FOI

# Genomics in Perspective

Science-Society-Security

ELISABET FRITHZ, ANDERS ALLGARDSSON AND PER STENBERG

FOI-R--4217--SE
ISSN 1650-1942 December 2015

Elisabet Frithz, Anders Allgardsson
and Per Stenberg

# Genomics in Perspective

Science-Society-Security

| | |
|---|---|
| Titel | Genomik i perspektiv |
| Title | Genomics in Perspective |
| Rapportnr/Report no | FOI-R--4217--SE |
| Månad/Month | December |
| Utgivningsår/Year | 2015 |
| Antal sidor/Pages | 30 p |
| ISSN | 1650-1942 |
| Kund/Customer | Försvarsmakten |
| Forskningsområde | 2. CBRN-frågor och icke-spridning |
| FoT-område | |
| Projektnr/Project no | A495116 |
| Godkänd av/Approved by | Mats Strömqvist |
| Ansvarig avdelning | CBRN-skydd och säkerhet |

Bild/Cover: WikiCommons

# Preface

Major scientific advances have been made in genomics recently, providing unprecedented technological abilities. This report addresses this critical field, along with the broader policy and security questions it raises. It is by no means intended to be a comprehensive review of this vast scientific field, but rather an attempt to give the reader an idea of the far-reaching opportunities and challenges genomics presents. The study retrospectively considers genomic developments in the last two decades and outlines what we believe to be the most important current and future issues.

In order to make the content accessible to a broad non-scientific audience we have tried to minimize, as far as possible, use of technical terms, and written in English. In particular, we hope that the study will reach decision-makers within the Swedish Armed Forces and other governmental services.

It is our ambition that this study will provide food for thought and contribute to the current debate on how society should best benefit from scientific progress while minimizing risks of its misuse.

# Sammanfattning

År 2000 presenterade den dåvarande amerikanske presidenten Clinton resultatet från arbetet med att kartlägga den mänskliga arvsmassan (DNA-sekvensen, genomet) i “the Human Genome Project, HGP”. Sedan dess har stora teknologiska och metodologiska framsteg gjorts inom genomiken och i dag har arvsmassan från en mängd organismer bestämts. Kunskap finns nu om genomsekvensen för såväl enkla organismer som virus och bakterier, men även om mer komplexa livsformer som maskar, svampdjur och insekter samt växter, reptiler och däggdjur. Även DNA från utdöda arter som mammuten har karaktäriserats. Landvinningarna i fältet har gjort det möjligt att på kort tid (dagar) bestämma DNA-sekvenserna för bl. a de samhällsfarliga virus som orsakar Ebola och MERS (Middle East Respiratory Syndrome) – en förutsättning för att snabbt kunna diagnostisera, smittspåra och isolera drabbade patienter.

Tillämpningarna av DNA-sekvensering och -analys får allt bredare genomslag i samhället. Exempelvis har amerikansk polis börjat använda DNA som ett verktyg för att förbättra sina gärningsmannaprofiler; genom rent DNA-baserade analyser kan man utröna misstänktas etnicitet och i viss mån även förutsäga deras kroppskontitution och anletsdrag. Genomiken kommer, genom analyser av genomiska och kliniska data, att lära oss mer om de underliggande faktorerna för exempelvis tumörer, ärftliga sjukdomar och hur vi reagerar på olika läkemedel. Sådana kunskaper kommer ge sjukvårdspersonal möjligheter till bättre prognoser, diagnoser och behandlingar av sjukdomar. Framstegen inom genomik öppnar också nya möjligheter till genetiska modifieringar av boskap och grödor, antingen genom direkta förändringar eller genom att kombinera genförändringar med traditionell avel.

HGP var det första “mega-projektet” inom de biologiska vetenskaperna, väl i paritet med större forskningsinitiativ inom bl. a partikelfysik och astronomi. I och med de mycket stora datamängder som genereras inom genomiken har fältet allt mer kommit att bli en “Big Data”-vetenskap som ställer allt högre krav på IT-infrastruktur. Kraven på beräknings- och lagringskapacitet ökar mycket snabbt och det har uppskattats att de data som genereras kommer att vara i samma storleksordning som Twitter och Youtube inom ett årtionde.

Ur ett säkerhetsperspektiv kan många av de teknologier och resultat som genereras inom genomiken klassificeras som produkter med dubbla användningsområden (dual-use). Samtidigt som många av de framsteg som görs inom genomik erbjuder stora möjligheter för förbättrad hälso- och sjukvård, kan de även missbrukas i antagonistiska syften. Genetisk information kan i sig användas för att misskreditera politiska ledare genom att ifrågasätta bland annat deras släkthistoria, mentala hälsa och förmågor. I kombination med tekniker för att förändra DNA-sekvenser (gene-editing) har genomiken potential att utveckla “etniska biovapen”. Sådana vapen kan vara genetiskt modifierade bakterier eller virus som är utvecklade för att angripa specifika genetiska sårbarheter eller karaktäristika på individ eller folkgruppsnivå. Även boskap och grödor kan angripas med sådana strategier.

Nedan har vi listat ett urval av framtida projektioner, tankeväckande möjliga applikationer och några av de frågor som associeras med dessa och som behandlas i denna rapport. Syftet med dessa nedslag är inte att ge läsaren en uttömmande bild av alla möjliga konsekvenser med genomik, utan att ge läsaren en känsla för de möjligheter som genomiken kan erbjuda och vilka utmaningar den ställer oss inför.

- **Sekvensering av alla människors DNA.** Kommer varje nyfött barns DNA att sekvenseras i framtiden? I dagsläget kan en människas genom sekvenseras på 24 timmar till en kostnad av ca 10 000 kr. I framtiden kommer DNA-sekvensering att vara ännu billigare och snabbare vilket ökar tillgängligheten för allmänheten.
- **Förutsäga sjukdomar och personliga drag.** I dagsläget är det möjligt att screena mänskliga embryon för genetiska sjukdomar som cystisk fibros och hemofili (blödarsjuka). I takt med att kunskapen om vår arvsmassa ökar kommer det bli allt lättare att förutsäga risken för att en individ ska drabbas av olika sjukdomar. Även personliga egenskaper och yttre kännetecken (t ex kognitiv förmåga, ögonfärg och ansiktsform).
- **Förändring av DNA som ett verktyg vid avel av växter och djur.** Traditionell avel är arbetsintensiv, tidskrävande och fokuserar främst på ekonomiskt viktiga grödor och boskap. Den snabba utvecklingen av vår förmåga att analysera och modifiera DNA från alla organismer – både vilda och domesticerade – kommer att möjliggöra ökad utvecklingstakt i avelsprogram. Genom att identifiera gener för önskade egenskaper (som t ex tolerans mot torka och höga temperaturer) och sedan överföra dessa gener till ekonomiskt värdefulla djur- och växtslag öppnas nya möjligheter.
- **Utveckling av nya livsformer.** Genom att kombinera kunskap om naturligt förekommande DNA och förmågan att skapa syntetiskt DNA kan man designa och skapa nya livsformer och även återskapa utdöda arter. Idag har syntetiska varianter av olika virus som polio och SARS skapats.

Exemplen som beskrivs ovan är framtida projektioner, men med tanke på den mycket snabba utvecklingen i fältet kan de komma att realiseras inom en snar framtid. Vår ökande förmåga att sekvensera och analysera DNA ställer dock än mer omedelbara frågor och det är uppenbart att utvecklingen i fältet kommer att få konsekvenser på många nivåer i samhället.

Exempelvis kommer hälso- och sjukvården möta patienter som upptäckt att de är genetiskt predisponerade för olika sjukdomar som t ex hjärt-kärlsjukdom, depression, Alzheimers sjukdom eller cancer. Kommer tidiga medicinska interventioner spara pengar för samhället och minska lidandet hos dessa patienter? Är det bra för människor att veta om vilka riskfaktorer de har?

Ska försäkringsbolag, arbetsgivare, polismyndigheten etc. ha tillgång till information om individers arvsmassa? Hur ska vi skydda vår genetiska integritet? Kan vi lagstifta mot genetisk diskriminering?

Ska militär, polis och rättsväsendet ha möjlighet att identifiera individer, bestämma släktskap, identifiera brottsoffer och profilera kandidaters lämplighet för olika roller inom t ex försvaret?

Den snabba utvecklingen inom genomik kan få långtgående konsekvenser i samhället och bör vara av intresse att följa för såväl politiker som företrädare för civila och militära myndigheter.

