YOUR NOTES ⬇

6.3 Manipulating Genomes

CONTENTS

6.3.1 DNA SEQUENCING

Sequencing Methods

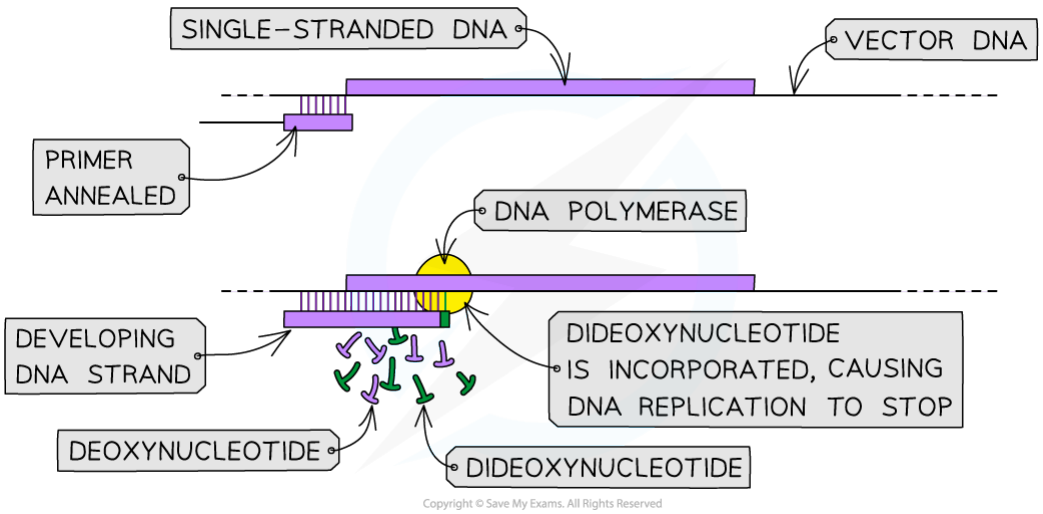
- DNA sequencing allows for the **nucleotide base sequence** of an organism's genetic material to be identified and recorded
- In the 1970s the **chain termination method** of sequencing was developed by Frederick Sanger and his colleagues
    - The chain termination method is also known as **Sanger sequencing**
- Advances in technology have enabled the development of **high-throughput sequencing** methods which allow scientists to **rapidly** sequence the genomes of organisms
    - The use of a method called **capillary electrophoresis** enables the chain termination method to be carried out in a high-throughput way (see below)
    - The newest high-throughput methods do not involve electrophoresis and are known as **next-generation sequencing** methods e.g. nanopore sequencing and pyrosequencing
- Most sequencing methods used are now automated rather than requiring manual interpretation
- The data obtained from sequencing can be entered into computers with specialised programmes that compare the base sequences of different organisms

6.3 Manipulating Genomes

YOUR NOTES ⬇

The chain termination method, or Sanger sequencing

- The **chain termination method** of DNA sequencing uses modified **nucleotides** called **dideoxynucleotides**
  - Dideoxynucleotides have a slightly different structure to the nucleotides (which can be referred to as deoxynucleotides) found within the DNA of organisms
- Dideoxynucleotides can pair with nucleotides on the **template strand** during DNA replication
  - They will pair with nucleotides that have a complementary base
- When DNA polymerase encounters a dideoxynucleotide on the developing strand it stops replicating, hence this method of sequencing is referred to as the **chain termination method**



Once the dideoxynucleotide is added to the developing strand DNA polymerase stops the replication of the developing DNA strand to produce a shortened DNA chain

6.3 Manipulating Genomes

YOUR NOTES ⬇

The chain termination method in action

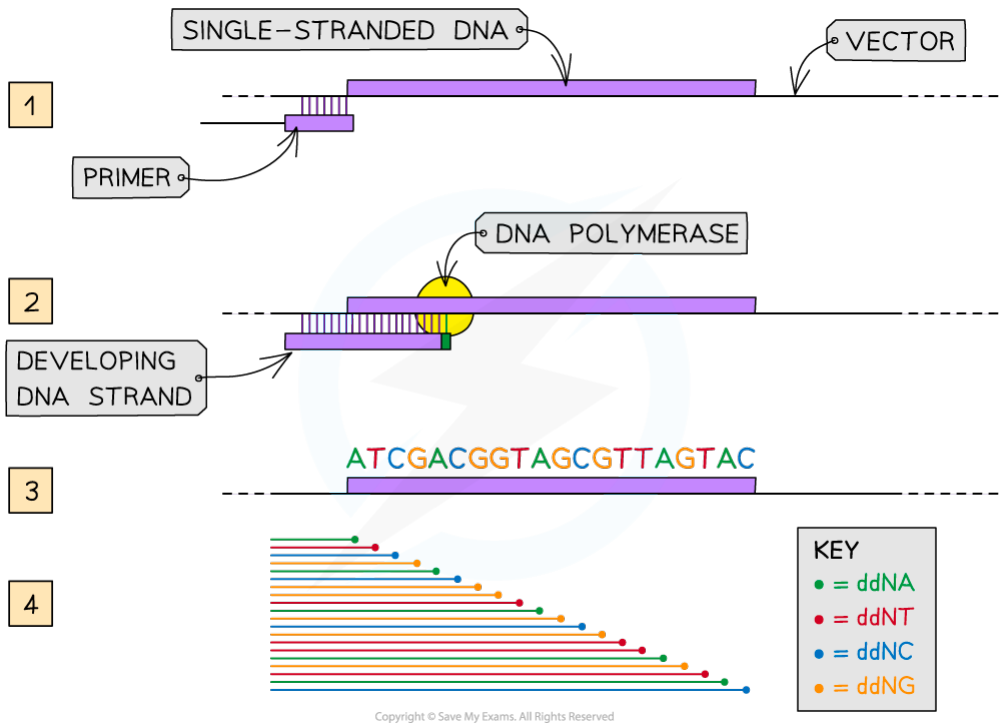
- Four test tubes are prepared that contain the DNA to be sequenced (in the form of a **single-stranded template**), **DNA polymerase**, DNA **primers**, free nucleotides A, C, T, and G, and one of the four types of dideoxynucleotide; either A*, C*, T*, or G* (you may notice that this process bears a strong resemblance to PCR, but with the addition of dideoxynucleotides, which are notated here with *)
- The test tubes are incubated at a temperature that allows the DNA polymerase to function
- The primer **anneals** to the start of the single stranded template, producing a short section of double stranded DNA at the start of the sequence
- DNA polymerase attaches to this double stranded section and begins DNA replication using the free nucleotides in the test tube
    - **Hydrogen bonds** form between the complementary bases on the nucleotides
- At any time, DNA polymerase can insert one of the dideoxynucleotides by chance which results in the **termination** of DNA replication
    - Because each of the test tubes only contains one type of dideoxynucleotide, it is possible to know what the terminal nucleotide of each fragment is (i.e. if the test tube contains A*, then researchers will know that the final nucleotide of every chain in that test tube is A)
- Because the point at which the dideoxynucleotide is inserted varies with every strand, complementary DNA chains of varying lengths are produced
    - These chains can vary in length from one nucleotide to several hundred nucleotides
- Once the incubation period has ended the new, complementary, DNA chains (also referred to as the **developing strands**) are separated from the **template DNA**
- The resulting **single-stranded DNA** chains are then separated according to length using **gel electrophoresis**
    - The gel will have four wells, one each for A*, C*, T*, and G*
    - A fragment that consists of only one nucleotide will travel all the way to the bottom of the gel, and every band above this on the gel represents the addition of one more base. E.g. If the band on the gel that travels furthest comes from the C* well, scientists can see that the first base in the sequence is C. If the next furthest band comes from the T* well, the second base in the sequence is T, and so on
    - This allows the base sequence to be built up one base at a time

6.3 Manipulating Genomes

YOUR NOTES ⬇

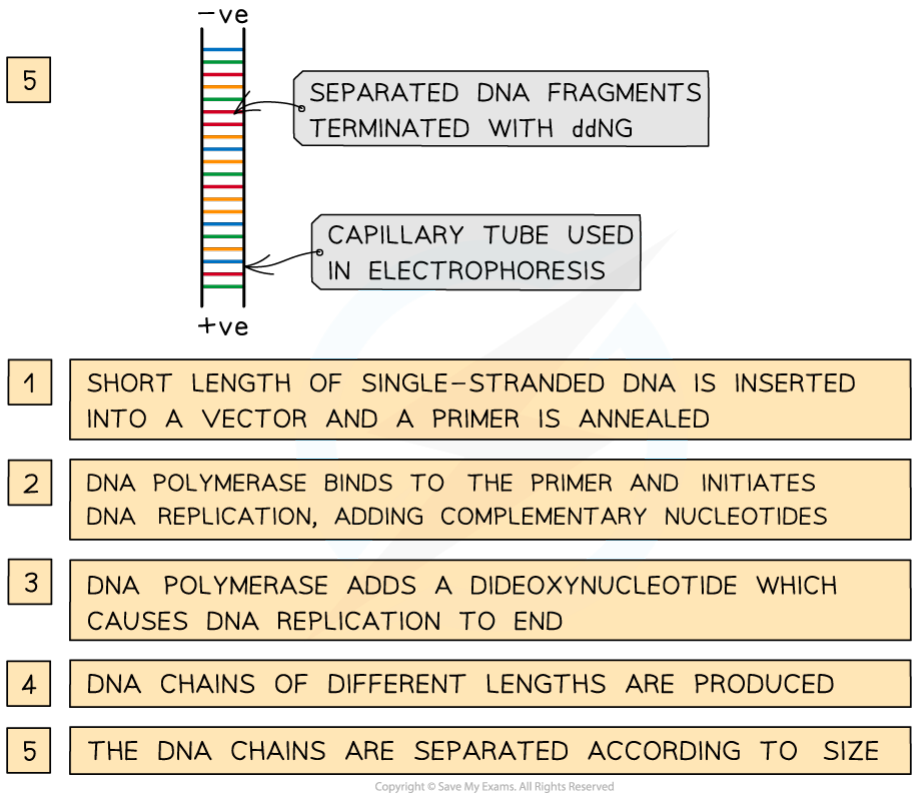
High-throughput sequencing

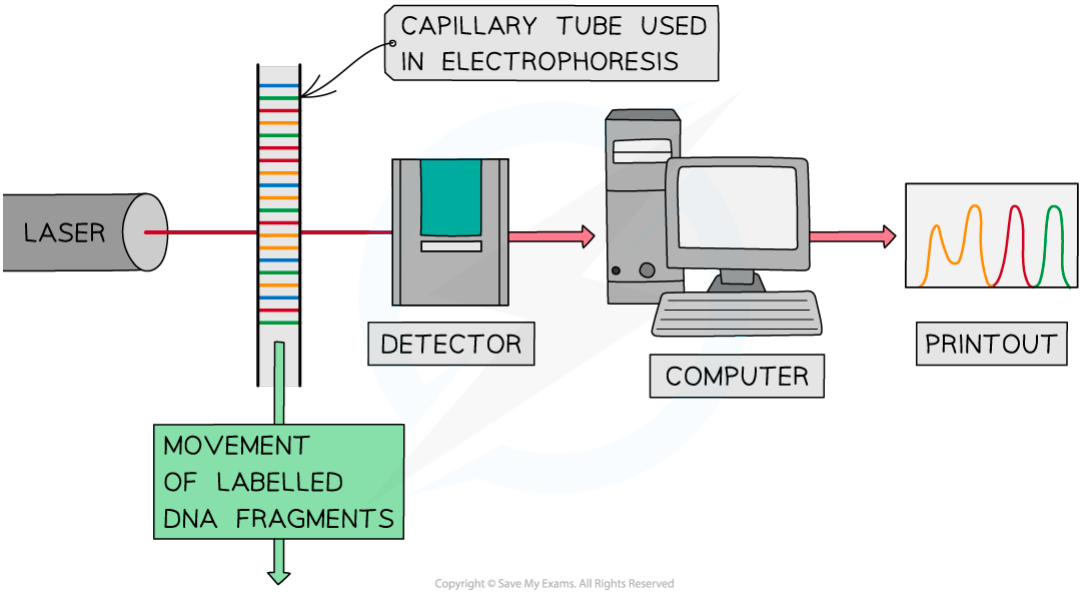
- High-throughput sequencing can also employ the chain-termination technique by using a different method of fragment separation
- Each type of dideoxynucleotide is labelled using a specific **fluorescent dye**
  - Dideoxynucleotides with an **adenine** base (ddNA) are labelled **green**
  - Dideoxynucleotides with a **thymine** base (ddNT) are labelled **red**
  - Dideoxynucleotides with a **cytosine** base (ddNC) are labelled **blue**
  - Dideoxynucleotides with a **guanine** base (ddNG) are labelled **yellow**
- The **single-stranded** DNA chains are separated according to mass using a specific type of **electrophoresis** that takes place inside a **capillary tube**
  - Known as **capillary electrophoresis**
  - This type of electrophoresis has a very **high resolution**. It is capable of separating chains of DNA that vary by only one nucleotide in length
- A **laser beam** is used to illuminate all of the dideoxynucleotides, and a **detector** then **r**eads the colour and position of each fluorescence
- The detector feeds the information into a **computer** where it is stored or printed out for analysis



