HL IB Biology

Your notes

Mutations & Gene Editing

Contents

Gene Mutations

Your notes

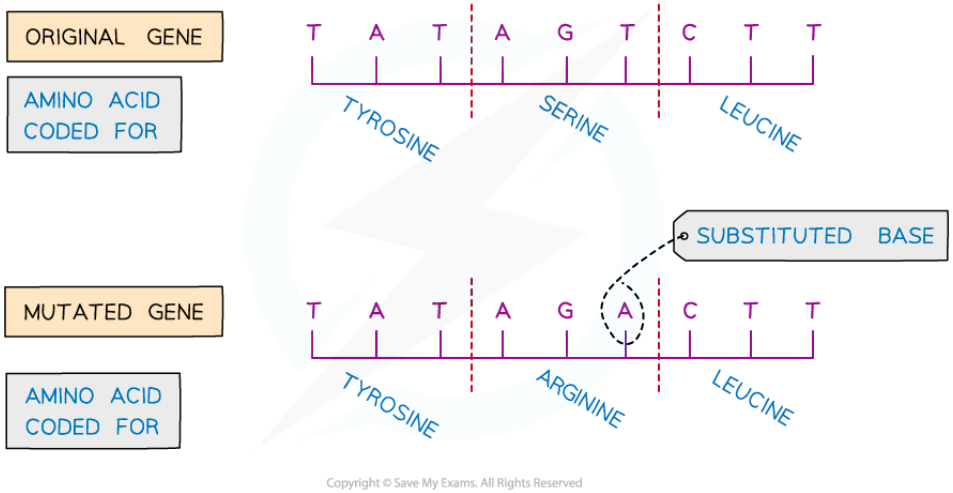
Gene Mutations

- A **gene mutation** is a change in the sequence of base pairs in a DNA molecule; this may result in a new allele
  - Mutations occur **all the time** and **at random**
  - There are certain points in the cell cycle when mutations are more likely to occur, for example, **copying errors** when DNA is being replicated (S phase of interphase)
- As the DNA base sequence determines the sequence of amino acids that make up a polypeptide, **mutations in a gene** can sometimes lead to a **change in the polypeptide** that the gene codes for
- Most mutations are **harmful** or **neutral** (have no effect) but some can be **beneficial**
- **Inheritance** of mutations:
  - Mutations present in normal body cells are **not inherited**, they are eliminated from the population once those cells die
  - Mutations within gametes are inherited by offspring, possibly causing genetic disease

Substitution mutations

- A mutation that occurs when a nucleotide base in the DNA sequence is **randomly swapped** for a different base is known as a **substitution mutation**
- A substitution mutation will only change the amino acid for the triplet (group of three consecutive bases) where the mutation occurs; it will **not have a knock-on effect** further along the gene/polypeptide

Substitution mutation diagram



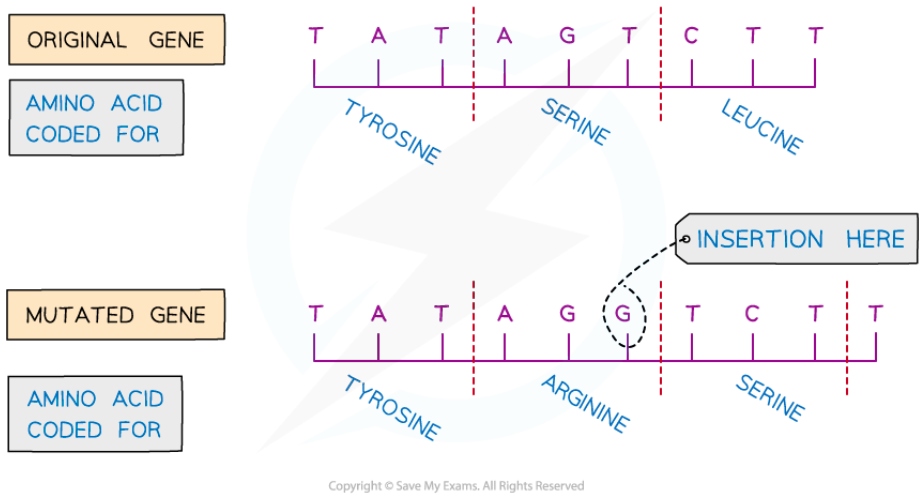
An example of a substitution mutation altering the sequence of amino acids in the polypeptide

