

The impact of genetic data on society and its effects

Massimo Calabrigo

12247382

Contents

1	Introduction	3
2	Introduction to the human genome	4
2.1	The structure of human genome	4
2.2	Genes and Genetic Diseases	5
2.3	Behavioral traits	6
3	Genetic Testing	7
3.1	History of genetic testing	7
3.2	Genetic testing in the age of AI	8
3.3	Genetic Counsellors and Direct-to-consumer companies	8
4	Economy and value of genetic data in the long term	10
4.1	Economic value of genetic testing	10
4.2	Major companies and lobbying	10
5	Gene Patents	12
5.1	Patents in Europe	12
5.2	Patents in US	12
5.3	Differences on EU-US patents	12
6	Genetic Literacy	13
6.1	Literacy	13
7	Genetic Discrimination	15
7.1	Disclosure of a person genome reveals information about its family	15
7.2	Genetic discrimination and its effects	15
7.3	Genetic discrimination in school	16
8	Governments abuse of genetic data	17
8.1	China: a huge male genetic databases is being built without citizens consent	17
8.2	China: Uighurs crime prediction based on genes	17
9	Genetic Data: differences with other types of personal data	19
9.1	Anonymization is not possible for genetic data	19

10 Current laws regulating genetic data and genetic testing	20
10.1 <i>GINA</i> and <i>ACA</i>	20
10.2 <i>GDPR</i>	20
10.3 Genetic Data: Problems in current legislations	21
11 Addressing genetic data abuse on the AI era	22
12 Bibliography	23

1 Introduction

"The sequence of the human genome would be perhaps the most powerful tool ever developed to explore the mysteries of human development and disease" - Leroy Hood, scientist working at the Human Genome Project [1] and creator of the first automated DNA sequencer [7]

The human genome project started in 1990 and lasted 12 years costing more than 3 billion dollars with the objective to sequence the whole human genome and use it to greatly improve medicine and disease prevention [1][3][5]. Not only now we are able to full sequence DNA for 600 dollars in less than 3 weeks [2], but we have also discovered that we can identify behavioral traits like self-directedness, cooperativeness, or self-transcendence [4].

This has profound social and ethical implications for the future of our data-based society; genetic data is a gold mine for AI algorithms and it's effects can range from cancer prevention [5] to insurance discrimination [6]. Our goal is to clearly state problems about how our society understands genetics in both legal and social/ethical contexts, and formulate some suggestions aimed at improving the current understanding, preparing our society to a future where genetic data will be as common as biometric data.

In section 2 we introduce some basics about genetics necessary to understand what information can be extracted from a person DNA, focussing the attention on the probabilistic nature of genetic testing results, which seems to be cause of much confusion [8]. Then, in section 3, we address "Genetic Testing", a set of techniques used to extrapolate information from genetic data that ranges from simple string comparisons to complex AI algorithms. We then proceed analyzing the economical aspects in section 4, in particular we will look at the growth of genetic testing from the Human Genome Project up to now, and in section 5 we will look at the discussion about patenting genes.

In sections 6 and 7 we will look at the social and ethical aspects: from the current awaness of the general public about genetics and it's effects, to the problem of genetic discrimination through the usage of advanced AI models.

In section 8 two disruptive examples of abuse of genetic data by both the Chinese and US governments are supplied, in section 9 we will see why genetic data can't be treated as other types of personal data protected by current legislations, and in section 10 current norms and regulations for US (GINA, ACA) and Europe (GDPR) are discussed.

Finally, we will end with some recomendations on how to address a society in which genetic data is commonplace are made in section 11, and both positive and negative aspects of genetic testing and data are discussed.

2 Introduction to the human genome

In this chapter we are going to introduce concepts as *human genome*, *chromosomes*, *genes*, *SNP* and *genetic traits*, in order to better grasp how information is encoded and how it can be extracted by algorithms.

2.1 The structure of human genome

The Human genome is a collection of 46 chromosomes, 23 inherited from the mother, 23 from the father. Chromosomes are contained inside the *nucleolus*, an inner organule inside the *nucleus* of every human cell, from neurons to cells in your mouth; even if every single cell has a full copy of the genome.

Each chromosome is composed by tangled chromatine and hystons, which are structures that enables information to be stored more densely, and chromatine is in turn composed by a double helix, also called *Deoxyribonucleic acid* or *DNA* 1.

The basic units of information are the 4 bases that composes DNA, namely *adenine* , *cytosine* , *thymine* or *guanine* (A,C,T,G)[5]. These bases are molecules which works like pieces of a puzzle, and links as pairs, linking the two helixes which composes DNA (See Fig. 1).