6.3 Manipulating Genomes

YOUR NOTES





The process of capillary electrophoresis. This is a high-throughput way of carrying out the chain termination method. Note that because this method is essentially still Sanger sequencing, it cannot be referred to as next-generation sequencing

6.3 Manipulating Genomes

YOUR NOTES ⬇

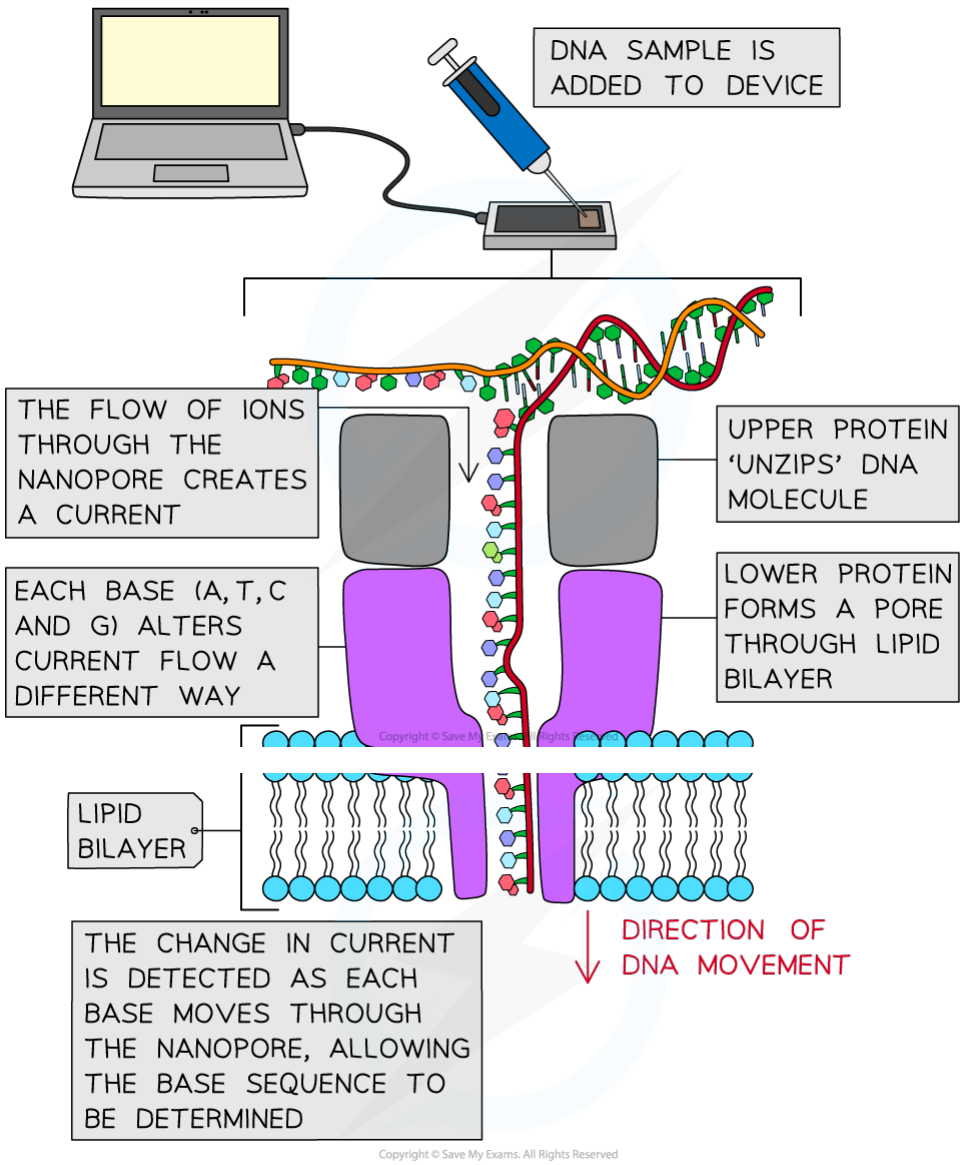
- The increase in speed enabled by high-throughput sequencing has allowed scientists to sequence and analyse the genomes of many organisms
- Scientists can determine the function of sections of DNA by 'knocking out' genes to see how this affects an organism
- Genes can be rewritten to alter their function, and then inserted into cells using genetic engineering techniques; this means that scientists can potentially design new molecules with huge potential for drug production (this branch of biology is known as **synthetic biology)**
- Genome sequence data can also provide information about evolutionary relationships

Next-generation sequencing

- Any method of DNA sequencing that has replaced the Sanger method is referred to as **next-generation sequencing** (NGS)
- Thousands to millions of DNA molecules can be sequenced at the same time (**in parallel**)
- NGS methods can be one thousand times **faster** than older methods of sequencing
- The reduction in time required for sequencing means that **costs are also greatly reduced**
    - NGS methods cost roughly 0.1% of the cost of chain-termination methods
- **Nanopore** sequencing is currently being developed by scientists
    - This method of sequencing will be extremely rapid and allow for sequence data to be obtained **outside the lab** and used for a range of applications

6.3 Manipulating Genomes

YOUR NOTES



A vertical section through a nanopore, showing how it can be used in next-generation DNA sequencing

6.3 Manipulating Genomes

YOUR NOTES

Exam Tip

Examiners may ask you which DNA strand the base sequence has been obtained for. In Sanger sequencing methods, it is the base sequence of the **developing/test strand** that is being identified, not the template strand that was initially provided. Due to the complementary nature of DNA sequences, once you know the base sequence of the developing/test strand you can automatically work out the sequence of the template strand according to base-pair rules!

Adenine pairs with thymine and cytosine pairs with guanine. So if a test strand had the sequence: ATGC then the template strand would have the sequence: TACG.

YOUR NOTES ⬇

6.3 Manipulating Genomes

6.3.2 COMPARING GENOMES

Comparing Genomes

- A **genome** contains **all of the genes within an organism**
- Advances in technology have allowed scientists to sequence the genes within an organism's genome
- Sequencing projects have read the genomes of a wide range of organisms from flatworms to humans
- Genome-wide comparisons can be made **between individuals** and **between species**

Sequencing DNA to determine protein sequences

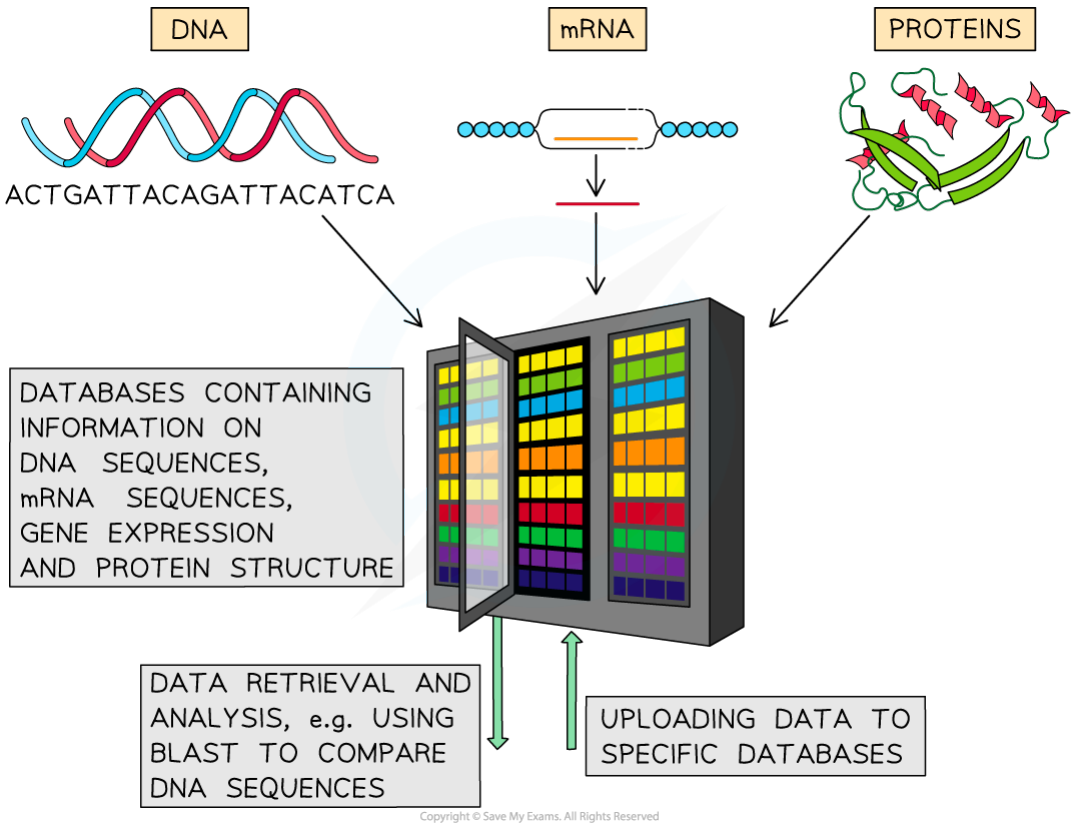
- The genetic code can be used to **predict the amino acid sequence** within a protein
- Once scientists know the amino acid sequence they can predict how the new protein will fold into its tertiary structure
- This information can be used for a range of applications, such as in **synthetic biology**

Bioinformatics

- Bioinformatics is a field of biology that involves the storage, retrieval, and analysis of data from biological studies
- These studies may generate data on **DNA sequences, RNA sequences,** and **protein sequences,** as well as on the relationship between **genotype** and **phenotype**
- High-power **computers** are required to create databases
- The large databases contain information about an organism's **gene sequences** and **amino acid/protein sequences**
- Once a genome is sequenced, bioinformatics allows scientists to make **comparisons with the genomes of other organisms** using the many databases available
- This can help to find the degree of similarity between organisms which then gives an indication of how closely related the organisms are
- This can be useful for scientists looking for organisms that could be used in experiments as a **model organism** for humans
    - E.g. The nematode worm *Caenorhabditis elegans* is an animal that has been used as a model organism for studying the genetics of **organ development, neurone development and cell death**. It was the first multicellular organism to have its genome fully sequenced and as it has few cells (less than 1000), and is transparent, it has been a useful model organism
- Bioinformatics has contributed to the study of genetic variation, evolutionary relationships, genotype-phenotype relationships, and epidemiology (see below)

6.3 Manipulating Genomes

YOUR NOTES



Bioinformatics allows for large amounts of sequence data to be instantly available to researchers across the globe

Genetic variation and evolutionary relationships

- The **genetic variation** within a species can be investigated
  - Many individuals of the same species have their genomes sequenced and compared
  - A species that has a high level of genetic variation will exhibit a **large number of differences** in base sequences between individuals
- The **evolutionary relationships** between species can be investigated by comparing the genomes of different species
  - Species with a small number of differences between their genomes are likely to share a **more recent common ancestor** than species with a large number of differences
  - The protein **cytochrome c** is involved in respiration, and so is found in a large number of species (including plants, animals, and unicellular organisms). For this reason it is especially useful for making comparisons between different species