Your notes

Insertion mutations

- A mutation that occurs when **a nucleotide (with a new base) is randomly inserted** into the DNA sequence is known as an **insertion** mutation
- An insertion mutation **changes the amino acid that would have been coded for by the original base triplet**, as it creates a **new, different** triplet of bases
  - Remember – every group of three bases in a DNA sequence codes for an amino acid
- An insertion mutation also has a **knock-on effect** by **changing the triplets** (groups of three bases) **further on in the DNA sequence**
- This is sometimes known as a **frameshift** mutation
- This **may dramatically change the amino acid sequence produced** from this gene and therefore the **ability of the polypeptide to function**

Insertion mutation diagram



An example of an insertion mutation

Deletion mutations

- A mutation that occurs when **a nucleotide (and therefore its base) is randomly deleted** from the DNA sequence
- Like an insertion mutation, a deletion mutation **changes the amino acid that would have been coded for**
- Like an insertion mutation, a deletion mutation also has a **knock-on effect** by **changing the groups of three bases further on in the DNA sequence**
- Like an insertion mutation, this is sometimes known as a **frameshift** mutation

- This **may dramatically change the amino acid sequence produced** from this gene and therefore the **ability of the polypeptide to function**

Your notes

Causes of Gene Mutations

- Mutagenic agents are environmental factors that **increase the mutation rate of cells**
  - **Radiation** can cause chemical changes in DNA, this includes:
    - **High-energy** radiation such as UV light
    - Ionising radiation such as X-rays, gamma rays and alpha particles
  - **Chemical substances** can also caused changes to DNA, examples include
    - Benzo[a]pyrene and nitrosamines found in tobacco smoke
    - Mustard gas used as a chemical weapon in World War I
  - Mutagens can also come from inside the cell such as particular enzymes that either break down DNA or produce substrates that are mutagenic
- Some mutations may be produced at **random**, this can happen most frequently during **DNA replication and repair** where errors in the nucleotide sequence are not detected by the proofreading process carried out by DNA polymerase
  - If the polymerase detects that a wrong nucleotide has been added, it will remove and replace the nucleotide before continuing with DNA synthesis
- **Most mutations do not alter the polypeptide** or only alter it slightly so that its structure or function is not changed
- As the genetic code is **degenerate** (more than one triplet code codes for the same amino acid) some mutations will not cause a change in the amino acid sequence

Randomness in Mutations

- Mutations can occur anywhere in the base sequence of a genome on all chromosomes in all organisms
  - This is how new strains of viruses or bacteria can come into existence
- Some locations of the genome are more likely to mutate than others
  - **Uncoiled DNA has a higher probability of encountering mutations** than DNA tightly coiled around a histone as it is more exposed
  - Many mutations occur in **non-coding regions of DNA** such as satellite DNA
  - **Mutation hotspots** are regions where mutations are more frequent. One hotspot is where the nucleotide cytosine (C) is followed by guanine (G) and is called a **CpG site**
    - When methylation occurs here, C can mutate into Thymine (T) in a substitution mutation
    - Where this occurs repeatedly it is known as a CpG island and is associated with particular cancers such as colorectal cancer

Intentional changes to base sequences

- No known mechanisms exist where cells are able to intentionally mutate or change their DNA base sequence
- Proofreading processes exist to change a mutation back into its original sequence but **no mechanism exists for making a deliberate change to a base or sequence of bases** with the purpose of changing a trait of the organism

Consequences of Gene Mutations


Your notes

Consequences of Base Substitutions

- Base substitutions are a mutation that occurs when **a base in the DNA sequence is randomly swapped for a different base**
  - A substitution mutation can only change the amino acid for the triplet in which the mutation occurs; it will not have a knock-on effect on the rest of the sequence
- A base substitution can result in **single nucleotide polymorphisms**, frequently called SNPs (pronounced "snips")
  - These represent a difference in a single DNA nucleotide. E.g. a SNP may replace the nucleotide cytosine (C) with the nucleotide thymine (T) in a certain stretch of DNA
- SNPs occur normally throughout a person's DNA
  - They occur once in every 300 nucleotides on average, which means there are roughly 10 million SNPs in the human genome
- SNPs are commonly found in the non-coding regions of DNA between genes
- They can act as biological markers, helping to locate genes that are associated with disease

