FACTFILE:
GCE BIOLOGY

GENE TECHNOLOGY

Rewarding Learning

Gene Technology

Learning outcomes

Students should be able to demonstrate knowledge and understanding of using DNA probes to locate a specific section of DNA:

- that a DNA probe is a short length of DNA with a known base sequence;
- that probes can be fluorescent and therefore useful as a marker; and
- that the probe will base pair with any complementary nucleic acid strands.

Students should be able to demonstrate knowledge and understanding that differences in nucleotide sequences can be identified by:

- assessing differences in nucleotide sequences as a measure of genetic variation; and
- different genetic markers, for example microsatellite repeat sequences (MRSs) and single nucleotide polymorphisms (SNPs).

Students should be able to demonstrate knowledge and understanding of microarray technology:

- it enables rapid gene expression profiling or gene sequence variation of thousands of genes in an individual;
- it involves DNA bound (hybridised) to a microarray chip;
- binding of the sample DNA to a probe on the chip can produce a signal that can be fluorescent or chemiluminescent; and
- digitally analysing the chip to calculate the strength of the signal produced at each spot with the strength of the signal representing the level of gene expression in the individual.

Students should be able to demonstrate knowledge and understanding of genetic fingerprinting and have an appreciation of its potential uses to include:

- using restriction endonucleases to cut chromosomal DNA and the subsequent separation of fragments according to size (using gel electrophoresis) to produce unique profiles; and
- using DNA probes to locate specific DNA fragments.

Students should be able to demonstrate knowledge and understanding of the stages involved in gene transfer:

- methods used to obtain donor DNA, for example using restriction endonucleases and reverse transcriptase;
- using DNA probes to locate DNA fragment within the desired gene;
- incorporating the donor genes into a 'vector', for example bacteriophages and bacterial plasmids;
- transforming recipient cells, for example *Escherichia coli, Saccharomyces cerevisiae*;
- using marker genes to identify transformed cells, for example antibiotic resistance and fluorescent marker genes; and
- producing large numbers of transformed cells.

Students should be able to demonstrate knowledge and understanding of the role of genetically modified (GM) microorganisms:
- the wide range of substances that GM bacteria produce; and
- using GM viruses to treat disease, for example killing human cancer cells and treating bacterial infections.

Students should be able to demonstrate knowledge and understanding of the role of transgenic organisms, such as improving desirable traits, by inserting genes into:
- animals to:
  - encourage faster growth rate and better food quality traits;
  - produce substances of medical and pharmaceutical value; and
  - use as models in human disease research; and
- plants to:
  - produce higher crop yields, increased variety or better food quality traits;
  - produce pest and disease resistant crops; and
  - cultivate GM crops that grow in unfavourable environments.

Students should be able to demonstrate knowledge and understanding of gene therapy:
- that absent or faulty genes may cause genetic diseases;
- that adding a functional copy of a gene into a cell to restore metabolism and eliminate disease;
- the advantages and problems of somatic-cell gene therapy;
- the types of vector that can be used in the gene transfer;
- the ethical and technical issues surrounding germ-line gene therapy; and
- understanding that gene therapy:
  - has successfully treated some diseases, for example haemophilia B; and
  - requires repeated rounds of treatment to manage symptoms of other diseases, for example cystic fibrosis.

Students should be able to demonstrate knowledge and understanding of the process of gene sequencing:
- genome of an organism as the complete DNA sequence (on one set of chromosones in diploid, eukaryotic organisms);
- genome sequencing as determining the order of nucleotides and so the genetic code;
- that the genomes of many organisms have been sequenced, including some that are extinct, for example Neanderthals;
- that knowledge of the genetic code allows primary protein structure to be determined (and that molecular modelling software can predict secondary, tertiary and quaternary structure);
- that the Human Genome Project organised to map and sequence 3 billion nucleotides in the human genome and to identify all the genes present (approximately 21000); and
- that sequencing has become more rapid and less expensive.

Students should be able to demonstrate knowledge and understanding that the inactivation or replacement of genes helps to understand gene and organism function:
- genes may be made inoperative by being disabled, removed or replaced with defective alleles (details of techniques not required);
- this technique provides clues to the role genes play in a cell or organism;
- the mouse (knockout and knockin mouse) as a model organism for the study of genes; and
- inactivation or replacement can be used to study the development of genetic disorders and drug therapies.