There are around 3.2 billion bases in the human genome [5], so a full genome sequence is roughly 725 MB of data stored as strings of letters from the alphabet "A,C,T,G" [6], and also it is interesting to note that about 1%-2% of the human genome is considered composed of *coding sequences*, that is sequences, also called *protein-coding genes* that codes protein production. The rest of the human genome is formed by highly repetitive sequences and is called *non-coding*, and it is still debated whether or not this part of the genome is useless or not, but recent discussion seems to point toward assigning some function to it. [31]

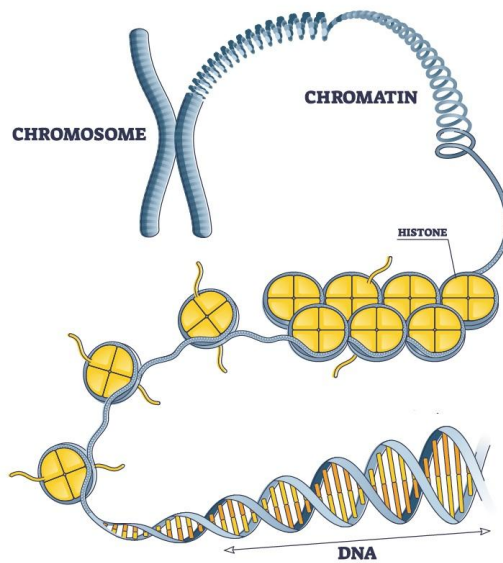


Figure 1: Chromosome structure. We can see that chromosomes are composed of chromatin and that histones allow chromatin to maintain a dense structure

2.2 Genes and Genetic Diseases

Genes are subsequences of DNA and it is estimated there are between 20,000 and 25,000 genes in the whole human genome. Genes are like cookbooks that can be read and copied, and most of them are used for protein production; these genes are copied by the *mRNA*, and then brought to the *ribosomes*, also called "factories of proteins", for protein production.

Although each cell contains a full complement of DNA, cells use genes selectively. For example, a neuron will use a subset of genes different from a liver cell.

Genetic diseases are mostly influenced by genes, that through proteins determine the efficiency of metabolization of food, how toxins are eliminated, . . . There are both harmful, and not harmful gene mutations:

- single-nucleotide mutation (SNM): Harmful changes in genes, most commonly misspelling. For example, progeria, a disease which causes the patient to age much faster, is caused by a SNM, which causes the gene that encodes the information about protein "lamin" to encode "lamin-a" instead, which in turn prevents the formation of the inner barrier of the *nucleolus* [5];
- single-nucleotide polymorphism (SNP): Naturally occurring changes in genes that are part of what distinguishes individuals, these changes are not harmful, for example genetic variants in a single gene account for the different blood types: A, B, AB, and O. A single individual can carry millions of SNPs.

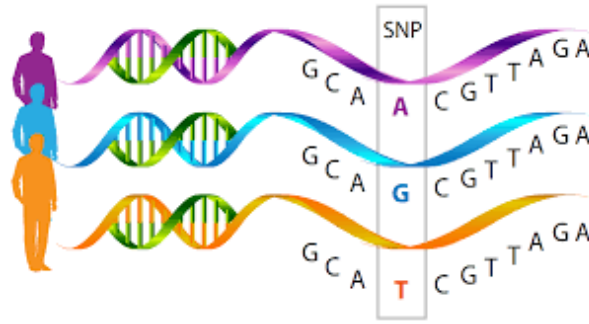


Figure 2: DNA mutations and polymorphisms. Changes in the DNA are usually not a threat, and may cause not noticeable effects. The most dangerous DNA mutations are the ones that affects genes responsible for protein production, like laminin-A which causes progeria [5]

2.3 Behavioral traits

Behavioral genetics is a strand of genetics that study the effect that genes have on people personality. Multiple studies that stated that the genetic part counted more than the enviromental part on the shaping of one's personality have been criticized, in fact, the scientific community seem to be more conservative [26].

A recent broad study gives a less conflictual result: behavioral traits around 30 to 60 percent heritable, depending on the type of trait [4].

Knowing that behavioral traits are heritable does not mean that your personality is inherited too. Surely a predisposition toward "aggressivness" can skew ones personality toward violence, but it also depends on the enviroment on which one is grown and lives; as the study says the heritability does not go higher than 60 percent. It is also important to remark that behavioral traits are not fixed in stone, and that major life events and persistent intervantion can change them as demonstrated by multiple psycological and neuroscientific studies [27].

3 Genetic Testing

Full Genome Sequencing, like the Human Genome Project [1] consists in reading the bases of every chromosome and storing it into a computer, enabling the application of different techniques and algorithms collectively called *Genetic Testing*. In this chapter we will take a look at the history of genetic testing and the figure of the genetic counsellor that is emerging in western democracies.