The effect of SNPs

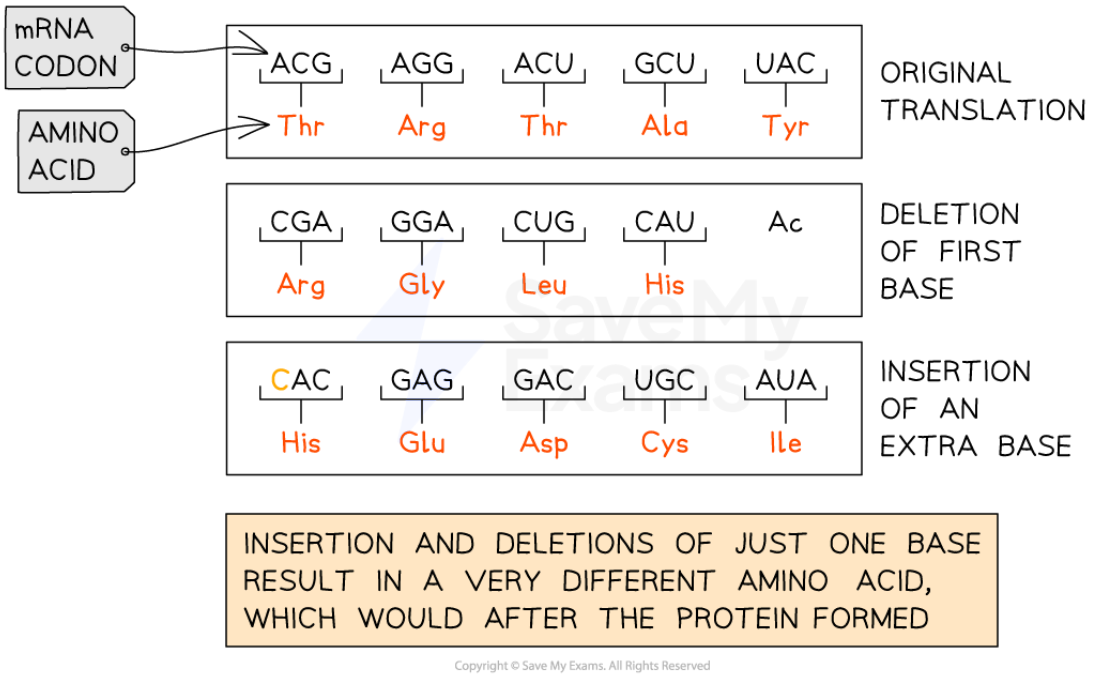
- Substitution mutations can take three forms which may or may not change the amino acid of a polypeptide chain:
  - **Silent mutations** – the **mutation does not alter the amino acid sequence** of the polypeptide (this is because certain codons may code for the same amino acid as the genetic code is degenerate)
  - **Missense mutations** – the **mutation alters a single amino acid** in the polypeptide chain (sickle cell anaemia is an example of a disease caused by a single substitution mutation changing a single amino acid in the sequence)
  - **Nonsense mutations** – the **mutation creates a premature stop codon** (signal for the cell to stop translation of the mRNA molecule into an amino acid sequence), causing the polypeptide chain produced to be incomplete and therefore affecting the final protein structure and function (cystic fibrosis is an example of a disease caused by a nonsense mutation, although this is not always the only cause)

Consequences of Insertions & Deletions

Your notes

- Insertions and deletions are two types of point mutations, which are mutations that involve a change in the DNA base sequence at a **single location**
  - An **insertion** occurs when an **extra** nucleotide is incorporated into the DNA sequence during replication
  - A deletion mutation occurs when a nucleotide is missed or **absent** from the replicated strand
- These mutations are often considered more harmful than substitutions, because they impact on the way the rest of the sequence is read by mRNA or the ribosome
- Insertions and deletions of nucleotides can also have the effect of a **frameshift mutation**
- This causes a complete change to the entire amino acid sequence of a protein after the mutation site and can cause the polypeptide to cease to function
- This happens because of the way the translated mRNA is read by the ribosomes
  - The mRNA is read in **codons (**groups of 3 nucleotides) so if an additional 1 or 2 nucleotides are added or removed, the sequence is '**shifted**'
  - The ribosome still reads the sequence of triplet codons along the length of mRNA which means the entire mRNA and resulting protein are completely different
- The result of frameshift mutation means the entire DNA sequence following the mutation will be incorrectly read. This can result in the **addition of the wrong amino acids to the polypeptide chain and/or the creation of a codon that stops the protein from growing longer**
- Although a frameshift mutation during translation is rare (10^{-5} to 10^{-7} per codon), the effects are generally catastrophic for the resulting protein
- The same can be said for large insertions and deletions of nucleotides of the DNA sequence

Consequences of Mutations Diagram



Insertion and deletion mutations can cause a frameshift in the sequence of bases which can drastically alter the amino acid coded for. The resulting protein may be very different to the original intended.

