

Produced by the Centre for Genetics Education. Internet: <http://www.genetics.edu.au>

Important points

- The completion of the mapping of the estimated 20,000 genes in the human genome was announced in April 2003
- The number of human genes is only double that found in the roundworm
- About 11,000 genes in which the genetic code has been sequenced have been mapped to either one of the numbered chromosomes 1-22 (autosome) or the X or Y sex chromosomes in the nucleus of the cell or to the mitochondria
- Of these genes, only about 380 have been found to be associated with a genetic condition when the information in the gene is changed in some way. It is only for such conditions where the change in the gene has been clearly identified that DNA direct genetic testing may be available
- Understanding the other factors that interact with the genetic information in the development of complex conditions is also an area of intense research that may pave the way for the development of preventive strategies
- Research is ongoing in trying to understand the role that the 'non-coding' DNA and the interactions between the genes play in the expression and control of the genetic information
- The challenges that remain are :
 - To gain knowledge about the structure of the genome and its importance and the development of new technologies
 - To translate the genome based knowledge into providing diagnosis and predictive testing for genetic conditions and new treatments for genetic conditions
 - To maximise the benefits and minimise the harms in implementing the genomic based knowledge and technologies. Issues will include the provision of predictive or presymptomatic testing for conditions for which there is currently no treatment, privacy, genetic testing, population screening and establishing and regulating the boundaries

The Helix of Humanity

25th April 2003 marked the 50th anniversary of the discovery of the helical structure of DNA by Watson and Crick in 1953, as published in the journal, *Nature*

A brief history of genetics and human genetics discoveries (Table 24.1)

- The field of genetics had been initiated by the work of Gregor Mendel who in 1865 proposed the basic laws of heredity based on the observations from breeding pea plants
- This was followed by a number of other hallmark discoveries such as Thomas Morgan's experiments in 1910 with the fruit fly (*Drosophila*) that revealed the existence of the sex chromosomes and confirmed that genes reside on chromosomes
- In 1944 Oswald Avery, Colin Mcleod and Maclyn McCarthy discovered that DNA, and not protein, was the hereditary material in most living organisms

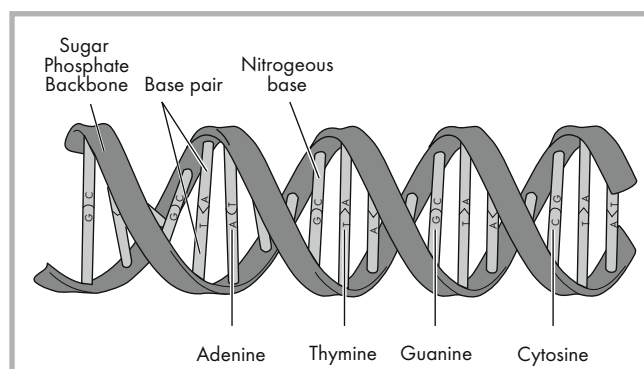


Figure 24.1: The DNA helical structure identified by Watson and Crick in 1953.

- It was the work of Watson and Crick however that heralded the field of the 'New Genetics' by determining that the DNA molecule must have a helical structure to enable the replication of DNA in cell division (Figure 24.1)
- Further research and technological advances progressed rapidly after the mid-1970s, so that the isolation and manipulation of the DNA was possible
- During the 1980s an increasing number of genes involved in human health, growth and development were located and their sequence of coded information described
- The further realisation of the potential for diagnosis and prediction of genetic conditions began with the initiation of the Human Genome Project (HGP) in 1990, established as an international research effort with the goal of producing a variety of biological maps of human chromosomes and determining the complete chemical sequence of human DNA: the substance that makes up genes. All the genetic material in the cells of a particular living organism is referred to as its genome. The human genome thus describes all the genetic material found in a human cell. For more information about the arrangement of this material into genes and chromosomes, see Genetics Fact Sheet 1
- The announcement that the human genome had been mapped was announced in April 2003 and thus set the stage for determining the function of the 20,000 or so genes in the human genome, a task that continues today

Table 24.1: The history of human genetics discoveries up to the 50th anniversary of the discovery of the DNA helical structure in 1953.

