Genomics, Big Data and Privacy:
Reflections upon the implications of direct-to-consumer genetic testing

Mariana Vitti Rodrigues

Abstract: This paper investigates epistemological and ethical implications of the growing availability of direct-to-consumer genetic testing for the science and society. Direct-to-consumer genetic testing is characterized as the genetic testing sold directly to consumers without any assistance from professionals. By offering empowerment and control, companies convince consumers to sequence their genome by granting the company access to their genetic data in exchange to results that are not always accurate. To which extent do consumers properly understand the results of their genetic testing? Are consumers aware of the possible implications for disclosing genetic information to private companies? I address these questions in two steps: firstly, I discuss epistemological implications of the development of Genomics, understood as a data-intensive science, by delving into the theoretical commitments of the concept of gene and the notion of circular causality. Secondly, I reflect upon the privacy risks of taking direct-to-consumer genetic testing in a data-rich world. Finally, I draw some conclusions on the possible consequences of direct-to-consumer genetic testing by suggesting key-concepts that may help to clarify the limits and scope of genetic testing.

Keywords: Direct-to-consumer genetic testing; Genomics; Big data; Circular causality; Privacy.

1. Introduction

> *[…] this push toward bio-engineered perfection strikes me as the wave of the future, one that will sneak up on us before we know it and, if we are not careful, sweep us up and tow us under.*
> *(Kass, 2003, p. 10)*

What if we could change ourselves to become better humans? What if we could trick nature to change our predestinations? What if we could take a magic pill to become stronger, more intelligent or to live longer? With the slogan "better than well", the transhumanism movement "[…] affirms the possibility and desirability of fundamentally improving the human condition through applied reason, especially by developing and making widely available technologies to eliminate aging and to greatly enhance human intellectual, physical, and psychological capacities"[1] (More, 2013, p. 3). The transhumanism movement advocates

---

[1] See also https://humanityplus.org/philosophy/transhumanist-faq/

the premise according to which the human species is developing and that we should accelerate this process of development with the use of new technologies: "[w]e can also use technological means that will eventually enable us to move beyond what some would think of as 'human'"[2] (More, 2013, p. 3)

The transhumanist slogan "better than well", together with the public curiosity on what makes us human and how we could possibly improve the human condition, opened the doors to the commercialization of biotechnologies to the general public. One of the consequences was the widespread growing of direct-to-consumer genetic testing, which is characterized as the genetic testing sold directly to consumers without any assistance from specialized counsellors. Companies that sell those products promise empowerment and control to their clients by disclosing the information stored in their unique genome, i.e., the entire set of DNA that makes a particular organism[3]. If one knows the human genome, s/he would be able to change the human nature by the unveiling of its blueprint, the basic set of information that makes us who we are. This idea is the backbone of scientific determinism that aims to understand nature in order to control it (Santos, 1998). Although scientific determinism has been challenged by the development of complex systems theory, systems biology and quantum mechanics, its core belief is hold in scientific circles as well as it is spread to the general public.

One cannot deny that the development of biotechnologies allows the identification of monogenic diseases, i.e., diseases which are highly probable to be caused by the expression of a single gene (for example, cystic fibrosis and Huntington's disease). However, as Koepsell and Covarrubias (2016, p. 350) exemplify, "[…] the genetics of the most prevalent diseases such as cardiovascular ailments, dyslipidemias, type-2 diabetes, schizophrenia, dementia, or Alzheimer´s are explained, although only partly, by the variation in hundreds of locations (*loci*) across the genome". Multigenic diseases challenge our understanding of the mechanism behind the genome echoing in our misunderstanding of the concept of gene.