Nyckelord:

DNA-sekvensering, genomik, genmodifiering

# Executive summary

Ever since US President Clinton presented the first draft of the human genome, in a special ceremony in the White House in 2000, there have been exceptional advances in genomics and associated technologies. Today, the genomic material (DNA) from a vast array of organisms on earth, ranging from the simplest viruses and bacteria through worms, sponges and insects, to plants, reptiles and mammals, have been subjected to intensive genomic sequencing efforts. Even the DNA of extinct species like the mammoth has been sequenced. Notably, dangerous viruses such as Ebola and MERS (Middle East Respiratory Syndrome) can be sequenced within days, allowing their rapid diagnosis, tracking and isolation of patients.

DNA sequencing and analysis are also being applied in other social sectors. For example, US police are already beginning to use DNA to decipher suspects' ethnic backgrounds and predict their appearance/facial features.[1,2] In addition, in the beginning of 2015 the Obama administration launched an extensive science and healthcare program called "the Precision Medicine Initiative", essentially based on advances in genomics. It is envisioned that by analysing genomic and clinical data we will learn more about the underlying causes of tumours, inherited diseases and responses to drugs etc., which will enable healthcare practitioners to predict, diagnose and treat disease more efficiently in the future. Advances in genomics have also opened new possibilities to improve human welfare by facilitating genetic modification of livestock and crops.

The Human Genome Project was the first mega-project in the biological sciences, similar in scale to the major initiatives seen in, for instance, physics. As genomics has evolved it has become a "Big Data science", demanding ever-increasing computer capacity for data storage and analysis, which will be comparable to capacities required by Twitter and YouTube within a decade according to some estimates.

From a security perspective, many of the technological breakthroughs and results generated within genomics have the potential of being of dual-use. While they offer great promises for improving human health and welfare, they could also be potentially misused for adverse or hostile purposes. Genetic information can in itself be used to discredit political leaders by raising doubts about their ancestry, mental health, capabilities, etc. Furthermore, in combination with "gene-editing" (rewriting DNA), genomics raises prospects of the potential development of "ethnic bioweapons", e.g. modified bacteria and viruses targeting specific genetic vulnerabilities of human individuals. In a similar way, livestock and crops could be targeted.

Below, we have listed a selection of future projections, thought-provoking applications, and associated questions, addressed in this report. These focal points are intended to give the reader a feeling for the opportunities that genomics is presenting, and the challenges it is posing, rather than exhaustively defining all of its potential consequences.

- **DNA-sequencing of all humans**. Will the DNA of every new-born baby be sequenced in the future? Currently, a human's total genome can be sequenced in 24 hours at a cost of US$ 1 000. In the future DNA sequencing will be even quicker, cheaper, and thus increasingly available to the general public.
- **Predicting disease and personal characteristics**. Currently it is possible to screen human embryos for genetic diseases such as cystic fibrosis and haemophilia. As our genomic knowledge expands, it will be possible to predict an individual's risks of developing increasing numbers of diseases as well as various cognitive abilities and personal traits.
- **DNA-sequencing and reshaping animals and plants.** Traditional breeding efforts have been labour-intensive, time-consuming and concentrated on economically significant

[1] http://www.forensicmag.com/articles/2015/01/first-dna-phenotyped-image-person-interest-double-homicide. Accessed 2015-12-02.
[2] http://www.forensicmag.com/articles/2015/09/defining-face-what-can-dna-phenotyping-really-tell-us-about-unknown-sample. Accessed 2015-12-02.

livestock and crops. However, the rapid advances in our ability to analyse and modify the DNA of any organism - domesticated or wild - will allow massive acceleration of breeding programs through the identification of valuable combinations of genes (e.g. for drought or high temperature tolerance), and populations of target organisms that carry them.

- **Engineering new life forms**. Combining knowledge of natural genomes with the ability to build synthetic DNA/genomes will also allow the creation of new life forms (forerunners are synthetic versions of polio, SARS and other viruses) and reconstruction of extinct species and pathogens.

The examples described above are future projections (which may be realised surprisingly soon, given the exponential rates of recent advances), but our increasing ability to sequence and analyse DNA efficiently is also raising more immediate issues. Although we still have much to learn about how to interpret the vast amounts of DNA-data generated, these developments are expected to have consequences at many levels of society. For example:

Health practitioners will see patients who have discovered they have genetic predispositions for various medical conditions, e.g. cardiovascular disease, depression, Alzheimer's disease or cancer. Will early medical consultation save money by reducing future care costs and suffering? Where do we draw the line considering ethics? Is it a good thing for people to know that they are "at risk"?

Should insurance companies, employers, the police, etc., be able to access information on an individual's genetic makeup? How should we be able to protect our genetic integrity? Can we legislate against genetic discrimination?

Of particular societal interest are the possibilities for the armed services, police and judicial services. Should they be able to identify individuals, determine family relationships, identify victims, and profile applicants for the armed forces to determine suitability, etc. (and, if so, who should set the criteria)?

Politicians, civil and military service authorities at all levels should be interested in following the progress of genomics, particularly the medical, ethical, judicial, social and security implications.

Keywords:

DNA sequencing, genomics, genetic modification

# Table of Contents

# 1 Introduction

The genetic information of an organism is encoded in the DNA inherited from its parents. DNA consists of strings of four molecules called nucleotides or bases (usually designated by the letters G, A, T and C) in specific sequences that carry genetic information (genes and various regulatory sequences). In conjunction with environmental factors, the genetic information determines the features or traits of all living organisms.

The size of a genome ranges from a few thousand nucleotides (in small viruses) to several billion nucleotides (in mammals); the human genome consists of ca. 3 billion nucleotides. The technological revolution in sequencing during the last two decades has made it easier and cheaper to sequence, or "read", the genetic code of living and even extinct organisms (see Figure 1 for the general workflow of a sequencing initiative). The scientific field of studying entire genomes is referred to as genomics.

Presently, a human genome can be sequenced in 24 hours at a cost of US$ 1 000. In the future, DNA-sequencing will be even faster and cheaper, becoming increasingly available to government agencies, healthcare services and the general public. The advances in DNA analysis will have consequences at many levels of society and are important to monitor.

One field being transformed by the technological breakthroughs in DNA sequencing is healthcare. Patients' DNA can be sequenced either through official agencies or increasing numbers of private sequencing companies. The sequences can then be used to diagnose genetic diseases or to predict a range of physical traits as well as disease-risks and ethnic ancestry. Genome analysis has also spurred interest outside healthcare: insurance companies, the armed services, and judicial services could all benefit from the ability to identify individuals, determine family relationships, identify victims and (for instance) soldiers with the optimal genetic composition for specific assignments.

Besides being able to read genetic information, we now have the technological ability to write and modify living organisms' genetic codes. This is providing the ability both to fix problems and custom design organisms according to our desires. The development of weapons targeting people with a specific genetic profile is a future possibility. Thus, genetic data can threaten more than personal integrity.

Clearly, the access to genomic information can benefit society, but this information is a double-edged sword and there are many legal and ethical issues to consider, among others, who should have access to that information and what should individuals or agencies be allowed to do with it? Politicians, civil and military service authorities need to be aware of both progress in genomics and (particularly) the medical, ethical, judicial, social and security implications.

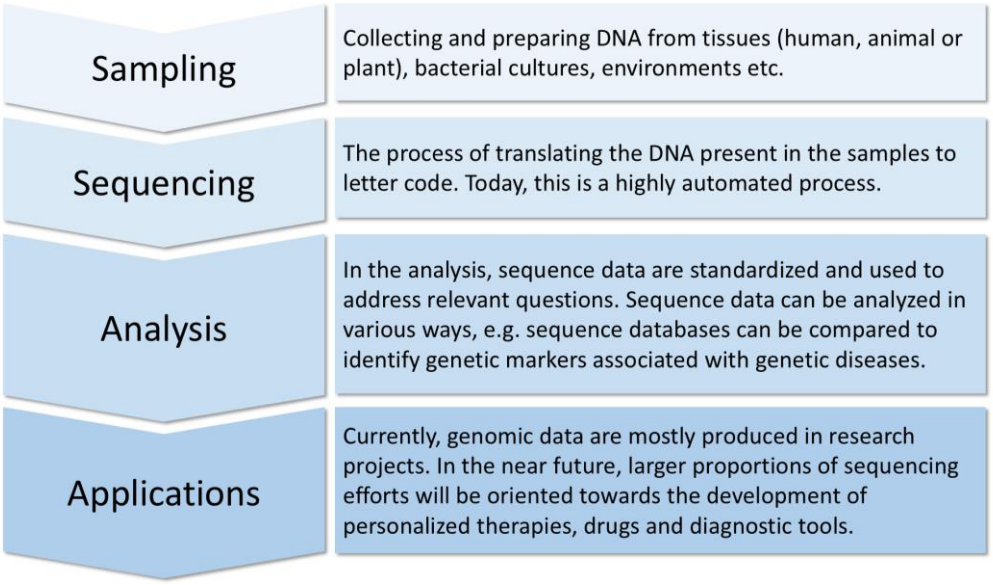


**Figure 1**. Flowchart describing the workflow of a genome sequencing initiative**.**

# 2 Reading and analysing human genomes

## 2.1 The Human Genome Project

The Human Genome Project (HGP) was a huge international research collaboration with the goal to sequence and map all human genes. It formally started in 1990 and was successfully completed in April 2003 at a cost of about 3 billion dollars. The project was coordinated by US agencies and universities, with contributions from laboratories in the UK, France, Germany, Japan and China. However, a parallel private initiative to sequence and map the complete human genome was started in 1998 by Celera Genomics, led by Craig Venter. For a while the two genome sequencing initiatives seemed to compete in efforts to finish first, but eventually they united in a common undertaking to obtain the first complete sequence of the human genome.