Genotype-phenotype relationships

- Genome sequencing can aid the understanding of **gene function and interaction**
- Genotype-phenotype relationships are explored by “knocking out” different genes (stopping their expression) and observing the effect it has on the phenotype of an organism
  - When an organism’s genome sequence is known, scientists can **target specific base sequences** to knock out

6.3 Manipulating Genomes

YOUR NOTES ⬇

Epidemiology

- Epidemiologists study the spread of infectious disease within populations
- The genomes of pathogens can be sequenced and analysed to **aid research and disease control**
  - **Highly infectious strains** can be identified
    - E.g. the Delta variant of SARS-CoV-2 (a well-known coronavirus)
  - The ability of a pathogen to **infect multiple** species can be investigated
    - E.g. Ebola can infect primates as well as humans
  - The most appropriate **control measures** can be implemented based on the data provided
  - Potential antigens for use in **vaccine production** can be identified

Genome comparison in action: The Human Genome Project

- A genome project works by collecting **DNA samples** from many individuals of a species. These DNA samples are then sequenced and compared to create a **reference genome**
  - More than one individual is used to create the reference genome as one organism may have **anomalies/mutations** in its DNA sequence that are atypical of the species
- The Human Genome Project (HGP) began in 1990 as an **international, collaborative research programme**
- It was **publicly funded** so that there would be no commercial interests or influence
- DNA samples were taken from multiple people around the world, sequenced, and used to create a reference genome
- Laboratories around the globe were responsible for sequencing different sections of specific chromosomes
- It was decided that the **data** created from the project would be made **publicly available**
  - As a result, the data can be shared rapidly between researchers
  - The information discovered could also be used by any researcher and so maximised for human benefit
- By 2003 the human genome had been sequenced to 99.9% accuracy
- The finished genome was over 3 billion base pairs long but contained only about 25,000 genes, a surprisingly low number
- Work is currently underway to sequence the human proteome and the human epigenome

6.3 Manipulating Genomes

YOUR NOTES ⬇

Applications of the Human Genome Project

- The information generated from the HGP has been used to tackle human health issues with the end goal of finding cures for diseases
- Scientists have noticed a correlation between **changes in specific genes** and the **likelihood of developing certain inherited diseases**
- For example, several genes within the human genome have been linked to increased risk of certain **cancers**
    - If an individuals BRCA1 and BRCA2 genes are mutated then they are substantially more likely to develop breast cancer
- There have also been specific genes linked to the development of Alzheimer's disease

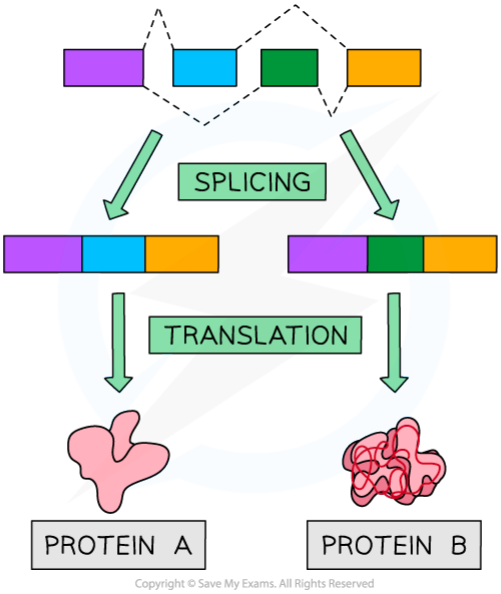
YOUR NOTES ⬇

6.3 Manipulating Genomes

6.3.3 NON-CODING DNA & REGULATORY GENES

Non-Coding DNA & Regulatory Genes

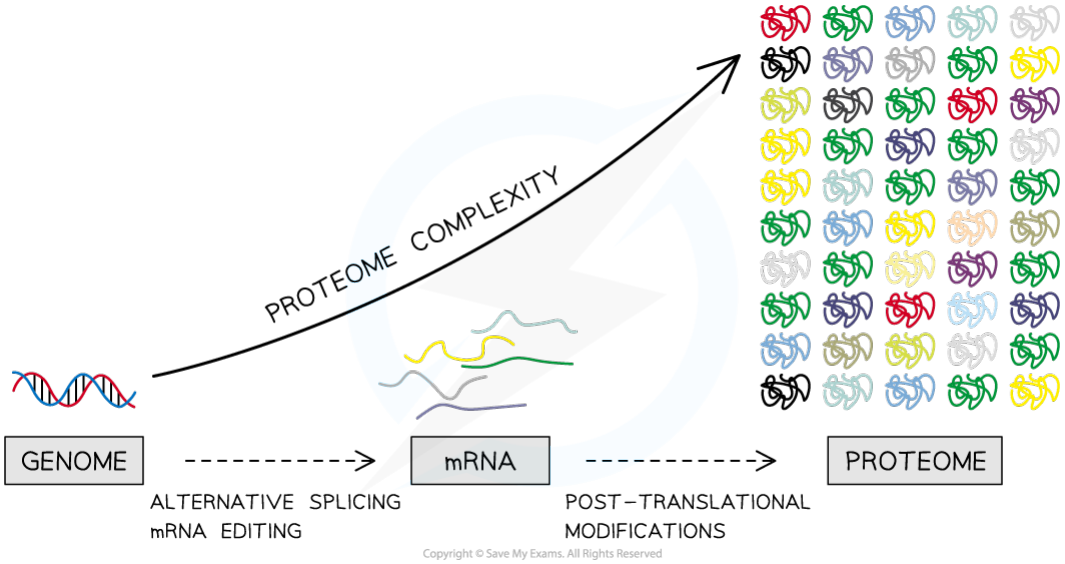
- It can be highly difficult to translate the Genome of **complex organisms** into their proteome
- Determining the proteome of humans is difficult as large amounts of **non-coding DNA** are present in human genomes
    - It can be very **hard to identify** these sections of DNA from the coding DNA
- The presence of **regulatory genes** and the process of **alternative splicing** in human genomes also affects gene expression and the synthesis of proteins
- The proteome is **larger** than the genome due to:
    - Alternative splicing
    - Post-translational modification of proteins (often takes place in the Golgi apparatus)



Alternative splicing allows for a single gene to produce multiple proteins

6.3 Manipulating Genomes

YOUR NOTES



The proteome is larger than the genome

Exam Tip

You are expected to know the definitions of genome and proteome for the exam.

YOUR NOTES ⬇

6.3 Manipulating Genomes

6.3.4 SYNTHETIC BIOLOGY

Synthetic Biology

Sequencing DNA to determine protein sequences

- The genetic code can be used to predict the **amino acid sequence** within a protein
- Once scientists know the amino acid sequence they can predict how the new protein will fold into its tertiary structure
- This information can be used for a range of applications, such as in **synthetic biology**

Synthetic biology

- Synthetic biology is a recent area of research that aims to **create new** biological parts, devices, and systems, or to **redesign** systems that already exist in nature
- It goes beyond genetic engineering, as it involves **large alterations to an organism's genome**. This new genome can cause a cell to **operate in a novel way**, not yet seen before
- The assembly of the new genome can be done using existing DNA sequences or using entirely new sequences
    - These new sequences can be designed and written (using special computer programmes) so that they produce specific proteins

Synthetic biology in action: producing artemisinin

- The most well-known use of synthetic biology is the commercial production of **antimalarial drug, artemisinin**
- Artemisinin was first isolated in China from the native plant *Artemisia annua*
- *A. annua* is difficult to cultivate, leading to an unstable supply of artemisinin at an ever-changing price often too expensive for those needing the drug most
- Scientists have constructed a DNA sequence for a whole new metabolic pathway containing genes from bacteria, yeast, and *A. annua.* This pathway results in the production of artemisinic acid, a precursor to artemisinin
- This pathway can be inserted into yeast cells which then produce artemisinic acid. The conversion of this precursor into artemisinin can then be carried out using an inexpensive process

YOUR NOTES ⬇

6.3 Manipulating Genomes

6.3.5 POLYMERASE CHAIN REACTION

Polymerase Chain Reaction (PCR)

- **Polymerase chain reaction** (PCR) is a common molecular biology technique used in most applications of gene technology
  - For example, DNA profiling (eg. identification of criminals and determining paternity) or genetic engineering
- It can be described as the **in vitro method of DNA amplification**
- It is used to produce **large quantities** of specific fragments of DNA or RNA from very small quantities (even just one molecule of DNA or RNA).
  - Using PCR, scientists can produce billions of **identical copies** of the DNA or RNA samples within a few hours, these can then be used for **analysis**

The requirements of PCR

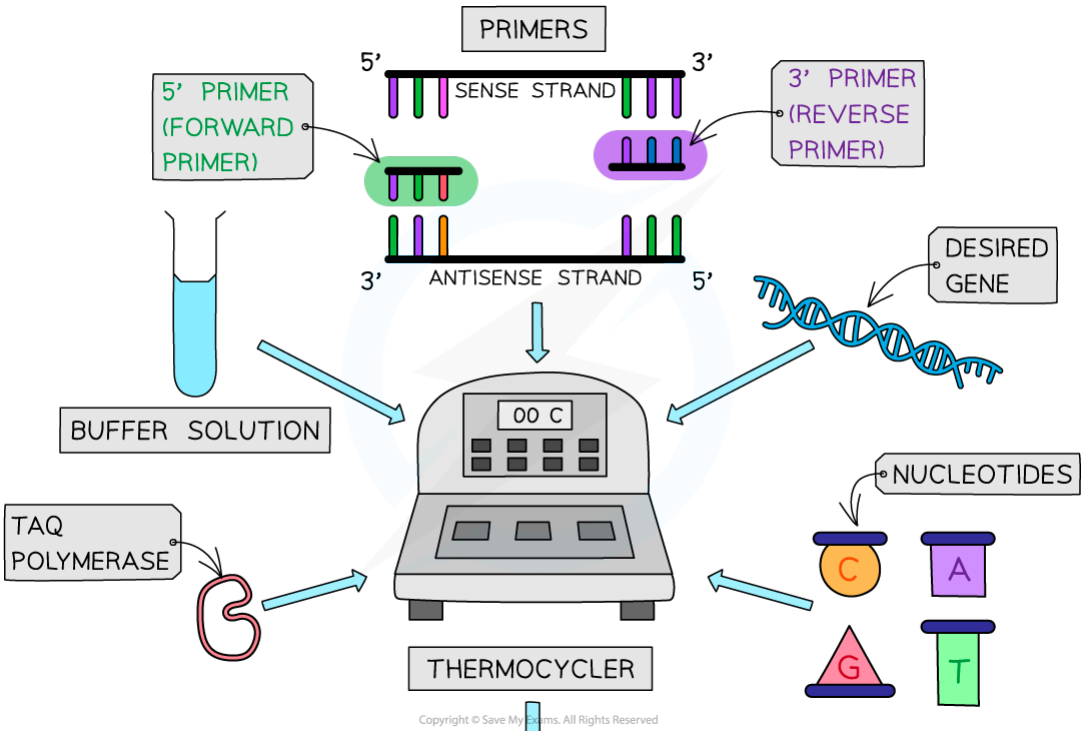
- Each PCR reaction requires:
  - **Target DNA** or RNA being amplified
  - **Primers** (forward and reverse) – these are short sequences of single-stranded DNA that have base sequences complementary to the 3' end of the DNA or RNA being copied. They define the region that is to be amplified by identifying to the DNA polymerase where to begin building the new strands
  - **DNA polymerase** – is the enzyme used to build the new DNA or RNA strand. The most commonly used polymerase is ***Taq* polymerase** as it comes from a thermophilic bacterium *Thermus aquaticus* which means it **does not denature** at the **high temperature** involved during the first stage of the PCR reaction and secondly, its optimum temperature is high enough to prevent annealing of the DNA strands that have not been copied yet
  - **Free nucleotides** – used in the construction of the DNA or RNA strands
  - **Buffer solution** – to provide the optimum pH for the reactions to occur in