The development of Genomics, allied with the advancement of genetic engineering, not only allows one to discover a relationship between a single gene and its phenotypic trait; it also enables the intervention in the genome by editing its parts. For example, the identification of the CCR5 gene as a marker for HIV immunity became famous after the

---

[2] https://humanityplus.org/philosophy/transhumanist-faq/
[3] https://history.nih.gov/exhibits/nirenberg/glossary.htm

scientist He Jiankui used CRISPR/Cas9[4] system to modify the expression of the CCR5 gene in embryos. Even if the mutation of the CCR5 gene is identified as a possible marker for HIV immunity (Allen et al., 2018), its expression, without the presence of mutation, may be related to possible immunity to other infectious diseases, such as Nile fever (Lim & Murphy, 2010) and influenza (Falcon, 2015). This case brings up the problem of identifying biological markers by detecting correlational patterns in large amounts of data that are not always well structured. Data-intensive sciences, such as Genomics, allow the discovery of correlational patterns between genetic markers and the corresponding biological functions. However, not all the traits can be correlated to a single gene. In which ways do these correlational patterns really hold for the biological functions? Is the advancement of new technologies for dealing with data improving our understanding of the genetic influence in human behavior?

The development of Genomics, in association to direct-to-consumer genetic testing, can accelerate decision-making processes without considering the possible scientific and ethical consequences. In the present paper, I challenge the naive belief in scientific determinism by investigating epistemological and ethical implications of the growing commercialization of direct-to-consumer genetic testing. I reflect upon two interrelated questions: (i) To which extent do consumers properly understand the results of their genetic testing? (ii) Are consumers aware of the possible implications for disclosing genetic information to private companies? In order to clarify those questions, firstly, I present some examples of direct-to-consumer genetic testing purchase alternatives that companies offer to potential clients. Secondly, I address some of the epistemological challenges involved in Genomics, understood as a data-intensive science. Thirdly, I address ethical implications of genetic testing in relation to privacy concerns. Lately, I provide some reflections upon possible ways to overcome such challenges via education of key-concepts such as evolution, adaptation, phenotype, privacy, and circular causality.

2. Direct-to-consumer genetic testing: empowerment and control

Your DNA is your blueprint. We store it so that we can run new tests for genetic predispositions as soon as they´re validated. (GoodCell.com)

---

[4] CRISPR/Cas9 is a system that allows scientists to edit parts of the genome by removing, adding, or changing sections of the DNA.

Direct-to-consumer genetic testing (DTC-GT) is characterized as the genetic testing sold directly to consumers without any assistance from professionals (Su, 2015, p. 359). Su (2015) categorizes three reasons that summarize why people choose to take those tests: (1) Identity seeking in order to identify paternity, ancestry and ethnicity; (2) disease risk-testing that complements health care; and (3) curiosity-driven testing aiming a better lifestyle. I argue that, by offering a better way of life or the unfolding of one´s true identity, companies profit on consumer´s misunderstandings and naive curiosity.

The main argument used by DTC-GT specialized companies is empowerment (Su, 2015). By taking a genetic test, one may know in advance the risks of having a given disease and, thus, the opportunity to change his/her habits in order to avoid his/her fate. Genetic tests give consumers the sense of "[…] enhanced power and being in control" (Su, 2015, p. 361). It mirrors the backbone idea of scientific determinism, according to which one should understand nature, in this case, his/her own genome, in order to control it. In the following, I give some examples by presenting advertisements that offer the opportunity to improve one´s life using DTC-GT.

The company AncestryDNA[5] offers the opportunity to tight the bonds with your ancestors by building up your family tree. Once you buy the company´s kit, you should simply spit into a tube and send it back to the company by mail. From doing so, you may get insights from your DNA to better understand your ethnicity and your family health history. Moreover, the company offers you the opportunity to decipher your genetic code in order to unveil how your family story was shaped by defining moments such as in suffrage history. Could you imagine you might be partially responsible for women´s right to vote?

Another example can be found on the website from the company called MyHeritage[6]. The company stresses that your DNA test allows you to have a ***powerful experience*** of discovering what makes you unique by learning where you really come from. In doing so, MyHeritage company believes to be ***enhancing*** the lives of its clients. It sounds amazing that one can really use genetic testing as a tool for human enhancement from the comfort of her own home.

Among their health products, MyHeritage offers testing of polygenic risk reports, such as the risk for female breast cancer caused by many different variants or testing the risk of hereditary BRCA cancers; in other words, breast cancer caused by the mutations of the

---

[5] <https://www.ancestry.com/>
[6] <https://www.myheritage.com/>

specific genes BRCA 1 and BRCA 2. They also offer their consumers the opportunity to test what they call the susceptibility to HIV-1 infection, by checking if one has a mutation on the CCR5 gene[7].

