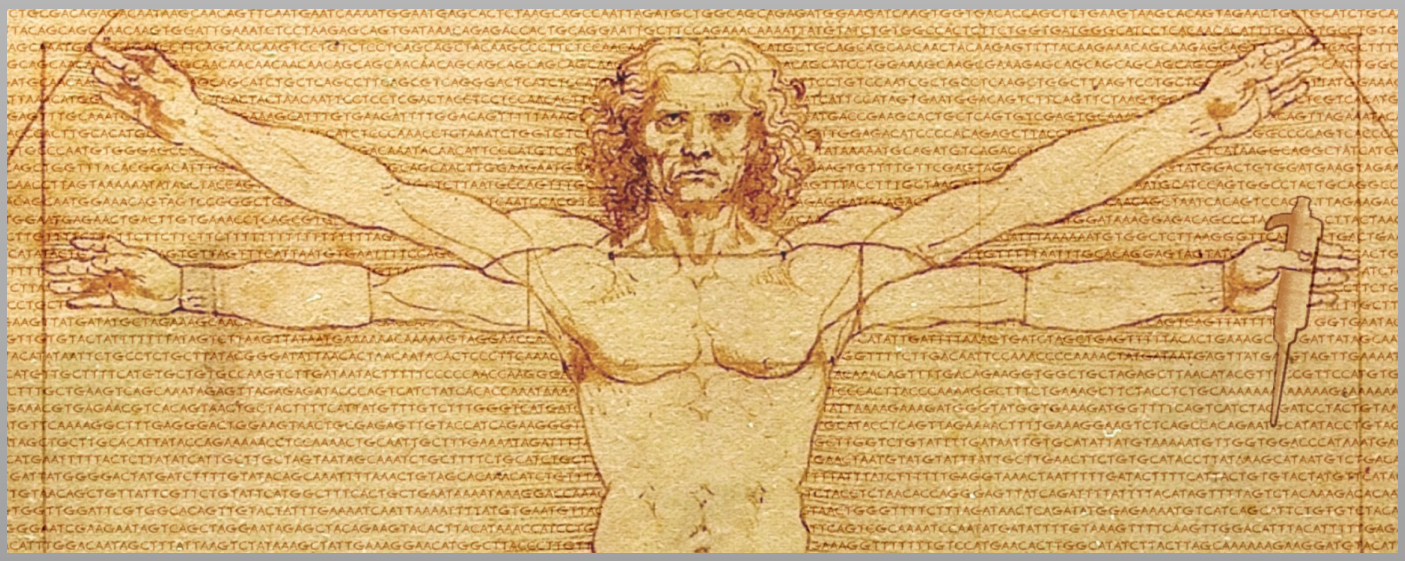
NCEA Biology Resource

STUDENT RESOURCE

# The Genomic Revolution



## The use of genomics in medicine, forensics, conservation, agriculture and consumer genetics, and its emerging impact on New Zealand society

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

Lints & Print | Genomics Into Medicine

This resource provides information about the scope of genomics research across a range of scientific disciplines and societal applications. It describes how modern technological innovations have fueled the emergence of vast amounts of DNA sequence data. This unheralded trove of information is now reshaping our understanding of how life has evolved, how organisms interact, and the mechanisms behind human evolution, susceptibility to disease and impact on the environment.

Human genomic data, gathered from an ever-increasing number of people, raises serious concerns over its use and presents profound ethical dilemmas that New Zealand society will need to grapple with. Misinterpretation of genetic data is also discussed, providing an important precedent to consider when addressing ethical questions about the use of human genomic information.

Cover picture: Leonardo da Vinci's drawing of Vitruvian Man (circa 1490). The ancient Roman architect, Marcus Vitruvius, believed that the beauty in nature's design was based in universal laws of proportion and symmetry. Vitruvian man reflects Leonardo's own fascination with proportion and the relationship of man to nature. We now know that life, with its 'endless forms most beautiful and most wonderful' (to quote Darwin) stems from the structural properties of DNA. The DNA sequence is from the *FOXP2* gene.

# Table of Contents



***Helpful hint:*** *In this resource,* **blue text hyperlinks** *(underlined) will take you to websites with additional information. The hyperlinks in* **maroon bold type** *lead to video or audio files (in some cases the video/audio will be found after a brief browsing of the linked page). If you are working from a printed version of this resource but wish to access these links, they are also referenced by number at the end of the resource as TinyURLs that you can type into a web browser. More source information about some of the images is also provided at the end of the resource.*

## Preview

This resource will introduce you to a lot of new information. However, it is not intended that you follow all the links or read the entire document (but major kudos if you do!). The purpose of the resource is threefold. First, it provides a fascinating overview of the breadth of genomics research. Second, it gives a background on the discovery of DNA and the technological advances that are now transforming our understanding of biology. Third, the resource highlights some of the pressing ethical questions now being raised by genomics and gene editing.

***For an overview of this genomics resource we recommend you watch the videos linked on this page.*** The link numbers are the same as those used elsewhere in the resource (although the link text might differ). The numbers make it easy to find and explore sections of the resource that will inform your thinking on the socio-scientific impacts of genomic research. The time required to watch all these videos is less than two hours.

As you know, variations in DNA sequence between people contribute significantly to our differences in **phenotype and disease risk**[7]. Indeed, there are many types of **variation**[116] that occur in our genome. Our understanding of this complexity has been enormously advanced by the **Human Genome Project**[102]. More recently, amazing technological progress in **Next-Generation Sequencing**[121] has opened the way to whole-genome sequence of millions of people and this is having a major impact on **precision medicine**[115]. Moreover, pilot programs in the **U.S.**[166] and **U.K.**[167] are exploring the costs and benefits of whole genome sequencing of newborns to identify potential risks of disease throughout life.

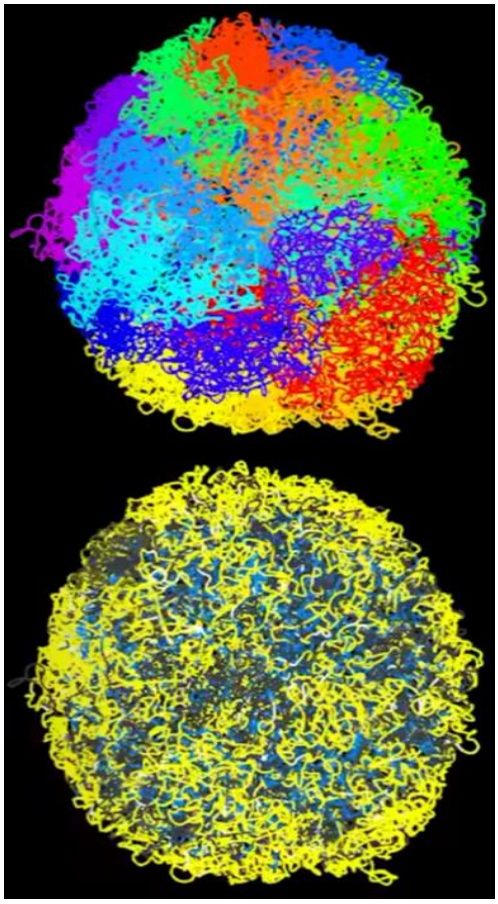
Scientists are making bold predictions for where this research will head within the next **decade**[8]. Beyond human health, genomics is making important contributions to **conservation biology**[50] and agriculture, in New Zealand and **overseas**[123]. Along with this increased genomic knowledge comes the possibility of specifically modifying organisms using **gene editing**[65] (even including the genetic manipulation of **human embryos**[88]).

Genomics is also providing fundamental new insights in anthropology, highlighting ancient relationships with our distant **Neandertal**[106] and Denisovan relatives. It has illuminated our understanding of the migration of Pacific peoples to **Aotearoa**[144]. Yet, we still know too little about Māori and other **Pacific peoples' genomes**[137]. This lack of knowledge runs the danger of perpetuating **racist inequities**[130] in health care.

Genomics also raises ethical questions in the criminal justice system. Can it help reduce racial bias and avoid **misleading interpretations**[191] about race and antisocial behaviour? When criminal behaviour might have a **genetic contribution**[210], how should the courts handle this information? With new understandings about gene-by-environment interactions, in part due to the **Dunedin Study**[198], what changes can society make to enhance equity and reduce antisocial behaviour?

Perhaps you now wish to follow up with a few readings. A topical place to start is by looking at how advances in DNA sequencing[26] have been central to understanding SARS-CoV-2[6] and the rapid response to the COVID-19 pandemic. A wealth of genomic sequence data has also provided crucial insights into the polygenic origins[22] of non-infectious human diseases such as cancer, cardiovascular disease and diabetes.

That data mostly comes from European populations. Indigenous peoples[149] are rightly concerned about the governance, use and storage of their genomic data. For Māori, a damaging controversy swirled around the MAOA gene[187], involved in gene-by-environment[199] interactions. Particular care must be taken to avoid harm and in future ensure that genomic medicine in Aotearoa/New Zealand reduces health inequities[153].



Stevens *et al.* (2017) *Nature*.
For full citation see p28.

A **genome**[1] in a nucleus. Top view, chromosomes are in different colours. In the lower view, active genes are in blue and less active genes are coloured yellow.

**Function follows form**[2] and so biologists are interested in how the chromatin structure of DNA in the nucleus regulates expression of genes in the cell.

## What is genomics?

Sydney Brenner[3], a geneticist awarded the Nobel prize for his work on the soil nematode *Caenorhabditis elegans*, liked to say that genomics is just a glorified new term for genetics. Genomics and genetics are concerned with DNA, genes and the acquisition and inheritance of genetic variations (mutations). However, the term genome relates specifically to the complete nucleic acid sequence of an organism, organelle or virus. Despite Brenner's comment, he championed the *C. elegans*, Fugu, *Mola mola*[4] and Human Genome Projects.

Juergen Berger / Max Planck Institute

Scanning electron micrograph (SEM) of an hermaphrodite *Caenorhabditis elegans* nematode worm - the first multicellular organism for which a complete genome sequence was determined.

We have two copies of the genome in each diploid cell in our bodies, as well as mitochondrial genomes in each cell (organellar genomes being present due to the **endosymbiotic**[5] origin of eukaryotes). In addition, we all likely carry viral genomes present in at least some of our cells. For some viruses, such as SARS-CoV-2[6] that causes COVID-19, the genome is comprised of RNA, not DNA. Sequencing of viral genomes to reveal the structure and likely function of proteins they encode represents a fundamental step in the process of developing candidate vaccines.

Genomics, then, is the study of the large amounts of DNA sequence information from an organism or population. The suffix -omics is also used to describe the study of other data-rich biological research approaches. Transcriptomics is the study of all the RNA transcripts found in an organism or cell, proteomics is the study of the

large numbers of different proteins present, and metabolomics identifies the complex small molecule metabolites found in a cell, tissue or organism (for example, in blood or urine). However, because of the simple and universal structure of DNA and its central role[7] in defining an organism's characteristics, it is genomics that has progressed most rapidly and now raises the greatest benefits and concerns for society. Major advances are on our doorstep, captured in ten bold predictions[8] for what the next decade of genomics has in store, including:

- The biological function(s) of every human gene will be known and non-coding regions of the genome largely understood.
- Generating and analyzing a complete human genome sequence will be as straightforward as carrying out a DNA purification.
- Studies involving analyses of genome sequences for millions of people will be regularly featured at school science fairs.
- Human genomics will enable social and scientific studies to move away from historic social constructs such as race.

## Applications of genomics

The life history and individual characteristics of all organisms and their responses to the environment are strongly influenced by genome-level DNA variation. Not surprisingly, nearly all aspects of biological research are therefore informed and impacted by genomics. A representatively diverse range of genomics-related studies includes research on:

- The personal hygiene of bees[9].
- The genetic modification of plants[10] in an effort to mitigate climate change.
- The evolutionary consequences of ecosystem disturbances caused by earthquakes and other geological processes[11].
- The political hierarchy and elite god-kings[12] of Neolithic Ireland.
- Damage to astronaut DNA[13] due to elevated radiation exposure when in space.
- The identification of performance-enhancing bacteria[14] in the gut of elite marathon runners.

Virgin Money London Marathon

Metagenomics reveals that elite long-distance runners, such as Mo Farah and Eliud Kipchoge, have increased *Veillonella* in their gut microbiome. These bacteria convert lactate to propionate, thereby increasing the body's resilience to exercise stress.

Perhaps you might not be a god-king, astronaut or world-class marathoner, but nevertheless the genome is already influencing your path in life. For example, genetic contributions to educational attainment[15] are now coming into focus. As with other complex traits and disease-risks, allelic variants of many genes, each with small positive or negative effects, cumulatively contribute to achievement in school. Summing the effects of these variants, inherited in a probabilistic way, produces a 'polygenic score'[16] for educational attainment. However, while such scores may be informative at the population level, they are not predictive at the individual level, in large part because there are also so many social and environmental factors that also contribute.


University of Auckland / Faculty of Medical & Health Sciences

Polygenic scores for educational attainment rest on genetic contributions to intelligence, self-control, conscientiousness and grit. Of course, environmental factors also have a strong influence in the path through tertiary study, celebrated at graduation.

The previous examples highlight the fact that modern genome sequencing technologies now allow biologists to tackle a huge array[17] of brand new interesting, challenging and sometimes esoteric questions. But thinking more generally, what are the main areas of human activity currently impacted by genomics?

There are six major intersections of genomics with New Zealand society – these occur in the context of the healthcare and criminal justice systems, conservation, primary industries, consumer genetics and reproductive genetics. Below, we consider the role of genomics in each of these six areas. The impact of the genomics revolution on medicine and law enforcement, and ethical concerns this creates, is considered in greater depth in the latter half of this resource.

## *Genomics and healthcare*

In 2001, two drafts of the three billion base pairs of the human genome were simultaneously published by competing public[18] and private[19] research teams. Francis Collins, then director of the U.S. National Human Genome Research Institute, described the genome sequence thus:

"*It's a history book - a narrative of the journey of our species through time. It's a shop manual, with an incredibly detailed blueprint for building every human cell. And it's a transformative textbook of medicine, with insights that will give health care providers immense new powers to treat, prevent and cure disease.*"

Nature: Eric Lander & Darryl Leja / Science: Ann Elliott Cutting

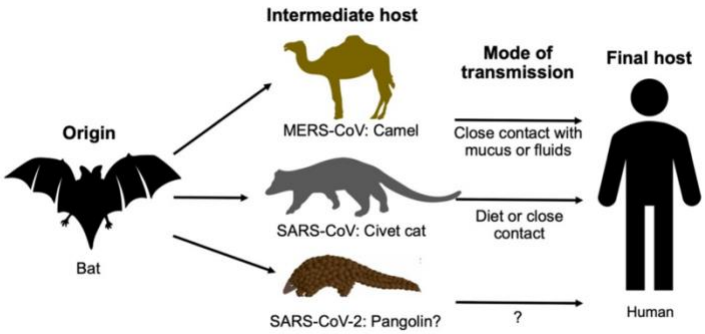
February 2001 covers of the world's two leading science journals. *Nature* published the International Human Genome Sequencing Consortium's genome draft and *Science* published Celera's genome draft.