3.1 History of genetic testing

Historically, the presence of genetic diseases was inferred using the inheritance law about genetic diseases inherited from the family [5], and the presence of symptoms that could be reconducted to that genetic disease or predisposition, for example multiple miscarriages, recurrent childhood death in the family or multiple family members suffering from cancer.

After the human genome was sequenced, and it became more affordable, also genetic testing was added in the form of matching the string representing a patient gene, and the string representing an healthy gene. With this technique we were able to look for specific patterns in the DNA and check for mutations in genes responsible for the production of some protein which in turn caused some specific disease.

In section 2 we saw the example of progeria, it is estimated that over 300 million people in the world suffer from some form of a rare disease [3], and the time necessary for a person to be correctly diagnosed is between 3 and 5 years in industrialized countries, greatly reducing life expectancy due to the fact these diseases gets worse with time if not treated.

Genetic testing can also be used to predict the predisposition of some types of cancer, which, alike genetic diseases, if caught in time it's impact on a person's life can be greatly reduced. [2] Another strand regards genetic testing on embryos and fetus which are optional in the EU and USA legislations [25]. Embryos stored with in-vitro techniques can be genetically tested to check for genetic diseases, predispositions toward diseases, probability of miscarriages, physical abnormalities, ...; in that case only the "healthier" embryo can be implanted on the mother womb. This technique only works on in-vitro embryos, but genetic testing can also be applied to fetus, which unbinds genetic testing from in-vitro techniques and allows it to be extended to natural pregnancies aswell. [2]

Another strand about genetic data usage is about precision medicine [2], which is a new school of thought which states that medicine should be tailored around an individual, because each person has different biological characteristics. Genetic testing links to precision medicine because it allows the emergence of new fields like "pharmacogenetics" that is assigning to each individual an exact dosage of medicines instead of the stands pills, based on both biometric and genetic data.

3.2 Genetic testing in the age of AI

As we have seen, historically genetic testing was able to use genetic data to discover predispositions and traits regarding possible diseases and health problems.

Usually determining the causal effect between a genetic mutation and a genetic effect followed the following process:

1. individuation of a common genetic mutation in patients sharing a certain predisposition or disease
2. laboratory tests and studies are conducted on a sample of patients to ascertain what exactly that gene mutation modifies with respect to a non-mutated gene
3. a study is published, together with an analysis of the causal relations between mutation and disease

Individuation of genetic traits like "aggressiveness" or "introversion" are much harder to analyze because genetic traits are much more complex, less understood and difficult to measure [4].

This is where artificial intelligence (AI) steps in, by feeding genetic data to large algorithms it will be possible to look for complex patterns, linking a specific behavioral trait to some genetic sequence without the need of uncovering the precise biological effects that that mutation causes to the organism in order to produce that effect.

Of course, this approach creates ethical problems, for example are we sure that the algorithms will learn the correct linkage between a genetic sequence and a predisposition, as it was amply demonstrated this is not always the case.

In the cases where AI uncover correct relationships we have to remember that behavioral traits are composed of an environmental and a genetic part, and that the latter determine from 30% to 60% of a person's trait [4], depending on the trait.

Modern AI represents a step forward for genetic testing, it can be used both to look for more complex cancer predispositions or genetic diseases, or to predict predisposition toward crime based on ethnical and geographical location by using genetic data together with biometric and web navigation data, its uses must be regulated in order to avoid discrimination as we will see in section 10, but it is also an instrument that could lead to saving many lives.

3.3 Genetic Counsellors and Direct-to-consumer companies

Genetic counselors are professional figures trained in genetics which helps a patient or a family interpret genetic family history and genetic data [5], their purpose is to present difficult to comprehend patterns and results about a person's genome into clear and easy to comprehend information, and to propose changes in the lifestyle in order to minimize the probability of contracting some disease the patient is predisposed toward or other

healthcare professionals in order to treat genetic diseases the patient may have. Genetic counselors are also tasked with reducing a patient anxieties toward the results of an exam, and to give the patient psychosocial tools to cope with these results, in fact, genetic counsellors follows a patient for multiple sessions and do not tell patients what to do, but take a non-directive stance and support the patient's decisions, so they act as intermediaries.

In the last session genetic counsellors gives patients a document containing the major topics discussed, and the patient can decide whether to share it with third parties or not.

Given that the exam results are not deterministic, but always involves some probability or predisposals, it is highly suggested for a patient to sequence and analyze its genome thought a genetic counsellor, otherwise that patient may not be able to understand the results, and feel emotional distress and economic consequences [5].