Your notes

Mutations in Germ & Somatic Cells


Your notes

- The effect of a mutation can vary depending on whether it occurs in a **germ cell or somatic cell**
  - A germ cells use meiosis to produce **gametes**
  - Somatic cells use mitosis to produce **cells all over the body** which can grow into tissues and organs

Germ cells

- If a **mutation occurs in a germ cell it can be passed on to the offspring** and next generation
  - Cells involved in inheritance of genetic information, eggs, sperm and zygote, are know as the **germ line**
  - A mutation that occurs in sperm cells could potentially affect the zygote of that offspring and **all cells developed from that zygote will contain the mutation**
  - A female that has inherited a mutation will contain the **mutation in the germ cells of their ovaries which will be passed onto future offspring**

Somatic cells

- **Somatic cell mutations are not inherited by offspring,** instead these mutations are associated with cancers
- Cancers demonstrate how important it is that **cell division** is **precisely controlled**, as cancers arise due to **uncontrolled mitosis**
- Cancerous cells divide repeatedly and uncontrollably, forming a **tumour** (an irregular mass of cells)
- Cancers start when **a mutation occurs in the genes that control cell division**
- If the mutated gene is one that **causes cancer** it is referred to as an **oncogene**
- Mutations are common events and don't lead to cancer most of the time
  - Most mutations either result in **early cell death** or result in the cell being destroyed by the **body's immune system**
  - As most cells can be easily replaced, these events usually have no harmful effect on the body
- The mutations that result in the generation of cancerous cells **do not result in early cell death or in the cell being destroyed by the body's immune system**

Mutations & Genetic Variation


Your notes

Mutations & Genetic Variation

- **Differences** exist between organisms of the same species
  - These differences are known as **variation**
  - Examples of variation include:
    - Coat colour in mammals
    - Body length in fish
    - Flower colour in flowering plants
- **Variation** results from small differences in **DNA base sequences** between individual organisms within a population
- There are several sources of these differences in DNA base sequences:
  - Mutation
  - Meiosis
  - Random fertilisation during sexual reproduction

Mutations

- The **original source of genetic variation is mutation**
- Mutation results in the **generation of new** alleles which can influence evolution of a species
- Mutations that take place in the **dividing cells of the sex organs** lead to changes in the alleles of the gametes that are passed on to the next generation
  - A new allele may be **advantageous, disadvantageous** or have **no apparent effect**
  - An advantageous allele is **more likely to be passed on** to the next generation because it increases the chance that an organism will survive and reproduce
  - A disadvantageous mutation is **more likely to die out** because an organism with such a mutation is less likely to survive and reproduce
- Mutations in a species are, in the long term, **essential for evolution by natural selection**
- Note that a mutation taking place in a body, or somatic, cell will **not be passed on to successive generations**, and so will have no impact on natural selection
- Mutation is the only source of variation in asexually reproducing species

NOS: Commercial genetic tests can yield information about potential future health and disease risk. One possible impact is that, without expert interpretation, this information could be problematic

- There are two types of genetic testing available
  - **Clinical or medical genetic testing**
    - This is carried out through **healthcare providers** such as doctors, nurse practitioners, or genetic counselors
    - Healthcare providers determine which test is needed, order the test from a laboratory, collect the DNA sample, send the DNA sample to a laboratory for testing and analysis, and importantly they **share the results with the patient and ensure understanding** of the test results and the implications to the individual and their families

- **Commercial genetic testing** provides genetic tests marketed **directly to customers**
  - The test kits can be bought online or in stores
  - Customers send the company a DNA sample and receive their results directly from the genetic company or lab
  - Commercial genetic testing provides people **access to their genetic information without necessarily involving a healthcare provider**
  - This can pose some problems
    - Commercially available genetic tests are **not scientifically validated** and can give **inaccurate results**
    - Unexpected information that a customer receives about their health, family relationships, or ancestry may be stressful or upsetting
    - People may make important decisions about disease treatment based on **inaccurate or misunderstood information from their test results**
    - Individuals often are **not provided with genetic counseling**