A famous company called 23andMe[8] encourages their consumers to take action to stay healthy. The company offers ancestry, health, and trait services. Apart from the health-related trait testing, it offers the opportunity to discover, from the analysis of one´s genome, the following traits: cilantro taste aversion; fear of heights; sweaty *versus* salty; wake up time; and ice cream flavour preference. If one is curious to know all about oneself, there are plenty of technologies available out there.

With the growing interest for such tests, companies also offer the opportunity for careful parents to sequence relevant parts of their children´s genome. In doing so, parents may exert control over the future of their children in order to allow them to become "better than well". The company Map*my*gene[9] offers the Inborn Talent Gene Test which aims to identify children´s natural talents and aptitudes that might not always be obvious at a young age, so that parents can provide better learning opportunities to their children by enhancing their inner talents. Another company, Orig3n[10], allows parents to discover if their children would be a picking eater, if they would easily learn new languages or to unveil which kind of sport would be the perfect match for them.

Furthermore, several companies, such as MyTrueAncestry[11], GEDmatch[12], and Genomelink[13], offer free of charge basic analysis of DNA by accessing the raw data taken from the results of other companies such as 23andMe, AncestryDNA and MyHeritage. Moreover, these companies offer online infrastructures where one can share the results of his/her genetic testing in order to compare them with the results of other consumers. By providing this kind of free service, companies increase their databases by collecting consumers data in an unprecedent fashion.

In 2018, direct-to-consumer genetic testing sales exploded mainly in the United States, after the politician Elizabeth Warren shared the results of her test to the nation (Molla, 2020). However, in 2019, sales started to decrease, and genetic testing suddenly became

---

[7] For more information, see <https://www.myheritage.com/health/disease-list>
[8] <https://www.23andme.com/>
[9] <https://mapmygene.com/>
[10] <https://orig3n.com/>
[11] <https://mytrueancestry.com/>
[12] <https://www.gedmatch.com/>
[13] <https://genomelink.io/>

unpopular (Molla, 2020). According to Molla (2020), sales may have been declining for three main reasons: (i) market saturation, as most of the Americans bought the tests, they would not buy it anymore; (ii) privacy concerns, although the companies ensure that the consumer's data will be safe, leaks, hacking and non-robust laws of personal data rise privacy concerns; (iii) limited utility, consumers started to realize that the results of their genetic testing are not as accurate as they imagined, as well as they become aware that some of the results do not add relevant information to one´s life.

To overcome such decline in sales, companies started to provide new tools for testing, novel products, highlighting new advantages for potential consumers. An example can be found in the company GoodCell[14]. The company´s webpage is "shining" and attractive, offering its users the irresistible headline: "Your cells today could save your life tomorrow". It continues "store them, for life", "store them, for your friends", "store them, for you family", "store them, for your kids", "store them, for yourself". The company, differently from the previous ones, offers you a personal biobanking. When you buy a kit from GoodCell, a trained phlebotomist takes a sample of your blood and the company stores it indefinitely, in order to unveil your bio-information at any time. By doing so, the company guarantees to update your health information as science progresses and figures out new mechanisms behind genetic diseases.

With the growing alternatives offered to DTC-GT consumers, epistemological and ethical challenges arise. In the following, I address the epistemological challenge by presenting the development of Genomics as a data-intensive science. Subsequently, I discuss the theoretical commitments of Genomics by investigating to which extent the use of the word "gene" is validated by the scientific development of Genomics.

3. Epistemological implications

In this section, I present the development of Genomics as a data-intensive science, challenging the allegedly success of the automation of scientific process by addressing theoretical commitments around the concept of the gene and the idea of complex phenotypes. Subsequently, I bring to the discussion the concept of circular causality to enrich the argument in favour of the recursive and dynamic account of the mechanisms behind the

---

[14] https://goodcell.com

genetic code. Afterwards, I discuss the epistemological consequences of scientific advancements to the interpretation of the results of DTC-GT.