Students should be able to demonstrate knowledge and understanding of the term pharmacogenetics:
- as the effect that the genotype has on an individual's drug response;
- understanding that personalised medicine deals with genetic differences specific to an individual patient, where certain patients can respond differently to treatment with certain drugs, for example codeine;
- understanding that using microarrays can identify genetic differences between individuals that result in different responses to drugs;
- that personalised medicine explores how treatment can be personalised to improve clinical outcome; and
- that the development of 'designer drugs' can be matched to an individual's genetic profile.

Students should be able to demonstrate knowledge and understanding of the social, legal, ecological and ethical issues of the benefits and risks of gene technology:
- the social and ethical implications of genetic modification and genome sequencing;
- the potential risks associated with genetically modifying organisms; and
- safety precautions currently used, for example using bacterial strains ill-adapted to the human physiology, 'suicide genes' and containment mechanisms.

Microarray technology

Consider the following problems:

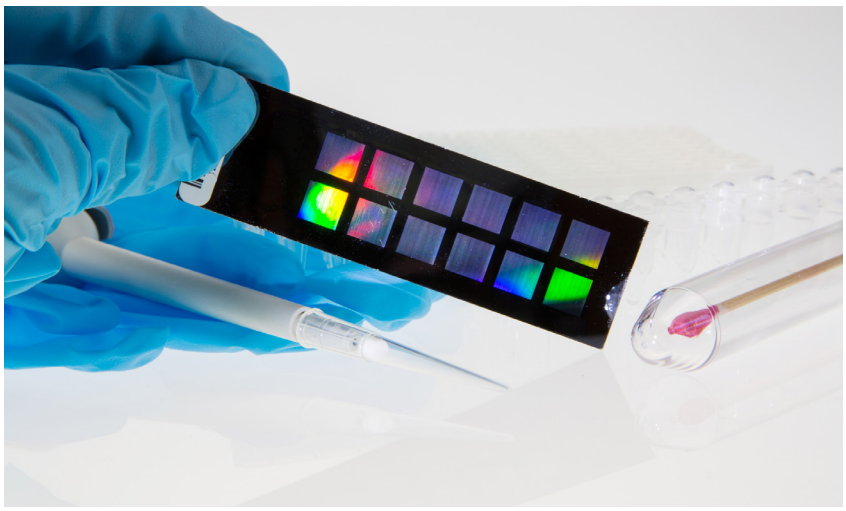
- In every population, variation exists in genetic base sequences (for example SNPs). If we know the possible base sequences for a region of a gene, how can we find out which base sequence an individual has, for example in a gene which affects how well they will respond to a certain drug?

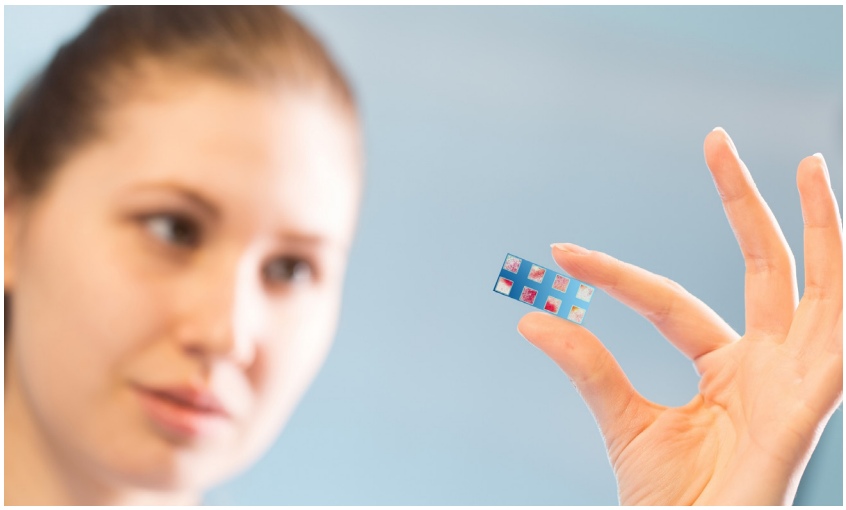
- Genes can be switched on (expressed) or switched off in cells and sometimes changes in which genes are switched on in a tissue can result in (or from) disease. For example how can we compare the genes being expressed in normal tissue with those expressed in cancerous tissue?