Your notes

Gene Editing (HL)


Your notes

Investigating Gene Function

- The **entire set** of **genetic material** of an organism is known as its **genome**
- Biologists now know the entire **human genome** (they have worked out all the genes that are found in humans)
- The **Human Genome Project** (completed in 2003) was the name of the international, collaborative research effort to determine the DNA sequence of the entire human genome and record every gene in human beings
- This was a very important breakthrough but following this scientists now want to know what each gene codes for and the effect of a gene on an organism; this will help us better treat and prevent disease

Gene knockout

- One way to find out the **effect and function of a gene** on an organism is to **remove it from the genome or make the gene unusable**; this technique is called **gene knockout**
- The organism that has had its genes "knocked out" is called a **knockout organism**
  - Common knockout organisms are laboratory mice
- Scientists create knockout organisms to study the impact of removing a gene from an organism, which often allows them to then learn something about that gene's function
- This is classed as a **genetic engineering technique**
- A genetic **library of knockout organisms** exists, such as for the bacterial species *Saccharomyces cerevisiae*, exists in order to
  - Understand the mechanism of action of a drug
  - Target specific biological processes or deficiencies
- Conditions that have been studied using gene knockout techniques include
  - Obesity
  - Diabetes
  - Cancer likelihood
  - Addiction
  - Cardiovascular disease

Examiner Tip

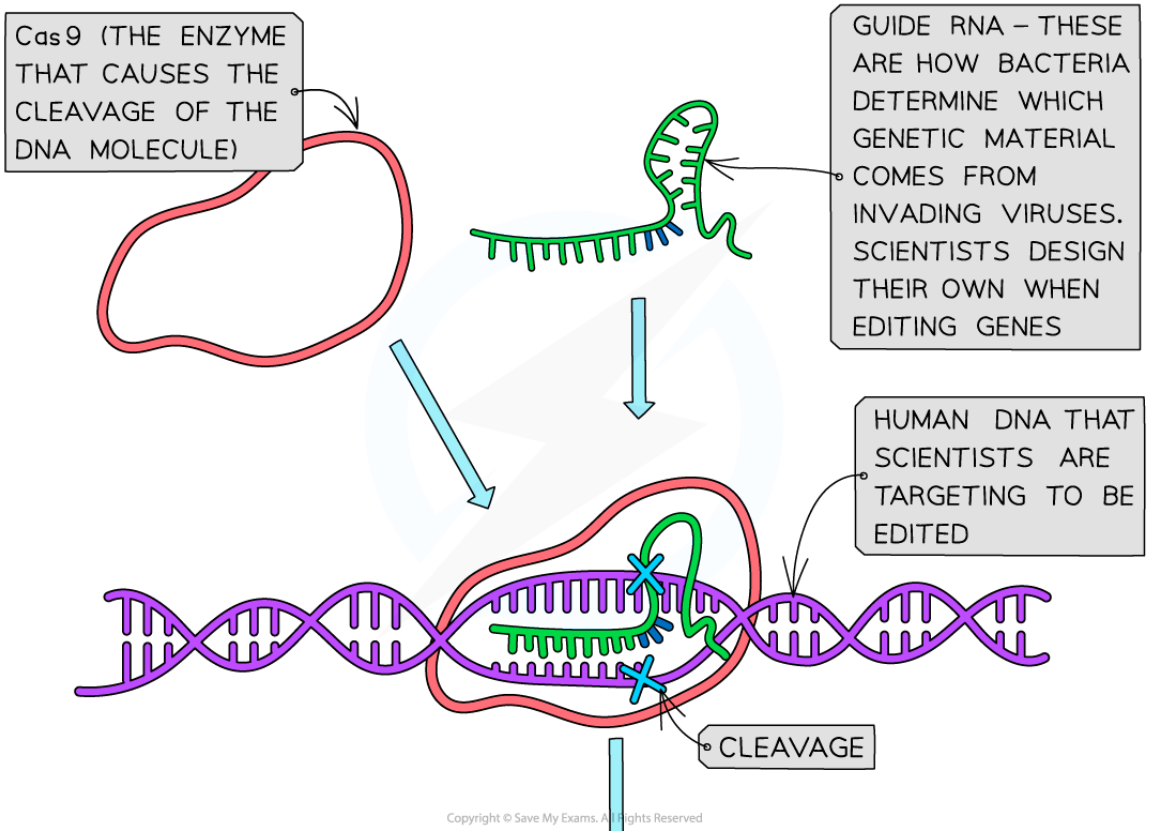
You are not required to know or understand the techniques involved in created knockout organism