The transformative impact of genomics on medicine continues to grow. Genomics research has identified many genes contributing to cancer, obesity, diabetes, heart, psychiatric and auto-immune diseases. Genomics has helped individuals with very rare monogenic (single gene) disorders, for example affecting skeletal development[20], identifying causative mutations and enabling the connection of New Zealand families affected with rare syndromes[21] to others in the international community.

However, in contrast to the Mendelian genetics taught in the classroom, prevalent non-infectious diseases are typically polygenic[22], involving 10's to 100's of genes that each contribute some proportion of disease risk. Advances in genomics now make it practicable to sequence an individual's genome to identify the genes that contribute to their particular disease risk and response to drug treatment, ushering in a new era of personalized, or precision, medicine.

Genomics has also made important contributions to fighting infectious disease. Amazing developments in the miniaturization and speed of DNA sequencing technology have enabled genomic surveillance of pathogens, such as ebola and zika[23], in real-time during disease outbreaks in tropical environments.

Most recently, the genomic sequence of the entire SARS-CoV-2 genome, released January 11th 2020, set the stage for tracing the zoonotic origin[24] and spread of coronavirus around the world, which has led to the most severe global pandemic since the 1918 "Spanish Flu".



Yi *et al.* (2020). *Int J Biol Sci.* Full citation on p28. (CC-BY-NC 4.0).

Bats are the primary reservoir for zoonotic corona-viruses that affect humans. In some cases an intermediate host is implicated. Pangolins may have been involved in the initial SARS-CoV-2 transmission to humans. SARS-CoV-2 -related coronaviruses, bats and pangolins coexist over an extensive geographical range, extending from China into Thailand[25].

Built on a long history of progress in DNA sequencing (see pages 8-12), genomic analysis of SARS-CoV-2[26] is now central to saving millions of lives in the current pandemic. The worldwide epidemiology of the virus, utilising genomic data, can be explored on an interactive website[27].

## *Genomics and the criminal justice system*

The use of DNA tests to solve crimes and contribute evidence to civil court cases has been a feature of innumerable TV shows and movies over the years. In many cases, DNA evidence can provide a high degree of confidence about an individual's involvement in a crime – or their innocence. Most forensic DNA testing does not involve genome sequencing, but this may change in the coming decade(s) as costs drop. Historically, New Zealand has been in the vanguard of nations utilising DNA tests as evidence in the legal system. In 1995, New Zealand was just the second country (after the United Kingdom) to establish a legislative framework for DNA sample collection and profiling for criminal justice purposes. The Criminal Investigations Bodily Samples Act (CIBS Act[28]) permits DNA collection from known individuals, by consent or by compulsion, and established a national databank to hold the resulting DNA profiles.

On average, the genomic DNA sequence of any two individuals in a population is reported to be 99.9% identical. However, with a genome size of three billion base pairs, this still amounts to several million nucleotide differences (and, of course, there are, equally, millions of differences between the two copies of the genome in each diploid cell). Indeed, an analysis of 2,500 human genomes from 26 populations reveals that a typical human genome differs from another (or from the reference human genome) at 4.1 to 5.0 million positions. Sampling of only a small proportion of these genetic differences is sufficient to unambiguously identify an individual (excepting identical twins) by matching their DNA profile to a forensic specimen. Thus, DNA constitutes a far more rigorous and reliable source of forensic evidence than other biological sample analyses (for example, of **hair structure**[29]) that, too often, have resulted in the miscarriage of justice.

The development of DNA profiling arose from the work of an English scientist, Alec Jeffreys[30], who immediately recognized the application of human genomic variation to forensics and civil court cases.

Standard DNA profiling techniques[31] in forensic analysis now mainly focus on microsatellite, or short tandem repeat (STR), typing. In New Zealand, this work is carried out by ESR[32], the Institute of Environmental Science and Research, which performs a variety of STR-based forensic techniques and has developed world-leading software (STRmix) capable of resolving individual DNA profiles from complex DNA mixtures. These techniques have enabled ESR to identify the perpetrators of some of New Zealand's most perplexing murders, including those of Maureen McKinnel[33] and Teresa Cormack[34]. Genomics is the next major leap in the forensics toolkit and recently has been used to identify the Golden State serial killer[35], responsible for at least 50 rapes and 12 murders in California between 1976 and 1986. However, the complexity and expanding availability of genomics data raises a host of thorny ethical issues, some considered later in this resource.

Abaca Press / Alamy

Joseph James DeAngelo Jr was responsible for robberies, rapes and murders in California between 1973 and 1986. In 2016, forensic DNA data was submitted to a public genetic genealogy database, GEDmatch permitting his identification.

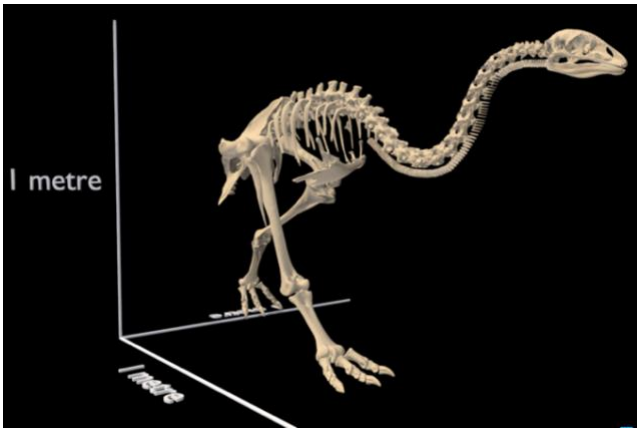
## *Genomics and conservation biology*

New Zealanders show a great deal of pride and concern for the beautiful and diverse habitats the country is blessed with. Human occupation of Aotearoa and the introduction of foreign species has nevertheless profoundly affected native fauna and flora, in many cases resulting in species extinction. The continued survival of many remaining native species hangs on a knife edge and faces increased uncertainty due to human population growth and pressures on land use. Against this backdrop, scientists are employing genomic techniques to get a better understanding of the biology and genetic diversity of endangered species, monitor the health of the

environment and work toward the elimination of predators.

Genomes of several iconic New Zealand species have been sequenced, including the kea, the North Island brown kiwi[36], kakapo[37] and tuatara[38]. These studies reveal some of the genetic changes that underlay the specialisation of these animals for their niche, but also highlight the limited genetic diversity in many existing populations - an important conservation issue.

Recently, international researchers have also succeeded in assembling about 85% of the little bush moa[39] genome from DNA extracted from a 700-year-old toe bone. This has revitalized discussions about the merits of attempting the de-extinction[40] of moa. As conservation biology is a chronically underfunded area, the costs of de-extinction efforts need to be weighed against the costs of ensuring the success of extant (living) species. Cultural views on such an effort would also need to be considered carefully.



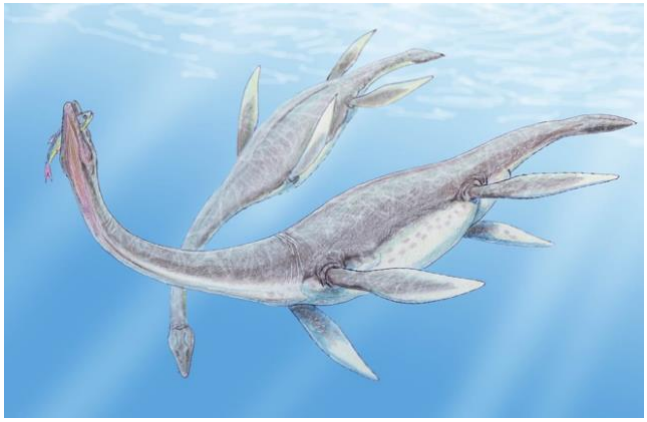

The Auckland War Memorial Museum Tāmaki Paenga Hira

An image from a 3D-rotatable[41] representation, from the Auckland Museum, of the skeleton of the little bush moa, New Zealand's smallest and most widespread moa species. An emu genome sequence was used as a scaffold to assemble the moa genome from the fragmented DNA retrieved from a toe bone.

Many New Zealand native avian species have become extinct or are threatened, in no small part because of brushtail possums, stoats and rats[42] (dogs and cats have not helped either). Advanced genetic techniques utilizing genomic information, such as gene drives[43], are being contemplated for the elimination of predators and other introduced pests in New Zealand. To meet the 2050 target of eradicating pests from New Zealand, the Department of Conservation[44] and other agencies/sectors will require a larger workforce with conservation genetics expertise. Internationally, gene drives are being explored as a way to control the spread of diseases, such as malaria[45], through insect vectors.

The genomic sequencing of some of New Zealand's famous avian fauna – the North Island brown kiwi and the little bush moa – was conducted by foreign research teams. However, genomic sequencing and other research on our native taonga species, and the conservation efforts these inform, should deeply involve Māori and would benefit from the wellspring of indigenous knowledge[46]. Recent genomic research on the tuatara, in collaboration with Ngātiwai[47], serves as an excellent model for this approach. As more young Māori and Pacific New Zealanders take up careers in research, the intersection[48] of indigenous knowledge and genomics will become more widespread.

Another important application of genomics to conservation is in environmental monitoring. Metagenomics[49] – sequencing of the genomes of many different (often) microbial organisms that coexist in an environmental sample - can provide key information about the health of a water or soil source. Surprisingly, just like a crime scene, organisms in their environment leave traces of genomic DNA wherever they go. These DNA traces, termed environmental DNA (eDNA[50]) can also be used to assess the species diversity of a habitat and gauge its health. This approach has been used to address one of the enduring challenges of scientific discovery - to confirm or refute the existence of the Loch Ness monster[51]. Current evidence suggests that it is not some sort of plesiosaur from the age of the dinosaurs, but instead from the more ancient lineage of anguilliformes![52].



Dmitry Bogdanov

Plesiosaurs were quite common Mesozoic era marine reptiles and have left behind a substantial fossil record spanning from about 200 million to 66 million years ago. Arguably, there have been reports of a 'water beast' in Loch Ness, or the River Ness, dating back to 565 A.D. Modern 'sightings' since 1933 have kept the interest of the public – and the occasional scientist.

## *Genomics in primary industries*

Humans have been genetically selecting for useful traits in crop species and domesticated animals for several thousand years. Genomics is transforming our understanding of the plant and animal domestication process, including for crops that are dietary staples of the world population, such as rice[53], maize[54] and wheat[55]. With the advent of sophisticated molecular tools to make directed changes to species' genomes, such as CRISPR[56], the production and diversification of new cultivated food sources is likely to increase dramatically in coming years.

Researchers in the New Zealand primary industry sector are making heavy use of genomic information to increase the efficiency of the selective breeding process and enhance desirable attributes. The assembly of the radiata pine[57] genome will enable the forestry industry to breed different lines of trees specifically suited for construction timber, paper or biofuel production. The sheep[58] genome is being studied to identify DNA sequence variants that contribute to disease resistance and meat quality factors. Metagenomics of the ruminant gut microbiome[59] is also highlighting ways to reduce methane emissions from farm animals, which contribute to global warming and account for nearly half[60] of New Zealand's greenhouse gas emissions.

Genomics is also being used in aquaculture efforts to establish farmed fish[61] species native to New Zealand waters, including tāmure (snapper; *Pagrus auratus*) and araara (trevally; *Pseudocaranx dentex*). A critical early step in this project was to assemble the snapper[62] genome, to identify useful genetic markers for the breeding process. In addition to the commercial benefit of providing sought-after fish for local and international markets, this aquaculture initiative will hopefully reduce some of the pressure on wild fish stocks.

The apiculture industry in New Zealand is also likely to benefit from the application of genomics approaches. Manuka[63] (aka tea tree; *Leptosperumum scoparium*), a taonga species esteemed by Māori for its medicinal uses, recently became the first New Zealand native plant for which whole genome sequence was obtained. The genome of the honey bee[64]. (*Apis mellifera*) was published in 2006 by an international consortium of researchers. This provides the foundational knowledge for efforts currently underway in New Zealand to use genomic information to improve bee stocks.

Noel O'Riley

A honey bee on a manuka blossom, the backbone of a $5 billion industry. Bees are predated upon by two social wasp species that were introduced to New Zealand in 1921 and 1945. In 2015, 20% of North Island beehives were lost to wasp attacks, making *Vespula* wasps a target for gene drive control.

The rapid accrual of genomic information, for native species, introduced pests and commercially important crops and animals, places an ever-increasing pressure on the New Zealand government, and the public it represents, to decide whether and how this information should be used in the genetic editing of organisms released into the environment[65].

## *Genomics and consumer genetics*

For many members of the public, their first and most intimate connection with genomics is through direct-to-consumer genetic testing[66] of ancestry and certain traits and health-related information (DTC genetics). The enthusiasm for this sort of information has prompted many people to use such a service without fully appreciating[67] the many ways their genomic data could be used.

There are now dozens of companies offering DTC genetics service, two of the biggest being 23andMe and Ancestry.com. New Zealanders have embraced DTC genetic testing at higher rates than many other nationalities, perhaps because the country's geographic isolation and recent colonial past prompts Kiwi's to seek their place in the larger story[68] of the world.

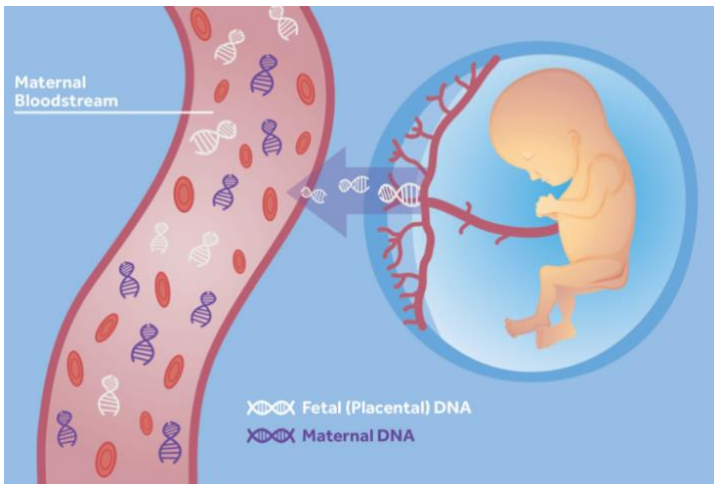
Direct-to-consumer genetic tests sequence only a relatively small percentage of the genome, but target informative polymorphic markers[69] and a few genes of interest known to contribute to particular traits and disease risks. Although not whole genome DNA sequence, DTC genetic test results can nevertheless be a modern-day

Pandora's box. Sometimes, these results provide unexpected surprises about parentage[70]. Consumer genetic testing also raises important issues about genetic privacy[71] and the storage, ownership and usage of the data acquired by these testing companies. Significant concerns have also been voiced about DTC testing's implications for Māori[72]. Later in this resource, we explore indigenous perspectives on genomics in more detail.