6.3 Manipulating Genomes

YOUR NOTES

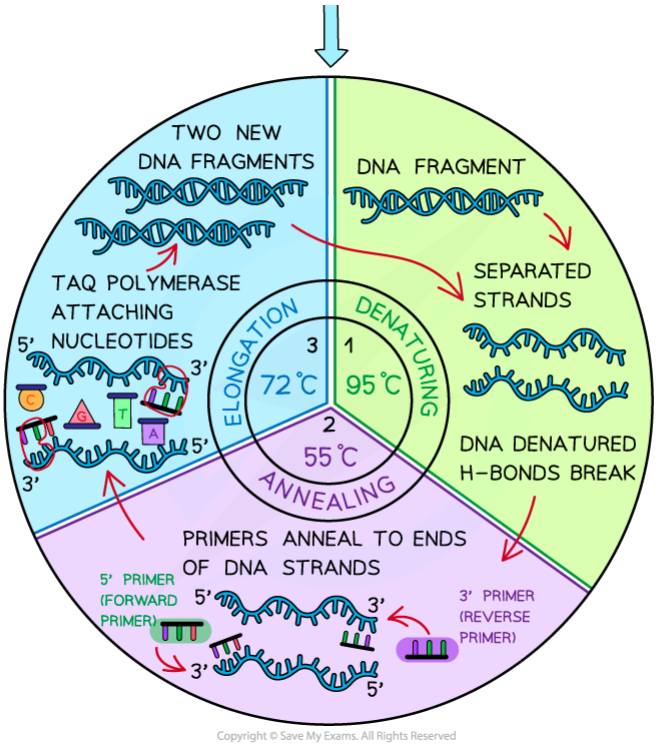
The key stages of PCR

- The PCR process involves **three key stages** per cycle
- In each cycle the DNA is doubled (so in a standard run of 20 cycles a million DNA molecules are produced)
- The PCR process occurs in a piece of specialist equipment called a **thermal cycler,** which automatically provides the **optimal temperature** for each stage and controls the **length of time** spent at each stage
- The three stages are:
    1. **Denaturation** – the double-stranded DNA is heated to 95°C which breaks the hydrogen bonds that bond the two DNA strands together
    2. Annealing – the temperature is decreased to between 50 – 60°C so that primers (forward and reverse ones) can anneal to the ends of the single strands of DNA
    3. **Elongation / Extension –** the temperature is increased to 72°C for at least a minute, as this is the optimum temperature for *Taq* polymerase to build the complementary strands of DNA to produce the new identical double-stranded DNA molecules



6.3 Manipulating Genomes

YOUR NOTES



The substances required for polymerase chain reaction to occur and the three key stages of the reaction

Exam Tip

It is important to know the three stages and the temperatures the reactions occur at during the different stages. You must also know why the *Taq* polymerase is used in PCR.

6.3 Manipulating Genomes

YOUR NOTES ⬇

6.3.6 ELECTROPHORESIS

Electrophoresis

- **Gel electrophoresis** is a technique used widely in the analysis of DNA, RNA, and proteins
- During electrophoresis, the **molecules** are **separated** according to their **size/mass** and their **net (overall) charge**
- This separation occurs because:
  - The **electrical charge** molecules carry:
    - Positively charged molecules will move towards the cathode (negative pole) whereas negatively charged molecules will move towards the anode (positive pole) eg. **DNA is negatively charged** due to the **phosphate** groups and thus when placed in an electric field the molecules **move towards the anode**
  - The **different sizes** of the molecules:
    - Different sized molecules move through the gel (agarose for DNA and polyacrylamide – PAG for proteins) at different rates. The tiny pores in the gel result in **smaller molecules** moving **quickly**, whereas **larger molecules** move **slowly**
  - The **type of gel**:
    - Different gels have different sized pores which affect the speed at which the molecules can move through them

DNA separation

- DNA can be collected from almost anywhere on the body, e.g. the root of a hair or saliva from a cup. After collection DNA must be prepared for gel electrophoresis so that the **DNA** can be **sequenced** or analysed for **genetic profiling** (**fingerprinting**)
- To prepare the fragments scientists must first increase (amplify) the number of DNA molecules by the polymerase chain reaction (PCR).
- Then **restriction endonucleases** (enzymes) are used to cut the DNA into fragments
  - Different restriction enzymes cut the DNA at different base sequences. Therefore scientists use enzymes that will cut close to the **variable number tandem repeat** (VNTR) regions
  - **Variable number tandem repeats** (VNTRs) are regions found in the non-coding part of DNA. They contain **variable numbers of repeated DNA sequences** and are known to **vary between different people** (except for identical twins). These VNTR may be referred to as ‘satellite’ or ‘microsatellite’ DNA

6.3 Manipulating Genomes

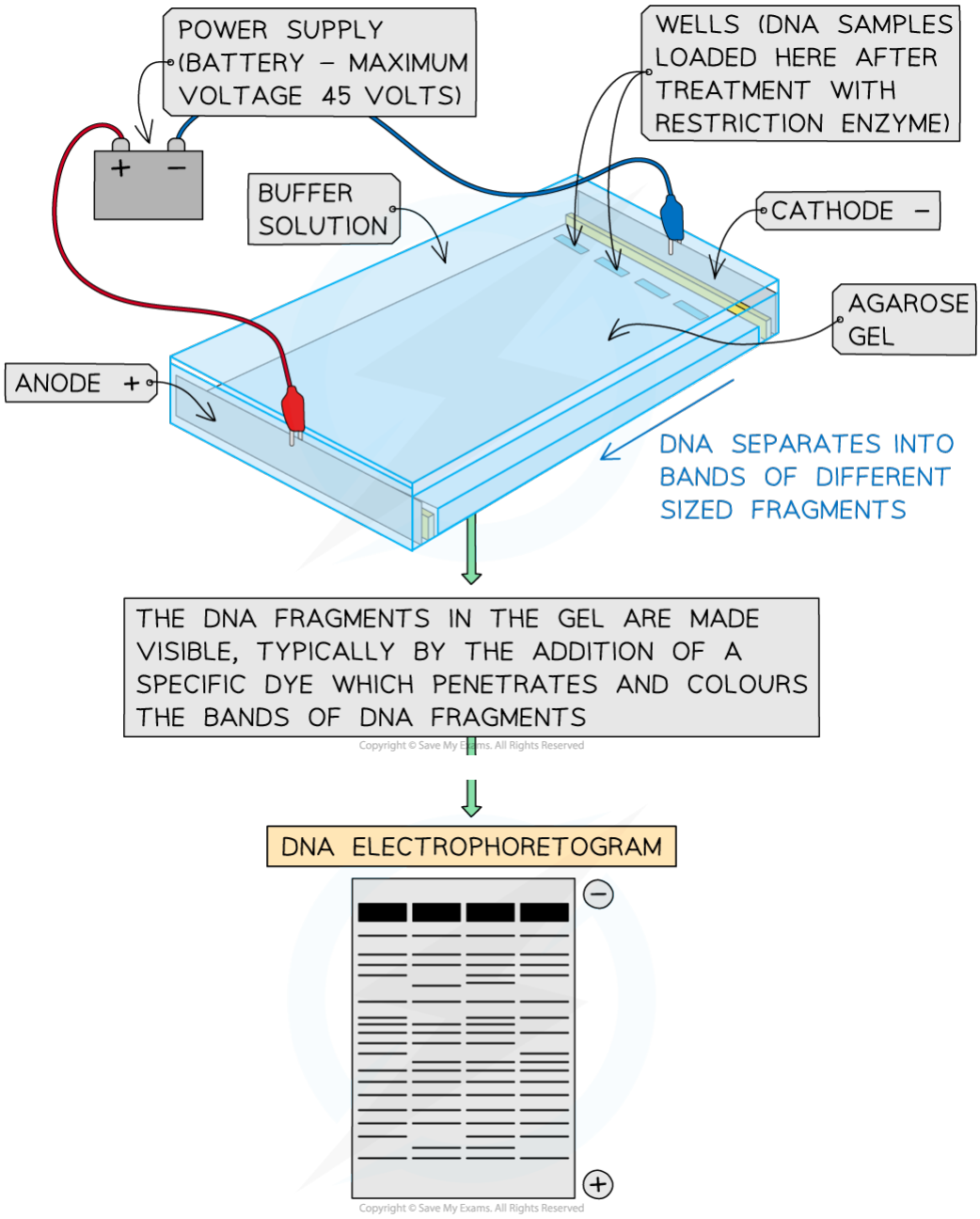
YOUR NOTES ⬇

Method

- To separate the DNA fragments in gel electrophoresis the scientists :
    1. Create an agarose gel plate in a tank. Wells (a series of groves) are cut into the gel at one end
    2. Submerge the gel in an electrolyte solution (a salt solution that conducts electricity) in the tank
    3. Load (insert) the fragments into the wells using a micropipette
    4. Apply an electrical current to the tank. The negative electrode must be connected to the end of the plate with the wells as the DNA fragments will then move towards the anode (positive pole) due to the attraction between the negatively charged phosphates of DNA and the anode
    5. The smaller mass / shorter pieces of DNA fragments will move faster and further from the wells than the larger fragments
    6. The fragments are not visible so must be transferred onto absorbent paper or nitrocellulose which is then heated to separate the two DNA strands. **Probes** are then added, after which an X-ray image is taken or UV-light is shone onto the paper producing a pattern of bands which is generally compared to a control fragment of DNA
- **Probes** are **single-stranded DNA sequences** that are **complementary** to the **VNTR** regions sought by the scientists. The probes also contain a means by which to be identified. This can either be:
    - A **radioactive label** (eg. a phosphorus isotope) which causes the probes to emit radiation that makes the X-ray film go dark, creating a pattern of dark bands
    - A **fluorescent stain / dye** (eg. ethidium bromide) which fluoresces (shines) when exposed to ultraviolet (UV) light, creating a pattern of coloured bands

6.3 Manipulating Genomes

YOUR NOTES



The process of electrophoresis

6.3 Manipulating Genomes

YOUR NOTES ⬇

Protein separation

- The different amino acids (because of the different R groups) determine the charge of proteins. The charge of the R groups depends on the pH and therefore buffer solutions are used during the separation of proteins to keep the pH constant
    - Proteins are prepared for electrophoresis by:
        - Denaturing (to break the disulfide bonds)
        - Then manipulating the proteins into rod shapes (which are negatively charged) to allow separation by size
- Gel electrophoresis can be used to show genotypes of individuals by separating polypeptide chains produced by different alleles
    - eg. The haemoglobin variants, α-globin, β-globin and the sickle cell anaemia variant of β-globin, have different net charges and therefore will separate out during electrophoresis to show the presence of the sickle cell allele

Exam Tip

Remember gel electrophoresis is the separation of molecules according to their size and charge (negatively charged DNA molecules move to the positive pole). Examiners like to ask questions about gel electrophoresis.