Gene Editing Techniques

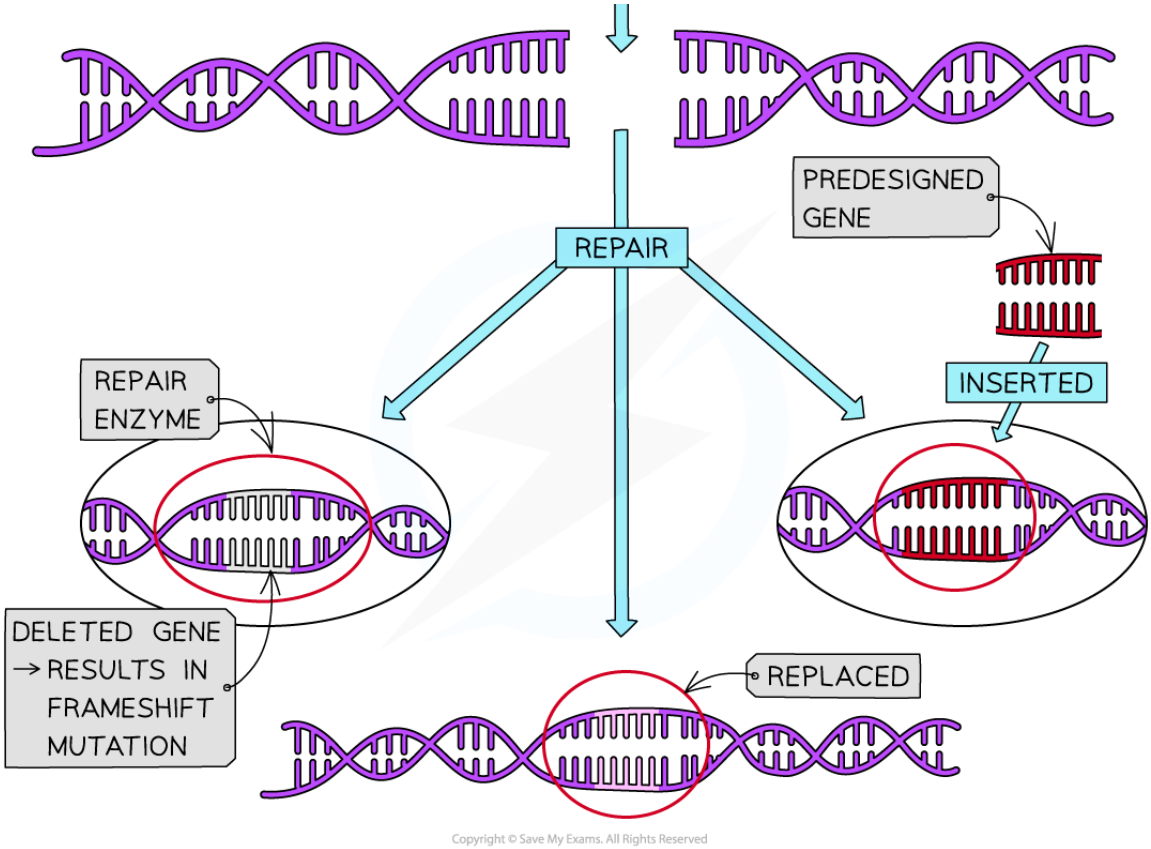
Your notes

- **Gene, or genome, editing** allows genetic engineers to alter the DNA of organisms by **inserting**, **deleting** or **replacing DNA** at specific sites in the genome known to cause disease.
  - It differs from genetic engineering in that it involves **modification of the existing DNA** of an organism rather than the insertion of DNA from another organism.
  - Note that the term 'genome' refers to **all of the DNA**, or genetic information, **found inside a cell**.
- Gene editing enables the scientists to be **more accurate** in their manipulation of the genome
- Older gene editing techniques include
  - Modifying viruses to insert DNA, e.g. into the gene causing a disease
    - This sometimes resulted in DNA being inserted into other genes causing unforeseen consequences
  - Liposomes (small spheres of lipid molecules) containing the normal version of a gene being sprayed into noses
    - This was only a short-term solution as the epithelial cells lining the nasal passageway were short lived
- Today scientists have developed **new gene editing techniques**, the most commonly used one being **CRISPR** (**C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats)
  - This technique involves using the natural defense mechanism bacteria (and some archaea) have evolved to cut the DNA strands at a specific point as determined by a guide RNA attached to an **enzyme (Cas9)**
  - Once cut scientists can then either insert, delete or replace faulty DNA with normal DNA
- Gene editing is involved in gene therapies (e.g. developing treatments for cystic fibrosis and sickle cell anaemia).
  - **Gene therapy** is the **treatment of a genetic disease by altering the person's genotype**
- As scientists learn more about the human genome (from the Human Genome Project) and the proteome, and have the technology to process large quantities of data through computational biology, they can gain a better understanding of which genes are responsible for genetic diseases and where they are located,and therefore what base changes need to occur to treat or cure the disease

Gene editing techniques diagram



Your notes



CRISPR is an example of a gene editing technique

Your notes

Examiner Tip

You are not required to know the role of the CRISPR-Cas system in prokaryotes. However, you should be familiar with an example of the successful use of this technology.

NOS: Certain potential uses of CRISPR raise ethical issues that must be addressed before implementation

- Genetic engineering practices raise a number of **ethical issues** that need to be considered
  - Issues may arise around consent of genetic data
  - Insurance companies may require data on genetic testing results
  - The need for legal control over how data is used, particular data involving human genomes
- Fortunately decisions based on the morality of genetic engineering tend to be made on a **world wide scale**
- **Ethics committees** exist which must approve all experiments carried out and they gain advice and guidance from **world leading experts**
- Countries have their own **laws** in place to **protect participants** of genetic technology research