The results from the Human Genome Project showed that the human genome encoded around 20 500 genes, fewer than the expected ca. 80 000 genes, and within the same range as in mice. The analysis of the human genome also showed that most of our DNA consists of repetitive DNA-segments of mostly unknown function.

The overall goals of the Human Genome Project were to provide foundations to understand how the human organism works, identify markers for genetic diseases, and hopefully obtain cues to facilitate the development of novel drugs.

## 2.2 Recent large-scale human sequencing projects

The Human Genome Project paved the way for many other initiatives, some of which are listed in Table 1. A shared ambition for all of them has been to broaden knowledge of the human genome and how it functions, and common goals are to link distinct genetic variations to risks of developing specific diseases or traits (i.e. association studies).The Human Genome Project also spurred numerous advances in sequencing technology and analysis.

The parts of the genome that do not contain genes (non-transcribed DNA) have also been addressed, notably in the ENCODE project. The results indicate that this so-called "junk-DNA" plays important roles in regulating the expression of human genes.

**Table 1.** Current and historical large-scale human sequencing initiatives

| Sequencing project | Objective/Scope | Initiative | Year |
|---|---|---|---|
| The Human Genome Project (HGP) | Sequencing and mapping all human genes | US-initiated. International collaboration | 1990-2000/2003 |
| HapMap Project | Cataloguing genetic similarities and differences in humans | Global | 2002-2010 |
| ENCODE | Determining the role of "junk-DNA", i.e. untranscribed DNA | US-initiated. International collaboration | 2003- |
| The Personal Genome Project (PGP) | Creating public genome, health, and trait data | US-initiated. International collaboration | 2005- |
| 1 000 Genome Project | Determining genetic variations and associations with disease | UK-initiated. International collaboration | 2008- |
| 100 000 Genome Project | Combining genomic sequence data with medical records | National Health Service (NHS), UK | 2013- |
| 1 000 000 Genome Project | Focusing on human health and disease | BGI, China | 2013- |
| 1 000 000 Genome Project | Focusing near-term on cancers and pharmacogenomics & longer-term on human health and disease. | Precision Medicine Initiative, USA | 2015- |

### Private enterprises

Both the cost and time needed to sequence a human genome have fallen dramatically since the days of the HGP (see below). Sampling DNA from an individual is also quite trivial; a simple mouth swab or saliva sample is enough. Hence, there is a growing market for DNA sequencing, and a number of private sequencing companies are now offering their services not only to scientists, hospitals and academic institutions etc., but also to the public.

A giant in this market is the China-based BGI (formerly Beijing Genome Institute), which has local offices and offers services globally. Currently, BGI is estimated to generate at least 25% of the genomic data in the world per year and is a flagship for future scientific Chinese endeavours. It is unclear whether BGI should be described as a private or a governmental organisation since it receives funding from both sectors. Its corporate homepage states that BGI "is the first citizen-managed, non-profit research institution" but also that it "Includes both private non-profit research institutes and sequencing application commercial units". BGI has participated in a number of major genomic projects. Examples of other large enterprises are Illumina (USA)[3] and DeCode (Iceland)[4].

[3] http://www.illumina.com/. Accessed 2015-11-20.
[4] http://www.decode.com/. Accessed 2015-11-20.

Selling "personalized genomics" sequencing and analysis services direct to the consumer has also proven to have commercial potential. For example, the homepage of the US-based genomic company 23andMe (recently established in Sweden), announces that "We provide genetic reports on your ancestry, family history and help you connect with your DNA relatives".[5] Numerous companies are selling similar genetic analysis and tests.

## 2.3 Extracting information from a human genome

### DNA profiling in forensic science

The major use of DNA analysis in forensic science is in DNA profiling. The technique was developed in the 1980s and is used to identify individuals through characteristics of their DNA. It is used in parentage testing, and to identify or place persons in specific locations in criminal investigations. Most markers used in human DNA profiling were identified well before completion of the Human Genome Project and it is generally believed that current and emerging sequencing technologies will become important assets in the forensic toolbox, enabling more rapid and thorough analysis of DNA. For example, recent technological advances now enable forensic analysis of genomic DNA from single cells, or even degraded mitochondrial DNA.

### Using DNA to estimate an individual's appearance

Through analysis of the variations in genes that determine physical traits, individuals' characteristics such as sex and colour of their eyes, skin and hair can be predicted. In addition, markers providing increasingly reliable predictions of more discriminative traits like freckling, presence of moles and hair texture are currently under development [3]. Combining such predictive capacities with data from genetic anthropology initiatives like the Genographic Project[6] enables the creation of virtual snapshots of people. At the time of writing, several companies have emerged that provide such profiles based solely on DNA profiles[7].

Such predictions are not yet as accurate and reliable as desired, but there are intense efforts to improve both the markers and predictive algorithms. As the ability to predict people's physical appearance has strong potential applications in security contexts, as well as forensics, this research is drawing attention from funding agencies such as the US Departments of Defence and Justice.[8,9]

### Metagenomics in forensic science

The human body contains billions of bacteria, viruses and fungi. The composition of this so-called microflora has been shown to be surprisingly unique for every person. Thus, by sequencing the microbial community of (for instance) rootless hairs or smartphone surfaces it is possible to generate metagenomic fingerprints (DNA-based fingerprints of the microflora) from which individuals' identities can be inferred with no need for human DNA [4, 5].

[5] https://www.23andme.com/ Accessed 2015-10-22.
[6] https://genographic.nationalgeographic.com/. Accessed 2015-11-19.
[7] https://snapshot.parabon-nanolabs.com/ and http://www.identitascorp.com/. Accessed 2015-10-20.
[8] http://science.iupui.edu/news/iupui-awarded-11-million-grant-develop-tools-predict-physical-appearance-dna. Accessed 2015-11-16.
[9] http://www.dtra.mil/Portals/61/Documents/dtra-sbir-phase-ii-awards-2009-2013.pdf. Accessed 2015-11-16.

## Understanding genetic diseases

Today, we understand the genetic backgrounds of some conditions where there is a straightforward correlation between DNA-variants and clinical symptoms. Examples of such conditions include the diseases sickle cell anaemia, phenylketonuria and cystic fibrosis.[10] However, we have limited understanding of more complex genetic disorders. The genetic bases of disorders such as asthma, cardiovascular diseases and diabetes are complex and involve multiple genes. Furthermore, these diseases are only partially determined by genetic factors, as the development of clinical syndromes is also heavily dependent on environmental factors. While inheritance patterns and risks of developing clinical symptoms (penetrance) are well characterized for some diseases, the inheritance and penetrance of complex disorders associated with numerous genetic components are difficult to determine.

> The Million Veteran Program (MVP) was launched by the US Department of Veteran Affairs to increase the understanding of medical risks and develop future treatments for diseases and conditions like post-traumatic stress disorder (PTSD). It is developing a biobank that links genetic data and information about lifestyle, military exposure and health information. So far, almost 400,000 people have enrolled in the program [1].