6.3 Manipulating Genomes

YOUR NOTES ⬇

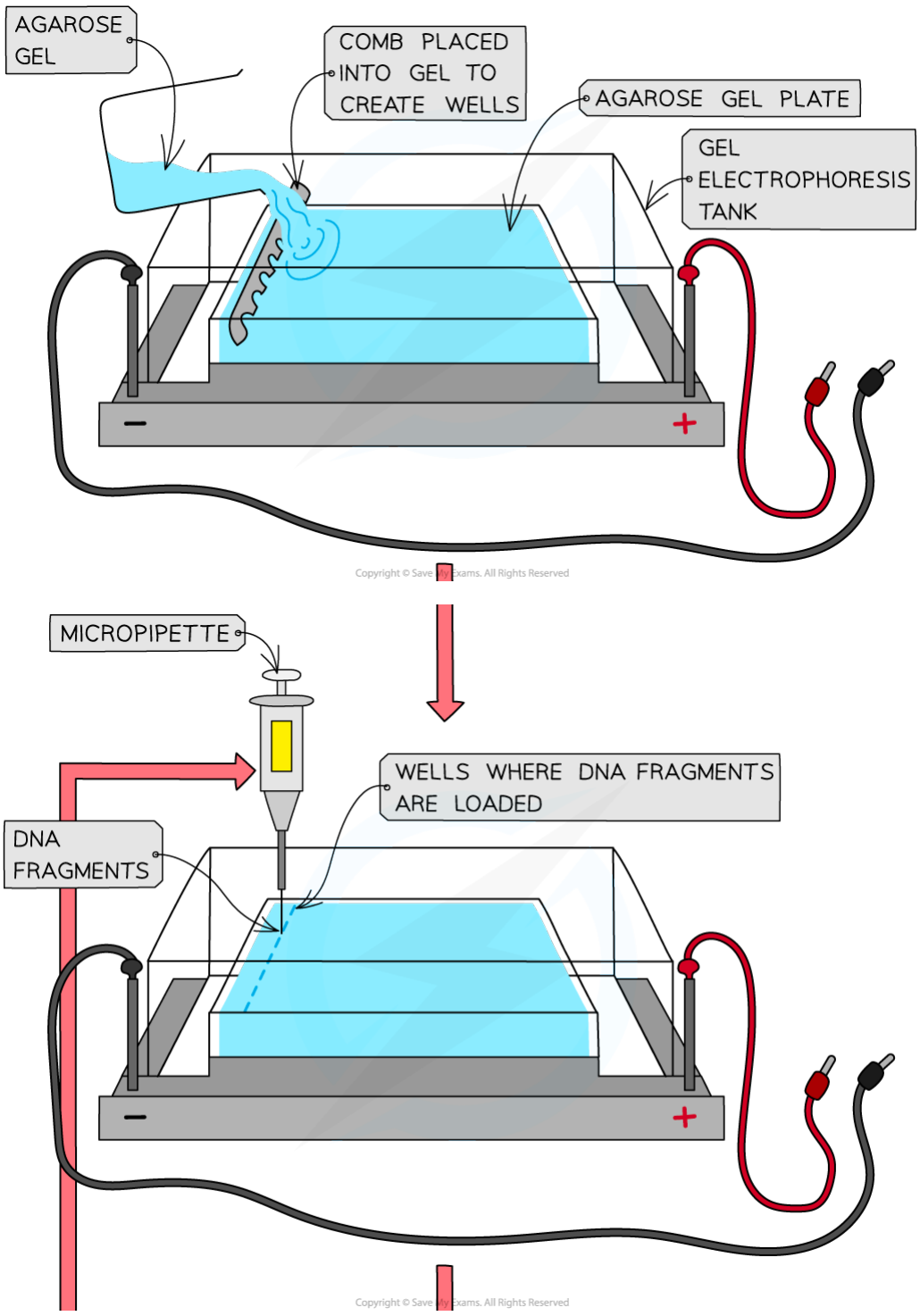
6.3.7 DNA PROFILING

Principles of DNA Profiling

- **DNA profiling** (genetic fingerprinting) enables scientists to **identify suspects for a crime** and **identify corpses** because every person (apart from identical twins) has repeating short non-coding regions of DNA (20 to 50 bases) that are unique to them, they are called **variable number tandem repeats (VNTRs)**
- The **number** of VNTR regions are inherited from your biological parents
    - The more closely related you are to a person the more likely the repeats have similar patterns
- The **length** of the VNTR regions are **unique** to each individual (apart from identical twins)
- When **DNA testing** occurs in forensic medicine and criminal investigations the image of these repeats in the DNA (indicated by a pattern of bars) creates a DNA profile or fingerprint. The profile is analysed to allow conclusions to be made (eg. who the suspects are)
- To **create a DNA profile** from the DNA being tested scientists complete the following in sequence:
    1. **Obtain the DNA**, which can be extracted from the root of a hair, a spot of blood or semen or saliva
    2. Increase the quantity of DNA by using **PCR** to produce **large quantities** of the required fragment of DNA from very small samples (even just one molecule of DNA or RNA).
    3. Use **restriction endonucleases** (different restriction endonucleases cut close to different VNTR sequences) to cut the amplified DNA molecules into fragments
    4. Separate the fragments using **gel electrophoresis**
    5. Add **radioactive or fluorescent probes** that are complementary and therefore bind to specific VNTR regions
    6. X-ray images are produced or UV light is used to produce images of the fluorescent labels glowing.
    7. These images contain **patterns of bars** (the DNA profile) which are then **analysed**

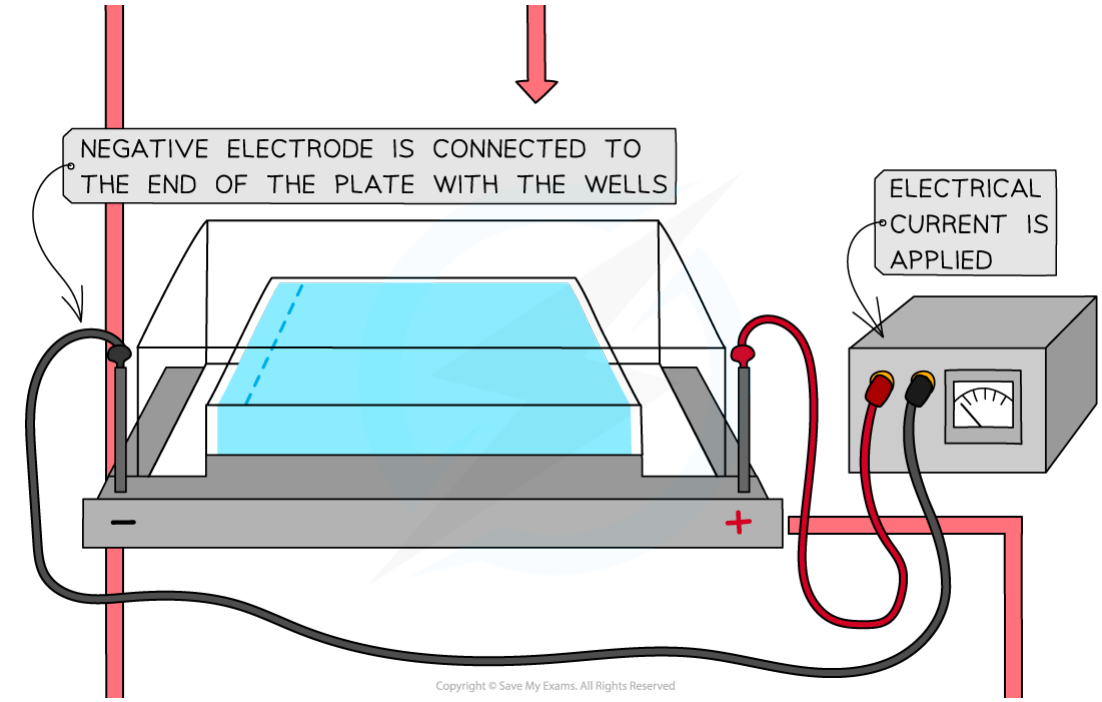
6.3 Manipulating Genomes

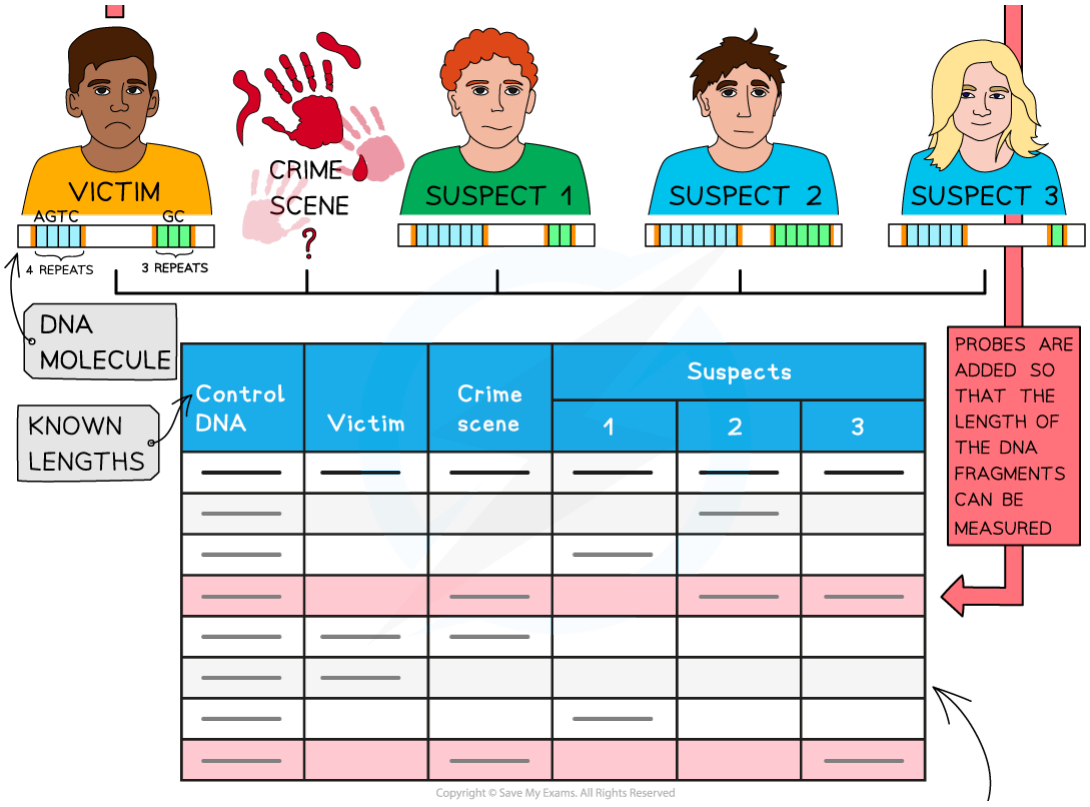
YOUR NOTES



6.3 Manipulating Genomes

YOUR NOTES

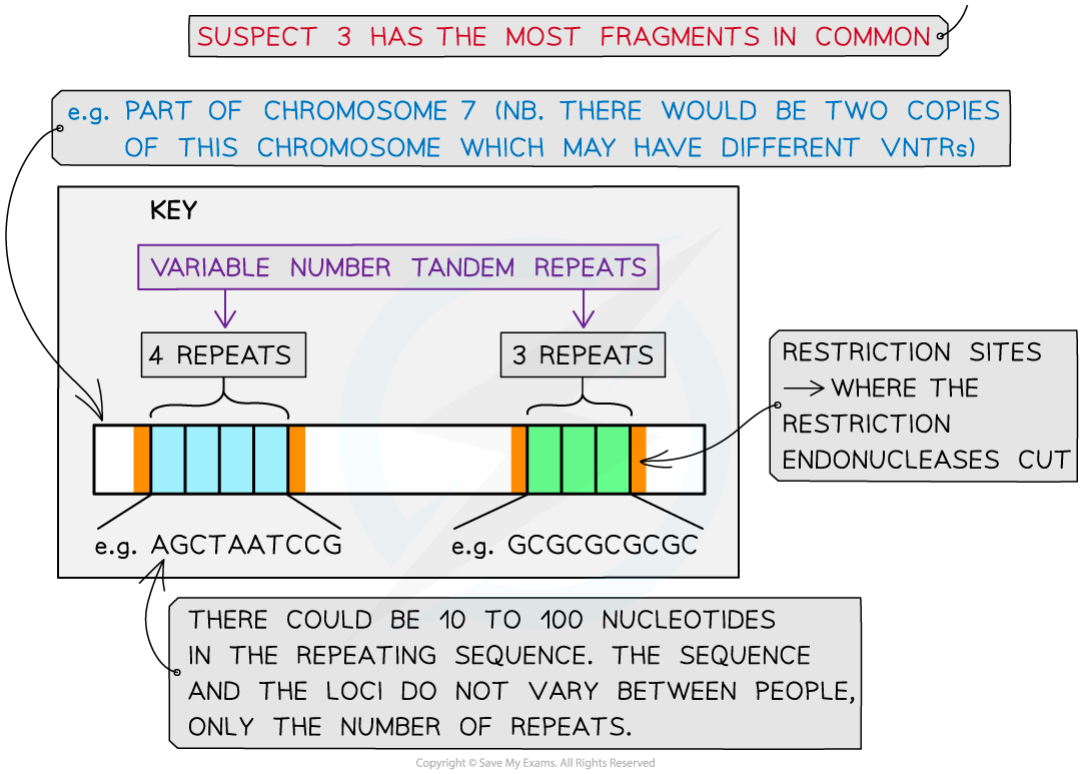




Control DNA	Victim	Crime scene	Suspects		
			1	2	3

6.3 Manipulating Genomes

YOUR NOTES



The separation of DNA fragments using gel electrophoresis. Gel electrophoresis can be used in DNA profiling where scientists separate the VNTRs (as these are unique to every person except identical twins)

Exam Tip

In the exam, you will be expected to interpret the results of gel electrophoresis experiments used to separate DNA fragments. For example, you will be given a few different genetic fingerprints and will have to match the victim to the crime or determine the parents of children. In these questions, you need to look for the most bands in common or a combination of parents' fingerprints that covers all the child's bands.