While these problems may seem quite different at first glance, variations on the technique of microarray technology can be used to solve each one.

A 'microarray' (sometimes called a DNA chip, or GeneChip®) can be thought of as a miniature spotting tile with thousands of 'spots' on it. It looks like a glass slide (some may be made of silicon), but it has thousands of DNA probes attached to it. One 'spot' contains many copies of a DNA probe, with a unique base sequence. Each spot contains copies of a different probe. The DNA probes are added to the chip robotically and information about each one and its location on the chip is held in a computer file.

Examples of microarrays are shown in the photographs below.





In order to solve the first problem outlined above (gene sequence variation), an individual's DNA is amplified by PCR, digested using restriction enzymes, made single-stranded and a label is added. The label may be fluorescent or chemiluminescent (where light is emitted as a result of a chemical reaction.)

[NB. These steps may occur in a different order, depending on the nature of the investigation.]

Once the labelled sample DNA has been prepared, it can be added to the microarray. The single-stranded DNA will only bind on the spots where complementary probes are located and because of the label, the chip can then be digitally analysed to indicate where binding has occurred. This information, along with the data on which probe is found where on the chip, can be used to determine an individual's base sequence for the gene of interest. If the gene of interest is one which is known to affect the way the individual responds to a certain drug, then we now know their genotype with regards to this gene and medication can be prescribed accordingly.

The following resources may be useful to teachers and/or students:

Animation of gene expression profiling

National Human Genome Research Institute factsheet on microarray technology

Excellent interactive virtual lab activity, where students complete an investigation of gene expression profiling in normal and cancer cells.

PDF produced by the Southwest Centre for Microsystems Education, with extensive but clear detail on microarrays.

Video produced by the Southwest Centre for Microsystems Education, with extensive but clear detail on microarrays.

Since thousands of probes can be used on the microarray, thousands of genes can be investigated at the same time using this technology. The technique can be used to investigate genotypes, including drug responses and disease alleles (such as BRCA1 mutations) and it also has uses in forensic analysis. In some cases, the use of microarrays has been superseded by sequencing, since the latter has become faster and cheaper. However, microarrays are currently less expensive than sequencing, so for large scale studies they are still the preferred option.

In order to solve the second problem (gene expression profiling), only small variations need to be made to the technique described above. When genes are expressed in a cell, transcription produces messenger RNA complementary to the gene. In order to analyse which genes are being expressed, all the mRNA is extracted from each tissue of interest (for example normal and cancerous tissue) and then reverse transcriptase is used to make a cDNA copy of each mRNA. These cDNAs are then labelled as before and can be added to the microarray and digitally analysed. In this way, researchers can see which genes are 'up-regulated' and which are 'down-regulated' in cancer cells, compared with normal cells.

Genetic Fingerprinting

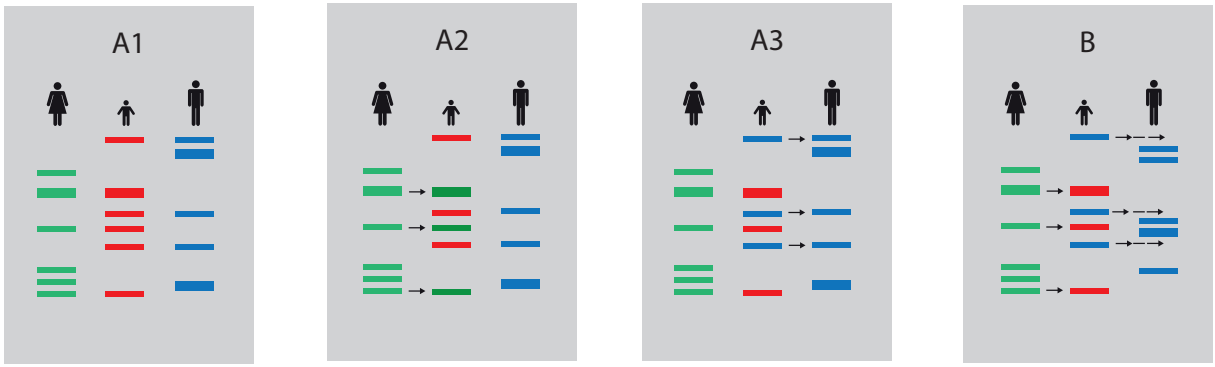
Genetic fingerprinting is a well-established technique for analysing and comparing the DNA of individuals. One method of producing a genetic fingerprint (profile) involves the following steps:

1. Sample DNA (which may have been amplified using PCR) is cut into smaller fragments using restriction enzymes.
2. The fragments are added to an agarose gel and separated by size, using gel electrophoresis.
3. The DNA in the gel is made single-stranded and copied onto a nitrocellulose sheet.
4. Labelled DNA probes are added to the plate and bind via base pair complementarity.
5. A detection system is used to visualise where binding of probes has occurred, using the label.