## Precision medicine and Cognitive Genomics

In January 2015, US president Obama launched a "Precision Medicine Initiative" with the goal to create more personalized healthcare in the USA.[11]

> "Tonight, I'm launching a new Precision Medicine Initiative to bring us closer to curing diseases like cancer and diabetes, and to give all of us access to the personalized information we need to keep ourselves and our families healthier." *From the State of the Union Address, January 20, 2015, by US President Obama.*

The US National Research Council (NRC) defines precision medicine as "the tailoring of medical treatment to the individual characteristics of each patient. It does not literally mean the creation of drugs or medical devices that are unique to a patient, but rather the ability to classify individuals into subpopulations that differ in their susceptibility to a particular disease, in the biology and/or prognosis of those diseases they may develop, or in their response to a specific treatment. Preventive or therapeutic interventions can then be concentrated on those who will benefit, sparing expense and side effects for those who will not".[12]

Oncology is a discipline where there are high hopes that this initiative will improve diagnosis and therapeutic options. Patients are already assigned to treatment based on the genetic profiles of their tumours. For example, the genetic alterations that drive a breast cancer may be similar to the genetic alterations of another person's lung tumour, so they might be sensitive to the same sort of medical treatment. Genetic profiling may also be highly valuable for elucidating why certain tumours sometimes become resistant to cancer therapy.

[10] https://www.socialstyrelsen.se/ovanligadiagnoser/fenylketonuri. Accessed 2015-10-07.
[11] https://www.whitehouse.gov/the-press-office/2015/01/20/remarks-president-state-union-address-january-20-2015. Accessed 2015-11-19.
[12] https://www.whitehouse.gov/files/documents/ostp/PCAST/pcast_report_v2.pdf. Accessed 2015-11-19.

### The Cognitive Genomics project (BGI)

A genomics project that has attracted much attention, as well as negative criticism, is the Cognitive Genomics Project, initiated and headed by researchers at BGI, China. The objective of the project is to identify the genes underlying human intelligence. This is a complex endeavour, which was recently considered too difficult by the scientific community. Although the genetic background is an important determinant of IQ, the genetic factors of intelligence are still unknown. [23]

In order to obtain enough material (i.e. genomes from very intelligent people), BGI is collaborating with King's College in London, UK, where researchers started collecting DNA in the 1970s.

In the Western world, such projects are controversial, and raise anxieties that they may foster pre-natal eugenic practices.

# 3 DNA-sequencing life

The genome sequence of an organism determines its molecular layout and heavily influences its biology. Consequently, knowledge of organisms' genomic sequences can be used to improve our understanding of properties like virulence and resistance to antibiotics. Genomic data are also highly valuable for tracing pathogens, in either natural outbreaks or following acts of bioterrorism.

Knowledge of the genomic layout of an organism also enables modifications that influence its biochemistry. This can be used to create modified organisms with novel properties designed for purposes such as drug production in bacteria, pesticide resistance in plants and models for human disease in laboratory mice. Currently, the genomes of several thousands of species (including some now extinct) from all taxonomic kingdoms (i.e. bacteria, archaea and eukaryotes) have been sequenced (Figure 2).

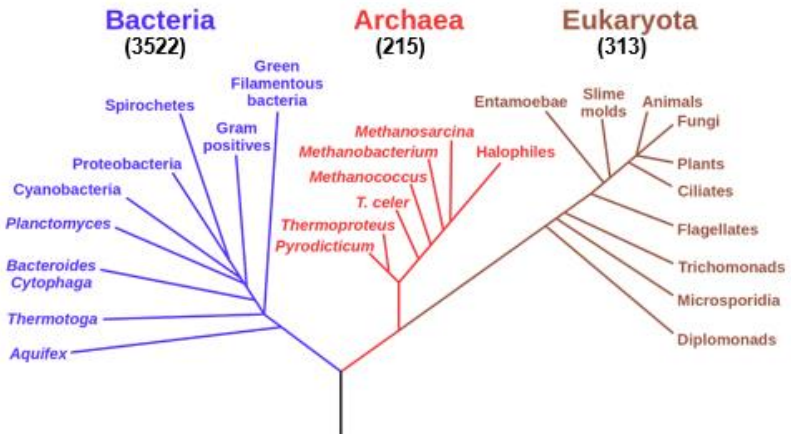


**Figure 2.** The phylogenetic tree of life (from Wikimedia Commons). Numbers of organisms in each "domain" (bacteria, archea and eukaryota) for which complete sequences are publicly available are shown in parentheses. The number of completely sequenced organisms (that are publicly available) are shown within parenthesis.[13] A few of the sequenced eukaryotic organisms were sequenced long after their extinction.

### Genomic sequences of microorganisms

[13] KEGG database, http://www.genome.jp/kegg/. Accessed 2015-11-12.

There are diverse uses for genome sequences. For example, sequences of infectious agents involved in disease outbreaks may provide both detailed information about the origins of the pathogens and convenient ways to track infections as they spread through communities.

Complete genome sequences of small viruses were first published in the mid-late 1970s, and the first bacterial genome was published in 1995 [6]. However, developments in sequencing technology during the last decade have truly revolutionized our ability to analyse microbial genomes, and numbers of published species' genomes are rapidly growing (see Figure 1 for current numbers of sequenced organisms).

In several recent viral outbreaks, such as the SARS (Severe Acute Respiratory Syndrome) outbreak in 2002-2003 and the recent Ebola and MERS (Middle East Respiratory Syndrome) outbreaks, access to genome sequences has been vital for understanding changes in the viral genomes over time. In the large European outbreak of what was initially believed to be EHEC (Enterohemorrhagic *Escherichia coli*), sequencing was pivotal to show that the pathogenic strain was a strain of EHEC that had acquired genes enabling production of Shiga toxin [7].[14]

### Metagenomic analysis of diverse environments

One of the emerging applications of genome sequencing is metagenomic analysis. In metagenomics, complex microbial samples can be analysed without culturing, thereby enabling identification of uncultivable species. Currently, this approach is used for purposes such as detecting pathogens in drinking water and analyses of microbial ecology. The species profiles generated in metagenomic analyses of defined environments can also be used to monitor changes in the environments' microbial composition, for instance to obtain early warnings of emerging infections in society. Notably, in the PathoMap project of Weill Cornell Medical College [8], surfaces in public areas in New York are sampled and sequenced to obtain profiles of the DNA present in urban environments. One of the goals of the project is to develop a pathogen warning system that identifies emerging microbial threats on a molecular level.[15]

# 4 Current sequencing technologies

### Sequencing faster and cheaper

During the 1990s and early 2000s the rate of technical development in DNA sequencing roughly matched the exponential development of computer power (doubling of the capacity of micro-chips of a given size every 18 months). However, around 2008 the rate of development further accelerated (Figure 3), and the generation of data by the sequencing community is now outpacing the development of data storage technology.

[14] http://www.cdc.gov/ecoli/general/. Accessed 2015-11-16.
[15] http://www.pathomap.org/. Accessed 2015-11-19.

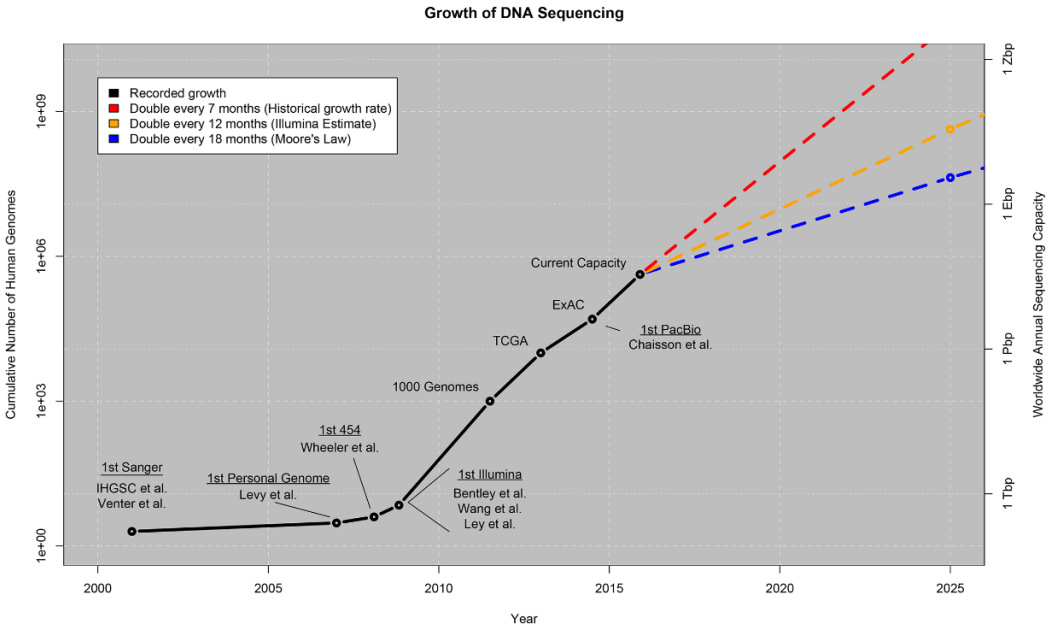


**Figure 3**. Growth of DNA sequencing. Figure and forecasts from [9]

## Short-read sequencing

The vast majority of DNA sequencing today involves cutting up a sample of DNA into huge numbers of short sequences of 100-400 basepairs (bp), which are then sequenced, generating "reads". A single machine run can produce several hundred millions or even billions of short reads. In order to transform these short sequences into a complete genome, the reads have to be mapped on a reference genome (i.e. genome of a closely related organism) or assembled *de novo*. *De novo* assembly (piecing together) of a large genome such as the human genome from short reads requires massive computer capacity and cannot provide the complete genome. This is because there are many repetitive regions in the genome that are longer than 100-400 bp (so many reads could originate from multiple locations in the genome).