1866 Gregor Mendel proposes basic laws of heredity based on pea plants	April 1953 Francis Crick and James Watson discover double helical nature of DNA	1977 Maxam, Gilbert and Sanger develop DNA sequencing	1985 First use of DNA "fingerprinting" in a criminal investigation	1992 US Army begins collecting blood and tissue from all new recruits as part of a "genetic dog tag" program to give better identification of soldiers killed in combat	1999 USA announce a 3 year mouse genome project First human chromosome sequenced: chromosome 22
1882 Walter Fleming (embryologist) discovers tiny threads in the nuclei of cells of salamander larvae that appeared to be dividing. These later turn out to be chromosomes.	1964 Charles Yanofsky and colleagues prove sequence of nucleotides in DNA correspond exactly to the sequence of amino acids in proteins	1978 First human gene cloned: insulin	1986 First automated sequencer developed Approval for first genetically engineered vaccine for humans, for hepatitis B	1993 First rough map of all 23 chromosomes produced Gene for HD cloned	2000 <i>Drosophila</i> (fruit fly) genome sequenced Chromosomes 5, 16 & 19 draft sequence Chromosome 21 sequenced
1883 Francis Galton coins the term <i>eugenics</i> referring to improving the human race	1969 First gene in a piece of bacterial DNA isolated. The gene plays a role in the metabolism of sugar	1980 Mapping human genome proposed using RFLPs (restriction fragment length polymorphisms)	1989 Creation of the National Centre for Human Genome Research (headed by James Watson) which would oversee the Human Genome Project (HGP) to map and sequence the genes in human DNA by 2005	1995 <i>H. influenzae</i> (virus) sequenced Microarray (CHIP) technology developed	2000 June "Working draft" of human genome sequence announced
1910 Thomas Morgan's experiments with the fruit fly (<i>Drosophila</i>) reveal some characteristics that are sex-linked: confirms genes reside on chromosomes	1970 Researchers at the University of Wisconsin synthesis a gene from scratch	1982 First genetically engineered drug approved: insulin	1990 Formal launch of the HGP First human gene therapy experiment performed on a 4 yr old girl with an immune deficiency	1996 <i>S. cerevisiae</i> (yeast) sequenced	2001 February Publication of initial working draft of the human genome published in Science & Nature by the two rival private and public groups
1926 US biologist Hermann Muller discovers X-rays cause genetic mutations in fruit flies	1973 First genetic engineering experiment: Insertion of a gene from an African clawed toad into a bacterium	1983 Genetic marker for the genetic condition Huntington disease (HD) located on chromosome 4	1990 Publication of Michael Crichton's novel "Jurassic Park" in which bio-engineered dinosaurs roam a palaeontological theme park: the experiment goes awry	1997 Cloning of "Dolly"	2002 Genome of mouse completed
1944 Oswald Avery, Colin MacLeod & Maclyn McCarthy discover DNA, not protein, is the hereditary material in most living organisms	1975 First call for guidelines governing genetic engineering	1985 Kary Mullis develops PCR (polymerase chain reaction) to rapidly reproduce DNA from a very small sample that enables genetic testing for health and other applications such as forensics and paternity testing	1991 First gene involved in inherited predisposition to breast cancer and ovarian cancer (BRCA1) located on chromosome 17	1998 <i>C. elegans</i> (worm) sequenced	April 25th 2003 Completion of the mapping of the genes in the human genome announced setting the stage for determining the function of the then estimated 30, 000 or so genes

Table 24.2: Goals and achievements of the Human Genome Project 1993 - 2003

Goals 1993 – 2003
Complete a map of human DNA to a resolution of 80, 000 bases Complete a map of the DNA of model organisms to a resolution of 80, 000 bases <ul style="list-style-type: none"> • bacterium <i>E. coli</i> • yeast <i>S. cerevisiae</i> • worm <i>C. elegans</i> • fruit fly <i>Drosophila melanogaster</i> • mouse
Status at 1998
Midpoint of mapping of human DNA reached Maps of several model organisms completed <ul style="list-style-type: none"> • Yeast 1996 • Bacteria 1997 • Worm 1998
Goals 1998 – 2003
Finish 1/3 of human sequence and working draft remainder Complete human sequence by April 2003 Complete sequence for remaining model organisms
Status in 2002
Draft of 90% of the human genome completed 2001 Drosophila genome completed 2001
Status in 2003
Human sequence largely completed and available at Ensembl http://www.ensembl.org/ <ul style="list-style-type: none"> • Genomes of many organisms eg mouse also available • All publicly available DNA variations listed • Links information to known disease states • Around 600, 000 hits/day