3.1. Genomics: a developing data-intensive science

The word *Genomics* was invented by Thomas H. Roderick in 1986 to characterize "[…] an activity, a new way of thinking biology" (Kuska, 1998, p. 1) in order to name a new journal which aim was to "[…] include sequencing data as well to include discovery of new genes, gene mapping, and new genetic technologies" (ibid.). According to Navarro et al. (2019, p. 2), Genomics can be characterized as a data science that emerged in the 1980s from the convergence of genetics, statistics, and large-scale datasets.

With the development of technologies that deal with massive amount of data, such as Big Data analytics, Genomics has developed as a data-intensive science with growing application. In the following, inspired by Laney (2001), Floridi (2012), Ideker et al. (2006), Navarro et al. (2019), and Leonelli (2019), it is presented the notion of Big Data analytics in the context of Genomics in terms of 3V´s (volume, velocity, and variety) and 4M´s (measurement, mining, modelling, and manipulating); challenges faced by the development of Genomics in the era of Big Data are also highlighted.

Although there is no consensus on an exhaustive definition of Big Data, Laney (2001) suggests an initial characterization of this term, which brings together 3 V's: volume, velocity, and variety. According to Floridi (2012), these three components refer to the technological challenge that consists in processing, analysing and storing, with the current computational power, a huge amount of varied data, not always well structured. In the literature of Big Data, other V´s have gained attention, such as value, variability, and veracity. In the present section, I focus on the initial 3V´s as presented by Navarro et al. (2019) in the context of Genomics.

In terms of volume, Navarro et al. (2019) stresses that Genomics is increasingly expanding the size of the available data generated by sequencers. In the scenario depicted by the authors, Genomics may soon surpass the volume of data in comparison to other data-rich disciplines such as earth sciences, astronomy, as well as social media (Navarro et al., 2019, p. 2). In terms of velocity, Genomics is increasing the speed of data generation as well as data analysis. As an example, they highlight the decrease of time spent to sequence the human genome that would soon take less than 24 hours, in comparison to "[…] 2 to 8 weeks by

current popular technologies and 13 years of uninterrupted sequencing work by the Human Genome Project" (Navarro et al., 2019, p. 3). The authors believe that, in the near future, Genomics may work with real-time processing. As a technical challenge, Navarro et al. (2019, p. 4) warn that the maximization of volume and rising velocity of Genomics may lead the field to "[…] reach a critical point at which cloud solutions might be indispensable for large-scale analysis".

The analysis concerning the variety of the data in the context of Genomics brings two aspects to the discussion: on the one hand, data seem not to vary too much because genetic data consist of a monolithic sequencing data (Navarro et al., 2019, p. 4). On the other hand, the variety of the data collected and stored in huge databases hides a complex aspect: its phenotypic data. According to the authors, the phenotypic data, because of its diversity, challenge attempts to standardization: "[…] phenotypic data can consist of such diverse entities as simple and unstructured text description from electronic health records, quantitative measurements from laboratories, sensors, and electronic trackers, and imaging data" (Navarro et al., 2019, p. 4). The challenge involved in the standardization of phenotypic data is also stressed by Leonelli (2015, p. 813): "[…] the paradox consists of the observation that, despite their epistemic value as 'given', data are clearly made", because "[t]he procedures involved in packaging data for travel involve various stages of manipulation, which may happen at different times and may well change the format, medium, and shape of data" (Leonelli, 2015, p. 816).

Besides the characterization of Genomics around Big Data 3V´s, Ideker et al. (2006) and Navarro et al. (2019) characterize Genomics in terms of 4M´s: measurement, mining, modelling, and manipulation. Roughly speaking, data measurement is undertaken by applying different technologies (such as machine learning). Once the measurement provides multivariate data, scientists need to use computational algorithms for mining the results in order to generate plausible hypotheses. Once those hypotheses are generated by the algorithms, the modelling process is undertaken in order to enable the establishment of predictions concerning "[…] complex, interconnected non-linear molecular systems" (Ideker et al., 2006, p. 1226). Finally, molecular manipulations are initiated in order to test the predictions enabled by the process of modelling and/or to develop new technologies (Ideker et al., 2006). In short, according to the 4M´s framework, Genomics can be characterized as a "[…] hybrid approach of combining data-mining and biophysical modelling" aiming to develop a probabilistic "health forecast" (Navarro et al., 2019, p. 5).