Currently, rather than investigating the whole genome, some of the most variable sections (containing MRS markers) are amplified by PCR and then analysed using the steps above.

Regardless of the technique however, the principle behind interpretation of the profile produced is the same: the probability of two people (other than identical twins) having the same profile when 10 or more markers are used is virtually nil and the sequences investigated are inherited in Mendelian fashion. This allows the profiles produced to be useful in crime scene analysis and paternity disputes and other situations where relatedness is to be proved or disproved, such as immigration disputes. It is also being used in conservation to establish species membership and assess genetic variation in populations. Commercially, organisations like The Kennel Club offer DNA profiling for dogs for identification and parentage queries.

The image below shows examples of genetic fingerprints which prove paternity (A1-A3) and disprove paternity (B).



Further information can be found from the following resources:

A video explaining the process and history of genetic fingerprinting

An interactive activity on genetic fingerprints

Factfile on how genetic fingerprinting can be used to investigate the ivory trade

Gene sequencing

An organism's genome can be defined as all of the genetic material (DNA) which the organism has in one set of chromosomes. Techniques have been used over the last twenty years to determine the order of the bases A, T, G and C in the genomes of many different species, a process known as genome sequencing. This information has been and continues to be very useful in numerous areas, including determining causes of disease and tailoring treatment, investigating evolutionary relationships, genetic testing and forensics. It is thought that DNA sequencing will eventually overtake microarray technology as the most useful genetic diagnostic tool (although associated costs mean that microarrays are currently the tool of choice for most laboratories). Knowledge of the sequence of bases in a particular gene enables the primary structure of the protein encoded by that gene to be worked out. Molecular modelling software can then be used to predict the ways the polypeptide chain might fold into secondary, tertiary and quaternary structures. (However, even computers struggle to deal with the many variables involved in this folding and this continues to be one of the big challenges for molecular biology.) See this article on The Science Behind Foldit for more information.

The first genomes to be fully sequenced were of viruses in the 1970s and 1980s. By 2000, milestones such as the first bacterial (*Haemophilus influenzae*), fungal (*Saccharomyses cerevisiae*), plant (*Arabidopsis thaliana*) and animal (*Caenorhabditis elegans*) sequenced genomes had been reached. As more and more species genomes were sequenced, so the body of knowledge about the genetic code grew. In addition, many previously held theories about evolutionary relationships and taxonomy have been revised in light of genome sequencing information.

DNA is a persistent molecule and scientists have been able to successfully extract it even from organisms which have been dead for many years. In 2010, the first draft of a genome sequence for Neanderthals (*Homo neanderthalensis*) was produced using DNA from fossilised remains of individuals who lived in Europe 40 000 years ago. The species is thought to have become extinct around 30 000 years ago and shares a common ancestor with *Homo sapiens*. The genome sequencing project has been able to contribute to knowledge of human evolution and migration patterns.

Publication of genomes for other extinct organisms can be exaggerated in the media. For example in 2014, the media reported that the woolly mammoth's genome had been sequenced. However in this case, the mammoth genome was fragmented and incomplete, meaning reports of potential 'de-extinction' were hypothetical at best.

The Human Genome Project

In 1990, an ambitious international project was undertaken to sequence the entire human genome (about 3 billion base pairs). Rather than one individual having their unique genome sequenced, a pool of several people contributed their genomes and an 'average' genome was produced. The project was completed in 2003.

Some findings on the human genome

- The HGP determined the number of protein-coding genes in the human genome to be around 21 000. The exons of these genes make up just 1-2% of the total genome.

- In addition, there are around 6000 RNA genes: genes which are transcribed into RNA which is not subsequently translated. Examples include transfer RNA and ribosomal RNA, but also many others for which functions are still not known. Some have been implicated in certain types of cancer and in Alzheimer's Disease.