Nowadays also private companies like 23AndMe [32] are offering both sequencing and genome data analysis, in this case a patient is given access to a web page which summarize pharmacogenetics, cancer predisposal, ancestry, allergies and other genetic-related traits.

The problem with direct-to-consumer companies is that the patient is given a lot of information and few instruments to interpret it, possibly leading to a wrong interpretation and an inadequate response. The problem of interpreting correctly genetic data is also compounded by the low level of genetic literacy in our society, of which we will talk about in section 6, that is people who are missing the basic concepts we explained in section 2.

4 Economy and value of genetic data in the long term

4.1 Economic value of genetic testing

As we can see in Fig. 3, direct-to-consumer companies are gaining more and more value, and the reduction in cost of genome sequencing is getting cheaper faster than the moore law.

Based on current estimates, the global genomic sequencing market is expected to grow from its previous 10.7\$ billion valuation in 2018 to 37.7\$ billion by 2026. Thus the genetic testing market is predicted to experience a compound annual growth rate of 19.1% for 2021-2026. [2]

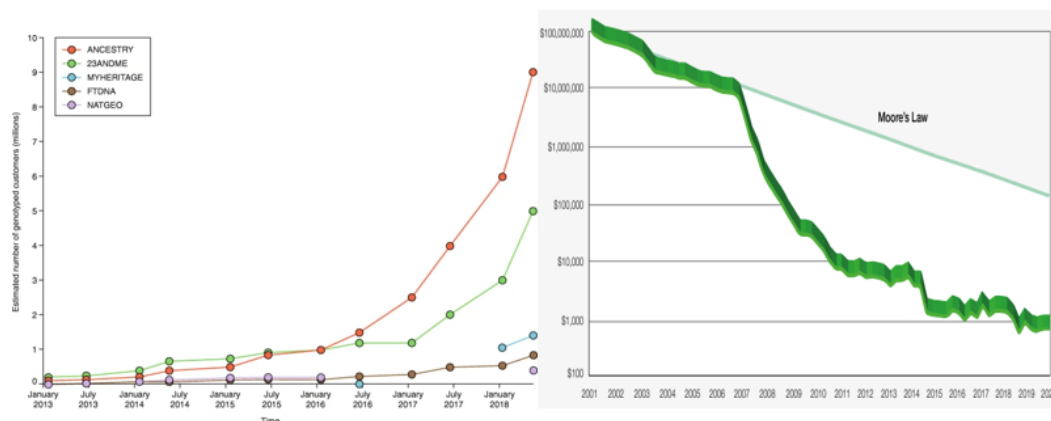


Figure 3: On the left economical growth of direct-to-consumer genetic sequencing companies, on the right reduction in cost for full genome sequencing from 2001 to 2020

4.2 Major companies and lobbying

Gordon Thomas Honeywell is a lobbying company, that represents the interests of big corporations in the US government. In a presentation in the HIDS conference in 2016, a Gordon Thomas Honeywell proposed that DNA databases should also come to the democracies of the world, praising saudia arabia and african dictatorships like Qatar for being so advanced in the development of mass genetic databases and describing parliamentary discussions and protests as "hurdles" [15]. The fact is that Gordon Thomas Honeywell represents Thermo Fischer, the second biggest DNA sequencing company by revenue [33].

Thermo Fischer produces DNA sequencing machines, and saw how much profitable was mass surveillance when in 2017 saw that more than 10 percent of their revenues came from China. [16]

Thermo Fischer is an example of the danger that can come when genetic testing tech-

nology is applied by a government in order to obtain a mass surveillance state. We will see more about its relationships with China in section 8.

5 Gene Patents

In this chapter we will talk about patents for genetic testing techniques. Broadly speaking, both in EU and US, patenting genes occurring in nature is forbidden, but it is possible to patent genetically modified genes at certain conditions.

5.1 Patents in Europe

In Europe genes patent have not been allowed since 1998, with the European Union directive 98/44/EC [14] the concept of *biological material* was defined as any material containing genetic information and capable of reproducing itself or being reproduced in a biological system. [11]

By Biotech Directives biological material that is isolated from its natural environment or produced by means of a technical process may be the subject of an invention, even if it previously occurred in nature. [11]

5.2 Patents in US

Until 2013 genes could be patented in the USA, but in the case of the *Myriad case*, the Supreme Court of the United States ruled that human genes cannot be patented in the U.S. because DNA is a "product of nature" [11].