### *Genomics & human reproductive genetics*

The first discoveries of human chromosome aberration, in 1958 and 1959, of trisomy 21 in Down syndrome and sex chromosome abnormalities in Turner and Klinefelter syndromes led to the emergence in the 1970's of prenatal karyotype testing of fetal cells obtained from amniotic fluid. The amniocentesis procedure does involve potential miscarriage risk and is now being replaced by non-invasive prenatal testing, NIPT[73]. Cell-free fetal DNA that enters the mother's blood stream can be obtained from a maternal blood sample and assessed for aneuploidy (fetal chromosome copy number alterations). Fetal medicine specialists have argued that NIPT should be introduced in New Zealand as a publicly-funded screening procedure, but such a move is not supported by some disability rights[74] advocates.

Non-invasive prenatal testing (NIPT) involves drawing a small amount of blood from the mother, just as in a standard blood test. Fragments of fetal DNA cross the placenta and into the mother. These can be assessed bioinformatically to discriminate between fetal DNA and the mother's circulating cell-free DNA.



Yourgene Health plc

Improvements in the sensitivity of DNA sequencing techniques and an emerging appreciation of the numerous small deletions and duplications we all carry in our genomes are driving the expansion of NIPT test panels. These chromosomal abnormalities are termed copy number variations, CNVs[75], and recent estimates indicate that a typical human genome has ~160 CNVs (and these average ~250,000 base pairs in size).

Genetic testing companies now view genome-wide NIPT as providing a market advantage[76]. This ratcheting up of the number of genes and conditions assessed through NIPT raises perplexing ethical questions[77] for families and society at large. At the extreme, these concerns bring to mind the awful history of eugenics. It is important to bear in mind that eugenic thought was not confined to Nazi Germany, underlying horrors there, but also influenced public policy in Britain[78], the United States[79] and New Zealand[80]. In New Zealand, these policies led to the establishment of 'psychopaedic' institutions such as the Templeton Centre[81], which was shuttered only relatively recently.

Similar issues come up when considering the ramifications of preimplantation genetic diagnosis (PGD) as part of *in vitro* fertilization IVF[82] testing. Many fertility clinics utilize PGD to allow prospective parents to choose the sex of their child and less scrupulous[83] practitioners are promoting eye and hair colour choice. Conceptually, it is not a big jump to the biopunk dystopia depicted in the movie Gattaca[84]. However, current research highlights the difficulties that would probably be encountered when trying to produce significant intelligence or physical gains[85], given these characteristics have complex polygenic origins.

The debate[86] on the ethics of this research continues. Even so, IVF children have been born after correcting for a fatal mito-chondrial disorder using an approach termed spindle nuclear transfer[87] (the maternal nucleus is swapped into a host egg prior to fertilization).

Most controversially, in 2018 a Chinese scientist, Jiankui He, modified a gene in human embryos using CRISPR and implanted these into women, resulting in the birth of the world's first genetically edited babies[88]. This prompted a huge outcry from the public and scientific community, ultimately resulting in a three-year jail sentence for the scientist. Concerningly, both intended and unintended consequences[89] of genetic editing can be handed down though the generations.

## A short history of DNA research through to the current genomic era

- *1865* Mendel publishes experiments establishing laws of genetic inheritance
- *1866* Ernst Haeckel proposes the nucleus contains hereditary factors
- *1869* Friedrich Miescher discovers DNA and names it nuclein
- *1902* Boveri and Sutton postulate heredity units are on chromosomes
- *1905 - 1910* Stevens and Morgan show chromosomes carry heritable information
- *1909* Johannsen uses the word "gene" to describe units of heredity
- *1919* Phoebus Levene shows that DNA and RNA are comprised of nucleotides
- *1928* Griffith finds a "transforming principle" that transfers bacterial traits
- *1944* Avery et al. demonstrate heritable change resulting from DNA transfer
- *1950* Erwin Chargaff develops rule of the relative abundance of nucleotides
- *1953* The structure of DNA as an antiparallel double helix is published
- *1956* Kornberg discovers the replicative DNA polymerase enzyme
- *1957* Crick & Brenner develop The Central Dogma: DNA→RNA→Protein
- *1958* Meselson and Stahl demonstrate DNA replication is semiconservative
- *1961* Nirenberg and Matthaei begin to crack the codon code
- *1971* The recombinant DNA era begins with researchers cutting and joining DNA
- *1977* Frederick Sanger develops an efficient DNA sequencing method
- *1981* The human mitochondrial genome is sequenced
- *1982* First recombinant DNA-based drug (human insulin) is produced
- *1983* Kary Mullis comes up with the idea for the polymerase chain reaction
- *1986* The first automated DNA sequencing machine is developed
- *2001* The first draft human genome sequences are published
- *2010* Svante Pääbo publishes the first ancient genome, of Neanderthals
- *2012* George Church and colleagues encode a book into DNA
- *2018* First genetically-edited human babies delivered

We have seen how genomics is now having a pervasive impact on society, but how has this come about? To appreciate how rapidly the genomic revolution has taken off, it is helpful to know a little about the history of DNA research to truly appreciate the recent technological developments described in the next sections. Depending on your inclinations, you might skim through this history, or learn a little bit more[90] about the key scientists and discoveries that ushered in the "Genomic Revolution".

Not long after Gregor Mendel conducted his famous experiments on peas (1856-1863), establishing the foundation of genetics and identifying recessive and dominant inheritance of traits, a Swiss physiological chemist, Friedrich Miescher, discovered DNA. In 1869 Miescher set out to study proteins inside the nuclei of white blood cells, obtained from pus-coated bandages. He found instead a phosphorus-rich substance with a chemical composition and properties unlike any known protein, which he termed nuclein (later called nucleic acid).

In 1905, Nettie Stevens deduced the chromosomal basis of sex determination and in 1910 Thomas Hunt Morgan showed in the fruit fly, *Drosophila*, that genes conferring traits are carried on specific chromosomes. However, Miescher's findings had fallen into obscurity and the connection of DNA to genes was unclear.

In 1919 a Russian biochemist, Phoebus Levene, discovered that nuclein was comprised of nucleotides with a phosphate-sugar-base structure, with deoxyribose being the sugar present in DNA and ribose the sugar in RNA. Levene proposed a tetranucleotide structure for DNA, in which the four nucleotides were linked, always in the same order.

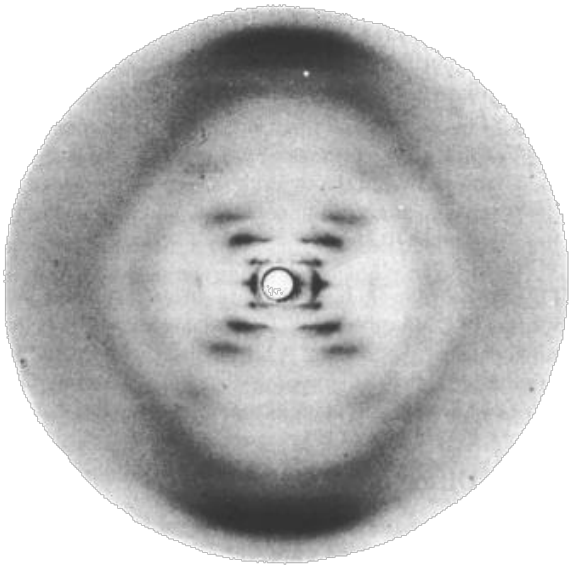
The simplicity of Levene's tetranucleotide structure did not lend itself to encoding the information required for genetic inheritance, so many researchers at the time favored proteins to be the genetic material as they were already known to be very complex. It was not until 1944, and the publication of a paper by Oswald Avery, Colin MacLeod and Maclyn McCarty, that DNA was established as the "transforming principle[91]" capable of causing heritable changes.

In 1950, inspired by reading Avery's work, Erwin Chargaff (an Austro-Hungarian-born American biochemist, who had fled to New York from Nazi Germany) began working on DNA. He discovered that, contrary to the prediction of Levene's tetranucleotide model, the composition of DNA

varies between species, but that in the DNA of any organism the number of A and T nucleotides is the same, and the number of G and C nucleotides is the same (Chargaff's rule).

Chargaff's rule, in conjunction with an X-ray crystallographic photograph of DNA taken by Rosalind Franklin Rosalind Franklin[92], who was working in the lab of Maurice Wilkins at King's College, London, provided the critical information needed for Oxford University scientists James Watson and Francis Crick to discover the structure of DNA, reported in their landmark 1953 paper[93] in the journal *Nature*. As Watson and Crick understatedly mentioned at the end of their *Nature* paper:

"*It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material*".



King's College London Archives/CC-BY-NC 4.0

'Photo 51' Franklin's X-ray diffraction photograph of b-form DNA – perhaps the most important image ever taken - and one of the more enigmatic (see **here**[94] for a refresher video on DNA structure). More details about the discovery of the double helical structure of DNA, including interviews with Watson and Crick and links to other resources, can be found at **HHMI BioInteractive**[95].

In 1958 two graduate students working at Caltech, Matthew Meselson and Franklin Stahl demonstrated the semiconservative replication of DNA, confirming Watson and Crick's conjecture through an **elegant experiment**[96].

Into the early 1960's scientific attention focused on the challenge of unlocking the code of life: namely, how the sequence of the 20 different amino acids that form proteins is specified by the four DNA bases and the order of nucleotides along an RNA. The first steps in deciphering the codon code were taken by Marshall Nirenberg and Heinrich Matthaei, at the U.S. National Institutes of Health, who showed that an RNA containing a repetition of uracil bases could be translated into a protein chain of phenylalanine residues.

The period of 1971-1973 ushered in the era of recombinant DNA research, first with Paul Berg, at Stanford University, using the restriction enzyme EcoR1 to cut bacteriophage lambda DNA and simian virus DNA (SV40) and join these together. Soon after, Stanley Cohen (at Stanford) and Herbert Boyer (at UCSF) cloned EcoR1 cut DNA fragments into plasmids and transformed these into bacteria, directly leading within just a few years to the birth of the biotechnology industry[97].

Recombinant DNA techniques were hugely useful in molecular biology and later were essential for the construction of DNA libraries used in the human genome project. However, a more immediately realized application was the development, in 1982, of the first recombinant DNA-based production of human insulin, which since its discovery in 1921 by Banting, Best, Collip and MacLeod had been purified from animal organs as a **life-saving treatment**[98] for diabetic patients.

The first genome sequence, the 3569 nucleotide-long RNA of bacteriophage MS2, was obtained in 1976 through a very laborious procedure. A major step forward in understanding genes and genomes came with the invention, in 1977, of far more efficient sequencing methods by Maxam and Gilbert and by Frederick Sanger. Ultimately, Sanger's method proved to be the one favoured. Soon thereafter, the first complete sequence of the 16,569 base pair human mitochondrial genome was published (in 1981), followed by the 48,502 base pair genome of bacteriophage lambda in 1982.

In 1983, Kary Mullis[99], working at Cetus Corporation, came up with the idea for the Polymerase Chain Reaction (PCR) in what is told as a Eureka! moment. PCR is a method for **DNA amplification**[100] and forms the basis of an essential step in many current high-throughput DNA sequencing procedures.

In 1986, Leroy Hood and colleagues at Caltech developed the first automated sequencer (notably mentioned in the 1993 film, *Jurassic Park*), which utilized Sanger's **dideoxy**[101] chain termination sequencing method, but now used a

combination of capillary gel electrophoresis and fluorescent labeled nucleotides (instead of slab gels and radioactive nucleotides).

In 1995, the first complete bacterial genome sequence was published (the 1,830,137 base pair genome of *Haemophilus influenzae*), in 1996 the first eukaryote genome was sequenced (the brewer's yeast *Saccharomyces cerevisiae*; a 12 million base pair genome) and in 1998, Sydney Brenner's favorite animal model, the nematode worm *C. elegans*, became the first multicellular organism to have its whole genome sequenced (101 million base pairs).

## The Human Genome Project & beyond

October 1st 1990 marked the official beginning of the Human Genome Project[102], with the first phase of research devoted to establishing a detailed genetic map of the human genome. The pilot phase of sequencing the yeast and worm genomes began in 1996 to establish methods needed to adequately cover and assemble large eukaryote genomes. High-throughput human genomic DNA sequencing commenced in 1999, involving the efforts of universities and institutes across North America, Europe and Asia (the International Human Genome Sequencing Consortium; IHGSC).

In parallel, a private biotechnology company, Celera Genomics, led by Craig Venter, set out to sequence the human genome using a faster but less methodical strategy. Ultimately, the IHGC and Celera[103] competitive efforts synergized, and these groups simultaneously published the first draft sequences of the human genome in February 2001.

A 'completed' human genome sequence, with over 99.99% accuracy over 92% of the genome, was announced in April 2003. Nevertheless, there were important gaps in the human reference genome due to difficulties in sequencing particular regions of DNA. These gaps have now been closed through the use of long-read DNA sequencing strategies by the T2T consortium[104] (T2T: telomere-to-telomere). In July 2021 the complete human genome sequence[105] was published, with the addition of 200 million base pairs closing existing gaps and identifying hundreds of new genes.

The chimpanzee genome was released in November 2004 and in 2010 a draft version of the Neandertal genome[106] was published, providing new insights into primate and human evolution.

Frank Vinken

Svante Pääbo admires a reconstructed Neandertal skull. Also termed Neanderthal, the name relates to the 1856 finding of fossil remains in the Neander valley (Tal or, at the time, Thal in German). The most recent of Neandertal fossil records date to about 30,000 years ago.

Eukaryote genomes are now being published regularly. Some of these can be accessed using the genome browser[107] from the Genomics Institute of the University of California, Santa Cruz. This short tutorial[108] on the browser is a useful introduction before you explore.