Today there are two major techniques for generating large quantities of short reads: Illumina and IonTorrent (see Table 2). Both require fragmentation followed by amplification of the original DNA. Sequence adaptors to each fragment then an enzyme (DNA polymerase) copies the library of sequences. In Illumina sequencing, incorporation of nucleotides during the copying process is monitored using fluorescent dyes (one colour for each of the four nucleotides) and image analysis. The image analysis is time consuming and computationally intensive. In IonTorrent sequencing the image analysis is replaced by use of an extremely sensitive sensor that can detect release of a single proton (which occurs each time a nucleotide is incorporated by DNA polymerase). In 2014 Illumina held about 70% of the market for sequencing machines.[16]

## Long-read sequencing

As mentioned above, significant drawbacks of the shorter reads produced by the Illumina and IonTorrent platforms are the problems associated with *de novo* assembly. Longer reads make it much easier to piece together genomes, but currently only one sequencing platform on the market yields significantly longer sequence reads. Pacific biosciences has developed a technique that does not require amplification of DNA and where the additions

[16] Zimmerman, Eilene (18 February 2014). "50 Smartest Companies: Illumina". *MIT Technology Review* (Massachusetts Institute of Technology). Accessed 2014-08-25.

of fluorescent nucleotides by DNA polymerase is measured in real time (without image analysis). This platform, the SMRT sequencing system, does not require amplification of DNA and the additions of fluorescent nucleotides by DNA polymerase are measured in real time (without image analysis). It can produce very long (more than 50 000 bp) reads, but only about 50 000 reads per machine run (compared to more than a billion in Illumina sequencing, see Table 2).

A novel technique that will soon be commercialized is Oxford Nanopore's MinION system, which generates similar numbers of sequences, of similar lengths, to SMRT sequencing, but without the need for a DNA polymerase. In this technique, single-stranded DNA is forced through small pores in a membrane by applying an ionic current through the pores (DNA is an acid and is negatively charged at neutral pH). The specific disruption of the ion flow through the membrane caused by passage of each nucleotide is then measured and translated into DNA sequence. Since no DNA polymerase is involved, the technique can sequence DNA very rapidly, up to 500 bp/second (ca. 25-fold faster than the Illumina system). The drawback with the current form of this system is that it creates more sequencing errors than the other techniques.

In 2003 the Chinese sequencing company BGI acquired California based company Complete Genomics, which had developed a new sequencing technology. US national security issues were raised at high political levels and several attempts to stop the acquisition were made before the deal was closed.

**Table 2**. Performance parameters of current sequencing platforms. Adapted from Wikipedia and [10].

| Method | Read length | Error rate | Reads per run | Time per run |
|---|---|---|---|---|
| **Illumina** | Up to 300 bp | 0.10% | Up to 6 000 000 000 | 1-11 days |
| **IonTorrent** | Up to 400 bp | 2% | Up to 80 000 000 | 2 hours |
| **SMRT** | Up to 50 000 bp | 13% | Up to 50 000 | 0.5-4h |
| **MinION** | Up to 80 000 bp | 20% | Up to 40 000 | 0.5-12h |

## Big data

A standard text-file of about three gigabytes is needed to store the sequence of a human genome, but this is not the only information from a sequencing project that should be stored. In addition, all individual sequence reads along with their quality values (indicating the estimated accuracy of the sequencing at corresponding positions) needs to be safely stored for future consultation, if necessary. This information is important if previous results are to be examined and validated. In a recent perspective paper in PLOS Biology, Stephens et al. [9] estimate that by 2025 the storage requirements for sequencing data will be on a par with, or even exceed, the needs for YouTube and Twitter. It will be challenging simply to store all these data (estimated amounts are 2-40 exabytes[17]). In order for these data to be of any use, they also need to be searchable and retrievable. Analysis of the vast amounts of data being produced will likely require significant advances in computer technology and algorithms.

[17] 1 exabyte = 1 million terabytes = 1 billion gigabytes

# 5 Editing genomes

CRISPR-Cas9 are components of a bacterial immune system that were recently adapted for use in a highly specific gene editing tool. The method can be used to introduce changes to DNA sequences at very precise locations in a genome [2].

The properties of an organism can be altered by modifying its DNA, and thus creating a genetically modified organism (GMO). GMOs have been created for decades (or even millennia, if hybridizations of agricultural crops are included), but with the modern tools of DNA manipulation the process is easier, cheaper, and more precise. In the early days of DNA editing, off-target effects caused by limited precision of the methods severely restricted its genomic-scale application. However, the current level of technology with methods such as TALENs and CRISPR-Cas9 enables high precision editing even of large eukaryotic genomes. This section includes a few examples of methods and applications that benefit from our increased ability to edit genomes' DNA sequences.

## 5.1 Genetic modification of livestock and crops

By transferring genetic material between organisms, it is possible to develop microorganisms, animals and plants that carry traits from other organisms (transgenic organisms). One of the most widespread commercial examples is the transfer of bacterial genes to plants that confer resistance to the herbicide glyphosate, allowing transgenic crops such as corn and soybeans (but not non-resistant weeds) to grow in fields sprayed with glyphosate. Unfortunately, somewhat analogously to the development of antibiotic resistance in bacteria, the evolutionary pressure of the massive global use of glyphosate has resulted in many common weeds developing glyphosate resistance.

In classical breeding, variants, strains or lines with desired traits (e.g. rapid muscle growth in cattle or cold tolerance in crops) are identified and crossed to generate offspring expressing those traits more strongly, or in new combinations, over multiple generations. Although thoroughly proven over the ages, this method of breeding is very time consuming. One way to accelerate the process, particularly in plant breeding, is to increase the mutation rate by applying mutagenic chemicals or ionizing radiation, then selecting individuals with the desired traits. A drawback with this, widely employed, strategy is uncertainty: the numbers and genomic positions of the mutations are essentially random, and any individuals displaying desired traits may carry unknown numbers of additional mutations in their genomes.

Modern genome editing methods are in many respects superior to classical breeding techniques, as their high precision allows specific modifications of selected genes and minimizes off-target effects. A recent example is the increase in domestic pigs' resistance to African swine fever, which is lethal to domestic pigs but does not kill African warthogs, by selective modification of a single gene of the pigs to resemble the warthog version.[18]

[18] http://www.theguardian.com/science/2015/jun/23/could-these-piglets-become-britains-first-commercially-viable-gm-animals. Accessed: 2015-10-25.

### Transgenic animals as models for human disease

In medicine and biomedicine, laboratory animals such as mice or guinea pigs are used as "models" to study human diseases or conditions in ways and at levels of detail that would not be possible with human patients. Today, transgenic animal models are available for a wide range of human conditions, including obesity, neurodegenerative disorders and various forms of cancer. The scientific and technological developments of the last decade, with increased knowledge of the genetics of human disorders and genomes of potential model animals, now allows the design of increasingly accurate and robust experimental animal-based human disease models.

## 5.2 Genetic pest control - biopesticides

### Transgenic crops

One of the most established applications of genetic pest control is in so-called Bt-crops. The name stems from the bacterium *Bacillus thuringensis*, which expresses a range of proteins that are naturally toxic to many insect species. When genes encoding these toxins are successfully transferred into economically important crops such as potatoes, cotton and rice, the plants express the toxins in their leaves, roots or stems, thereby killing any parasitic insects. Unfortunately, the development of resistance is increasing, rendering the plants again susceptible to insect attacks. However, such transgenic crops are widely used in many countries, including the USA, India, Brazil and Argentina. In the European Union approximately 40 strains of pest-controlling plants have been approved, but use of GMO crops is low in most member countries (an exception being Spain).[19]

### RNAi – genotype specific pesticides

A more recent development in the field of biopesticides is the use of RNA interference (RNAi) or "silencing" to kill pests.[20] In this process modified RNA molecules inhibit the expression of targeted genes in the pests, hence species-specific agents that kill pests but do not harm beneficial organisms can be constructed. Furthermore, as RNAi constructs are applied to the outside of protected plants and are not integrated into their genomes, the treated plants are not considered GMOs in regulatory frameworks. A drawback with the method is that the required levels of specificity are difficult to obtain, demanding extensive research and validation efforts to ascertain that unwanted interactions do not occur.