The Court decided that because nothing new is created when discovering a gene, there is no intellectual property to protect, so patents cannot be granted. Prior to this ruling, more than 4300 human genes were patented. The Supreme Court's decision invalidated those gene patents, making the genes accessible for research and for commercial genetic testing. [9][11]

Differently from Europe, the *Myriad case* also banned patents on isolated biological material stating that "nothing new and useful is created when a gene is located/discovered or simply isolated" [11].

5.3 Differences on EU-US patents

Europe and USA legislations are not much dissimilar, in both cases patents duration is capped at 20 years, and in both legislations artificial DNA constructs (cDNA) and sequences altered by humans remain patentable, as they are not naturally found in nature. [11]

USA patents legislation is slightly more strict than EU, banning patents on genes recreated in isolated environment make it so some patents could be valid in Europe but not in USA.

6 Genetic Literacy

In this chapter we will see how the general public and the judicial systems are aware of the basic genetics concepts we introduced in section 2.

6.1 Literacy

In 2022 a study was conducted for estimating genetic literacy [8], that is the knowledge about genetic-related concepts. The population comprised 2023 individuals, that were asked to state both their perceived familiarity with concepts like "chromosome", "gene", "mutations", and also some specific questions about the genome (See Fig. 4).

The results showed that genetic literacy has not yet been achieved, but this study also noted that confronted with their results from 2013 with the same set of questions, showed a 10% increase in accuracy, hinting that genetic literacy is slowly increasing.

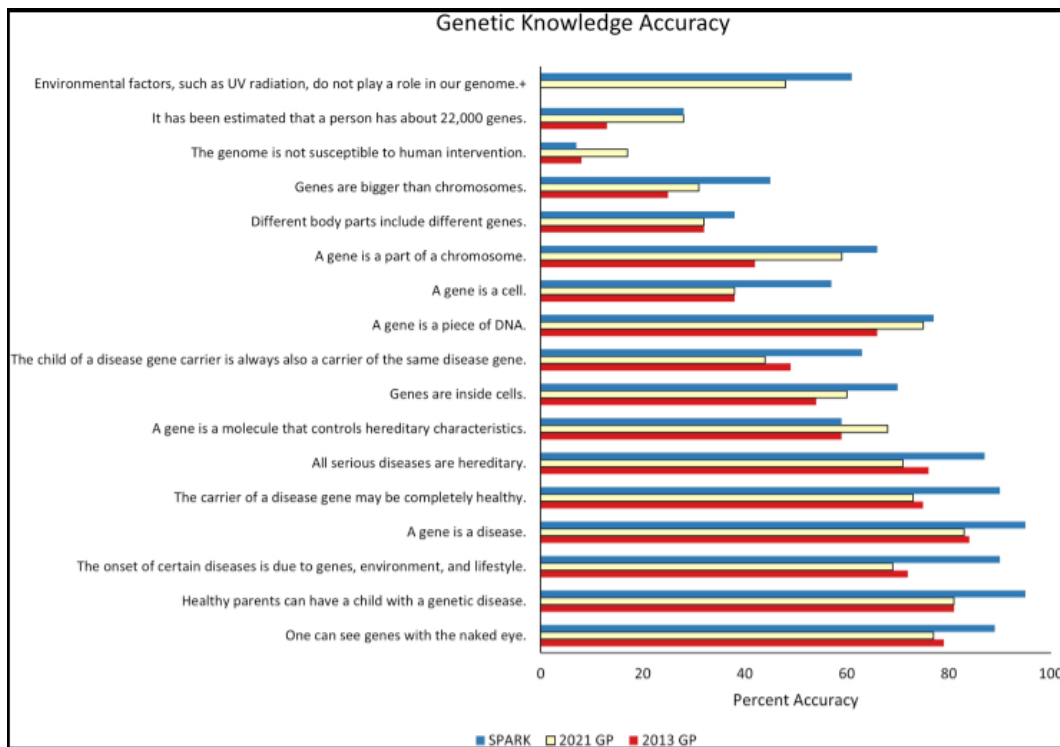


Figure 4: The questions ranges from hard questions like "How many genes are estimated to be present in the human genome?" to easy questions like "Are genes diseases?". We can see the aaccuracy is around 45% for the easy questions, and 80% for the hard questions

Genetic literacy will be essential in the future in the judicial system. Because ge-

netic data can give a predispositions toward traits like "aggressivness", that could lead a judge to impose harsher than necessary conditions to the defendant or create other unjust situations based on ignorance about genetics from the judge and lawyers sides [6], also, ignorance about genomics from the part of legislators has been signaled as the main barrier toward the global implementation of genomic medicine [22], which includes *pharmacogenetics* and many disease prevention techniques.

From studies about genetic literacy [8] and cases in the judicial systems [6] it emerges the lacking of knowledge about genetics, which is both a problem for the application of new medical practises [22] and the right of not being discriminated.