The pace of genomic research is ever- increasing and biologists have now entered the era of 'Big Data Science'. The Darwin Tree of Life Project[109] plans to sequence the genomes of all 70,000 eukaryote species found in Britain and Ireland. The Vertebrate Genome Project[110] has begun sequencing genomes of the 71,657 known mammal, bird, reptile, amphibian and fish species, recently completing the genomes of the Canada lynx, platypus and kakapo[111]. On an even grander scale, the Earth Biogenome Project[112] hopes to sequence genomes of the 1.5 million known eukaryote species on earth, over the next ten years.

A step-change in healthcare and our understanding of human natural variation is just around the corner. Genomics England[113] has recently entered into an agreement with sequencing company Illumina, to deliver 300,000 genomes over the next five years. On the continent, 22 European countries have committed to the 1+ Million Genomes Initiative (1+MG[114]) to give cross-border access to one million sequenced human genomes, as well as clinical and diagnostic information. In the USA, the National Institutes of Health (NIH) is funding the All of Us[115] program, in which one million volunteers will share their health information and DNA for whole genome sequencing.

The T2T consortium's pangenome project (mentioned above) aims to sequence, to the highest level of accuracy and completeness, the genomes of 350 individuals chosen as representative of populations from around the world. Together, these large projects will dramatically refine precision medicine, identifying health impacts of the many different types of variation[116] in the genome, such as SNVs, SNPs, CNVs, indels, deletions, duplications, inversions and translocations.

Looking further into the future, how will we store all this new sequence information alongside the growing exabytes of other digital data? In 2011 George Church demonstrated the feasibility of encoding of a book into a DNA sequence. Numerous research groups around the world are now actively exploring DNA as a digital storage medium, taking advantage of its ultra-high density, ultra-long-life and universal format. In the long-term, DNA-based digital information storage[117] may help address humanity's insatiable need for storage capacity.

But you might be wondering, if the Human Genome Project took 13 years to complete and the cost associated[118] with generating just the first human genome sequence was between $US 500 million and $US 1 billion dollars, how is the scale of these new projects even possible?

The precipitous decline in the raw cost of sequencing a human genome (in healthcare, there are significant additional costs associated with the bioinformatic analysis of data). Note the log scale on the Y-axis. For reference, the rate of change in the computer industry, in terms of miniaturization of transistors on computer chips, is given by Moore's law.

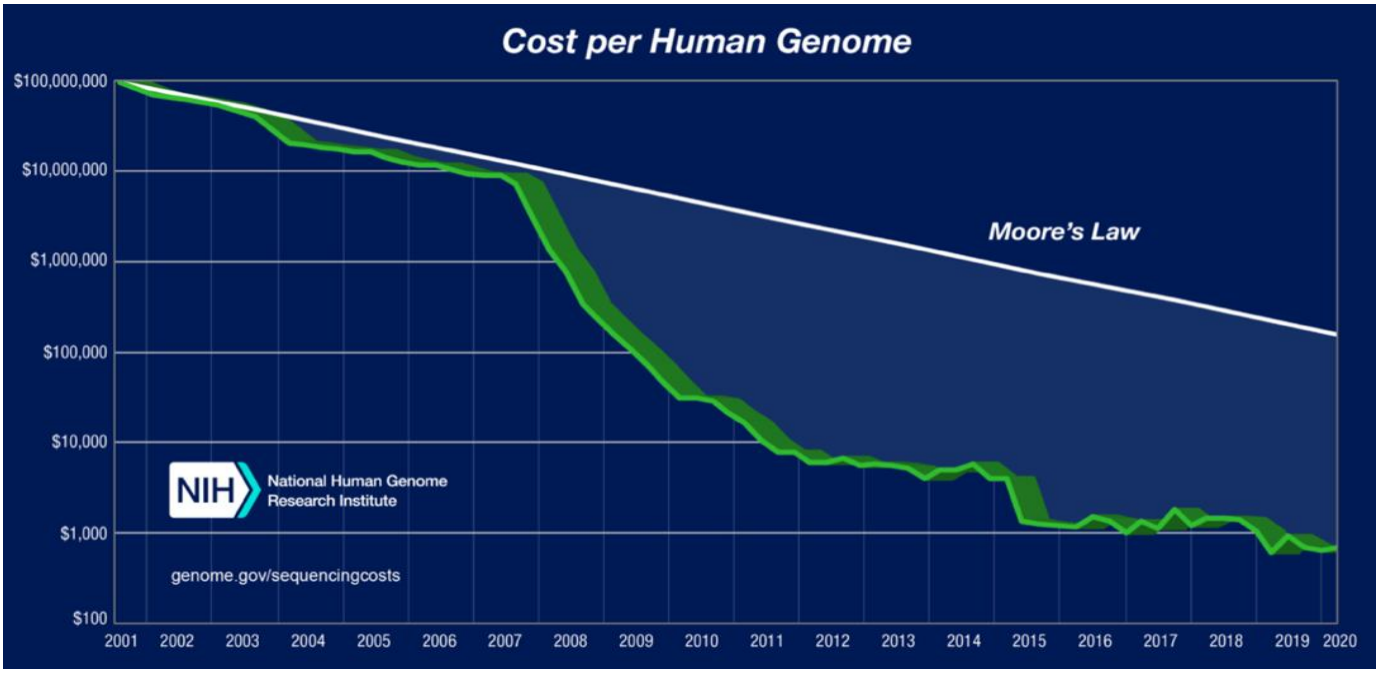
## Next-generation sequencing

If you are interested about the interface of technology and biology, the answer to the above question is that next-generation sequencing (NGS) techniques have rapidly driven down the cost of sequencing and vastly increased sequencing throughput. A human genome can now be sequenced for less than $1000 and major players in this area are touting advances toward the $100 genome[119].

Here, we briefly contrast two NGS sequencing platforms as examples of the transformative technologies now available. The majority of genome sequencing performed so far has utilized the Illumina sequencing-by-synthesis (SBS) platform. Although there are differences in the chemistry involved, this approach is the great-grandchild of Sanger sequencing, with the automated fluorescence-based sequencing developed by Lee Hood being a conceptual grandparent.

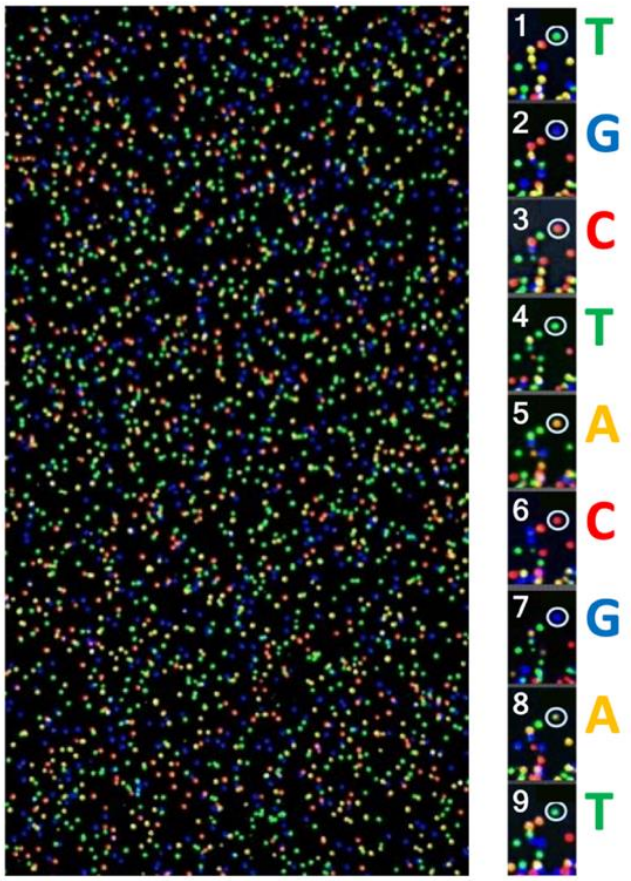
The quantum leap of the Illumina method (explained in this LEGO analogy[120]) is that gel electrophoresis is not required to identify fluorescently labeled DNA fragments of different lengths (to infer the order of nucleotides). Instead, each new fluorescent nucleotide added to the growing DNA strand is imaged at a fixed spot on a glass slide (formed by amplification of a DNA fragment), before adding the next fluorescent nucleotide.

The whole process is carried out in a massively parallel fashion, as approximately ten million clusters, in patterned flow cells[121], are present on



National Human Genome Research Institute, USA

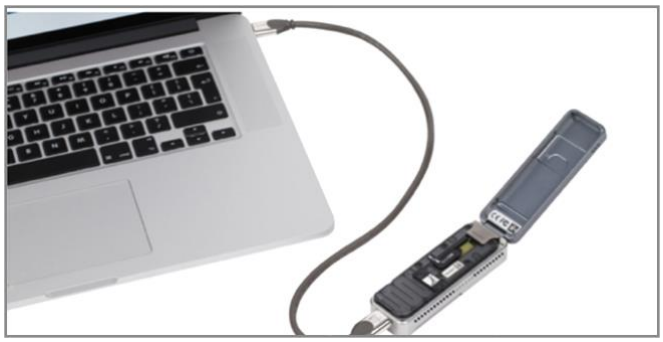
each square centimeter of slide. Although the DNA fragments sequenced on the Illumina platform are small (150-300 base pairs in length) the number of fragments that can be sequenced simultaneously is huge. The current state-of-the-art Illumina sequencing machine, a NovaSeq 6000, is capable of producing 20 billion sequence reads simultaneously. This is sufficient to sequence, in a single two-day sequencing run, forty-eight 'complete' human genomes.




Marcel Dinger

A small section of an earlier generation SBS slide (before patterned flow cells were introduced) showing fluorescent clusters of clonally-related DNA fragments on a glass slide. In the right-hand column, the identity of each sequential base in a cluster is read (base called) from the images taken across consecutive cycles of fluorescent nucleotide addition.

A radically different sequencing method has been developed by Oxford Nanopore Technologies[122] (ONT). Instead of sequencing-by-synthesis, ONT devices read out small fluctuations in an electric current as a single DNA strand is passed through small protein pores (100-3000 in number) embedded in a membrane. An algorithm identifies which nucleotide has passed through the pore at each point in time, based on these changes in current.



Oxford Nanopore Technologies

An Oxford Nanopore Technologies MinION device connected to a laptop, highlighting its portability. This is not even the smallest of ONT's sequencing platforms. In addition to being able to read long DNA sequences, nanopore technology is also able to directly sequence RNA and identify methylation and other epigenetic modifications of DNA, something not possible using the SBS approach.

In contrast to the stop-start nature of the Illumina SBS process (with an image taken after each nucleotide addition), DNA strands move through the nanopores at a rate of 400 bases/second. These strands can be very long, with single sequence reads of over one million bases in length being reported. Although the accuracy of nanopore sequencing is much lower than Illumina sequence reads, long sequence reads are helpful in closing gaps in genome sequence data. In terms of the breadth of applications, it is the portability of ONT sequencing platforms that is unmatched, facilitating genomics on African farms[123] and in tropical jungle[124], arctic[125], and extraterrestrial[126] environments.

Massimo Delledonne

An adult 'nanofrog' (to scale) from Tanzania, a subject of nanopore based sequencing in efforts to conserve such species.

## How will genomics affect New Zealand society?

You now have a good perspective on the many applications of genomics, its history and the technological progress that is driving this area of research forward. However, there are costs and benefits associated with the genomics revolution. In terms of ethical issues, perhaps the biggest conundrums relate to the impact of genomics on reproductive decisions. The path that leads from using genetic techniques to maximize the best life opportunities for a prospective baby could potentially lead down the slippery slope to society

sanctioning aspects of the philosophy of eugenics. However, in terms of *broadly-felt* societal impacts this may be an issue that the New Zealand public won't need to face until the latter half of the 21st century (or even further in the future).

Of more immediate concern, the healthcare and criminal justice systems are two areas where controversial impacts of genomics on the New Zealand public are already upon us. We are going to take a deeper dive into these two areas to explore the potential of genomic information to do both good and harm, raising some thorny issues that you may already have been contemplating.

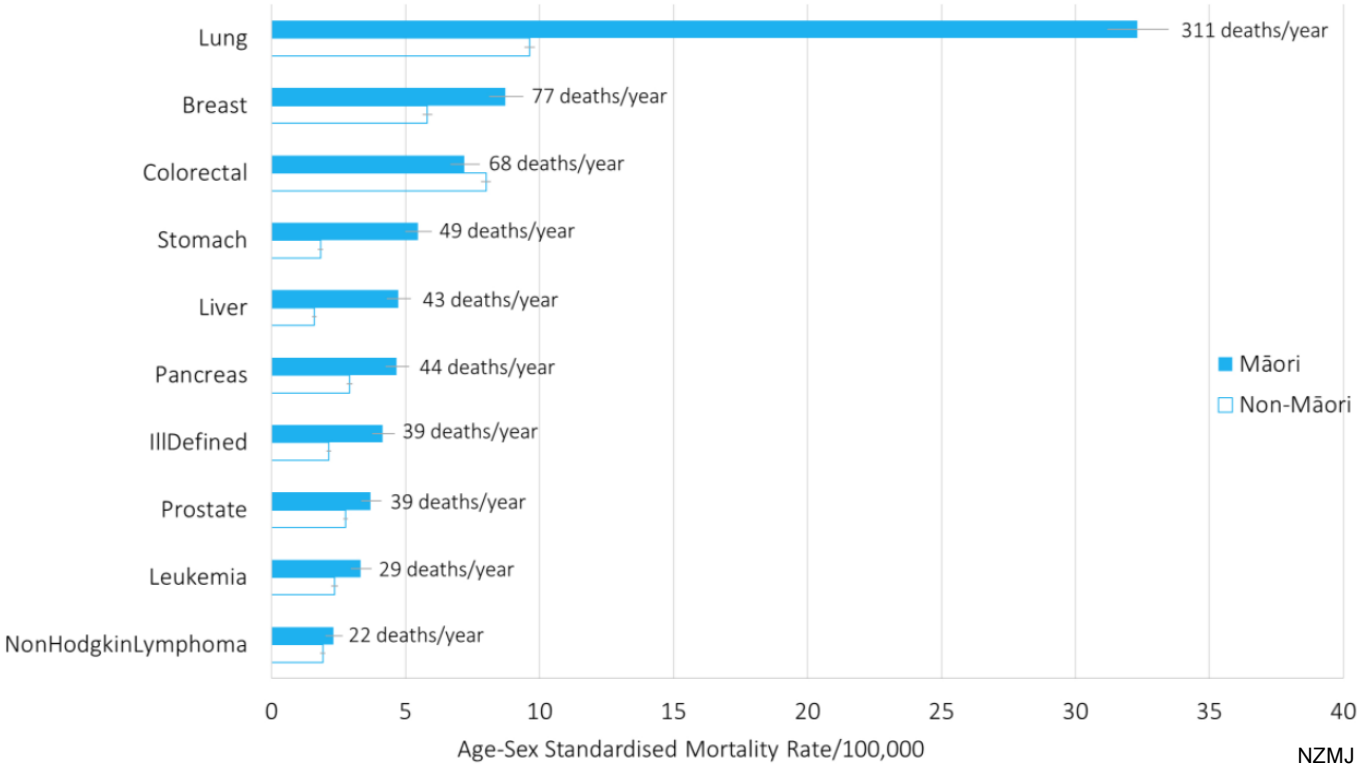
## Health equity & indigenous genomics

### *Health disparities in New Zealand*

Worldwide, there are systemic **health inequities**[127] between demographic groups within populations. In New Zealand this is apparent in the fact that Māori and Pacific peoples have, on average, a seven-year shorter life expectancy[128] than non-Māori, with even greater disparities reported for some provinces. Expressed another way, based on the rate of increase of life expectancy over time across the whole New Zealand population, this seven-year life expectancy gap corresponds to more than three decades worth of health improvement. The causes of this life expectancy difference between Māori and non-Māori are many, but a contributing factor is found in the way that different forms of racism[129] within the health care system impact individuals and communities.