One of the main challenges brought by Big Data analytics is that the analysis of massive amount of data does not necessarily guarantee the unveiling of causal relationships. In the context of Genomics, Koepsell and Covarrubias (2016, p. 350) highlight that "[…] the identification of genetic variants responsible for complex multifactorial diseases will require the study of hundreds and thousands of unrelated individuals only to render a modest association between genetic variations and the disease". Although Big Data analytics may help the understanding of the monogenic diseases' mechanisms, the complexity involved in many phenotypic expressions challenges the full automation of Genomics, as a data-intensive science.

Another challenge, posed by Leonelli (2019, p. 2), concerns the theoretical commitments of data-intensive sciences: "[w]e need to acknowledge that no data are 'raw' in the sense of being independent from human interpretation". By arguing against the possibility of "raw" data, Leonelli challenges the hope in the complete automatization of scientific practice through the implementation of Big Data analytics. As Leonelli (2019, p. 5) argues "[a]utomated data analysis is an exciting prospect for biological discovery. Far from making human judgement unnecessary, the increasing power of computational algorithms requires a proportional increase of critical thinking".

In the following, I explore theoretical commitments of the concept of gene, highlighting possible epistemological consequences for the interpretation of the results of direct-to-consumer genetic testing.

3.2 Theoretical commitments: the genome and the concept of gene

The genome is not the organism
(Keller, 2014, p. 2425)

Throughout the literature in Philosophy of Biology, there are different characterizations of the concept of gene (Keller, 2000, 2014; Moss, 2003; El-Hani et al. 2013; Griffths et al., 2013). In this section, I present, inspired by Moss (2003), two of them: Gene-P and Gene-D, exploring the development of the gene concept along with the advancement of Genomics.

According to Moss (2003), the concept of Gene-P, or phenotypic gene, stands for an instrumental characterization of the gene. A Gene-P has a strict causal relationship to its

phenotypic traits, being exemplified by the famous expression "genes for": gene for breast cancer, gene for Alzheimer, gene for sugar consumption (Charney et al., 2012).

Although it is well known by the scientific community that a single gene does not bear such strict causal relationship between its expression and its trait markers, the notion of gene-P is an important and relevant tool for statistical analysis. In other words, the notion of an instrumental gene, Gene-P, is applied as a model to make predictions of possible development of a specific phenotypic aspect from the presence of a given genotype. Again, this is a simplistic model that depicts a deterministic version of the relationship between genotypes and phenotypes. The problem does not lie in the application of a deterministic concept of gene in order to make predictions, but it starts when one conflates the idea of Gene-P or "gene for" with the material-based version of the gene, or Gene-D.

The concept of Gene-D, or molecular gene, stands for the material base of the gene, comprehending a realistic version of the DNA structure (Moss, 2003). The gene was conceptualized as a material entity in its molecular properties from the discovery, claimed by Francis and Crick (1953), of the double helix structure of the DNA. According to Keller (2014, p. 2425), "[g]enes were now concrete material entities; protein makers rather than trait makers, carriers of molecular information required to string together the sequence of amino acids to construct a poly-peptide". The general idea was that the DNA contains information that, once transcribed and translated by molecular processes, makes proteins that build up an organism.

The mechanisms involved in the informational process of making proteins from the instructions carried by the gene were explained by the Central Dogma proposed by Francis Crick (1958). According to the Central Dogma, there is a one-direction causal structure where the information contained in the DNA makes RNA, which, in turn, makes proteins (Figure 1): "[...] once information has passed into protein [from RNA], it cannot get out again" (Crick, 1958, p. 153). Apart from the discussion about the concept of information adopted by Crick, the important aspect to be considered here is the one-way directionality of the process of making protein from DNA instructions.
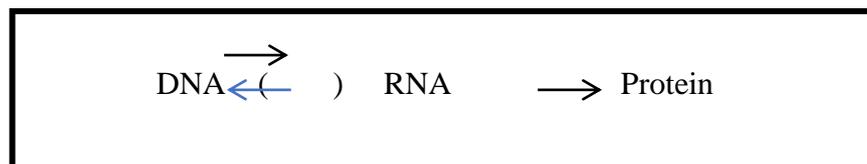

Figure 1 – The Central Dogma one-way flow of information

Diagram adapted from Hoffmeyer (2008, p. 58)

The Central Dogma disseminated the idea that the DNA is composed by "units", called "genes", that hold the instructions to generate proteins which make who we are. In this scenario, the concept of gene (gene-D) was conceived, roughly, as a unit directly associated with a biomarker; this unit *acts* upon the system generating proteins (Keller, 2014, p. 2424). Consequently, our genome was conceived as a blueprint of our organism.