A study published in 2019 in the *Legal Issues Journal UK* proposes to both promote education among the general population and introduce genomic education for lawyers and judges as part of their professional development programmes, also in their law degree. [6]

7 Genetic Discrimination

In this section we will see the discriminating effects genetical testing and genetic data could have on society, and remind that genetic discrimination is already here using the example of a boy which was banned from school due to a predisposal toward a genetic disease.

In section 8 we will take a look at genetic discriminations performed by governments at a massive scale which are already present in both US and China.

7.1 Disclosure of a person genome reveals information about its family

The predictive nature of genetic information allows prediction of genetic traits for the extended family, or even a whole community (we will say that genetic data holds the property of *estensibility*), for this reason it raises unique social issues [5].

Sharing genetic information with one own's family has a series of social consequences: unaffected spouses may view their partners differently, siblings of children with special needs may see themselves ignored by the parents or [5].

Genetic testing performed on a person could also reveal predisposal toward genetic diseases which affects also family members, hence posing the problem of whether to reveal this information or not.

Genetic Testing can also affect a community at large, in the past genetics have been used to discriminate based on race or ethnicity, with the result that already marginalized communities would see genetic practises such as prenatal diagnosis screening as an instrument to seclude them even more, hence refusing them and regating themselves also the positive healthcare effects.

Some other community, like the deaf community, have a different view on what they treat as a positive or a negative genetic trait, and would feel discriminated if deafness would be simply categorized as negative trait. [6]

7.2 Genetic discrimination and its effects

Genetic discrimination occurs when people are treated differently by their employer or insurance company because they have a gene mutation that causes or increases the risk of an inherited disorder. Fear of discrimination is a common concern among people considering genetic testing. [10]

A worrisome usage of genetic data in the commercial setting is its usage for hiring processes. Given that the DNA results are probabilistic, above all the behavioral ones, it isn't a given that the best person for a role would be selected. Moreover, this fatalistic view it imposes on the society could have unexpected consequences on the people psy-

chology, imposing constraints on capable people that don't have the best gene selection, hence it would increase discrimination without really giving the certainty of having selected the best candidate for a job.

Also another source of discrimination sprouts from the fact that the genetic traits linked to anti-social behavior are highly inheritable [6], which means that a state could store that data and use it to predict the likelihood of an individual committing a crime, acting as the environmental factor which will only further increase the likelihood of that person being a criminal, and becoming the main source influencing this outcome.

7.3 Genetic discrimination in school

Genetic data could also be used for augmenting streaming in schools, which is a process that consists in examining students in order to group them in classes with similar abilities, and a lot of literature argues that streaming would not really improve learning, but just creates disadvantages for many students. [12]

An example of genetic discrimination in school already happened, a boy in california was expelled because he has a genetic predisposition to cystic fibrosis [13]. The boy was expelled because two sibling with cystic fibrosis were already present in the boy middle school, and given that kids with the inherited lung disease can't be near each other because they're vulnerable to contagious infections, the school direction decided to expell the boy even if he didn't have the disease, but was just predisposed toward it.

8 Governments abuse of genetic data

8.1 China: a huge male genetic databases is being built without citizens consent

Chinese government is building a huge DNA database containing only male DNA, which is already 80 million strong and its stated objective is to catch criminals and the police says that they obtain DNA samples from voluntary subjects. [18]

Some officials within China, as well as human rights groups outside its borders, warn that a national DNA database could invade privacy and tempt officials to punish the relatives of dissidents and activists. Rights activists argue that the collection is being done without consent because citizens living in an authoritarian state have virtually no right to refuse.

There is some opposition to this program: Wang Ying, an official from Beijing, said that when the technology reached a certain scale, the government needed to protect the rights of users “in a timely manner”, however due to the highly centralized structure of the chinese communist party and the demand of a minimum threshold of DNA samples from the central authorities, it is to be expected that local party leaders would simply ask for the material, without further explanations as Mr. Jiang, a computer scientist that lives in Beijing says: “all our information is with them already” [18].

8.2 China: Uighurs crime prediction based on genes

Uighurs are a muslim minority group in China, from 2015 the chinese government held them in an iron grip, with the objective of assimilating them, reducing by force the frequency of revolts, and instead of using the army and violence, China chose to use modern technology and to create a mass surveillance state. [16]

The mass surveillance state was enabled by modern AI tools, in particular computer vision but also genetic data, in fact between 2016 and 2017, China used *Thermo Fischer* machines to fetch the genetic data of 36 million people, in order to use it as material for machine learning algorithms.