Cultural insensitivities in health care have been an **ongoing problem**[130] for over a century. The New Zealand government passed the Tohunga Suppression Act in 1907 to curtail Māori usage of rongoā[131], including traditional medicinal use of plants, herbs and other natural products.
Nowadays there is a far greater appreciation by 'Western' medicine of the value of indigenous medical knowledge and there are efforts world-wide to harness traditional information to identify novel compounds of pharmaceutical interest (this partly motivated the Manuka genome sequencing project, noted earlier). Indeed, the Ministry of Health now funds a small number of rongoā Māori providers nationwide.

Colonization introduced new diseases to indigenous peoples, decimating their populations in New Zealand, Australia, the Pacific Islands, North America and elsewhere (and this was a backdrop and impetus to the Tohunga Suppression Act). Today, it is tobacco, alcohol, high-calorific diets, sedentary lifestyles and other




Cancer mortality as an example of just one aspect of health disparity in New Zealand. Of the ten most commonly diagnosed cancers among Māori, nine have a higher rate among Māori than non-Māori. Other health disparities that affect Māori and Pacific New Zealanders the most include obesity, diabetes and cardiovascular disease.

social determinants of health[132] that represent major contributors to disease risk across all sectors of New Zealand society, but which have hit Māori and Pasifika communities particularly hard. In addition to different cancers, Māori are two times more likely to die from cardiovascular disease[133] than non-Māori and have a two-fold greater prevalence of diabetes[134] and three-fold greater likelihood of lower limb amputation as a consequence of diabetes complication. Indigenous peoples[135] worldwide, including Māori and Pacific peoples[136], are also more likely to suffer the adverse consequences of COVID-19 infection.

## *Indigenous genomics*

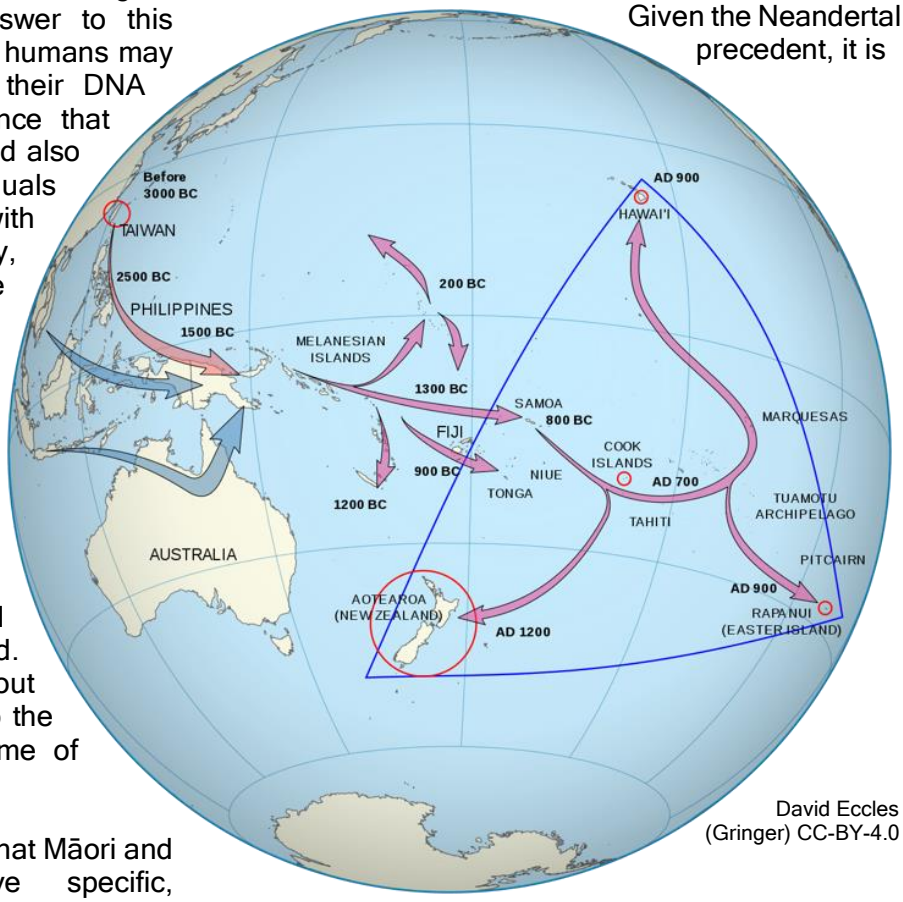
What is the place of genomics in health care for Māori and Pacific peoples? In many ways, genomics is the epitome of modern biomedical science and far removed from rongoā Māori, which has a holistic outlook incorporating rakau rongoā (medicinal preparations from native flora), mirimiri (massage) and karakia (prayer).

Can Māori and non-Māori benefit *equally* from humanity's newly developed understanding of our species' genome? The answer to this question is presently no. Although humans may generally be 99.9% identical in their DNA sequence, it is the 0.1% difference that underlays our individual identity and also our risk of disease, both as individuals and as members of groups with differing ancestral origins. Initially, the vast majority of human genome sequence information was obtained from European and North American Caucasian subjects. More recently, intensive sequencing of Asian peoples has added to this knowledge base. However, sequence data for indigenous peoples are extremely under-represented[137] in human genome databases, Māori and Pacific populations included. Therefore, crucial information about DNA variation and how it relates to the likelihood, progression and outcome of disease is missing for such groups.

What evidence is there to suggest that Māori and Pacific populations will have specific, consequential, DNA variations? Life histories of our remote ancestors are reflected in our genome through the process of natural selection. For example, the emergence of pastoral farming practices[138] in the Neolithic around 8-9 thousand years ago spurred the spread of DNA variations conferring lactase permanence[139] as a beneficial trait. Other DNA variations have resulted in changes to skin pigmentation[140] as human populations moved out of Africa into environments with lower incident UV light exposure.

Many genomic signatures of ancient life histories can be traced back to the long and rich evolutionary history of human interbreeding among divergent populations. Interbreeding between the ancestors of modern Europeans and Neandertals, 50-60 thousand years ago, has left a trace of ~2% of Neandertal DNA in the modern human genome. This legacy[141] may contribute to the risk of certain diseases, including skin keratosis, obesity, depression and COVID[142]. Genomic DNA from another archaic hominin, Denisovans[143], estimated to have occurred 50-15 thousand years ago, has similarly left a trace in the genome of modern humans.

Among 120 modern populations, traces of Denisovan DNA are most readily detected in Papuans, Aboriginal Australians and Bougainville Islanders, followed by Māori and Hawaiians. Given the Neandertal precedent, it is




David Eccles (Gringer) CC-BY-4.0

The diaspora of peoples across the Pacific Ocean to the 'Polynesian triangle', with approximate timings of arrival. The genomic impact of these migratory histories is still poorly understood.

not unreasonable to predict that DNA sequence variants associated with the Denisovans may contribute to disease risk among indigenous peoples of Oceania (despite these and other variants being potentially adaptive[144] in the past). In addition, the genomes of tūpuna (ancestors) will likely have acquired unique variants as well as shifted the frequency of common variants as a consequence of repeated founder effects and genetic drift during their diaspora across the Pacific, settlement of diverse geographic regions and, more latterly, exposure to the pressures of colonization (including introduced disease and reproduction with early European explorers, whalers and settlers).

SING Aotearoa

Summer interns at an annual week-long SING Aotearoa workshop that is open to Māori interested in the technical, cultural and ethical issues associated with genomic research.

## *Causes for concern and optimism*

Indigenous peoples have suffered a long history of abuse, neglect and dishonesty by the institutions of colonialism. Therefore, it is no surprise that indigenous groups harbor well-justified concerns, if not deep-seated fears, over the potential for genomic data to be misused. In the past, medical/science professionals from the (Caucasian) majority often regarded their research subjects as inferior, especially when from indigenous, minority or other disadvantaged groups. The infamous 40 year-long Tuskegee Study[145] on syphilis in African American men and a clandestine parallel study in Guatemala[146] are two tragic examples. Although the Tuskegee study ran into the 1970's, it clearly violated fundamental tenets of ethical research[147], such as voluntariness, informed consent and beneficence. These were laid out in the 10 principles of the 1947 Nuremberg Code[148], the development of which represents a momentous landmark in the history of the ethics of medical research.

To protect against such abuses of the Nuremberg code, in relation to indigenous peoples and genomic research, it is essential that there is full and honest engagement with indigenous communities. Moreover, concerted international efforts[149] are underway, such as the Summer Internship for Indigenous peoples in Genomics (SING[150]) programme, to build genomics expertise in these communities so that indigenous peoples can lead their own genomic research in ways that harmonize with their cultural beliefs and practices. In New Zealand, a major avenue for this training is via SING-Aotearoa[151].

There is a clear recognition[152] by New Zealand medical researchers and clinicians that genomics should be used to reduce pre-existing health (hauora) inequities in the nation (a more detailed discussion can be found here[153]). Government-funded projects are underway to develop the medical genomics infrastructure in New Zealand, characterize genetic variation in the New Zealand population and foster close collaboration between genomics researchers and Māori healthcare providers. This will enable primary care practitioners to make real-time genomically-informed decisions that put the needs of Māori front and centre[154]. The Genomics Aotearoa Variome Project[155] is guided by Māori experts in whakapapa (genealogies) who are versed in oral tribal histories and knowledgeable about the waka hourua (voyaging canoes) and tūpuna (ancestors) who travelled in them to Aotearoa. Knowing the affiliation of the descendants of the tūpuna who made these voyages ensures that the Variome Project is representative of the genomic diversity of Māori.

## *Māori cultural perspectives on genomics*

Nevertheless, tensions exist in reconciling Māori perspectives with practices foundational to international genomics research, such as rapid pre-publication sharing of genomic data (the Bermuda Principles[156]) and views on how genomic data should be regarded, obtained,

stored and protected. The deep cultural importance of whakapapa[157] to Māori intersects with the fact that human genome data does not just pertain to the individual being sequenced, but is shared with whānau (family), hapū (clan or descent group) and iwi (tribe). Therefore, in contrast to western/European notions, where emphasis is put on individual freedom and consent, Māori place more onus on collective consent[158], as do other indigenous peoples.

More broadly, whakapapa can be considered as extending to all lifeforms, in accordance with Māori cosmogony (origin stories), with Ranginui (sky father) and Papatūānuku (earth mother) being the primordial parents from whom all other gods, plants, animals and humans descend. In a sense then, we can draw a parallel between the thread of whakapapa and the DNA present in the genome of all organisms, again related by descent. Other core concepts are that of mauri – a life essence (although also applied to inanimate objects) and wairua – the non-physical spirit, that exists beyond death and is distinct from the body and the mauri. In this Māori worldview, DNA (and any human tissue sample) holds whakapapa, mauri and wairua and, as such, is a taonga (a precious object, resource or entity). It is therefore tapu, having a sacred or special nature demanding its protection.

Nicola Coburn

Rangitāne iwi members bury the nearly 730 year-old remains of a tupuna at the Wairau bar in 2016. The remains of tūpuna were removed by archaeologists, scientists, and even members of the public, beginning in 1939. The repatriations help to address this injustice. The iwi gave permission for researchers to sequence DNA obtained from bone, before their return to the ancestral burial site.

These beliefs and principles also inform nuanced opinions of Māori on gene editing[159] (i.e. using CRISPR). As with any group of people, not all Māori hold the same views on genomic research and some groups, such as the Rangitāne o Wairau iwi[160], have been more actively engaged in this area than others.

Māori have also voiced continuing discomfort over the acquisition and use of their genome data, as illustrated by one commentator's perspective[161] that concludes thus:

"*All biological samples including DNA and Genomes of species that whakapapa to Ranginui and Papatuanuku contains whakapapa, wairua and mauri, therefore are tapu and are a taonga. These genetic taonga should be recognized in Article 1 of the Treaty of Waitangi. If recognised as a Taonga, this would also give all New Zealanders protection that no one in the future will claim ownership of their DNA and genome data.*

*Often DNA and Genomes are digitized and turned into computer readable code. Despite a different appearance, this digital data is a taonga that still contains wairua, mauri and whakapapa. Therefore it is tapu. This is digital colonisation and Data Sovereignty concerns need to be considered in addition to customary Māori rights and beliefs*."

Being taonga and tapu, Māori wish to maintain stringent governance over the biological samples taken in genomics research, as well as the data produced. This has important implications for where the data is stored. Cloud computing and storage of genomics data (often in the petabyte range) is seen as inappropriate, as the servers that physically hold such data are not located within Aotearoa. Collaboration of New Zealand-based genomics researchers with international colleagues is also more difficult, as the sharing of genomics data could place a copy of taonga DNA sequence on a foreign computer.

Different individuals, Māori and non-Māori alike, will have personal views on questions of the privacy and control of genomic data. Should DTC ancestry testing or basic research on the genome of an individual require only that person's consent, or the collective consent of related persons? How should smaller benefits for the greater good be weighed against larger benefits at the individual or whānau level? For example, if an individual is concerned about their genetic risk of a particular disease, should they need to have the collective consent of more distant relatives (for example, at hapū or iwi level) before signing up for genomic testing? In the case of gout[162] and

stomach cancer[163] (also see link #150), for which there is increased genetic risk among some Māori and Pacific whānau, broader engagement and understanding has lessened the stigma associated with disease.

