different factors can influence your epigenetics, including diet, pollution, radiation, exercise, toxins (smoking), and even age. To help us study the effect of these factors in people, identical twins are used. As identical twins share the same DNA and normally the same epigenetic landscape when first born, we can study any changes over time as they age. This has been extremely important in linking many factors such as diet, and how it truly effects the expression of our genes (and the genes of our children).

In science, animal models are often studied as they have a shorter lifespan and are similar in genetics to humans. A common model for mammals is the mouse. One important study which looked at the effects of diet on epigenetics and the epigenetics of offspring was the ‘Agouti mice’ experiment. Pregnant mice were fed different diets to change their levels of DNA methylation. When DNA methylation was disrupted, their offspring (pups) were born with blonde fur and developed obesity. This was because the gene called **“*agouti”* was not correctly turned OFF by DNA methylation,** resulting in unhealthily high levels of *agouti* gene expression. When the same mother was fed a healthy diet which allowed DNA methylation to occur normally, their offspring were born with brown fur and had healthy body weights as the *agouti* gene was correctly marked with DNA methylation.

The reason we study epigenetics is that it has a major effect on which genes are expressed and when. Defects in epigenetic regulation can cause many diseases, such as cancer. As epigenetics is changeable, it is a great target for therapies for these diseases.

**Key Definitions**

* **Amino acid:** A simple organic compound containing both a carboxyl (-COOH) and an amino (-NH2) group. Amino acids are translated from a codon.
* **Central Dogma:** The central dogma of molecular biology explains the flow of genetic information from DNA, to portable mRNA to make a final product, a protein.
* **Codon:** A unit of three DNA bases which refers to a specific amino acid.
* **DNA:** Deoxyribonucleic acid. It is a self-replicating material that is present in nearly all living organisms as the main constituent of chromosomes. It is the carrier of genetic information.
* **DNA base:** A unit of DNA. There are 4 bases: Adenine (A), guanine (G), thymine (T), and cytosine (C). They are read in threes as a codon to encode for amino acids.
* **Epigenetics:** The study of changes in organisms caused by modification of gene expression rather than alteration of the genetic code itself.
* **Gene:** A unit of heredity which is transferred from a parent to offspring and is held to determine some characteristic of the offspring. A distinct sequence of nucleotides forming part of a chromosome.
* **Gene expression:** Gene expression is the process by the information from a gene is used to generate a protein. First the DNA is copied into RNA, which is then converted into Proteins. Increased gene expression means more RNA copies, and therefore more proteins produced from the gene’s DNA.
* **Histone:** Any of a group of basic proteins found in chromatin. DNA wraps around histones so it fits into the cell nucleus.
* **Protein:** Proteins are large biomolecules and macromolecules that comprise one or more long chains of amino acids.
* **RNA:** Ribonucleic acid, a nucleic acid present in all living cells. Its principal role is to act as a messenger carrying instructions from DNA for controlling the synthesis of proteins.