Also, China contributed to a public catalogue of genes “1000 Genomes Project”, which is supposed to contain only DNA from willingly and informed patients, with other 2000 Uighurs DNA [16].

In 2017, one week before the start of the DNA collection program, a Thermo Fischer researcher, Dr. Zhong Chang, said at a conference in Beijing that the company could help in identifying specific ethnic related genetical traits which could be found on Uighurs and tibetans [18], but in 2019 Thermo Fischer was forced by the US government to cease selling medical equipment to China as part of the sanctions that US imposed on China

for the Uighurs concentration camps, however recently it was discovered that Xinjiang police force is still buying DNA machines from the US company. [17]

9 Genetic Data: differences with other types of personal data

In this section we will see what contraddistint genetic data from other types of personal data. We saw the first difference in section 7, that is genetic data reveals not only information about a person, but an entire family. A second relevant difference regards the impossibility to anonymize genetic data. [34]

9.1 Anonymization is not possible for genetic data

In a 2013 study by *LAURA L. RODRIGUEZ* published on *Science* [34], it was found that it was possible to use genetic data available in public genetic databases combined with more general information like age and state of residency to link a sequenced anonymized genome in a public database to an individual with high probability, so data that is granted to research institutions under the clauses of anonymization, specified by both US and EU regulations [9][10][6], is not really anonymous and could be used to identify a person, and consequently identifying also its extended family. We will say that genetic data holds the property of *un-anonymization*

Given that the current legislation allow a person to anonymize the data [6][9], but not always to delete the data itself, for example in the case where data is used for reseach purposes, the anonymized data isn't really anonymous, even after removing the additional information.

The dangers of letting genetic data roaming free, even in its anonymized form, are only compounded by data breaches being so common: there have been 4,145 publicly disclosed breaches that exposed over 22 billion records in 2021 [21], and once the genetic data is out, even if it is anonymized, it can be used to track a single person nonetheless and then it can be used to predict behavioral traits or genetic diseases of that person or extended family.

10 Current laws regulating genetic data and genetic testing

10.1 *GINA* and *ACA*

In USA we have 2 laws protecting genetic data: *GINA* (Genetic Information Nondiscriminatory Act) and *ACA* (Affordable care act). *GINA* prohibits an employer to seek an employee genetic material, but the former is allowed to ask the latter for its genetic material for "voluntary wellness programs", and although the employee can refuse, compliance is often rewarded with benefits for participants such as reduction in health insurance cost, and non-participants are punished. [6] In addition, given that health insurances paid by private companies to an employee often extends to the employee's family, also the partner can be asked for its genetic material.

ACA prohibits adjustment of premiums based on genetic material, however *ACA* allows the usage of geographical data of residence and age, making a loophole to be exploited due to the correlation between geographical area of residence and genetic information. [29]

Currently the judicial system doesn't seem to be aware of the importance of genetic data [8], as we can see on the court case in which an insurance company denied to pay for breast-ovarian cancer treatment to a customer on the excuse that that type of cancer was "not an illness, but a genetically based condition" [6].

10.2 *GDPR*

Europe provide the most advanced data protection law up to now: the General data protection regulation (*GDPR*).

The General Data Protection Regulation of the European Union, states in the preamble that "The protection of natural persons in relation to the processing of personal data is a fundamental right".

Privacy is also considered a fundamental right in both European Convention of Human Rights 1953 (Article 8) and the European Charter of Fundamental Rights 2000.

Even if *GDPR* is the most advanced regulation protecting genetic data, it is not devoid of flaws.

Article 9 is the most problematic because it lists a number of exceptions on the handling of genetic data with the aim of supporting research, for example it is possible for EU genetic data to be sent outside the member states, even to countries that does not adhere to the same standards of the GDPR in terms of privacy when "the requirements in this Regulation relating to transfers subject to appropriate safeguards, including binding corporate rules, and derogations for specific situations are fulfilled" (Recital 107 of the *GDPR*) [23].

Another problem of Article 9 is the possibility of member states to "maintain or introduce further conditions, including limitations". [23]

For example in UK The Health Service (Control of Patient Information) Regulations 2002, require that in England and Wales, data be disclosed without patient consent for purposes, such as: to diagnose risks to public health, recognise trends in such risks, control and prevent the spread of risks and monitor and manage the delivery, efficacy and safety of immunization programmes. [6]

10.3 Genetic Data: Problems in current legislations

GDPR tries to both protect citizens privacy and enable research, the problem here being that *GDPR* assumes that anonymized data can be used for research without the risk of a person being identified by that data, hence protecting privacy, but as we have seen this is not the case for genetic data. [34] Also, data breaches are very common [21] and can be used to infer information about one's extended family [5][10], making "genetic data used for research" and "privacy" are mutually exclusive.