Nevertheless, some individuals would prefer not to know of their genetic risks. Relatedly, how should incidental findings[164] be handled in a situation where particular information that was *not* sought and of uncertain significance, but that could conceivably impact an individual or family's health?

Provocatively, similar conundrums about the transfer of genomic knowledge play out on a larger population scale. Māori should have kaitiakitanga (guardianship) of their taonga human genomic data, but some research findings may be directly applicable to other Oceania populations (perhaps even including through shared Denisovan genomic contributions). What rights should those groups have to Māori genomic data?

Ultimately, we all share whakapapa and novel variants or altered allele frequencies associated with disease in Māori may represent valuable information for understanding mechanisms of disease that would benefit diverse human populations worldwide. Other nations, such as those involved in the European 1+MG project, are leveraging cross-border sharing of genomic and clinical data to benefit their populations. The advances in knowledge these studies bring will also surely benefit all New Zealanders. How do we reconcile the need to protect our taonga human genomic data with notions of reciprocity for the 'greater' good?

How New Zealand society handles these and other socio-scientific dilemmas posed by medical genomics could impact the likelihood of achieving health equity for Māori and Pacific peoples in the future.

## Genomics and the criminal justice system

We are now going to explore how genomic data poses serious ethical questions for forensics, policing, sociology and law. In doing so, we will also draw on recent international developments as these could signal trends that other nations, including New Zealand, might follow.

Mylius (GFDL 1.2) Wikimedia Commons

Lady Justice (often blindfolded or with eyes closed) symbolises an impartial arbiter of the law, passing judgement without explicit or implicit bias. A question for society is whether genomics will help level the scales, or tip them further out of balance.

A trend that governments of many nations have followed is to gather genetic information for an ever-increasing proportion of their population. Hypothetically, how might such a trend occur in New Zealand? Recall, the raw cost of human genome sequencing has fallen dramatically over the past two decades and the $200 genome may be in the not-too-distant future. In fact, because of the predictive health advantages of having genome sequence data from birth (e.g. knowing from the outset a person's risk for childhood obesity, cardiovascular disease, cancer,

addiction, schizophrenia or Alzheimer's), population-wide genomic testing may become economically desirable for the healthcare system. If (perhaps when) such time comes, this might be incorporated into the standard practice of taking a blood sample for newborn metabolic disorder screening, the Guthrie test[165]. In the United States[166] and United Kingdom[167], pilot programs have been initiated to explore how parents, and society at large, will cope with the information obtained by whole genome sequencing of newborns as a population screening approach. Understandably, this possibility is evoking quite strong positive[168] and negative[169] sentiments. Two very important concerns are how to protect New Zealand families from currently possible genetic discrimination (for example, by the health/life insurance industry[170]) and whether the government will use genetic data obtained from newborn population screening in other contexts, such as in the criminal justice system.

Eric Sheler - USAF Photographic Archives

The Guthrie test is routinely taken with newborns and involves a small heel prick and collection of blood spots on a card. The sample is then tested for a small panel of rare but potentially fatal diseases.

## *Genomics in forensics and policing*

The use of Guthrie test card DNA by police is tightly regulated and used as a last resort, but has come under recent scrutiny[171]. In particular, parents of children at the centre of criminal investigations (either as victims or perpetrators) were not informed of the potential for forensic usage of their children's Guthrie cards. As quoted in this linked news piece, a civil rights lawyer observed:

"*The police will come back and argue that this is for a very important purpose, we're identifying a deceased person and something like that……the trouble with that is that once you go down that path as a reason for basically over-riding informed-consent provisions, where does it stop?*".

This is perhaps especially problematic for Māori who view DNA as taonga and tapu, as the provision of a newborns' blood sample would be considered tākoha, a gift accompanied by special responsibilities to look after it. In these cases of Guthrie card DNA used for a purpose the parents were unaware of, it is also possible to draw a parallel to the case of Henrietta Lacks[172] and to situations that have angered indigenous peoples, such as the Havasupai[173] of Arizona and Nuu-chah-nulth[174] of Canada, who participated in research studies on diabetes and rheumatoid arthritis that afflicted their respective tribes, only to find their DNA was later used for other purposes without consent.

It is reasonable to imagine that the demands of forensic investigation and policing on the one hand, and privacy and civil liberties on the other, would come into increasing conflict if newborn DNA was routinely sequenced under the guise of societal health benefits.

Changes in policing strategies and investigative approaches in New Zealand will be influenced by standard operating procedures overseas. Some governments, most notably China, are already making use of genomic analysis to keep tabs on their population. The Chinese government initially applied large-scale genetic surveillance to the minority and predominantly Muslim population of Uighurs[175] in Xinjiang province. Moreover, efforts are underway to broaden the reach[176] of the national forensic DNA database by collecting Y-chromosome information from 10% of its male population (i.e. around 70 million boys and men). Because of the pattern of Y-chromosome inheritance this information will be sufficient to construct links to the entire male population of 700 million.

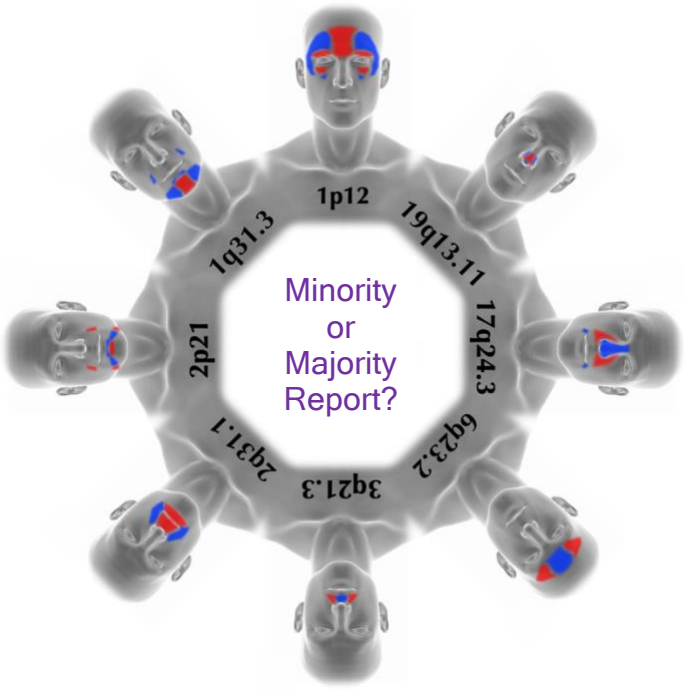
New Zealand is unlikely to follow China's example in the foreseeable future, but might be swayed by trends in its partner countries of the Five Eyes intelligence alliance; namely Australia, Canada, the United Kingdom and the United States. The U.K. and U.S.A. have amassed millions of samples in their own forensic DNA databases. In the UK, the national forensic database holds data from both suspects and convicted criminals and has genetic information on an estimated 1 in 8 males between the ages of 15 and 50 years. In the United States, state and federal databases contain genetic profiles of more than 17 million arrestees, immigrant detainees and asylum seekers.

Furthermore, assorted public and private databases accessible to authorities contain

genetic data on tens of millions of patients, consumers (through DTC genetics) and research participants. The rich information and large kindreds[177] identifiable through DTC genetic testing companies were used to identify the Golden State Killer (as mentioned earlier), as well as solving many subsequent cold-cases through forensic genealogy[178].

That a large proportion of individuals within a society already may be now identifiable through the information in such databases raises serious genetic privacy issues[179]. Indeed, this has turned the conversation in some quarters to debate the merits of police having a universal DNA database for all citizens of a country. These calls for a universal DNA database currently focus on just a handful of forensic genetic markers that would not reveal other sensitive patient medical information. On the positive side of the argument, bioethicist James Hazel proposes that a universal DNA database[180] would help to reduce racial profiling, discrimination and wrongful conviction. Such a database for law enforcement might also help reduce criminal behaviour by changing the risk-reward calculus[181] of prospective offenders.

For law enforcement, the temptation would be to include in a universal database those genetic markers that can be used to reconstruct a facial profile[182] (and other phenotypic traits).




Superstock / Universal Images Group (for head/neck model)

SNPs at the chromosomal positions shown (e.g. 1p12) associate with protrusion (red) or depression (blue) in discrete facial locations relative to an averaged face. The image is based on a genome-wide analysis of the genetics of facial shape[183].

Facial phenotyping from DNA data is an active area of research in Chinese[184] and American[185] labs. Perhaps universal DNA databases will ultimately end up sitting alongside sophisticated genome-to-phenome (and vice versa) prediction approaches, as just some of the suite of tools in the surveillance state armamentarium (admittedly a dystopian view).

### *Genomics impacts on sociology & law*

The forensic genealogy approaches mentioned above depend on whakapapa and so raise a red flag for Māori who, as a group, are already over-represented in the national "known person" databank. Provided unconscious bias in policing[186] is mitigated, a universal DNA database could lessen the disproportionate impact on Māori of familial searches.

Unconscious bias is pervasive. In the most culturally fraught episode of genetics research in New Zealand, the 'warrior gene' controversy[187], media coverage played on just such a bias in the minds of the New Zealand public (a more detailed account of the controversy and its implications can be found here[188]). The media storm swirled around the work of two researchers who were studying allele frequencies of the monoamine oxidase A (*MAOA*) gene in Māori and non-Māori participants, motivated by the association of MAOA activity with risk of addiction. This is an important public health issue as epidemiological data reveals that Māori have a ~3-fold increased lung cancer mortality rate compared with non-Māori (as shown in the chart on page 13) and also a 3-fold greater likelihood of adult tobacco use (population health data, including tobacco use, can be found on the Ministry of Health data explorer data explorer[189]).

The researchers found that Māori had a higher frequency of an allele associated with low MAOA activity than did non-Māori (it should be noted that this low *MAOA* allele is even more frequent in Chinese). Reduced MAOA activity in the brain results in increased levels of important monoamine neurotransmitters, such as serotonin, dopamine and epinephrine, providing the neurochemical substrate of *MAOA*'s complex contributions to behaviour.

However, in the 1990's, the *MAOA* gene had also been implicated in aggressive behaviour and criminality in a large Dutch family (a rare mutation resulting in Brunner Syndrome). Many subsequent studies, including in animal models, have confirmed a role for *MAOA* in modulating

aggressive behaviour and the gene became referred to in the press as 'the warrior gene'.

Unfortunately, the scientists involved in the New Zealand study speculated that the increased frequency of the low *MAOA* activity allele in the Māori population (based on a very small sample size) could relate to a hypothetical adaptive advantage of warrior-like attributes during the spread of peoples across the Pacific (for example, in competition for island resources).

Photosport

A cherished sporting tradition, seen by many New Zealand kids, perhaps also shapes subconscious perception of the nature of Māori from a young age. Moreover, the most frequently used haka by the All Blacks, Ka Mate, refers to an episode in the conflict between two powerful iwi. Rugby league sends a similar message with the New Zealand Warriors.

The colloquial name for the gene fed into racial stereotypes of Māori having a war-like heritage and public perceptions fueled by the over-representation[190] of Māori in the criminal justice system. Products of New Zealand art/culture also contribute, such as the acclaimed book/movie *Once Were Warrior*s. However, as eloquently put by Moana Jackson[191], Māori once were gardeners (and many still are).

The relationship of genetic variations to brain and behaviour are still poorly understood and the effects of MAOA, via the neurotransmitter levels it modulates, are likely to be diverse and dependent on social and environmental contexts. The simplistic 'warrior gene' hypothesis did considerable harm, including to the way that many Māori view genetic and genomic research.

An equally speculative but less incendiary Just-So-Story[192] for the increased frequency of the low *MAOA* allele in Māori can be entertained. The MAO-A enzyme reduces neuro-modulator levels, including that of dopamine which is involved in risk-reward behaviour. A study of Han Chinese responses to an economic task found that individuals with the low activity *MAOA* allele are *less* likely to take longshot risks compared to those with the high activity allele. Making the assumption this is generalizable to real-life situations of survival, one imagines the high activity *MAOA* allele and a tendency to take impulsive longshot risks would be selected *against* in the early peopling of the Pacific.

In fact, weighing and mitigating risks associated with island resources is critical. Evidence points to agricultural risk-management practices on Hawaii, dating to at least 1450AD. Historically, tūpuna would also face situations involving the decision of whether to remain on an island or to voyage - both options entailing risk. Economic experiments[193] reveal that, under conditions of risk, carriers of the low activity *MAOA* allele more frequently make optimal decisions.

Therefore, tūpuna on their waka hourua would have been skilled in risk assessment and optimal decision-making, and as expert oceanic wayfarers would likely have arrived on Aotearoa's shores in comparatively good health, contrary to some artistic depictions of the nineteenth century.

Greg Semu's *The Raft of the Tagata Pasifika*[194] references Théodore Géricault's 1819 painting, *The Raft of the Medusa*[195] (depicting survivors of the French shipwreck, *The Medusa*) and is also intended as a parody of Goldie and Steele's 1898 painting *The Arrival of the Maoris in New Zealand*[196]. The portrayal of Māori in Goldie and Steele's *The Arrival* has been criticised for having an implicit racial bias and greatly underestimating their prowess as maritime wayfarers.

*The Arrival* (2015) diptych: Greg Semu[197]

The 'warrior gene' controversy thus provides a cardinal lesson in the dangers of seeking simple explanations to connect complexities of the genome to complexities of human behaviour. Nevertheless, one lesson that has come out of the extensive international studies on the *MAOA* gene, and particularly the famed **Dunedin Study**[198], is that any *MAOA* contribution to aggressive/antisocial behaviours and even anger proneness in infants is moderated through gene-by-environment interactions[199], GxE).

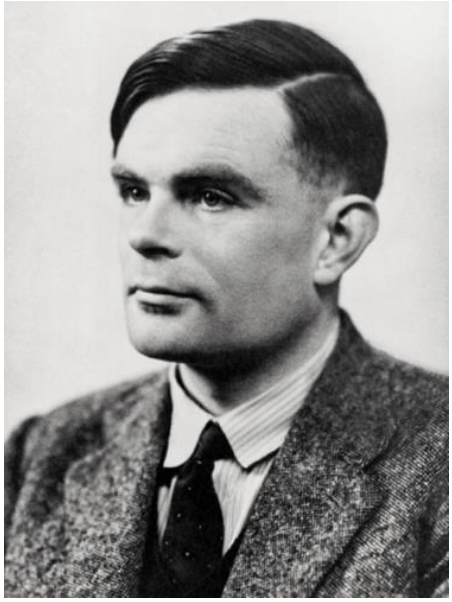
The molecular mechanisms of this are coming into view, as *MAOA* gene expression is regulated by a long non-coding RNA (lncRNA) that is itself regulated by epigenetic mechanisms such as DNA methylation and chromatin modification. Thus, early life stress (such as experience of abuse) result in epigenetic changes that lead to reduced *MAOA* expression. These effects may be compounded in individuals with a genotype conferring low MAOA activity, increasing the risk of antisocial behaviour in later life.