The Human Genome Mapping Project 1993-2003

The Human Genome Organisation (HUGO) was formed in Switzerland in 1988 to facilitate international scientific collaboration and funded from 1990 largely by public sector funding from the USA and the Sanger Institute in the UK. The project cost approximately \$3 billion.

More funding was used by biotechnology companies in a race to analyse the human genome and to commercialise many of its outputs. The work was carried out at 20 sequencing centres in China, France, Germany, Japan, the UK and the USA. All made their data freely available via the internet as soon as it was produced. In addition, the United Nations' Educational, Scientific and Cultural Organisation (UNESCO) promoted the continued involvement of developing countries in the project's activities.

Objectives and goals

There were two major objectives in the Human Genome Project:

1. To develop detailed maps of the location of genes in the human genome and the genomes of several other well-studied organisms: bacteria, yeast, nematode, fruit fly (*Drosophila*), mouse and *Arabidopsis thaliana*, a rapidly growing plant that has a small genome. It was anticipated that knowledge gained by the study of genomes of other organisms would assist in the analysis of the human genome

2. To determine the sequence of the coded information contained in the DNA of the various genomes studied and from this to identify all of the estimated 20,000 human genes. This information is in the form of chemical 'bases' that are described by the letters A, T, G and C (see Genetics Facts Sheets 1, 4 & 5)

The Human Genome Project successfully completed all the major goals in two 5-year plans, 1993-98 and 1998-2003, in which human DNA sequencing was the major emphasis. The schedule to complete the full sequence by 2003 to coincide with the 50th anniversary of the discovery of the structure of DNA by Watson and Crick in 1953, was 2 years ahead of previous projections (Table 24.2).

In the course of completing the human DNA gene sequence, a 'working draft' was announced in June 2000 and formally published by the two groups involved in February 2001. The publicly funded international group, called the HGP, published their draft sequence in the journal *Nature*, on February 15, 2001 and the private consortium, Celera Genomics, published their sequence in the journal *Science* on February 16, 2001. Celera genomics did not continue on to participate in the final completion of the gene sequencing.

The plan also included goals for

- Sequencing technology development; for studying human genome sequence variation
- Developing technology for functional genomics
- Completing the sequences of several other organisms such as bacteria, yeast, the fruit fly *Drosophila melanogaster* and the mouse, to be used as models for research
- Studying the ethical, legal, and social implications of genome research; for bioinformatics (the collection, organisation and analysis of large amounts of biological information such as DNA sequence, using networks of computers and databases
- Training of genome scientists

The future

Many of the goals of the HGP have been achieved but much remains to be done: the identification of the genes in the human genome is just the beginning.

The genes' normal functions, the extent to which they vary from person to person and the role of this variation in causing or contributing to developmental, growth and health problems will remain unclear for some time and may take another 50 years, maybe more. Clarifying these issues will provide an ongoing role for human genome research.

In the publication in the journal *Nature* to celebrate the 50th anniversary of the DNA helical discovery, Dr Francis Collins and his colleagues outlined what lies ahead in harnessing the benefits and minimising the harms generated by the Human Genome Project.

They envisioned that the future holds three 'Grand Challenges', all of which are intertwined and dependent and they likened each challenge to the three floors of a building. The foundations of the building are the findings of the HGP. These challenges are:

(a) Genomics to Biology

The work will continue to elucidate the structure and function of a number of genomes. The next area where there will be rapid developments is called proteomics – the study of the gene products and how they work and interact with each other, the genes and other components in the cells. The study of DNA variations in different genomes and how they have contributed to evolutionary variation.