The conception of our genome as a blueprint motivated the Human Genome Project (HGP), that, ironically, challenged the very concept of gene. The HGP, undertaken between 1990 and 2003, was described as "[…] an inward voyage of discovery"[15]. The aim of the HGP was to map and sequence the entire human genome, i.e., all the genes of the species *Homo sapiens.* As an outcome, it was discovered that we had far more non-genic DNA that it was expected, meaning that most part of our DNA is not composed by protein-maker units. If a gene is a protein maker, everything in the DNA that does not make proteins was conceived as "junk DNA". As Keller (2014, p. 2425) explains: "[o]nce the sequence of the human genome became available, it soon became evident that sequence information alone would not tell us 'who we are', that sequence alone does not provide a 'complete set of genetic instruction of the human being'".

Furthermore, the development of Genomics challenged the one-directionality claimed by the Central Dogma proposed by Crick. The investigation of non-genic DNA unveiled the *reactive* aspect of our genome. As Keller (2014, p. 2427, my highlights) points out, "In addition to providing information required for building and maintaining an organism, the genome also provides a vast amount of information enabling it to *adapt* and *respond* to the environment in which it finds itself". The author summarizes the aftermath of the gene concept as the shift from gene action to reactive genomes (Keller, 2014), proposing a recursive structure to explain the relationship between organism and environment:

> [...] many of the problems with the traditional perspective derive from the core belief in the separation of genes from soma, of genetics from development, and I agree that what is needed is a theoretical perspective that replaces the linear causal structure that is supposed to take us from genotype to phenotype by one that incorporates both the fundamental circularity (or recursivity) of living systems and their *a priori* inseparability from the environments in which, and out of which, they take their form (Keller, 2014, p. 2424).

---

[15] https://www.genome.gov/Funded-Programs-Projects/NHGRI-Genome-Sequencing-Program.

From a reactive and recursive characterization of the genome, the idea of "genes for" loses part of its epistemic value. There are far more processes, in different levels of analysis, that, in interaction with the environment, may affect an organism´s properties. To explore the recursive aspect of the interaction organism-environment and to enlarge the scope of our analysis, I present, in the following, the concept of circular causality (Lewis et al., 1999; Haselager & Gonzalez, 2002), reflecting upon the epistemological consequences of a recursive characterization of the gene concept in the context of genetic testing.

3.3 Circular causality: the recursive and dynamic process of complex systems

The concept of circular causality was developed by the cybernetic movement in the 1950s (Haselager & Gonzalez, 2002, p. 5). According to Haselager & Gonzalez (2002, p. 6, authors' highlights), the process of circular causality can be characterized by two main properties: (i) "[…] the effect of a cause affects its own cause, changing it and being changed by it" and (ii) "[…] there is an *collective interaction* among the basic elements, in the microscopic layer, which allows the emergence of a pattern in the macroscopic layer" that may "[…] causally constrain the behaviour in the microscopic layer". The notion of circular causality sheds light to the recursive aspect of the mechanisms´ of the genome. The genome does not simply give instructions to make proteins – and to make us who we are –, but it may also be changed by reacting to the organisms' interaction with the environment.