More than protecting the genetic data itself, it is the usage that should be regulated. Given the availability of mass genetic data sharing through cloud computing [6], data breaches [21], and possible usage of genomic data to identify traits of the extended family [5][10] it is almost impossible to ensure genomic data privacy.

Also, given the importance that big AI systems have, it becomes difficult for a person to state if it is being discriminated, because genomic data can go around for research purposes and through data breaches, commercial AI models could be trained on that data without the person being aware its genomic data is being used for such a purpose.

11 Addressing genetic data abuse on the AI era

Genetic data is fast becoming more and more important due to reduction in costs of genome sequencing and the increasing power of AI [2], but many people doesn't seem aware of that yet and this can cause emotional or economic distress in case a person chooses to entrust its genetic data to a private company, which assumes their customer have some knowledge about genetics and can interpret genetic information by themselves [32].

Fortunately the public unawareness toward genetics is decreasing. Genetic literacy has been improving for the last decade [8], also, new professions like the *genetic counsellor* are becoming more rooted, and will be tasked with translating difficult genetic information to patients.

Genetic data is very different from other types of personal data because it holds the properties of *un-anonymization* [34] and *estendibility* [5] that the other types of data are missing, making current privacy protecting legislations unfit to completely regulate it without modifications. In fact, we suggest to "organize interdisciplinary teams to draft specific legislations" [6], in order to understand better how to regulate all applications of genetic information, focussing on the ones driven by AI.

Making genetic data as private as other types of data is almost impossible due to it's properties, so we suggest to focus more on imposing constraints on the information we extract from it through genetic testing, more than trying to limit the circulation of that type of data toward research-only environments. [6]

Genetic information has many applications on the medical field: from *prenatal testing* [25] to *disease prevention* [5], but it can be used also to increase discrimination, either based on predispositions toward diseases [13] or by applying *behavioral genetics* and trying to predict crimes based on the genome of a single person [18] or even a whole population [18], enacting so *genetic discrimination* practises.

For these reasons it is necessary to "train relevant stakeholders on basic genetic findings" [6], and more generally raising genetic awareness to the general public.

12 Bibliography

References

- [1] The Human Genome Project
- [2] An Economic Analysis of the value of Genetic Testing - World Economic Forum 2021
- [3] I have a rare disease. This is my hope for the future of medicine - World Economic Forum
- [4] Uncovering the complex genetics of human character - Nature
- [5] Understanding Genetics: A New York, Mid-Atlantic Guide for Patients and Health Professionals - Genetic Alliance
- [6] GENETIC DATA MISUSE: RISK TO FUNDAMENTAL HUMAN RIGHTS IN DEVELOPED ECONOMIES - 2019 - Legal Issues Journal UK
- [7] Leroy Hood - Wikipedia
- [8] Understanding changes in genetic literacy over time and in genetic research participants - PubMed
- [9] Genetic Discrimination - US National human genome Research Institute
- [10] Genetic Discrimination and GINA - MedlinePlus
- [11] Patenting of gene and protein sequences: an EU and US perspective - Josie Miller
- [12] The effects of streaming in the secondary school on learning outcomes for Australian students - A review of the international literature - Olivia Johnston
- [13] DNA Got a Kid Kicked Out of School—And It’ll Happen Again
- [14] Directive 98/44/EC - EUR-Lex
- [15] Gordon Thomas Honeywell presentation: Here they come
- [16] China Uses DNA to Track Its People, With the Help of American Expertise - The New York Times
- [17] China Still Buys American DNA Equipment for Xinjiang Despite Blocks - The New York Times
- [18] China Is Collecting DNA From Tens of Millions of Men and Boys, Using U.S. Equipment - The New York Times

- [19] DNA-Sample Collection From Immigration Detainees - Federal Register
- [20] Whole Genome Sequencing cost 2023 - 3 Billion blog
- [21] Over 22 billion records exposed in 2021 - Security Magazine
- [22] Global implementation of genomic medicine: We are not alone - PubMed
- [23] GDPR article 9
- [24] Should you get your genome sequenced? - Stanford Healthcare Innovative Lab
- [25] Prenatal Genetic Testing - Cleveland
- [26] Is Personality Genetic? - VeryWellMind
- [27] Can you change your personality? - University of California
- [28] The impact of discrimination - American Psychological Association
- [29] Gene–environment correlations across geographic regions affect genome-wide association studies - Nature
- [30] Chromatin
- [31] Non-coding DNA
- [32] 23AndMe
- [33] Top 10 DNA-Sequencing companies - BioSpace
- [34] Identifying Personal Genomes by Surname Inference - Science