- **International committees** have been formed to discuss and debate the issues raised and make recommendations to governments, scientists so that any policies made have considered the ethics involved
  - One such committee is The International Commission on the Clinical Use of Human Germline Genome Editing
- The **World Health Organisation** (WHO) plays an important role to create guidance for best practice and issue guidelines
- The challenge is to ensure that all policy makers and countries work together to **coordinate regulations**
- These international regulations are applied to all gene editing processes including **CRISPR**


Your notes

Conserved Sequences (HL)


Your notes

Conserved Sequences

- A **conserved sequence** is a section of DNA or RNA that shows **minimal mutations** over time
  - A conserved sequence tends to be identical or similar across a species or a group of species
- Sequences that show **little to no mutations over long periods of evolution are called highly conserved sequences**
- Examples of such sequences include those that lead to **DNA replication, transcription and translation, and proteins involved in cellular respiration**
  - Sequences for DNA helicases, tRNA and ribosomes
  - Sequences for the respiratory proteins cytochrome c and ferredoxin
- A number of **hypotheses exist for the mechanisms that lead to conserved sequences and high conserved sequences**
  - One is that the functional requirements of the gene are those that the organism cannot survive without
  - The other is that some sequences of DNA are subject to slower mutation rates

Functional requirements

- Conserved sequences and highly conserved sequences **exist within genes that code for proteins that are essential for an organism's survival** such as transcription and translation
- If these processes were unable to take place, due to mutations in the genes that code for essential proteins, then **the cell would cease to survive**
- Therefore the **functional requirements of the cell maintain the conserved sequence** and minimise mutations, as any mutations created would not be passed on to future generations
- We can hypothesise that **natural selection maintains conserved sequences** by necessity and and does not let any mutations pass to future generations

Slower mutation rates

- This hypothesis suggests that **certain sections of gene sequences are less prone to mutations and the mutation rate is much slower** than in other areas of the genome
  - Mutation rate is **how many changes there is to the DNA sequence over time**
  - It can be measured as
    - The number of base pair changes in a single gene at each generation or cell division
    - Or, the number of base pair changes in the whole genome per generation
  - **DNA repair and proofreading mechanisms are very active in coding regions of the genome**, and within genes that have high functionality
    - It is thought that areas where there are lower mutation rates do not have lower mutation but that they are spotted and corrected more frequently and so do not show up in sequenced DNA
  - Error correcting and proofreading is less active in areas of non-coding DNA so here higher rates of mutation are found