- The remainder, often erroneously referred to as 'junk' DNA, was found to have a range of characteristics; for example large proportions make up gene introns, structural DNA (heterochromatin) and sections called 'transposable elements' which can move around in the genome.

The completion of the HGP was referred to at the time as 'the end of the beginning' and this has been borne out by further examination of the genome. The complexity of the human genome, containing overlapping genes, genes within genes and pseudogenes, leaves many aspects of our genetic code still to be unravelled. However by 2013, ten years after the completion of the project, significant advances had been made in determining the molecular basis of many diseases and tailoring treatments, as detailed in the table showing Quantitative Advances Since the Human Genome Project (HGP)

[NB. While the Human Genome Project was declared complete in 2003, in fact some small sequences within the genome remain unsequenced. These include some of the heterochromatic DNA (which is known to contain many repetitive sequences and few genes) but also some euchromatic DNA. These euchromatic gaps could not be sequenced via the techniques used for the rest of the genome (which involved splicing DNA fragments into bacteria before sequencing; these sequences are known to be 'toxic' to bacteria and were automatically deleted.) In 2014 a new technique was reported to have successfully sequenced many of these gaps.]

Faster, cheaper sequencing

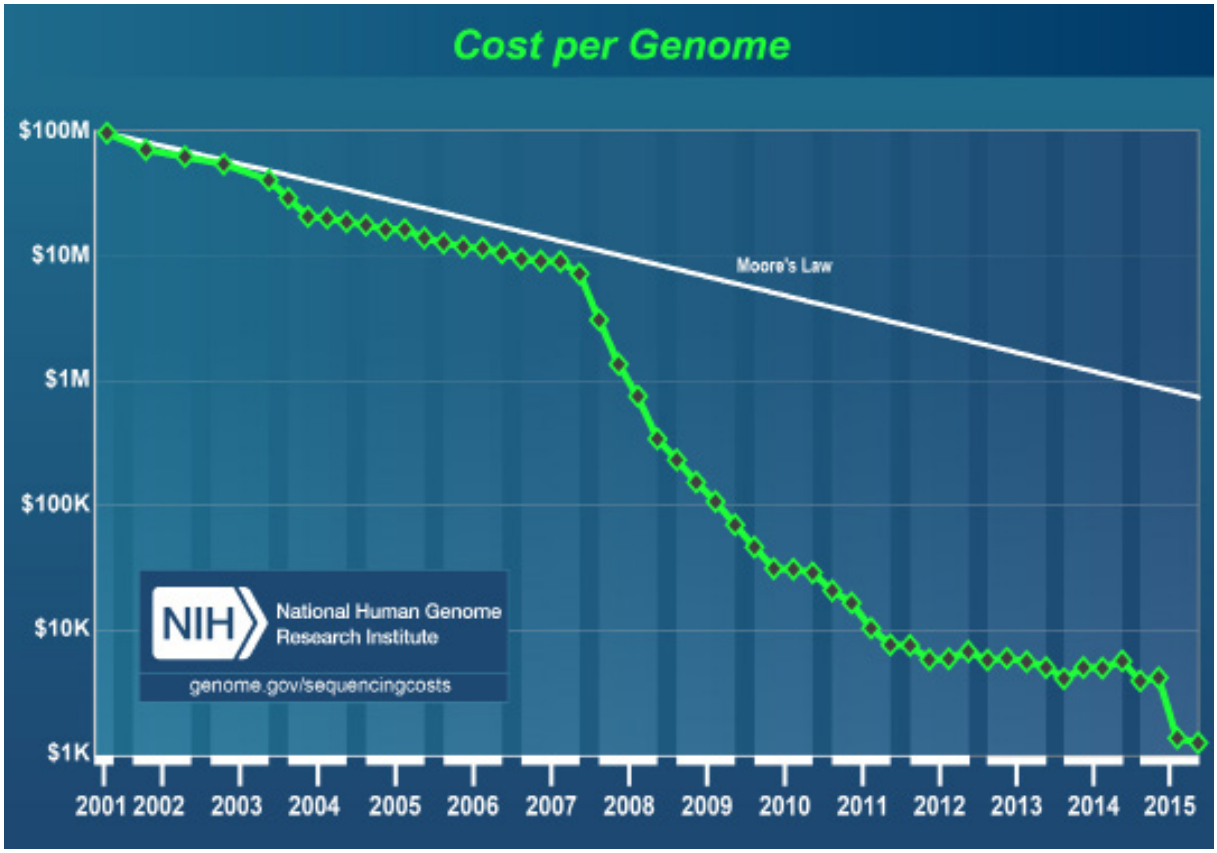
Huge advances in sequencing technology since the advent of the HGP have significantly reduced the costs involved and the time required to sequence a genome. This has been particularly notable since 2008 when so-called 'next-generation sequencing' began to be used. The table below summarises the consequences of these advances. (Costs are estimated in US$, since most of the laboratories carrying out this work are in the USA. $1000 is approximately £650.)