A large study[200] of 900 violent crime offenders in Finland has added further support to the association of *MAOA* with violent crime, and implicated another gene, cadherin 13 (CDH13), which also impacts serotonin signaling in the developing and adult brain. More recently, researchers have begun to examine polygenic contributions to antisocial behaviour[201] (a research paper based, in part, on results of the Dunedin Study can be viewed here[202]). Polygenic risk scores for antisocial behaviour will become more robust as the larger genomic datasets mentioned on page 10 become available. We also might expect that, by analogy with cancer, background variation[203] in the genome that confers a low polygenic risk score for antisocial behaviour might also *offset* risk associated with carrying an allele for low *MAOA* activity.

The future will bring a much deeper understanding of polygenic GxE interactions and the risk of criminal behaviour. Epigenetic changes due to early life stressors are, like genetics, beyond the control of a person who might come before the criminal justice system. Should a risk score that captures both genetic and epigenetic influences be taken into account during sentencing? If one is developed, should a future convict's polygenic risk score for recidivism be considered when up for parole?

Genomics and epigenomics are now posing these and many other profound questions for society about how the criminal justice system should adjust to this new information. Most notably: to what extent do we have free will and are morally and legally culpable[204] for our actions?

Although a false equivalence, it is illustrative to consider how New Zealand's laws and public perception have changed with respect to homosexuality. In 1840, the maximum sentence for same-sex male intercourse was the death penalty, but this was moderated such that in 1893 any sexual activity between males carried a maximum sentence of life imprisonment. Only in 1986 was consensual sex between males decriminalized.



Granger NYC / The Granger Collection

Alan Turing[205], one of the most consequential intellectual figures of the 20th Century, elected to undergo chemical castration in 1952 rather than go to prison for admitting to a homosexual relationship. He died two years later at the age of 42. The British Government posthumously apologised in 2009.

Many LGBTQ people regard their sexuality as not chosen as an act of free will, but instead an ineluctable part of them. Recent data indicates an influence of genetics on sexuality[206], although only accounting for a small percent of same-sex sexual behaviour, with environment playing the greater role[207]. The *in utero* environment is a likely major contributor, at least for gay males, due to the fraternal birth order effect (FBOE[208]). Thus, GxE interactions might influence the development of particular brain regions[209], predisposing an individual to homosexuality.

Of course, in contrast to consensual homosexuality, violent crime is an act that can result in death or lasting physical and psychological harm. So the courts will continue to impose significant punitive sentences for such

offences. Nevertheless, in some cases a predilection for violent behaviour can be traced to genetic and environmental (epigenetic) factors that together affect the development and function of the psychopathic brain[210].

Wikimedia Commons

American axe murderer Lizzie Borden, whose slaying of her father and step-mother in a town near Boston in 1892 captivated the nation and was immortalized in the children's skipping rope rhyme:

> *Lizzie Borden took an axe*
> *And gave her mother forty whacks.*
> *When she saw what she had done,*
> *She gave her father forty-one.*

Conversely, even in individuals with a family history of violent murder, such as the kindred of Lizzie Borden (including the neuroscientist in the video linked above), the risk of violent behaviour may be lessened by a supportive early environment, arguing for social support networks in society to preemptively mitigate psychopathy.

*How do you think New Zealand society and our criminal justice system should respond to the coming tide of socio-genomic knowledge?*

# Technical glossary

**Allele/allelic**: A variant version of a gene. In typical classroom Mendelian genetics discussed as two versions, one dominant, one recessive. In reality most genes have many more than two alleles and these exhibit far more blended interactions than the textbook dominant-recessive case.

**Amniocentesis**: A procedure in which a hollow needle is inserted through the abdominal wall and into the uterus to obtain amniotic fluid (used to assess the sex and chromosomal status of the fetus).

**Aneuploidy**: The presence of abnormal numbers of chromosomes in a cell (either too many or too few).

**Anguilliformes**: The taxonomic order of ray-finned fish to which eels belong.

**Apiculture**: Technical term for beekeeping.

**Bacteriophage (e.g. Lambda)**: A virus that infects bacteria (often shortened to phage). Bacteriophage means 'bacteria eating'. Have a massive impact on global carbon cycling by killing about 20% of the bacteria in the oceans each day.

**Cadherin (e.g. CDH13)**: The name of this superfamily of proteins/genes comes from 'calcium-dependent adhesion'. They regulate many aspects of cell-to-cell contact and communication.

**Capillary gel electrophoresis**: A method to separate charged molecules (such as DNA) by their ability to move along a fine tube of a porous gel (often a cross-linked polyacrylamide matrix) in an electric field.

**Chromatin**: The macromolecular complex of the genome, including chromosomal DNA, proteins (of which, histones are a major contribution) and RNA.

**CNV/copy number variation**: CNVs are a type of structural variant, of between 1000 and five million base pairs in length, involving alterations in the number of copies of specific regions of DNA, which can either be deleted or duplicated. A CNV may include from zero to many tens of genes. Several thousand CNVs are known, comprising over 10% of the genome and contributing to human variation.

**CRISPR**: Originally a mechanism used by bacteria to protect themselves from bacteriophages, this is now used as a genetic engineering tool to make highly selective changes to DNA sequence in a genome.

**Denisovan**: An extinct archaic human (sub)species that inhabited Asia during the Lower and Middle Paleolithic, about 200,000 to 50,000 years ago.

**Dideoxy chain termination**: The basis for the DNA sequencing method developed by Frederick Sanger. Incorporation of a dideoxynucleotide by DNA polymerase during DNA strand synthesis causes the process to stop as the dideoxynucleotide lacks a free hydroxyl group on the 3′ carbon of the sugar, necessary for the addition of the next base.

**DNA libraries**: A large collection of DNA fragments, together perhaps covering a whole genome, each of which is stored in a DNA vector (either a plasmid or a bacteriophage genome) and therefore able to be replicated in bacteria.

**DNA scaffold**: A framework for the alignment of a large number of DNA sequencing 'reads'. In the case mentioned in this resource, the known organization of the Emu genome was used as a scaffold to figure out how to assemble small fragments of DNA sequence in order to produce a rough Moa genome sequence.

**Dopamine**: An important neurotransmitter and neuromodulator that is particularly involved in brain circuits for reward and motor behaviours.

**DTC genetics**: Direct-to-consumer genetics. Most often provided by companies advertising genealogy services, such as Ancestry.com and 23andMe. Genetic information is provided to consumers without the involvement of health care professionals.

**Endosymbiotic**: A symbiotic relationship between organisms where one lives inside the other. Proposed by Lynn Margulis to be the mechanism that gave rise to the evolution of eukaryotes, through the acquisition of endosymbionts that evolved to become mitochondria and chloroplasts.

**Epigenetics**: Means 'above' or 'on top of' genetics. It is the study of those changes in gene function that are heritable but not attributed to alterations of the DNA sequence. Epigenetics is particularly concerned with how non-genetic influences (e.g. environmental factors) affect gene expression.

**Epinephrine**: Like dopamine, a monoamine neuromodulator. Epinephrine (also called adrenaline) is a hormone secreted by the adrenal gland and regulates the activity of the sympathetic nervous system, responsible for the body's 'fight or flight' response.

**Exabyte**: A unit of digital storage equivalent to one million terabytes or 1 billion gigabytes.

**Founder effect**: A reduction in genetic variation (a bottleneck) that results when a small subset of a large population is used to establish a new population. The latter may be very different from the original in terms of its genotypes and phenotypes (for example, including disease risk).

**Gene drive**: A genetic engineering technique (often based on CRISPR) that results in preferential inheritance of a specific allele or suite of genes throughout a population.

**Genetic drift**: A change in the frequency of an existing gene variant (allele) in a population over generations due to chance differences. Therefore genetic drift exerts its strongest effects in small populations.

**GxE/ gene-by-environment**: Where the effect of genes depends on the environment and/or the effect of the environment depends on the genotype. Two different genotypes may respond to environmental circumstances in different ways.

**Hominin**: A member of the zoological "tribe" Hominini (family Hominidae, order Primates), of which only *Homo sapiens sapiens* exists today. Extinct hominins include - in order of appearance - *Australopithecus*, *Homo erectus*, *Homo habilis*, *Homo sapiens neanderthalensis* and *Homo sapiens denisova*.

**Indels**: Indels are smaller scale versions of CNVs, where the insertions and deletions (hence, indels) are defined as being <1000 bp in length.

**Inversions**: Chromosomal inversions are (sometimes very large) rearrangements that occur if two breaks occur in one chromosome, with the region between the breaks rotating 180 degrees before rejoining with the two end fragments.

**Karyotype**: A karyotype is an individual's collection of chromosomes, or the picture of those chromosomes from which the number of chromosomes and sex of the individual can be determined.

**Keratosis**: A growth on the skin involving the structural protein keratin which is produced by keratinocyte cells within the epidermis.

**lncRNA/long non-coding RNA**: lncRNAs are RNA transcripts that are longer than 200 nucleotides in length and do not code for any proteins. Over recent years, these have become recognized as key regulators of tissue-specific gene expression. There are between 30,000 and 60,000 human lncRNAs.

**MAO-A/monoamine oxidase A**: A mitochondrial enzyme that catalyzes the deamination of the monoamine neurotransmitters dopamine, serotonin and norepinephrine. Low MAO-A activity can lead to elevated levels of these neurotransmitters.

**Metagenomics**: Metagenomics is the study of genomes from a mixed community of organisms. Very often, metagenomics refers to the study of microbial communities, either in the environment or within larger organisms.

**Methylation**: Broadly speaking, the attachment of a methyl group ($CH_3$) to a molecule. In the context of genomics, methylation of DNA and histones plays an important role in epigenetic regulation of gene expression.

**Microbiome**: The aggregate of all the microorganisms in a particular environment (including the body or a part of the body). Bacteria, archaea, fungi, algae, and small protists are included in the microbiome, but there is debate over whether the term should include plasmids and viruses (phages) which are not considered as living.

**Microsatellite**: A microsatellite (or short tandem repeat: STR) is a tract of repetitive DNA sequence of two to six base pairs in length, that is repeated typically 5–50 times. Microsatellite repeats have a high mutation rate and occur at thousands of locations across the genome, making them useful markers for forensic DNA fingerprinting.

**Monoamine neurotransmitters**: A class of neurochemicals that contain one amino group connected to an aromatic ring. Examples are dopamine, norepinephrine and serotonin.

**Monogenic**: Pertaining to a single gene. Monogenic disorders represent only a small minority of human non-infectious diseases. Most phenotypic traits and important human diseases are highly polygenic.

**NIPT**: Noninvasive prenatal testing. In contrast to amniocentesis, NIPT is carried out by obtaining a small sample of maternal blood. This contains circulating traces of fetal genomic DNA that can be tested for certain genetic abnormalities.

**Pangenome**: In metagenomics, the pangenome is the entire gene set of all strains of a species. In this resource, the human pangenome project refers to the T2T consortium's effort to fully sequence the genome of 350 individuals that represent human diversity across the globe.

**PCR/polymerase chain reaction**: A technique that uses sequential cycles of DNA polymerase activity to exponentially amplify small traces of DNA into amounts that can be used for sequencing. Although used widely in molecular biology laboratories, two prominent applications include forensic analysis and sequencing of archaic DNA.

**Petabyte**: A unit of digital DNA storage (approximately 1000 times smaller than an exabyte and 1000 times larger than a terabyte).

**PGD/preimplantation genetic diagnosis**: When genetic profiling is performed on the oocyte polar body or an early-stage embryo (cleavage stage or blastocyst) during assisted reproductive procedures.

**Phenylalanine**: One of 20 common amino acids that are incorporated into proteins during the process of translation. Phenylalanine incorporation is specified by the codons UUU and UUC, so an RNA homopolymer of uracil residues is translated by ribosomes into a polypeptide string of phenylalanines (The basis for a key finding by Nirenberg and Matthaei when unraveling the codon code).

**Plasmid**: A small circular extrachromosomal DNA molecule in the cytoplasm of a bacterium, capable of replicating independently of the cell's chromosomes.

Plasmids have been used extensively in molecular biology to propagate pieces of foreign DNA (for example, in DNA libraries). Plasmids also carry genes for antibiotic resistance and can be transferred between bacteria.

**Polygenic score**: Most human traits and disease risk is polygenic in origin. A polygenic score sums the estimated effect of many genetic variants on an individual's phenotype. Each of the trait-associated alleles typically contributes only a small amount to the overall score.

**Polymorphic markers**: A polymorphic genetic marker is an alternative DNA sequence or allele that occurs a a frequency of greater than 1% in the population. These can take several forms, the most common being single nucleotide polymorphisms (SNPs) and short tandem repeat polymorphisms (STRPs)

**Precision medicine**: A model of medicine where medical decisions, practices and treatments are tailored to a subgroup of patients based on genetic evidence, instead of a one-drug-fits-all model.

**Recombinant DNA**: DNA that has been formed artificially by combining fragments of sequence from different organisms. For example, this could include human genomic DNA sequences joined with plasmid DNA to create a human genomic DNA library that can be stored in bacteria.

**Restriction enzyme (e.g. EcoR1)**: A class of enzyme produced by bacteria, that cuts DNA at a specific short sequence (a recognition site, typically 4-8 base pairs in length). Frequently used in the production of recombinant DNA libraries (see above).

**SBS/sequencing-by-synthesis**: An umbrella term for many DNA sequencing techniques that, like the original Sanger technique, require DNA polymerase activity to synthesize DNA during the sequencing reaction. This approach notably contrasts with the nanopore-based sequencing approach.

**Semiconservative replication**: During cell division, each strand of the double helix acts as a template for synthesis of a new strand, resulting in replication of the original DNA sequence (a mechanism confirmed by Meselson and Stahl's famous experiment).

**Serotonin**: A monoamine neurotransmitter (like dopamine and epinephrine). Has multifaceted functions throughout the body. In the brain it modulates neural circuits involved in mood, cognition, reward, learning and memory.

**SNP**: Single Nucleotide Polymorphism. A sequence variation at a single base pair in the genome, for which the frequency of that particular variant in the population occurs at greater than 1%.

**SNV**: Single Nucleotide Variation. The less frequent version of the above - A sequence variation at a single base pair in the genome, for which the frequency of that particular variant occurs in fewer than 1% of the population.