Knowledge about the structure of the genome and its importance

While the discovery of new genes and knowledge about their function is an obvious benefit, the organisation of the genes within the genome is also important.

Questions such as: is it important for genes to be located on a particular chromosome, or in a particular order to function, will hopefully be answered. In addition, knowledge about epigenetic processes such as ‘imprinting’ of genes as described in Genetics Fact Sheet 15 where the expression of a gene is determined by whether it is passed down from the mother or the father, will be important information generated.

Development of new technologies

The Human Genome Project has been the catalyst for enormous advances in the development of technology. For example, the polymerase chain reaction, or PCR, invented in 1985 by a biochemist, Dr K. Mullis, enables a very small amount of DNA to be copied many times over.

Thus only very small samples of DNA are required for testing, such as the amount of DNA in several hair roots or in the cells scraped or washed from the inside of a person’s cheek. This is also very important in forensic science when only a small sample of tissue or blood may be available for study.

(b) Genomics to Health

Translating the genome-based knowledge into health benefits includes

Diagnosis and predictive testing for genetic conditions

Genes contain coded information that directs the body to carry out particular tasks. Changes in this information, called mutations, prevent the correct message being issued to the cells and may result in a genetic condition. Once a gene is isolated and its correct sequence defined, it is possible to determine if a person has the correct copy of the gene or the faulty version that may result in a particular genetic condition.

This information may be used to diagnose a condition in an individual or during a pregnancy (prenatally) (see Genetics Fact Sheet 17) or it may be used to test an individual prior to any symptoms being present.

Sometimes a variation in the gene sequence does not make the gene faulty but increases an individual’s susceptibility to develop a condition when triggered by environmental factors. Susceptibility genetic testing is called ‘predictive’ testing for a genetic condition (see Genetics Fact Sheet 21).

New treatments for genetic conditions

There are many conditions that are identified as being genetic because of their pattern of inheritance within families but doctors do not know the biological basis of the condition. This is clearly illustrated in Huntington disease (see Genetics Fact Sheet 44). The discovery of the gene which is faulty in those individuals affected with the condition has led to the discovery of a previously unknown protein called *huntingtin* which is obviously important in neurological function.

Such knowledge provides hope for the design of treatments based on correcting the malfunctioning gene or protein (see Genetics Fact Sheet 27).

Other treatments will involve the development of drugs guided by the genes and their products. This new field of pharmacogenetics (see Genetics Fact Sheet 25) has already yielded a number of drugs used in breast cancer treatment for example. The determination of an individual’s genetic make-up and how it interacts with certain drugs foreshadows an era of personalised medicine.

(c) Genomics to Society

This challenge recognises that there will be both benefits and harms engendered by the current and future findings generated by the Human Genome Project. About 3% of the total budget for the Human Genome Project was allocated to support research, discussion and proposals about the ethical and social issues generated by the project, and these concerns will require further consideration as more genetic information is generated. The issues, discussed in more detail in Genetics Fact Sheet 21, are only a few which must be addressed by society as a whole as well as the policy-makers in government.

Predictive or presymptomatic testing for conditions for which there is currently no treatment

Genetic counselling (see Genetics Fact Sheet 3) is strongly recommended to assist in informed decision-making in this area.

Some will view having testing to see if he or she has a particular faulty gene for a condition that may not develop for some years, as beneficial, although there may not be an effective treatment. Examples of this include having presymptomatic genetic testing for Huntington disease and predictive testing for young-onset forms of Alzheimer disease (see Genetics Fact Sheets 44 & 45). Others may not see such testing as beneficial at all.

Genetics Fact Sheet 23 discusses the dilemmas for families faced with these choices.

Privacy

It will be important to ensure that access to personal genetic information is only on the basis of informed consent. Such information is likely to be of interest to third parties such as employers and insurance companies. Equal weight should be given to an individual’s right to know, or not to know details of his or her personal genetic information.