* The '$1000 genome' has been a stated aim of genome research for several years and in January 2014 a company called Illumina announced that it had developed the technology to make it a reality. Its machines are able to sequence 5 genomes in a day, but in practice only a few laboratories around the world can take advantage of the economies of scale required to realise the $1000 per genome cost (since ten machines are required, at a cost of $10 million.)

	HGP begins (1990)	HGP completed (2003)	2015
Time required to generate a human genome	6-8 years	3-4 months	<1 day*
Cost of generating a human genome	$1 billion	$10-50 million	$1000*

Source: www.genome.gov

The graph below indicates the rate at which the cost of genome sequencing has fallen. Note the logarithmic scale on the y-axis. (Moore's Law is a long-standing trend, whereby computing power doubles every two years, contributing to falling costs for associated processes. If a technology 'keeps up' with Moore's Law, it is thought to be performing very well.)



Source: www.genome.gov

Further information on genome sequencing can be found at the following websites:

The website of the National Human Genome Research Institute (USA)
www.genome.gov

Produced by the Wellcome Genome Campus, Cambridge, UK. Section on 'Methods and Technology' may be of particular use
www.yourgenome.org

Social & ethical implications of gene sequencing

The advent of direct-to-consumer gene sequencing products (such as the 23andme kit, currently available from at least one high street retailer in the UK) has important social implications. Such kits enable consumers to obtain a report on various aspects of their genome, including whether they carry genes which may increase their risk of developing diseases later in life and how they metabolise various common medications. There are concerns about the release of this information to individuals, with little or no guidance on the nature of risk for example, or on the alleles which an individual may test positive for. Furthermore, there are concerns over who else might request access to the information; for example doctors, family, insurance providers and potential employers.

In the USA trials on genome sequencing of newborns are underway, and the current UK health minister has expressed an interest in doing the same. If this becomes a reality, who would read and interpret the genetic results? Would parents have an obligation to tell their children the results? What if parents preferred not to know which diseases their children may have an increased susceptibility to? Who owns the genome data, once it is sequenced? These issues are ones which todays A-level students may well encounter in the not-too-distant future.

Further information:

An article on ethical issues surrounding genome sequencing (2012)

Time magazine articles, mostly on parental perspectives of sequencing their children's genomes (2012)

Comment article on prenatal genome sequencing (2014)

Pharmacogenetics

Pharmacogenetics is the tailoring of drug treatments to individuals, based on their genotype (also referred to as pharmacogenomics). Both microarray technology and gene sequencing have made major contributions to advances in this field, by providing information on an individual's genetic makeup.

Due to genetic variation, people can respond differently to certain drugs. This can mean a drug that is very useful in treating certain conditions in some people can be ineffective or even dangerous in others. Alternatively, it can mean that the dose required to successfully treat a condition can vary with genotype.

Amongst the human population, there is variation in the response of individuals to the painkiller codeine; for example individuals can be classified as 'poor', 'intermediate', 'extensive' or 'ultra-rapid' metabolisers of codeine, depending on which alleles they possess for codeine metabolism. The different alleles result in different levels of activity of an enzyme involved in the conversion of codeine to morphine. High levels of morphine can be very dangerous.

The effective dosage for each classification group is different; if an 'ultra-rapid' metaboliser was given the normal dosage, the toxic effects could prove life-threatening. In contrast the same dosage may be completely ineffective in a 'poor' metaboliser.

In some ethnic groups, such as Africans and African-Americans, there is an increased incidence of the 'ultra-rapid' metaboliser genotype. An individual's genotype with regard to the alleles involved can be determined via microarray, so that a safe and effective dose can be prescribed. (NB. The enzymes involved, called cytochrome P450 enzymes, influence the metabolism of many other drugs besides codeine.)