[19] http://ec.europa.eu/food/dyna/gm_register/index_en.cfm Accessed 2015-11-18

[20] Antonio Regalado (August 11, 2015). "The Next Great GMO Debate" *MIT Technology Review* (Massachusetts Institute of Technology). Accessed 2015-11-24.

### Gene drives – genetic modification of wild populations to limit their ability to spread disease

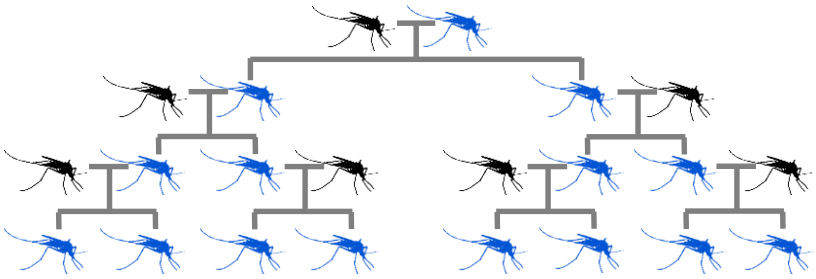

**Figure 4.** Gene drives (blue) always ends up in all offspring, even if only one parent is a carrier. This results in the drive eventually spreading through the entire population, given enough generations.

The term gene drive is based on the notion of so-called selfish genes. These are genes that show biased inheritance, i.e. inheritance at frequencies higher than 50%, and thus faster than normal spread through a population. Diverse genomic alterations can be potentially spread in a population via this strategy, including additions, disruptions, or modifications of genes. A proposed application for gene drives is to reduce the propensity of wild populations of mosquitoes or other harmful organisms to spread diseases[11]. In addition, if genes involved in the reproductive cycle were manipulated, either to reduce the reproductive capacity or to bias the progeny sex distribution, a gene drive could be used to cause a population crash[21], which could be potentially used (for instance) as an improved means of pest control at military camps.

### Genetically modified microorganisms

Genetically modified microorganisms (GMMs), a subset of GMOs, are extensively used as tools in biotechnology. Common applications include uses of modified bacteria and yeast cells in the manufacture of numerous products, e.g. proteolytic enzymes used in laundry detergents and drugs such as antibiotics. Even complex synthetic pathways can be isolated and transferred, as illustrated by promising results of attempts to transfer the ability to synthesize morphine from the opium poppy to modified yeast strains[12] [13].

### Gene therapy

Gene therapy is essentially a way to treat disorders caused by genomic errors by introducing, into the cells of a patient, DNA or RNA that compensates for or specifically corrects the errors. Even though gene therapy is a promising treatment option for a number of diseases (including inherited disorders, some types of cancer, and certain viral infections), the technique has raised both patient safety and ethical concerns. Consequently, although approximately 2000 clinical trials of gene therapy techniques have been performed or approved since the U.S. Food and Drug Administration first approved clinical experiments in 1990, gene therapy is still considered highly experimental, and until 2012 when the European Commission approved the drug Glybera (Alipogene tiparvovec), no gene therapy drug had been approved in the USA or Europe. Glybera treats the rare genetic disease lipoprotein lipase deficiency by delivering a functional copy of the gene encoding the protein lipoprotein lipase to muscle cells, which then produce a functional protein.

[21] http://wyss.harvard.edu/staticfiles/newsroom/pressreleases/Gene%20drives%20FAQ%20FINAL.pdf. Accessed 2015-10-27

# 6 Synthetic genomes

The feasibility of creating entirely synthetic organisms in the laboratory was first demonstrated by the synthesis of polio virus and bacteriophage X174 genomes, as reported in 2002 [14] and 2003 [15], respectively. Bacterial genomes are bigger and more complex. Consequently, the first synthesis of the entire genome of a replicating bacterium — the 1.08 Mb genome of *Mycoplasma mycoides* JCVI-syn1.0 by the J. Craig Venter Institute — was not reported until 2010 [16]. Eukaryotic genomes are even more difficult to synthesize, mainly due to their size. Nevertheless, the synthesis and integration of a modified variant of chromosome III of the yeast *Saccharomyces cerevisiae* was recently published [17].

While it is still far easier to read than write DNA, the technology for synthesizing DNA is rapidly developing. Although the synthesis of a large genome is still a major undertaking, smaller examples such as viral genomes can be synthesized rapidly, cheaply and reliably, with several commercial actors providing gene synthesis services.

### Designing biological machines

The emerging discipline synthetic biology aims to design novel organisms and redesign existing organisms to perform novel functions. Possible applications for such organisms include detection of explosives and chemical agents [18] as well as disease diagnosis and therapy. In another recently developed application (tested in a mouse model) a probiotic strain of *E. coli* was designed to selectively colonize tumour tissue, and produce specific marker molecules that can be detected in the urine when metastasis is present [19].

### DNA as an information storage medium

The major function of DNA is to store information and a physically minute amount can contain vast amounts of data. Consequently, as the technology for synthesizing, reading and storing DNA evolves and becomes more accessible, researchers are exploring the potential for storing digital data using DNA. According to a study published in 2013, the estimated information storage density of DNA is approximately 2.2PB/g DNA [20].

# 7 Implications

## 7.1 Legal, ethical and social aspects

### Genetic integrity

In 2013 the identities of a large number of individuals who had participated in the 1000 Genomes Project were uncovered, although they had been guaranteed anonymity, in a test of the possibility of de-anonymizing seemingly anonymous genomic data by researchers at the Whitehead Institute of the Massachusetts Institute of Technology (MIT). The strategy used was called “surname inference” and, briefly, involved use of other open access genomic databases where other participants had voluntarily displayed their full identities (including names). Comparison of specific DNA sequences, for example from the Y-chromosomes of male participants, between the anonymous and non-anonymous individuals enabled inference of long-distance genetic relationship between participants in the two groups i.e. first , second and third degree cousins etc. Since surnames are also inherited with little deviation, the researchers could use other publicly available records to track down individuals in the seemingly anonymous group simply by also knowing their age and state of residence [21].

The results of this "surname inference" experiment sparked discussion regarding the extent to which researchers can protect the genetic integrity/privacy of individuals participating in genome sequencing projects. Particular concern has been raised about ongoing projects where genomic data are analysed in combination with health care records of the participants.

Another way to tackle the problem was exemplified by the Personal Genome Project (PGP) initiated by George Church, a genetics professor at Harvard.[22] In the PGP all data, including genomes, environments, traits and medical conditions of the participants are open and accessible to researchers worldwide. So far around 2000 participants have volunteered to make their data accessible, although they run the risk of being identified.

Genetic privacy is not the only ethical problem that investigators face when analysing human genomes. Other dilemmas include: whether or not the results should be presented to the participants; whether they would be able to cope with the knowledge that they may have higher than average risks for illnesses such as heart failure; whether parents should be able to see their children's results, including risks for illnesses later in life; and what should be done if there is an unexpected paternity profile. To summarize, the ethical dilemmas of medical genomics can be condensed into three categories: (1) informed consent, (2) handling of the data, and (3) the return of the results.

### Genetic discrimination

The benefits that genomics may bring to medicine and public health should be weighed against the risks for misuse in society. One apprehension is that insurance companies might exploit genomic information in order to categorize individuals into different risk groups based on genetic profiles. However, one should bear in mind that most, if not all, of us carry genes that predispose us to specific medical conditions, but genes do not entirely dictate our destiny, our life-style and environment also influence our outcomes. Nevertheless, certain genetic variations are connected to serious medical conditions and could therefore provoke various forms of genetic discrimination, including reluctance of insurance companies to provide cover, and pressure from society on parents to abort foetuses with certain genomic profiles.

### Regulations - Who owns your genes?

A much-debated American lawsuit in 2013 concerned a patent held by Myriad Genetics for the human gene variants BRCA1 and BRCA2, associated with predisposition for breast and ovarian cancer, in 2013.[23] Holding a monopoly, Myriad Genetics could charge $3000 for a lab test on these genetic variants, a price far exceeding the actual cost of the service. Critics of Myriad's gene patenting also pointed out that the monopoly both hindered research on those genes generally and inhibited development of more accurate and precise tests. Myriad Genetics countered that invalidating the patent would limit incentives for future investments in biotechnology.