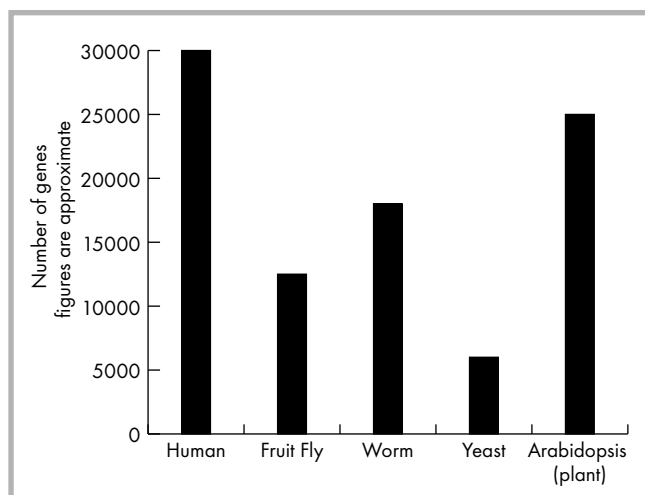


Figure 24.2: Comparison of number of genes between humans and other organisms

The Australian Law Reform Commission and the Australian Health Ethics Committee completed their report on the protection of human genetic information in Australia, *Essentially Yours*, in 2003 for the Federal Government available from (<http://www.alrc.gov.au>). This is a comprehensive coverage of all the issues related to privacy generated by the Human Genome Project.

Genetic testing and screening

It will be important to ensure that all genetic testing or screening is only carried out on the basis of voluntary, informed consent. Associated education and counselling services are highly recommended to avoid misunderstanding or discrimination.

The very existence of testing gives rise to the following questions:

- Should society ever impose testing on people who are at high risk of developing a condition?
- Should genetic information be provided to family members who did not seek the information for themselves but who may be at risk of developing, or passing on, a genetic condition?

Establishing the boundaries

The developments in technology enable the manipulation of the genetic material including the potential to clone a human. Society will need to keep abreast of these developments and determine if it should be applied, how it should be applied and in what situations (see Genetics Fact Sheet 23).

The human genetic code

The draft DNA sequence of the code contained in the genes announced in 2001 covered about 90% of the human genome, in which most of the genes are situated. The total number of chemical bases or letters in the genetic code was found to be about 3,164.7 million, making up the estimated (at that time) 30,000 genes.

In 2001, more than 50% of these genes were of unknown function. The final sequence covers almost 100% of the gene-containing human DNA sequence. The 'missing parts' are contained in about 400 gaps and remain unable to be deciphered by current technology.

The number of 'letters' in the genes is very variable: there are 3,000 per gene on average but the largest known gene is *dystrophin* (2.4 million letters or bases) in which mutations cause various forms of muscular dystrophy (Genetics Fact Sheet 41). Nevertheless, the order of letters in the genetic code of almost all genes is exactly the same in all people, regardless of race.

The total number of genes in the human genome is now estimated to be around 20,000. This figure is much lower than previous estimates of 80,000 to 140,000 and is not much more than in the genome of a very simple plant called *Arabidopsis* (25,706 genes) and is only double that found in the roundworm (Figure 24.2).

Only 25% of the human sequence identified is contained within genes (called the coding DNA) and only about 2% of the genetic code contains information to produce proteins. Therefore 50%-75% of the sequence is non-coding DNA or the 'string' between the genes.

Much of this non-coding DNA is made up of repeated sequences of 'letters'; this phenomenon is used in DNA fingerprinting for forensic and biological relationship testing as discussed in Genetics Fact Sheet 20.

Human DNA differs from the DNA sequence of other organisms by having a greater proportion of repeated sequences of letters in their non-coding DNA although it appears that during the last 50 million years, humans have largely stopped accumulating these repeats.

While the number of genes in humans and other organisms is surprisingly similar, humans have an average of three times as many kinds of proteins as the fly or worm. This is because the genetic code of the human genes can be read in a variety of ways, producing on average three proteins per gene compared to one protein per gene in the worm.