Another branch of pharmacogenetics involves cancer genotypes. Knowledge of the specific mutations which have occurred in the cancerous cells and which genes are being up- or down-regulated can be invaluable in determining the most appropriate chemotherapy treatment. Again, microarray technology and gene sequencing can be used to investigate this.

'Designer Drugs'

Traditionally, approaches to drug development involved investigating a range of chemicals to see which were useful in treating a disease in large numbers of patients. However, with increased knowledge of the genetic variation underlying diseases and drug responses, researchers can now explore treatments which may only work in a subgroup of patients, but which may be highly effective in that subgroup. Herceptin, a cancer drug, is an example of a treatment which is effective only in a genetic subgroup of cancer patients.

The development of the 'designer drug' approach will make adverse drug reactions and side effects less common and should ensure that patients are prescribed the right treatment at the right dosage first time round. Many therapeutic drugs now carry pharmacogenetic information on their labelling, specifying which genetic subgroup they are best suited to.

Further information:

Dr. Francis Collins, Director of the National Human Genome Research Institute, on personalised medicine, via the yourgenome YouTube channel

Paper activity looking at sequences from a cancer gene, to pinpoint mutations

Therapeutic Genetic Modification

1. The use of genetically modified viruses to treat cancer

In 2015, a study was published which had used genetically modified herpes (cold sore) virus to treat malignant melanoma. The genetic modifications to the virus were as follows:

a) It could no longer make a protein which normally enables the virus to reproduce inside healthy cells. Hence, it was unable to cause cold sores. However, since cancer cells themselves produce this protein, the virus was only able to reproduce inside cancer cells.
b) The viral gene which normally inhibits antigen-presentation on virus-infected cells was disrupted, so that the infected cells become more 'visible' to the immune system.
c) It had a human cytokine gene inserted into it, which triggers an immune response to the infected cells.

The trial had some success and research continues into viral therapies for other cancers, including bowel and prostate cancers.

2. The use of GM viruses to treat bacterial infections

Viruses have been genetically modified to improve the efficacy of antibiotics against bacterial infections. Some antibiotics target bacterial DNA, but bacteria have a defence mechanism which helps them repair DNA targeted by these antibiotics. The GM viruses disrupt this DNA repair action, allowing the antibiotic to get to work. In one study, mice treated with both GM virus and antibiotic had an 80% survival rate compared with a 20% for those treated with antibiotic alone.

3. Gene therapy: update

While gene therapy was first developed in 1972, it has had only limited success in treating genetic diseases. However, many scientists have continued to work on the technique, refining and developing it in light of previous findings and in 2011 it was reported that it had been used to successfully treat haemophilia B. Haemophilia B is a sex-linked blood clotting disorder, caused by a recessive allele found on the X-chromosome. The usual treatment involves injection with Factor IX, the clotting factor which is

not produced in sufferers. In the gene therapy trial, patients were injected with viruses which acted as vectors for a healthy copy of the gene for Factor IX. The viral vector delivered the healthy gene to cells in the liver, where the gene then successfully directed the synthesis of Factor IX. In 2015, the researchers continued to report the growing success of the treatment.

The gene therapy trial was an international effort; with a team in Memphis designing the viral vector, a team in London treating the patients and a team in Philadelphia monitoring their immune responses. The success of the trial has been put down to a number of factors, which were possible due to the teams learning from the problems encountered in other trials; for example the virus used as a vector was one which does not normally infect humans and therefore it was presumed that none of the patients would already have immunity to it. In addition, steroids were given to the patients, which acted to suppress their immune responses to the injected virus. The virus was also selected since it is one which rarely inserts itself into the host's chromosomes, therefore reducing the chances of serious side effects such as cancer, which have occurred in other gene therapy trials.

Gene therapy has been less successful in the treatment of cystic fibrosis where, despite ongoing research, progress has been slow. One of the reasons is that the target cells in the lungs are regularly renewed and so repeated doses of gene therapy need to be given. In addition, the viral vectors used in the past have induced immune responses. Recent studies have had modest success with a monthly treatment involving liposome vectors and it is thought that in 5-10 years, gene therapy will at least be part of an effective treatment for CF, if not an outright cure.

Some resources on gene therapy

Article – What is gene therapy?

Article – Is germline gene therapy ethical?