In its judgement, the US Supreme Court ended the practice of patenting human genes, which had been possible for 30 years, declaring that naturally occurring human genes cannot be subject to legal patents.

Ever since the 1980s, genes and genetic elements have been subject to patenting in Sweden. However, current patenting standards for genes are stricter than those of 30 years ago. Today, sequencing a gene or genetic element is no longer thought to be sufficiently inventive to warrant a patent, some kind of innovative added value is also required.

[22] http://www.personalgenomes.org/ Accessed 2015-11-18

[23] http://www.genomicslawreport.com/index.php/category/badges/myriad-gene-patent-litigation/. Accessed 2015-11-18.

### Genetic privacy

The Genetic Information Non-discrimination Act (GINA), enacted in the USA in 2008, is a law that protects the privacy of genetic information for individuals. The first case brought to trial under this legislation, in June 2015, concerned a grocery distributing company in Atlanta, Georgia, that had tried to identify which of its workers had repeatedly left faeces in the company´s warehouse. Two suspected employees had been asked to give cheek swabs. The following DNA tests showed that neither of them was the perpetrator. Soon after the event the employees sued the company for breaching their rights under GINA, and the court awarded them $2.25 million for their suffering.

In Sweden, the law of genetic integrity (2006:351)[24] provides similar protection: for example, an employer cannot ask an employee for genetic information, and it cannot be included in standard insurance agreements.

### Commercial genetic tests

Over-the-counter genetic tests are increasingly advertised in Sweden and abroad. Criticism has been raised concerning both the ethics and accuracy of the tests' predictions of risks for future disease. The end-user is often left with difficult interpretations and choices. Many companies also retain customers' genomic sequences for scientific purposes, as detailed in their terms of service. Regulating such marketing is challenging as much of it is internet-based and inherently transboundary, thus Swedish authorities are advocating international harmonization of e-commerce regulations.

## 7.2 Dual-use aspects

### Misuse of genetic information

Sequencing is becoming cheap and today a single cell provides enough material for sequencing a genome. Indeed, as already mentioned, a single hair or partial fingerprint is enough to identify individuals. As also already mentioned, databases (governmental or private) storing people's genetic information may be less secure than imagined or hoped. Thus, acquiring genetic information from others is becoming progressively easy, which clearly poses security risks. From obtained sequences it is increasingly possible to reconstruct individuals' ancestry, relatedness and traits, like eye-colour and susceptibility to certain diseases. Much of this information could be easily used to discredit individuals or harm them by exploiting identified susceptibilities.

By recording 30 000 facial parameters using 3D scanning combined with genome sequencing entrepreneur Craig Venter is trying improve computer models that can recreate a face solely from DNA information.

[24]Svensk författningssamling, Lag (2006:351) om genetisk integritet m.m., https://www.riksdagen.se/sv/Dokument-Lagar/Lagar/Svenskforfattningssamling/sfs_sfs-2006-351/. Accessed 2015-11-24

### Availability

Our abilities to read, write, modify and design genomes have opened up almost endless possibilities for medicine and biotechnology industries. However, as these technologies become easier to use they no longer require large expensive laboratories. In fact, basic genome engineering can now be performed at home with only modest investments (used research equipment is readily available on the open market). This has led to a Do-It-Yourself (DIY) Biology culture where non-experts work together to produce genetically modified organisms for various purposes.[25] These activities outside traditional research facilities are very difficult to oversee and regulate. Although most members of the DIY community are working on benign projects, some may not be, and risks of accidental releases of GMOs are likely to be much higher than from traditional research facilities.

BioBricks are publicly available DNA sequences with well-defined functions. They can be combined into biological circuits that will perform specific functions when introduced into an organism such as *E. coli*. Today more than 20 000 BioBricks are available to synthetic biologists.

Many commentators predict that synthetic biology will soon be competing with the IT industry in terms of investments, revenue and attraction of creative minds. Furthermore, many countries in Asia and the Middle East that have not traditionally invested much in scientific research are now investing heavily in synthetic biology research.

The possibility of producing synthetic organisms and designed pathogens has been discussed for more than a decade, but in the last few years the possibility has become a reality. The first organisms that do not share a common ancestry with all other life forms have already been produced and organisms (e.g. pathogens) once extinct can now be revived if their genome sequence is known. Since most research is still performed in regular research facilities funded by governments, all data is in general deposited in open access databases, which facilitates both legitimate and (potentially) malicious research.

### Targeted pathogens

The feasibility of producing ethnic weapons has also been discussed for many years, and the technology to produce such weapons is becoming available. We still lack sufficient data on the genetic composition of populations to make such weapons with the necessary specificity, but this may change as more human genome sequences become available through the major sequencing efforts already described in this report. Furthermore, it is already theoretically possible to design synthetic organisms to target a specific sequence variant known to be present in an individual. A prerequisite for such targeting is knowledge of that individual's DNA. Interestingly, some state actors are reportedly already collecting DNA from prominent persons.[26]

In animal models, bacteria and viruses designed to recognize and target specific cell types, e.g. tumours, for destruction are currently being developed. Indeed, a number of organisms and synthetically produced DNA constructs have been designed and tested that can target and kill specific pathogens, such as weeds and insect pests, while leaving other organisms unharmed, according to reports published in the last year. Thus, technology that could potentially be used for making ethnic weapons is maturing.

[25] More information can be found at http://diybio.org/ and https://en.wikipedia.org/wiki/Do-it-yourself_biology.
[26] https://wikileaks.org/plusd/cables/08STATE30340_a.html. Accessed 2015-11-24.

Synthetic DNA constructs that are relatively harmless for a general population, but readily transmittable could in theory be made. These could be encoded with a switch that activates them and transforms them into a harmful version only when in contact with a specific host or when a specific trigger is supplied.

As the agricultural industry learns how to modify crops to increase yields and resistance, it is a short step to use the same technology to create super-weeds. A weed that outcompetes important crops and is resistant to pesticides could seriously impact food production.

Detecting artificially modified or synthetically produced harmful organisms will be very difficult and require advanced sequencing and analytical tools. Pathogens with functional parts that are not found in nature would not give matches to databases with known sequences of known functions. Today, we mostly rely on specific markers to detect naturally occurring pathogens. However, it would be very easy for anyone modifying an organism to remove or sufficiently alter targets of the markers we rely on for detection. This would render any pathogen undetectable with current screening techniques. However, so far surprisingly little investment has been made in new detection techniques.

### The avian influenza case

Influenza A viruses have been circulating in human and animal populations for a long time. Seasonal influenza outbreaks, which mostly occur during wintertime, annually kill about half a million people worldwide. These viruses readily mutate and can sometimes jump between species. In rare cases this leads to global health disasters such as the 1918 influenza pandemic that killed between 50 and 100 million people. More recently the subtype H5N1 or the avian influenza virus that normally infects birds has been infecting humans, resulting in high mortality rates. Fortunately, H5N1 does not transmit well between humans. However, in 2012 two scientific papers were published that described experiments showing exactly which mutations were needed to enable airborne transmission of this virus between ferrets. The aim of these studies was to identify mutations we must be able to detect, and counter, if they occur naturally and potentially cause a pandemic. However, they triggered an immediate public debate about the wisdom of making such results publicly available as they could easily be used to produce a H5N1 virus that could potentially be transmitted between humans. Clearly, risk/benefit analyses involving scientists, policy-makers and the general public are important when conducting and publishing such research [22].

### Mitigating novel threats

Since our abilities to create synthetic organisms and modify existing organisms are rapidly increasing, while the scope for monitoring associated activities remains limited, there is an urgent need to discuss and develop means to handle future outbreaks resulting from deliberate or accidental releases of such organisms at all levels of society. Not only do we need to develop regulations and ethical guidelines, we also need to develop sequencing techniques and analytical tools to detect and evaluate organisms that do not have matches in any databases. Likely, tight surveillance and custom designed counter measures will be required.

# 8 Future perspectives

Some possible developments within genomics conceivable to occur within a couple of decades, and factors that may influence them, are listed below.

## 8.1 Opportunities

**Curing disease.** Most individuals will likely soon be offered routine sequencing. Consequently, many diseases will be preventable before they emerge by altering lifestyle or medication. It will be possible to permanently cure many diseases by fixing the genome before or after birth using gene therapy. It will be possible to tailor drugs to individuals and primary health care will become more automated.