Table 24.3: Comparison of the count of mapped genes with known genetic code by chromosomal or mitochondrial location with the count of associated genetic conditions (<http://www.ncbi.nlm.nih.gov/Omim/mimstats.html> March 2007)

	Autosomal	X-Linked	Y-Linked	Mitochondrial	Total
Genes where the DNA sequence identified	10733	498	48	37	11316
Gene with known sequence associated with a genetic condition	353	32	0	0	385

Genes and genetic conditions

At March 9 2007, 11,316 genes in which the genetic code is known have been mapped to either one of the numbered chromosomes 1-22 (an autosome), the X or Y chromosome (sex chromosomes) in the nucleus of the cell or to the mitochondria (Table 24.3).

Of these over 11,000 genes, however, only 385 have been found to be associated with a genetic condition when the information in the gene is changed in some way. It is only for these conditions that DNA direct genetic testing may be available (see Genetics Fact Sheet 22).

Limitations and opportunities gained from the Human Genome Project

The determination of the entire DNA sequence contained in the human genome will not enable geneticists to look at a person's DNA sequence and predict everything about their appearance, behaviour and other characteristics.

Many genes interact to produce a particular characteristic and the analysis of genes in isolation is not going to provide the whole answer. In addition, while an individual's genetic make-up makes a considerable contribution to their health, growth and development, appearance and behaviour, environment also plays a major role. This is not just the physical environment such as diet and climate but also education, housing and access to high quality health services.

We are all more than the sum of our genes.

Nevertheless, enormous opportunities for treatment and the diagnosis of genetic conditions will be generated from the Human Genome Project. Finding out this information from genetic testing can free families from the uncertainties of an unexplained illness, may enable the early detection, treatment or prevention of a current or future health problem or may be used to prevent the condition affecting further children in the family.

Along with this freedom comes the burden of choice and, sometimes, knowledge. Genetic testing should only ever be undertaken on an informed basis so that the advantages and disadvantages can be weighed up. Some individuals with a family history of a debilitating condition that starts in middle age would rather not know if they have inherited the faulty gene when they are in their twenties and still healthy. Yet others would rather know so that they can make life decisions appropriate for them.

Genetic conditions are family health problems so that a diagnosis and the resulting decisions are often 'family affairs'. Some family members, however, may not want to share this information: others in the family may therefore never find out about their risk and be aware of opportunities for genetic testing, or condition prevention and early treatment. There may also be interest in the genetic information by others outside the family: employers and financial institutions for example may want to know the results of genetic tests that could predict an applicant's future health.

Other Genetics Fact Sheets referred to in this Fact Sheet: 1, 2, 3, 15, 17, 20, 21, 22, 23, 25, 27, 41, 44, 45

Information in this Fact Sheet is sourced from:

Collins FS, Green ED, Guttacher AE & Guyer MS. (2003). A vision for the future of genomics research. *Nature* 422; 835-47

The Australian Law Reform Commission and the Australian Health Ethics Committee 2003. *Essentially Yours* [online]. Available from: <http://www.alrc.gov.au>. [Accessed June 2007]

The National Centre for Biotechnology Information (USA) [online]. Available from: <http://www.ncbi.nlm.nih.gov/> [Accessed June 2007]

The National Centre for Biotechnology Information (USA) *Genes and genetic disorders located to chromosomes* [online]. Available from: <http://www.ncbi.nlm.nih.gov/Omim/mimstats.html> [Accessed June 2007]

The Sanger Institute. Ensembl [online]. Available from: <http://www.ensembl.org/> [Accessed June 2007]

U.S. Department of Energy Office of Science, Office of Biological and Environmental Research, Human Genome Program. *Major events in the Human Genome Project and Related Projects* [online]. Available from: <http://www.ornl.gov/hgmis/project/timeline.html>. [Accessed June 2007]

Watson JD & Crick FHC. (1953). Molecular structure of nucleic acids: A structure for deoxyribonucleic acid. *Nature* 171, 737.

Edit history

June 2007 (6th Ed)

Author/s: A/Prof Kristine Barlow-Stewart

Acknowledgements this edition: Gayathri Parasivam

Previous editions: 2004, 2002, 2000, 1998, 1996

Acknowledgements previous editions: Mona Saleh; Bronwyn Butler; Prof Graeme Morgan; Prof Ron Trent; Merran Cooper