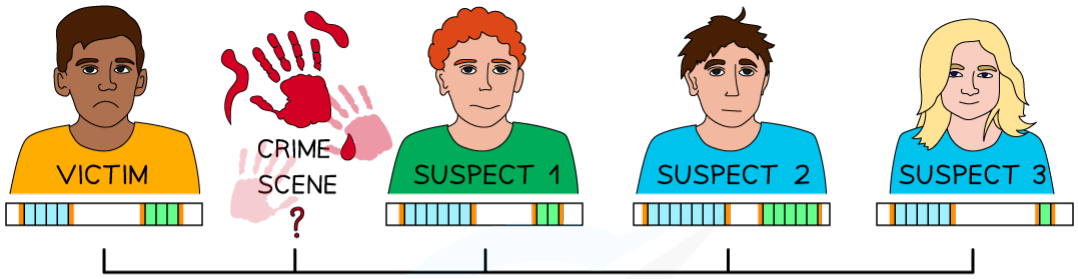
6.3 Manipulating Genomes

YOUR NOTES ⬇

Uses of DNA Profiling

Forensic medicine / criminal investigations

- DNA profiling has been used by forensic scientists to **identify suspects** of crimes
  - Samples of body cells or fluids (eg. blood, saliva, hair, semen) are taken from the crime scene or victims body (eg. rape victims)
  - DNA is removed and profiled
  - The profile is compared to samples from the suspect (or criminal DNA database), victim and people with no connection to the crime (control samples)
  - Care must be taken to avoid contamination of the samples
- DNA profiling can also be used in forensics to **identify bodies** or body parts that are unidentifiable (eg. too badly decomposed or parts remaining after a bomb blast)



KNOWN LENGTHS

Control DNA	Victim	Crime scene	Suspects		
			1	2	3

Using DNA profiling in criminal investigations. Suspect 3 has the most fragments in common with the crime scene DNA so it is likely that they are the culprit.

6.3 Manipulating Genomes

YOUR NOTES ⬇

Other applications

- DNA profiling (along with Next Generation DNA sequencing) can be used to identify individuals that are at **risk of developing particular diseases**, as research shows that certain VNTR sequences are associated with an increased incidence of particular diseases eg. cancers and heart disease
- DNA profiling can be used to **determine familial relationships** for paternity cases (to suggest who the father is) or immigration cases (to determine if the family are related)
- It can also be used in **species conservation** to help scientists with captive breeding programmes to reduce chances of inbreeding

6.3 Manipulating Genomes

YOUR NOTES ⬇

6.3.8 GENETIC ENGINEERING

Principles of Genetic Engineering

- **Genetic engineering** is a term usually used to refer to the manipulation of the DNA sequences of an organism
- The key feature of the genetic code that makes this possible is that it is **universal**, meaning that almost every organism uses the same four nitrogenous bases – A, T, C & G. There are a few exceptions
    - Additionally the **same codons code for the same amino acids in all living things** (meaning that genetic information is transferable between species)
- Thus scientists have been able to artificially change an organism's DNA by combining lengths of nucleotides from **different sources** (typically the nucleotides are from different species)
- The altered DNA, with the introduced nucleotides, is called **recombinant DNA** (rDNA)
- If an organism contains nucleotide sequences from a different species it is called a **transgenic** organism
- Any organism that has introduced genetic material is a **genetically modified organism** (GMO)
- The mechanisms of **transcription** and **translation** are also **universal** which means that the transferred DNA can be translated within cells of the genetically modified organism
- Genetic engineering is being used in the new field of science called **synthetic biology**
    - This is an area of research that studies the design and construction of different biological pathways, organisms and devices, as well as the redesigning of existing natural biological systems

6.3 Manipulating Genomes

YOUR NOTES ⬇

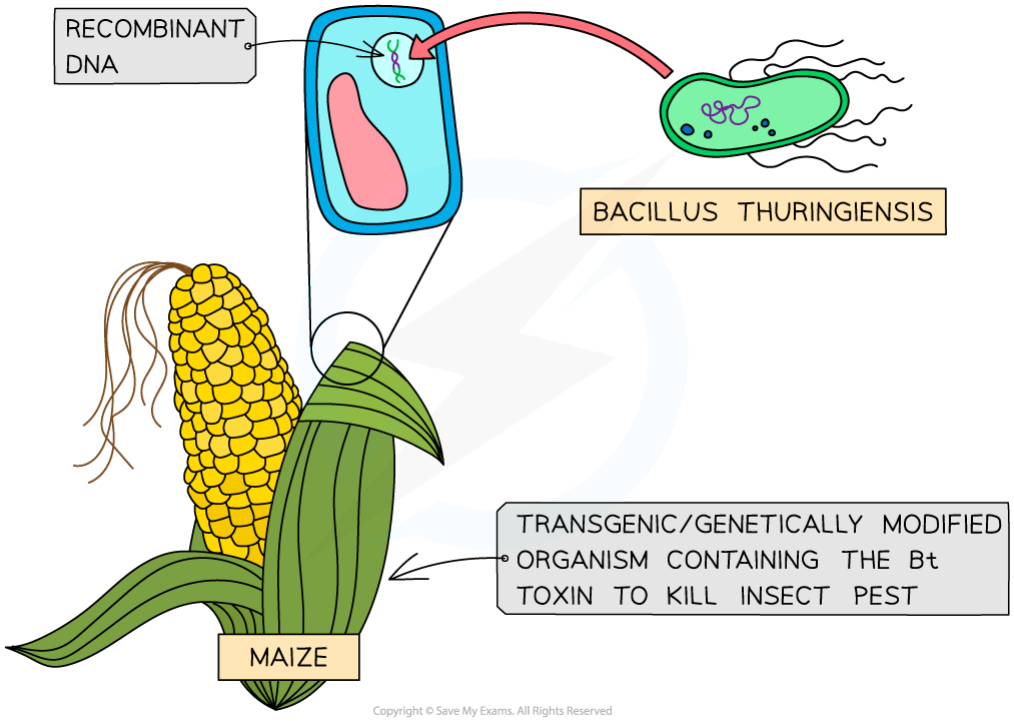


Illustration of a maize plant that has recombinant DNA (DNA from Bacillus thuringiensis)

Recombinant DNA technology

- This form of genetic engineering involves **the transfer of fragments of DNA** from **one organism/species into another** organism/species
- The resulting genetically engineered organism will then contain **recombinant DNA** and will be a genetically modified organism (GMO)

Uses of genetic engineering

- Some of the key uses of genetic engineering include:
  - Genetic modification of **crops** to increase crop yield through resistance to drought, disease, pesticides and herbicides; or to provide increased nutritional value (e.g. golden rice)
  - Genetic modification of **livestock** to give disease and pest resistance and increased productivity
  - Genetic modification of **bacteria** to produce medicines e.g. insulin. Additionally bacterial can be modified to decompose toxic pollutants or carry out large scale chemical production

6.3 Manipulating Genomes

YOUR NOTES ⬇

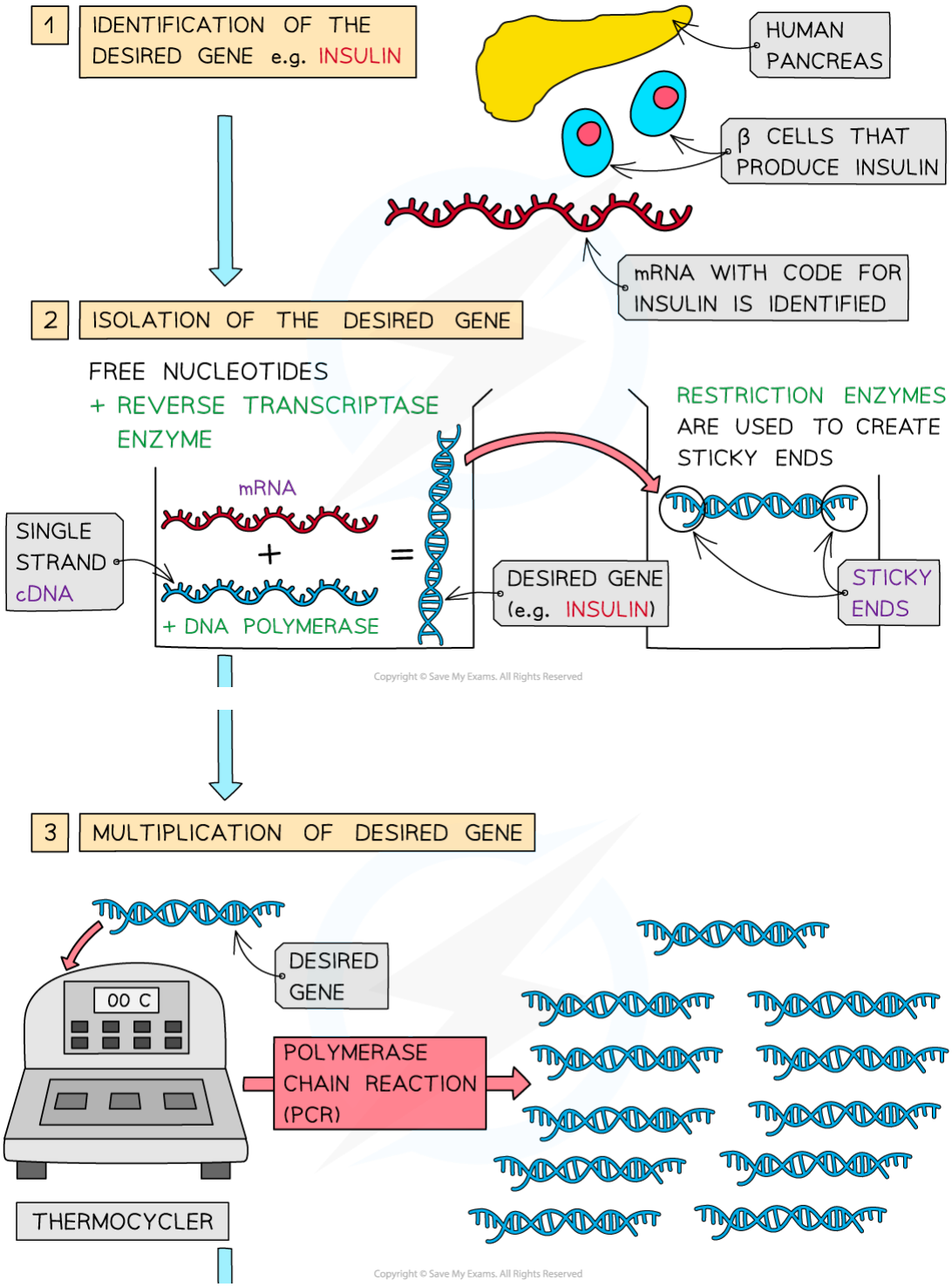
6.3.9 GENETIC ENGINEERING TECHNIQUES

Genetic Engineering Techniques

- In order for an organism to be genetically engineered the following steps must be taken:
  - **Identification** of the **DNA fragment** or **gene**
  - **Isolation** of the desired DNA fragment (either using restriction enzymes, a gene machine or reverse transcriptase)
  - **Multiplication** of the DNA fragment (using polymerase chain reaction – PCR)
  - **Transfer** into the organism using a **vector** (e.g. plasmids, viruses, liposomes)
  - **Identification** of the cells with the new DNA fragment (by using a **marker**), which is then cloned
- Genetic engineers need the following 'tools' to modify an organism:
  - **Enzymes**
    - **Restriction endonucleases** – used to cut genes at specific base sequences (restriction sites). Different restriction enzymes cut at different restriction site
    - **Ligase** – used to join together the cut ends of DNA by forming phosphodiester bonds
    - **Reverse transcriptase** – Used to build double stranded DNA from single stranded RNA
  - **Vectors** – used to deliver DNA fragments into a cell
    - **Plasmids** – transfer DNA into bacteria or yeast
    - **Viruses** – transfer DNA into human cells or bacteria
    - **Liposomes** – fuse with cell membranes to transfer DNA into cells
  - **Markers** – genes that code for identifiable substances that can be tracked
    - **Florescent** markers e.g. green fluorescent protein (GFP) which fluoresces under UV light
    - **Enzyme** markers e.g. β-glucuronidase (GUS) enzyme which transforms colourless or non-fluorescent substrates into products that are coloured or fluorescent
    - **Antibiotic resistance** marker genes – The required gene sequence is inserted into a gene for antibiotic resistance. This inactivates the antibiotic resistance gene and therefore means that successfully transformed bacteria will be wiped out if exposed to the antibiotic. A **replica plating** method is then used to isolate the successfully transformed bacteria