**Predicting traits**. As more data become available, we will get better estimates of genetic variability among individuals. Many physical traits (like facial structures) as well as some abilities and behaviours are very strongly determined by our DNA. It will be possible to predict many such traits in humans solely from their DNA sequences. Furthermore, as our understanding of genomic controls increases we will be able to use other information, e.g. DNA-methylation data (obtained using modified sequencing approaches) to predict ages, source tissues and the health status of samples, which is impossible using solely DNA sequence data. Such predictions could be used by employers, e.g. the military to select suitable people for specific assignments.

**Detection/Identification.** As sequencing in the field becomes routine, soldiers and emergency personnel can be rapidly screened for exposure to hazardous organisms in air or water. It will also be easy to identify remains from victims or casualties directly on site, especially if databases of the personnel's genomes are kept.

**Reshaping humans.** These techniques will not be limited to fixing problems or curing disease. They will also be used to modify human traits (or even introduce new ones) in the same way that we improve crops and livestock. It will be increasingly possible to enhance adults and design babies.

**Improving food production.** The yields and nutritional value of crops and livestock will be increased by modification and improvement of current traits, and/or introduction of new traits. They will also be less sensitive to climate changes, require less space, consume less resources and be more resistant to pests.

**Improving the environment.** A significant amount of biofuels will be produced by genetically modified/enhanced microorganisms that will simultaneously sequester $CO_2$ from the atmosphere. Synthetically designed microorganisms will enable efficient clearance of toxic waste and provide cleaner sources of chemical products.

## 8.2 Considerations

**Not everything is encoded in the genes.** Most complex traits such as our personality have both genetic and strong environmental components. We will not therefore be able to design or predict all types of traits in humans or other organisms without fully controlling the environment.

**The environment is a complex system.** It is very difficult to predict long-term consequences of actions such as releasing synthetic organisms into the environment or allowing offspring to inherit genetic modifications of humans.

**Limited understanding.** We still have far from complete understanding of single cells' complex processes, and even less complete understanding the molecular processes of entire organisms'. We are unlikely to develop the ability to design complex traits or

networks in the near future. However, optimization through simulated evolution on an industrial scale may provide substantial steps towards such goals.

**Legal/ethical issues.** Public opinion and legislation can heavily influence the speed and funding of developments within this field. For example, GMO research in Europe is falling far behind the rest of the world because of very strict legislation and negative public opinion. Similar obstacles to stem cell research have recently been raised in the USA. It is therefore difficult to predict rates of development, but most technological advances will probably be tested and eventually applied somewhere, since global regulation and control systems are unlikely to be agreed and introduced in the near future. Thus, as biology and medicine move away from discovery and into design and engineering phases, economic interests, intellectual rights, patents and large corporations will probably control much of the development.

**Handling of data/samples.** Vast amounts of data need to be safely stored for future generations, which is already posing problems. Processing and interpreting data are becoming increasingly complex tasks as larger functional networks are considered. Similarly, complex new privacy and security issues will be raised as we will be able to predict traits, personality and disease risks from DNA data increasingly accurately.

**Biological warfare.** It will become possible to design very specific pathogens that infect subpopulations or individuals (humans, livestock or crops), but are both harmless to others and virtually undetectable. These could include synthetic organisms with very complex programmed responses, such as capacities to detect complex stimuli and respond by triggering tailored attacks. The only defence will likely be to specifically design countermeasures to specifically designed threats after detection. This will require tight surveillance and very short response times.

# 9 Concluding remarks

Clearly sequencing technologies and their applications will continue to evolve rapidly. Sequencing projects will become increasingly large. Some projects and the acquired data will be accessible to the public while others will likely be classified by companies or the military. Within five to ten years we could have sequenced most humans and representatives of a large fraction of all other species. We will likely also see applications of sequencing technologies that will initially seem like science fiction, but will quickly become routine. Ideas that are already being evaluated include (among many others) constructing automated primary healthcare units, installing sequencers at home to keep track of our health and dietary needs, use of DNA sequences as real-time secure biometrics, tracking origins of food products, and monitoring pathogens and pests in both water and air. In addition, as the value of the sequencing market grows, more private enterprises will enter it, further increasing the speed of development.

We strongly believe that genomics coupled to the emerging genome editing technologies has the potential to heavily transform our society in areas ranging from medicine and forensics to security and warfare. Tightly linked to the technological development is the need for continuous development of legislation and ethical debates on a global level as technology and data become accessible to everyone. We hope that politicians, civil and military service authorities at both national and international levels will follow and adapt to the rapid progress within the field of genomics.

# 10 Acknowledgements

We would like to thank Anders Lindblad, Elisabeth Bemm, Mats Forsman and Andreas Sjödin for ideas and constructive criticism during the writing of this report.

# 11 References


1. Gaziano, J.M., et al., Million Veteran Program: A mega-biobank to study genetic influences on health and disease. J Clin Epidemiol, 2015.
2. Yang, X., Applications of CRISPR-Cas9 mediated genome engineering. Mil Med Res, 2015. 2: p. 11.
3. Fagertun, J., et al., Predicting facial characteristics from complex polygenic variations. Forensic Sci Int Genet, 2015. 19: p. 263-8.
4. Lax, S., et al., Forensic analysis of the microbiome of phones and shoes. Microbiome, 2015. 3: p. 21.
5. Tridico, S.R., et al., Metagenomic analyses of bacteria on human hairs: a qualitative assessment for applications in forensic science. Investig Genet, 2014. 5(1): p. 16.
6. Fleischmann, R.D., et al., Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science, 1995. 269(5223): p. 496-512.
7. Mellmann, A., et al., Prospective genomic characterization of the German enterohemorrhagic Escherichia coli O104:H4 outbreak by rapid next generation sequencing technology. PLoS One, 2011. 6(7): p. e22751.
8. Afshinnekoo, E., et al., Geospatial resolution of human and bacterial diversity with city-scale metagenomics. Cell Systems, 2015. 1(1): p. 72-87.
9. Stephens, Z.D., et al., Big data: Astronomical or genomical? PLoS Biol, 2015. 13(7): p. e1002195.
10. Karlsson, E., et al., Scaffolding of a bacterial genome using MinION nanopore sequencing. Sci Rep, 2015. 5: p. 11996.
11. Gantz, V.M., et al., Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. *Proc Natl* Acad Sci USA 2015.
12. DeLoache, W.C., et al., An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. Nat Chem Biol, 2015. 11(7): p. 465-71.
13. Fossati, E., et al., Synthesis of morphinan alkaloids in Saccharomyces cerevisiae. PLoS One, 2015. 10(4): p. e0124459.
14. Cello, J., A.V. Paul, and E. Wimmer, Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. Science, 2002. 297(5583): p. 1016-8.
15. Smith, H.O., et al., Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. Proc Natl Acad Sci USA, 2003. 100(26): p. 15440-5.
16. Gibson, D.G., et al., Creation of a bacterial cell controlled by a chemically synthesized genome. Science, 2010. 329(5987): p. 52-6.
17. Annaluru, N., et al., Total synthesis of a functional designer eukaryotic chromosome. Science, 2014. 344(6179): p. 55-8.
18. Antunes, M.S., et al., Programmable ligand detection system in plants through a synthetic signal transduction pathway. PLoS One, 2011. 6(1): p. e16292.
19. Danino, T., et al., Programmable probiotics for detection of cancer in urine. Sci Transl Med, 2015. 7(289): p. 289ra84.
20. Goldman, N., et al., Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. Nature, 2013. 494(7435): p. 77-80.
21. Gymrek, M., et al., Identifying personal genomes by surname inference. Science, 2013. 339(6117): p. 321-4.
22. Fauci, A.S. and F.S. Collins, Benefits and risks of influenza research: lessons learned. Science, 2012. 336(6088): p. 1522-3.
23. Plomin, R and I.J. Deary, Genetics and intelligence differences: five special findings. Mol Psychiatry, 2015. 20(1): p. 98-108

FOI, Swedish Defence Research Agency, is a mainly assignment-funded agency under the Ministry of Defence. The core activities are research, method and technology development, as well as studies conducted in the interests of Swedish defence and the safety and security of society. The organisation employs approximately 1000 personnel of whom about 800 are scientists. This makes FOI Sweden's largest research institute. FOI gives its customers access to leading-edge expertise in a large number of fields such as security policy studies, defence and security related analyses, the assessment of various types of threat, systems for control and management of crises, protection against and management of hazardous substances, IT security and the potential offered by new sensors.

FOI
Defence Research Agency
SE-164 90 Stockholm

Phone: +46 8 555 030 00
Fax: +46 8 555 031 00

www.foi.se