**Spindle nuclear transfer**: An assistive reproductive micromanipulation technique that allows transfer of the cell spindle apparatus, with maternal chromosomes attached, into an unfertilized oocyte that has its nuclear material removed. Used in three-parent IVF when the mother carries a lethal mitochondrial disorder (the recipient oocyte's mitochondria are normal, avoiding the lethality).

**STR/short tandem repeat**: See the above entry under Microsatellite.

**Telomere**: A region of repetitive DNA sequence that occurs at the ends of linear chromosomes. Telomere repeats have a role in protecting chromosomal integrity. Repeats shorten over the course of many cell divisions and this may contribute to the aging process (search online for the Hayflick Limit).

**Translocations**: A genetic change in which a piece of one chromosome breaks off and attaches to another, often occurring in a reciprocal fashion and observable in a person's karyotype.

**Trisomy**: Where an extra copy of a chromosome is present in the cell nuclei, causing developmental abnormalities (most commonly chromosome 21, causing Down syndrome). As with translocations (see above) observable in a karyotype.

**Uracil**: One of the four nucleobases of RNA, it is a demethylated form of thymine and used as a replacement in RNA for the positions that thymine occurs in the corresponding DNA sequence.

**Variome**: The variome is the whole set of genetic variations found in a population or group. Genomics Aotearoa has a major ongoing project to define the variome of Māori, as this will provide essential information for precision medicine in New Zealand.

**X-ray crystallography**: An approach to identify the structure of complex biological macromolecules when in a crystalline form, by examining the diffraction pattern of X-rays that hit regularly-spaced atoms.

**Zoonotic**: A disease that normally exists in animals, but that can be transmitted to humans. COVID19 / SARS-CoV-2 being a notable recent example.

## Referenced TinyURL links:

Below, we have listed TinyURLs that correspond to the links in the text. They have all been verified as safe and functional (as of November 2021). However, if you wish to inspect the primary URL that the Tinyurl refers to (good practice if you do not trust a link sent to you), then between the "http://" and the "tinyurl," type preview. For example, applying this to the first TinyURL listed below gives:

https://preview.tinyurl.com/3DGenomeOrganisation

The TinyURL website reveals that this TinyURL redirects to:

https://www.cam.ac.uk/research/news/visualising-the-genome-researchers-create-first-3d-structures-of-active-dna

This can also be a helpful strategy if one of the links in this resource gets broken. You might be able to find the intended page by exploring the web domain the TinyURL was directing you to.

1. https://tinyurl.com/3DGenomeOrganisation
2. https://tinyurl.com/Salk3Dchromatin-video
3. https://tinyurl.com/SydneyBrennerObit
4. https://tinyurl.com/FuguMolamola
5. https://tinyurl.com/Endosymb-video
6. https://tinyurl.com/BadNewsCoronaPackage
7. https://tinyurl.com/HMXgenome-video
8. https://tinyurl.com/BoldGenomicPredictions
9. https://tinyurl.com/HygienicBees
10. https://tinyurl.com/CarbonHungryTrees
11. https://tinyurl.com/GeologicalGenomics
12. https://tinyurl.com/IrishGodKings
13. https://tinyurl.com/AstroDNA
14. https://tinyurl.com/MarathonMicrobiome
15. https://tinyurl.com/PolygenicEducation
16. https://tinyurl.com/PGRS-intro
17. https://tinyurl.com/Agrigenomics-video
18. https://tinyurl.com/HGPgateway
19. https://tinyurl.com/CeleraRace
20. https://tinyurl.com/RobertsonVideo
21. https://tinyurl.com/CFZsyndromeVideo
22. https://tinyurl.com/PolygenicScoreIntro
23. https://tinyurl.com/EbolaZikaSeq
24. https://tinyurl.com/ZoonoticSarsCoV2
25. https://tinyurl.com/CoV2BatPangolin
26. https://tinyurl.com/Covid19genomics
27. https://tinyurl.com/ExploreCovidClades
28. https://tinyurl.com/CIBSactNZ
29. https://tinyurl.com/HairInjustice
30. https://tinyurl.com/AlecJeffreysBio
31. https://tinyurl.com/ProfilingMethod
32. https://tinyurl.com/ForensicsESRbooklet
33. https://tinyurl.com/McKinnelMurder
34. https://tinyurl.com/NZcoldcasehistory
35. https://tinyurl.com/GEDmatchGSK
36. https://tinyurl.com/NthIslandBrownKiwiGenome
37. https://tinyurl.com/KakapoGenomics
38. https://tinyurl.com/TuataraGenomics
39. https://tinyurl.com/LittleBushMoaGenomics
40. https://tinyurl.com/DeExtinctionDilemma
41. https://tinyurl.com/Moa3Dskeleton
42. https://tinyurl.com/NZPests4GeneDrive
43. https://tinyurl.com/NZgenedrives
44. https://tinyurl.com/DOCandGenetics
45. https://tinyurl.com/DrivenMossie
46. https://tinyurl.com/IndigenousConservation
47. https://tinyurl.com/NgatiwaiTuataraGenome
48. https://tinyurl.com/WhakapapaGenomes
49. https://tinyurl.com/StreamMetagenomics
50. https://tinyurl.com/eDNAofSea
51. https://tinyurl.com/NessieSearch
52. https://tinyurl.com/LochNessReport
53. https://tinyurl.com/BasmatiRiceGenome
54. https://tinyurl.com/MaizeEvoGenome
55. https://tinyurl.com/WheatbixGenomix
56. https://tinyurl.com/CrisprCAS9video
57. https://tinyurl.com/RadiataPineGenome
58. https://tinyurl.com/OvineGenome
59. https://tinyurl.com/Methanogenomics
60. https://tinyurl.com/LivestockGreenhouseGas
61. https://tinyurl.com/NZaquaculture
62. https://tinyurl.com/SnapperPodcast
63. https://tinyurl.com/ApisSociogenomics
64. https://tinyurl.com/HoneybeeSociogenomics
65. https://tinyurl.com/GeneEditingNZvideo
66. https://tinyurl.com/DTCgeneticsIntro
67. https://tinyurl.com/CDCcautionDTC
68. https://tinyurl.com/KiwiDTCsurprises
69. https://tinyurl.com/DTCexperience
70. https://tinyurl.com/DTCparentageSurprise
71. https://tinyurl.com/DTClegal
72. https://tinyurl.com/DTCimplicationsMaori
73. https://tinyurl.com/PGDandNIPT
74. https://tinyurl.com/NIPTdilemmasNZ
75. https://tinyurl.com/CNVsCommon
76. https://tinyurl.com/CNVs4NIPT
77. https://tinyurl.com/NIPTethics
78. https://tinyurl.com/BritishEugenicsHistory
79. https://tinyurl.com/AmericanEugenicsHistory
80. https://tinyurl.com/EugenicsInNZ
81. https://tinyurl.com/TempletonCentreVideo
82. https://tinyurl.com/IVFethics
83. https://tinyurl.com/FertilityInstitute
84. https://tinyurl.com/GattacaTrailerVideo
85. https://tinyurl.com/PolygenicIVFbabies
86. https://tinyurl.com/DesignerDebate
87. https://tinyurl.com/SpindleTransfer
88. https://tinyurl.com/HeCRISPRbabyJustification
89. https://tinyurl.com/EditingMistakes
90. https://tinyurl.com/DNAhistorytimeline
91. https://tinyurl.com/TransformingPrinciple
92. https://tinyurl.com/DoubleHelixDiscovery
93. https://tinyurl.com/Nature1953WatsonCrick
94. https://tinyurl.com/bFormDNAstructureVideo
95. https://tinyurl.com/BioInteractiveDNAstory
96. https://tinyurl.com/MeselsonStahlexpt
97. https://tinyurl.com/BoyerCohenBiotech
98. https://tinyurl.com/BantingBestCollipMacLeod
99. https://tinyurl.com/KaryMullisPCR
100. https://tinyurl.com/ThermalCyclePCR
101. https://tinyurl.com/dideoxyvideo
102. https://tinyurl.com/HGPoverviewVideo
103. https://tinyurl.com/HGPsummary
104. https://tinyurl.com/T2Tpangenome
105. https://tinyurl.com/CompleteAtLast
106. https://tinyurl.com/NeandertalandFOXP2
107. https://tinyurl.com/UCSCbrowserGateway
108. https://tinyurl.com/GenomeBrowserShortTutorial
109. https://tinyurl.com/70000genomesUK
110. https://tinyurl.com/VGP-71657
111. https://tinyurl.com/VGP-Jarvis
112. https://tinyurl.com/EBGinitiative
113. https://tinyurl.com/GenomicsEngland300k
114. https://tinyurl.com/Europe1MG
115. https://tinyurl.com/AllofUsGenomics

116. https://tinyurl.com/VariantsVideo
117. https://tinyurl.com/DNAinfoStorage
118. https://tinyurl.com/FirstGenomeCostAnalysis
119. https://tinyurl.com/BGIseq100dollars
120. https://tinyurl.com/LegoIllumina
121. https://tinyurl.com/SBSpatternedFlowCells
122. https://tinyurl.com/ONTintroVideo
123. https://tinyurl.com/BoykinCassava
124. https://tinyurl.com/DelledonneNanofrogs
125. https://tinyurl.com/ArcticSeq
126. https://tinyurl.com/RubinsSpaceSeq
127. https://tinyurl.com/SystemicHealthInequities
128. https://tinyurl.com/LifeExpectancyNZers
129. https://tinyurl.com/NZhealthcareRacism
130. https://tinyurl.com/NZhealthRacismVideo
131. https://tinyurl.com/RongoaHealing
132. https://tinyurl.com/SocialDeterminantsNZ
133. https://tinyurl.com/NZcardiostats
134. https://tinyurl.com/NZdiabetesstats
135. https://tinyurl.com/CovidIndigenousRisk
136. https://tinyurl.com/MaoriPasifikaCovidRisk
137. https://tinyurl.com/KeoluFoxIndigenousGenomics
138. https://tinyurl.com/PastoralGenomicChange
139. https://tinyurl.com/LactasePermanenceVideo
140. https://tinyurl.com/SkinPigmentationVideo
141. https://tinyurl.com/NeandertalDiseaseRisk
142. https://tinyurl.com/NeandertalCovidLink
143. https://tinyurl.com/DenisovanContributions
144. https://tinyurl.com/MatisooSmith
145. https://tinyurl.com/TuskegeeStudyVideo
146. https://tinyurl.com/GuatemalaStudyVideo
147. https://tinyurl.com/ResearchEthicsVideo
148. https://tinyurl.com/NurembergCode1947
149. https://tinyurl.com/HavasupaiResearchers
150. https://tinyurl.com/SINGmembers
151. https://tinyurl.com/AotearoaSING
152. https://tinyurl.com/GenomicsEquityAotearoa
153. https://tinyurl.com/HauoraEnhancingGenomic
154. https://tinyurl.com/RakeioraGenomics
155. https://tinyurl.com/AotearoaOmics
156. https://tinyurl.com/BermudaPrinciplesHGP
157. https://tinyurl.com/KP-whakapapa
158. https://tinyurl.com/IndigenousCollectiveConsent
159. https://tinyurl.com/TeMataIraGuidelines
160. https://tinyurl.com/PacificMigrationToWairauBar
161. https://tinyurl.com/TaongaDNAperspective
162. https://tinyurl.com/PacificGoutRisk
163. https://tinyurl.com/HDGCvideos
164. https://tinyurl.com/IncidentalDilemmas
165. https://tinyurl.com/NZnewbornTesting
166. https://tinyurl.com/USnewbornWGS
167. https://tinyurl.com/UKnewbornWGS
168. https://tinyurl.com/NewbornWGSyes
169. https://tinyurl.com/NewbornWGSno
170. https://tinyurl.com/NZinsurance
171. https://tinyurl.com/GuthrieCardForensics
172. https://tinyurl.com/HLethics
173. https://tinyurl.com/NavajoMoratorium
174. https://tinyurl.com/NuuChahNulthGenetics
175. https://tinyurl.com/ChinaThermofisher
176. https://tinyurl.com/GeneticStateControl
177. https://tinyurl.com/ErlichFamilyTree
178. https://tinyurl.com/ColdCaseGenealogy
179. https://tinyurl.com/ForensicGenealogyAudio
180. https://tinyurl.com/UniversalDNAdatabase
181. https://tinyurl.com/CrimePunishmentCalculation
182. https://tinyurl.com/FacialGenetics
183. https://tinyurl.com/SNPsFacialShape
184. https://tinyurl.com/DNA2faceChina
185. https://tinyurl.com/DNA2faceUSA
186. https://tinyurl.com/DNApolicingMaori
187. https://tinyurl.com/ControversyMAOAgene
188. https://tinyurl.com/RaumatiHookMAOA
189. https://tinyurl.com/DataExplorerNZMoH
190. https://tinyurl.com/DemographyIncarcerationNZ
191. https://tinyurl.com/MoanaJacksonVideo
192. https://tinyurl.com/SJGouldJustSo
193. https://tinyurl.com/RiskRewardMAOAalleles
194. https://tinyurl.com/GregSemuRaftPasifika
195. https://tinyurl.com/Medusa1819
196. https://tinyurl.com/GoldieSteeleArrival
197. https://tinyurl.com/GregSemu
198. https://tinyurl.com/DunedinStudyVideo
199. https://tinyurl.com/GxEinfantsMAOA
200. https://tinyurl.com/CDH13andMAOA
201. https://tinyurl.com/AntisocialPolygenics
202. https://tinyurl.com/DunedinPolygenicAntisocial
203. https://tinyurl.com/GeneticBackgroundRisk
204. https://tinyurl.com/GeneticsEpigeneticsCulpability
205. https://tinyurl.com/UKapologyTuring
206. https://tinyurl.com/SameSexBehaviourGenetics
207. https://tinyurl.com/PolygenicGxEsexuality
208. https://tinyurl.com/FBOEsexuality
209. https://tinyurl.com/NeuralUnderpinningsSexuality
210. https://tinyurl.com/BrainPsychopathy

## Additional image credits:

*Full citation for 3D genome image on pg.2*:
Stevens, T., Lando, D., Basu, S. et al. (2017) 3D structures of individual mammalian genomes studied by single-cell Hi-C. Nature 544, 59-64 . https://doi.org/10.1038/nature21429".

*Full citation for SARS-CoV-2 zoonotic schematic on pg.4*
Yi Y, Lagniton PNP, Ye S, Li E, Xu RH. (2020) COVID-19: what has been learned and to be learned about the novel coronavirus disease. Int J Biol Sci; 16(10):1753-1766. doi:10.7150/ijbs.45134. https://www.ijbs.com/v16p1753.htm