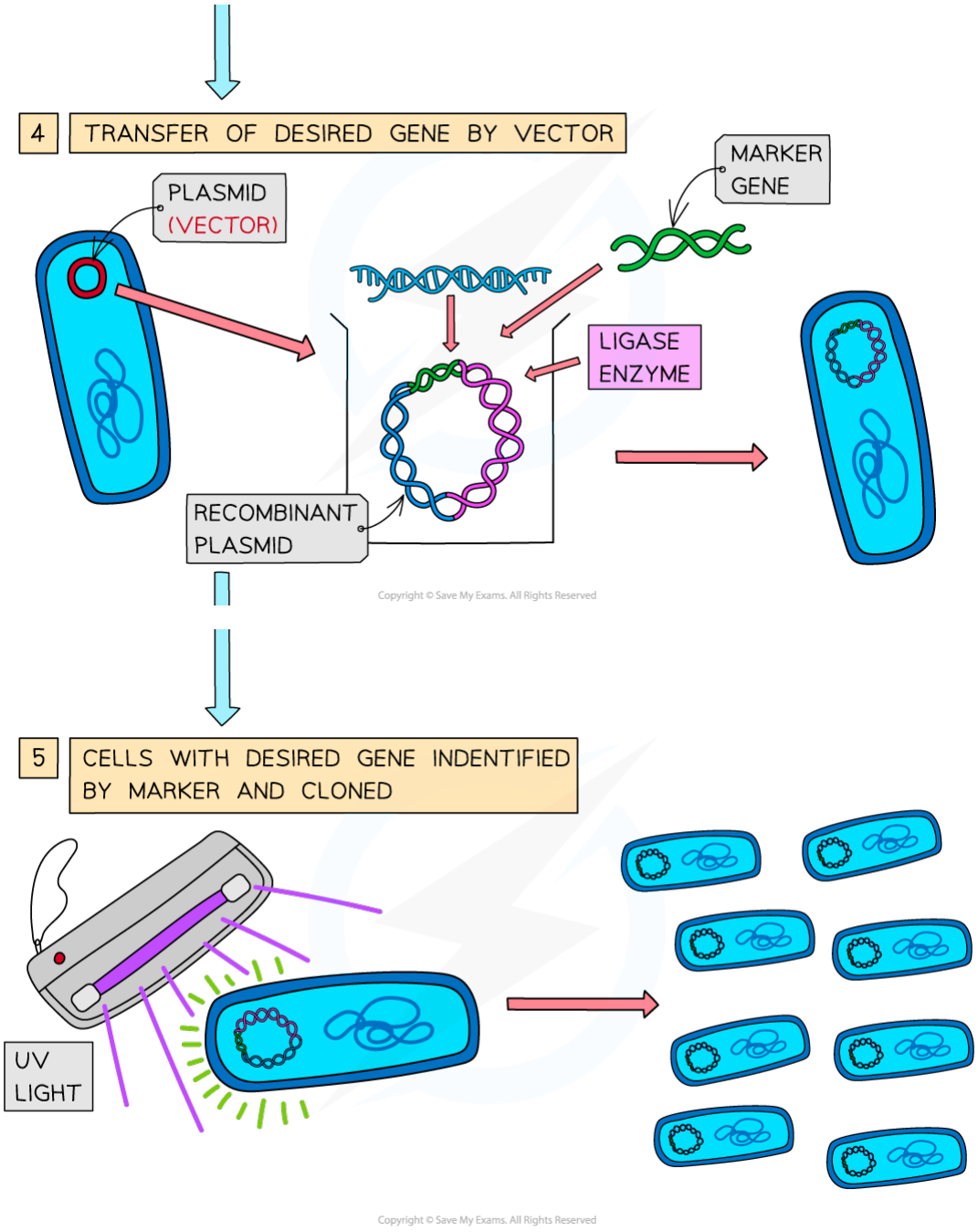
6.3 Manipulating Genomes

YOUR NOTES



6.3 Manipulating Genomes

YOUR NOTES



An overview of the steps taken to genetically engineer an organism (in this case bacteria are being genetically engineered to produce human insulin)

6.3 Manipulating Genomes

YOUR NOTES

Exam Tip

When answering questions about genetic engineering you should remember to include the names of any enzymes (**restriction endonucleases**, **reverse transcriptase**, **ligase**) involved **and** mention that markers (genes which can be identified) and vectors (transfer the desired gene) are also used.

6.3 Manipulating Genomes

YOUR NOTES ⬇

6.3.10 USES OF GENETIC ENGINEERING

Uses of Genetic Engineering

- When a biotechnology company genetically modifies an organism in a specific way they will **patent** the genetic modification
    - A patent gives the owner the legal right to **prevent others from replicating** their invention for a limited time period

GM microorganisms

- Recombinant proteins can be generated using **microorganisms** such as bacteria, yeast, or animal cells **in culture**
    - They are used for research purposes and for treatments (eg. diabetes, cancer, infectious diseases, haemophilia)
- Most recombinant human proteins are **produced using eukaryotic cells** (eg. yeast, or animal cells in culture) rather than using prokaryotic cells, as these cells will carry out the **post-translational modification** (due to the presence of Golgi Apparatus and/or enzymes) that is required to produce a suitable human protein
- The **advantages** of genetic engineering microorganisms to produce recombinant human proteins are:
    - More **cost-effective** to produce large volumes (i.e. there is an unlimited availability)
    - **Simpler** (with regards to using prokaryotic cells)
    - **Faster** to produce many proteins
    - **Reliable** supply available
    - The proteins are engineered to be **identical** to human proteins or have **modifications** that are **beneficial**
    - It can solve the issue for people who have **moral** or **ethical** or **religious** concerns against using cow or pork produced proteins

6.3 Manipulating Genomes

YOUR NOTES ⬇

The production of human insulin

- In 1982, **insulin** was the first recombinant human protein to be approved for use in **diabetes** treatment
- Bacteria plasmids are modified to include the human insulin gene
  - Restriction endonucleases are used to cut open plasmids and DNA ligase is used to splice the plasmid and human DNA together
- These recombinant plasmids are then inserted into *Escherichia coli* by transformation (bath of calcium ions and then heat or electric shock)
- Once the transgenic bacteria are identified (by the markers), they are isolated, purified and placed into fermenters that provide optimal conditions
- The transgenic bacteria multiply by binary fission, and express the human protein – insulin, which is eventually extracted and purified
- The advantages for scientists to use recombinant insulin are:
  - It is **identical to human insulin** unless modified to have different properties (eg. act faster, which is useful for taking immediately after a meal or to act more slowly)
  - There is a **reliable supply available** to meet demand (no need to depend on the availability of meat stock)
  - **Fewer ethical**, **moral** or **religious** concerns (proteins are not extracted from cows or pigs)
  - **Fewer rejection problems** or **side effects** or **allergic reactions**
  - **Cheaper** to produce in large volumes
  - That it is useful for people who have **animal insulin tolerance**

GM plants and animals

- Although plants and animals have been genetically engineered to produce proteins used in medicine, the main purpose for genetically engineering them is to meet the global demand for food
- The benefits of using genetic engineering rather than the more traditional selective breeding techniques to solve the global demand for food are:
  - Organisms with the **desired characteristics** are **produced more quickly**
  - **All organisms** will contain the desired characteristic (there is no chance that recessive allele may arise in the population)
  - The desired characteristic may **come from a different species/kingdom**

6.3 Manipulating Genomes

YOUR NOTES ⬇

GM crops

- Crop plants have been genetically modified to be:
  - **Resistant to herbicides** – increases productivity / yield
  - **Resistant to pests** – increases productivity / yield
  - **Enriched in vitamins** – increases the nutritional value
- Scientists have genetically modified crops such as maize (to be resistant to insect attacks) and rice (to produce β-carotene to provide vitamin A)
- GM crops could reduce the impact of farming on the environment due to there being less need to spray pesticides (eg. less beneficial insects being harmed)

Insect resistance in soya

- Soya bean plants are susceptible to a number of insect pests that cause billions of dollars of damage every year
  - E.g. the fall army worm and soybean podworm
- In response to these large losses in revenue, a biotechnology company has genetically modified the already herbicide-resistant variety of soybean (Roundup Ready™, RR1) by inserting a gene for the **Bt toxin**
  - This gene is taken from the bacterium *Bacillus thuringiensis*
- Soya plants modified with the Bt toxin gene produce their own insecticide
  - When an insect ingests parts of the soya plant, the alkaline conditions in their guts activate the toxin (the toxin is harmless to vertebrates as their stomach is highly acidic), killing the insect
- After 11 years of testing and development, the new variety of soybean (INTACTA RR2 PRO™) was introduced in Brazil in 2013
- Note that insect populations have developed resistance to the genes for Bt toxin, reducing its effectiveness as a means of protecting crops

GM livestock

- Some farmed animals have been genetically modified to grow faster
  - It is rarer for animals to be modified for food production due to **ethical concerns** associated with this practice
- Scientists have also genetically modified livestock to produce pharmaceutical drugs in a process known as **pharming**
  - These "**biopharm"** sheep and goats have been genetically modified to produce a number of useful human proteins in their **milk**
    - E.g. the human blood protein known as AAT in sheep milk
    - E.g. the human protein antithrombin (stops blood clotting) in goat milk

6.3 Manipulating Genomes

YOUR NOTES ⬇

GM salmon

- In 2015 AquaAdventure Salmon was approved by the US Food and Drug Authority (FDA) for human consumption
- This salmon has been genetically modified (GM) to **grow more rapidly** than non-GM salmon as a result of growth hormone being produced in the salmon throughout the year, instead of just in spring and summer. The producer, therefore, has a product to sell in half the time, which increases their yield
- Scientists combined a **growth hormone** gene from a chinook salmon with the promoter gene from an ocean pout, a cold-water fish. The ocean pout fish can grow in near-freezing waters, thus the promoter gene ensured the growth hormone was continually being expressed
- To prevent the GM salmon from reproducing in the wild, all the salmon are female and sterile

6.3 Manipulating Genomes

YOUR NOTES ⬇

GM pathogens

- Many **animal and plant pathogens** have been studied using the techniques of genetic engineering
  - Pathogens can be modified to shed light on their metabolism, drug resistance as well as how it causes damage to its host
  - The development of effective vaccines and drugs can also be aided by this research
- Adenoviruses can be genetically altered to act as vectors in **gene therapy**
  - These viruses are ideal vectors as they are not cell-specific or species-specific; they can infect the cells of many mammals
  - Specific genes are removed from the virus so that it can not replicate once inside host cells, creating space for the insertion of other desired genes

Tobacco mosaic virus (TMV)

- The gene that codes for the hormone **TMOF** can be inserted into the cells of crop plants via a genetically modified tobacco mosaic virus
- Modified TMV is sprayed onto the surface of the crops where it can **invade the plant cells**
- The host plant cells **transcribe the gene** to produce the hormone TMOF
  - TMOF inhibits the production of the enzyme trypsin within insect pests
  - It has no negative effects on the host plant
- The leaves of GM crops that have been exposed to the GM virus can be collected after harvest and ground into a powder to create a **repellent spray** against mosquitoes etc.

6.3 Manipulating Genomes

YOUR NOTES ⬇

Arguments Against the Use of GMOs

- Biotech companies **charge farmers more money for GM seeds** vs non-GM seeds to try and make back the money they have invested in their product
  - Seeds can not be kept from GM crops to regrow the crop the following year because GM crops do not "breed true"
  - Buying seeds year upon year can be a major struggle for farmers in developing countries
- Many people object to the use of GMOs in **food production** due to a **lack of long-term research** on the effects on human health
  - It is unknown whether it will cause allergies or be toxic over time (although there has been no evidence to suggest this would occur to date)
- Some state that without appropriate labelling the consumer cannot make an informed decision about the consumption of GM foods and so **choices are being made for them**
- Organic farmers have complained that the pollen from GM crops may **contaminate nearby non-GM crops** that have been certified as organic
- Environmentalists are concerned about the **reduction in biodiversity** for future generations
  - Crops with less genetic diversity are more vulnerable to extinction
  - GM crops may become weeds or **invade the natural habitats** bordering the farmland
- Herbicide-resistance genes could transfer to weed plants resulting in "**superweeds**"
- GM crops that produce toxins may cause harm to non-target species like the Monarch butterflies)
- The antibiotic-resistance genes that are commonly used as **marker genes** in genetic engineering could transfer to pathogenic organisms that would then be untreatable with antibiotics – "**superbug**"
- Tampering with viral genomes could result in a completely **novel animal virus** that can affect humans or cause existing ones to become **more harmful** to the host
  - This is only an issue if the pathogens are able to escape the lab and enter the wild
- Over time **mutations may occur in the inserted genes** that cause them to have unwanted effects in organisms

6.3 Manipulating Genomes

YOUR NOTES ⬇

6.3.11 GENE THERAPY

Gene Therapy

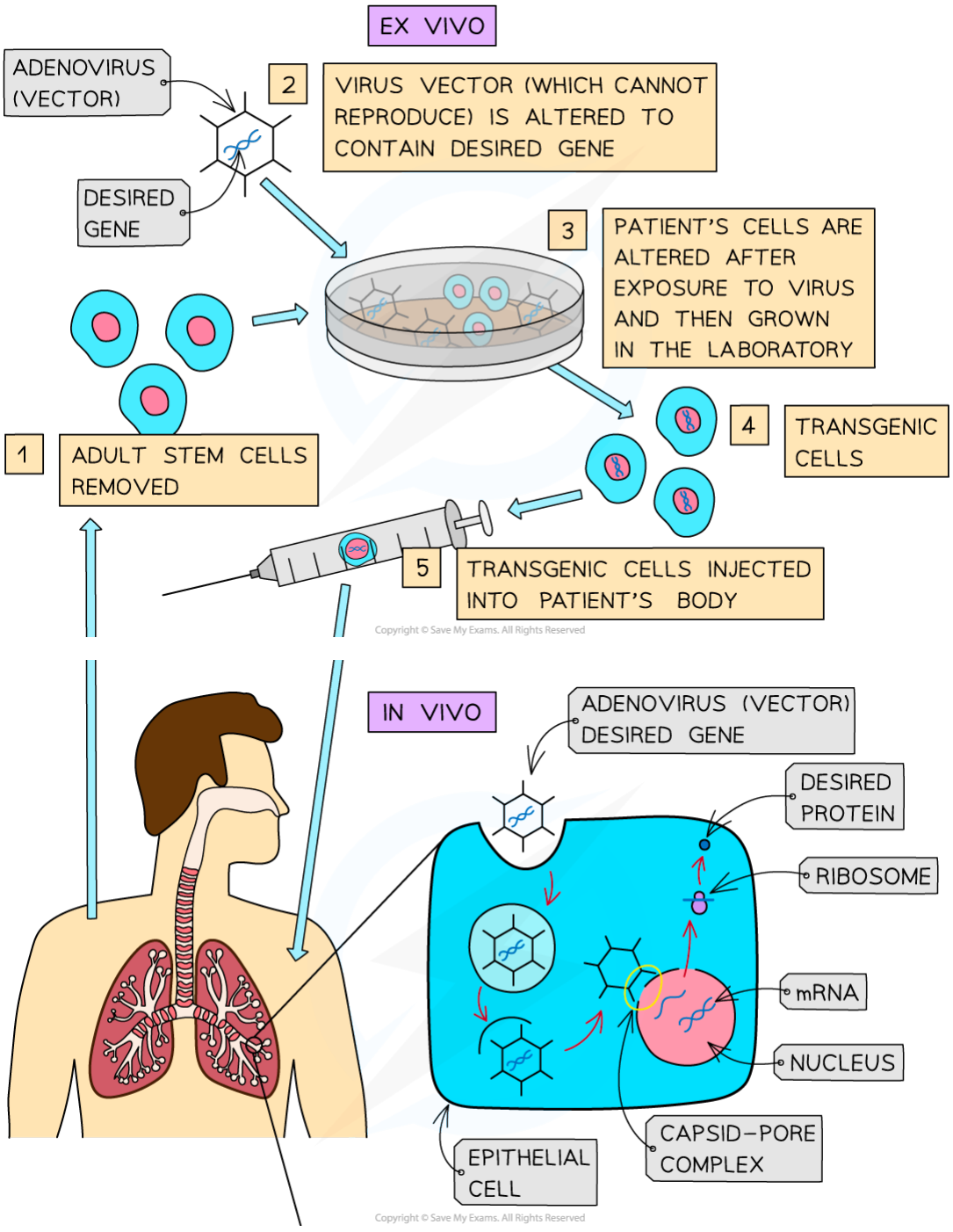
- **Gene therapy** involves using various mechanisms to **alter a person's genetic material** to treat, or cure, diseases
- As scientists gain a better understanding of the human genome, and therefore the location of genes that cause genetic disorders, the possibilities of gene therapy being able to **replace a faulty gene**, **inactivate a faulty gene** or **insert a new gene** are growing
- Experimental techniques are being used to treat and research treatments for genetic diseases such as **severe combined immunodeficiency** (**SCID**), Leber congenital amaurosis – a rare form of blindness, β-thalassaemia and haemophilia B
- Most gene therapies are still in the clinical trial stage because scientists are having difficulty finding delivery systems that can transfer normal alleles into a person's cells and how to ensure the gene is correctly expressed once there
    - **Vectors** are currently used as the delivery system
        - **Viruses** (eg. retroviruses and lentiviruses) are the most commonly used vectors as they have the mechanisms needed to recognise cells and deliver the genetic material into them
        - **Non-viral vectors** are also being researched (eg. **liposomes** and **'naked' DNA**)

Types of gene therapy

- Currently all gene therapies have targeted and been tested on **somatic** (body) Changes in genetic material are **targeted to specific cells** and so will not be inherited by future generations (as somatic gene therapy does not target the gametes)
- Often the effects of changing the somatic cells are **short-lived**
- There are two types of **somatic gene therapy**:
    - **Ex vivo** – the new gene is inserted via a virus vector into the cell **outside the body**. Blood or bone marrow cells are extracted and exposed to the virus which inserts the gene into these cells. These cells are then grown in the laboratory and returned to the person by an injection into a vein
    - **In vivo** – the new gene is inserted via a vector into cells **inside the body**
- There is the potential for new genetic material to be inserted into **germ cells** (cells involved in sexual reproduction eg. gametes or an early embryo)
- However, this is illegal in humans as any changes made to the genetic material of these cells is potentially **permanent** and could therefore be **inherited** by future generations

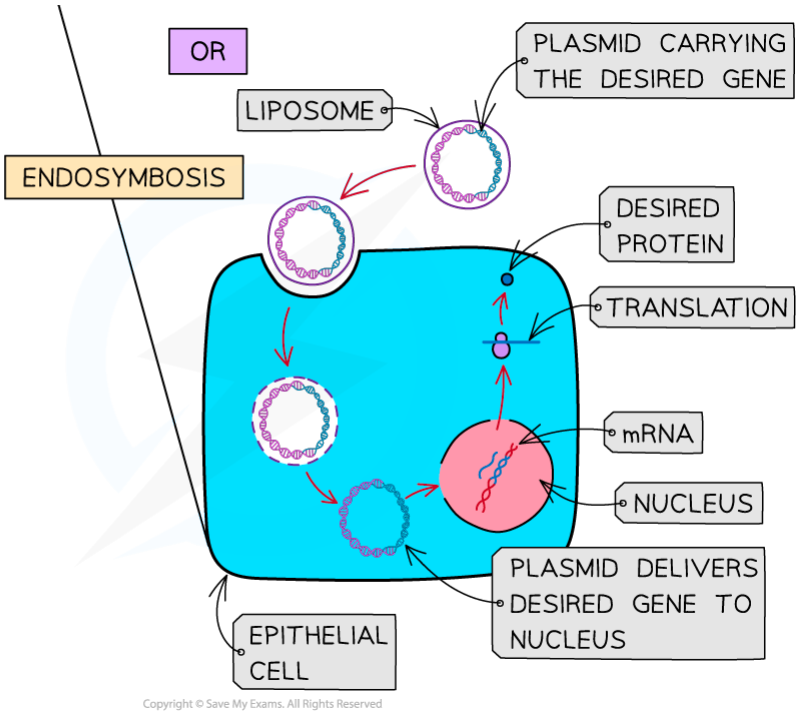
6.3 Manipulating Genomes

YOUR NOTES



6.3 Manipulating Genomes

YOUR NOTES



The two types of somatic gene therapy – ex vivo and in vivo. Two of the methods scientists are using to alter a patient's genetic material to treat or cure diseases.

6.3 Manipulating Genomes

YOUR NOTES ⬇

Severe combined immunodeficiency (SCID)

- **Severe combined immunodeficiency** (SCID) is caused by the body's inability to produce **adenosine deaminase (ADA)**, an enzyme that is key to the **functioning of the immune system**.
- Without this enzyme children can die from common infections and therefore need to be keep isolated often inside plastic 'bubbles'
- To treat SCID scientists have used **ex vivo somatic gene therapy.** During this therapy, a virus transfers a **normal allele for ADA into T-lymphocytes** removed from the patient and the cells are then returned via an injection
- This is **not a permanent cure** as the T-lymphocytes are replaced by the body over time and therefore the patient requires **regular transfusions** every three to five months to keep their immune systems functioning
    - Originally **retroviruses** were used as the vectors, however these viruses insert their genes randomly into a host's genome which means they could insert the gene into another gene or into a regulatory sequence of a gene (which could result in cancer)
    - Initial treatments did cause cases of **leukaemia** in children, so researchers switched to using lentiviruses or adeno-associated viruses as vectors.
    - **Lentiviruses** also randomly insert their genes into the host genome however they can be modified to not replicate, whereas **adeno-associated viruses** do not insert their genes into the host genome and therefore the genes are not passed onto the daughter cells when a cell divides.
        - This is an issue with short-lived cells like lymphocytes but has not been a problem when used with longer living cells such as liver cells

Inherited eye diseases

- An example of a group of inherited eye diseases that causes blindness due to damage to the light receptors in the retina are **Leber Congenital Amaurosis**. It begins to affect children from birth and by their 20s or 30s the person is totally blind.
    - There is no cure for these diseases
- Using **in vivo somatic gene therapy,** doctors injected into the retina **adeno-associated viruses** that contained the normal alleles of one of the genes that caused damage to the photoreceptors (there are at least 18 known mutated genes causing this group of diseases).
    - All patients that have had the injections have shown improvement in their eyesight

Exam Tip

Remember that gene therapy in somatic cells is not permanent, whereas in germ cells it is.