The following diagram withdraw from Lewis' book on complexity (Lewis, 1992, p. 13) depicts aspects of circular causality that plays an important role in the emergent, dynamic and recursive process of complex systems:
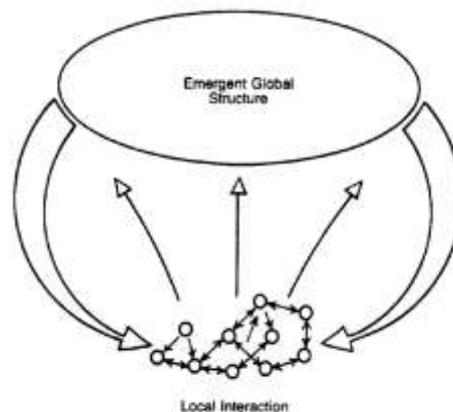
From the above picture, we can interpret the genome as being represented by the set of the lower-connected small circles. These small circles may represent parts of the genome that play different roles, for instance, by giving instructions to make proteins (Gene-D), to regulate the process of transcription, and so on. The emergent process, in our example, may be represented as the phenotypic expressions of an organisms' genome. Besides this process, agent-agent and agent-environment interactions also play a relevant role in the organism's constitution that may, in turn, influence its genome by affecting gene expressions and gene regulations. In sum, the idea of a recursive process that involves but is not reduced to the genome enriches the theoretical perspective about the concept of gene.

The epistemological outcome of this section can be summarized by Keller´s statement: "The genome is not the organism" (Keller, 2014, p. 2425). Furthermore, one should be aware of the difference between the molecular concept of the gene as a protein-maker (the Gene-D) and its instrumental concept (the Gene-P). This elucidation may shed light to different approaches of the concept of gene contributing to the discussion of the theoretical commitments involved in Genomics as a data-intensive science. Concerning DTC-GT, besides the epistemological challenges brought by the conceptualization of the gene, ethical aspects are at stake, which will be discussed in the following.

4. Ethical implications of Direct-to-consumer genetic testing

Direct-to-consumer genetic testing may have some undesirable consequences that can be exemplified by the following two personal reports. Ellen (2017) shared her disappointment about genetic testing to *The Guardian*: "Most people, like myself, have a low understanding of genetic variants, what phrases such as 'higher risk' or 'probability' actually mean or how to interpret our results correctly". Mahdawi (2018) realized that, in order to know that her Neanderthal heritage was associated with a reduced tendency to sneeze after eating dark chocolate, she disclosed her genetic information to private companies in exchange to not so revealing results.

Caulfield, the Canadian research chair in health and law policy, stresses that "Ancestry testing can be a fun recreational science. However, the marketing is based on a

scientifically inaccurate notion of the value of biological heritage" (Park, 2019, p. 613). In terms of genetic risks and predispositions, Caulfield highlights that "[…] the body of evidence tells us that genetic risk information does not have a meaningful impact on behaviour" (Park, 2019, p. 613). Taking a genetic testing and interpreting its results does not necessarily impact one's habit changing. As an example, Ellen (2017) indicates that "[…] people with conditions such as Huntington's disease in their family decide not to go ahead [and take the genetic test] because a diagnosis would change nothing for them".

Being alone in the inward journey through one's DNA, instead of empowerment and control, may bring psychological consequences and interpretative misunderstandings to the average of not scientifically well-educated consumers. The lack of professional counselling may cause severe anxiety when one discovers some (maybe not accurate) undesirable news. Furthermore, it may propitiate future unwelcome surprises concerning the disclosure of personal data. In the following, it is discussed ethical consequences of DTC-GT concerning privacy.

4.1. Testing oneself, sharing one´s results: privacy risks

Privacy concerns expressed by authorities, physicians, scholars, and consumers have grown with the development of new technologies such as electronic health records and genomics allied with the rapid power of computation and Big Data analytics. Four possible implications brought by the disclosure of genetic information are discussed in the following: (i) the possibility of growing surveillance of citizens by the government but also by private companies; (ii) the possibility for discrimination about one´s genetic condition that may lead to an increase in the health insurance bill; (iii) the increasing of anxiety levels generated by undesirable results about future risks; and (iv) the emergence of the need for a broader concept of privacy.

A famous case, which exemplifies the use of genetic information by government surveillance (i), is the capture of a serial killer, in 2018, enabled by the analysis of his relative's genetic information from the results of DTC-GT published in an open database (Park, 2019, p. 612). The disclosure of genetic information through DTC-GT, and through sharing its results in public platforms, allows the information to be used by authorities, health insurance and private companies (for example, a company may check the genetic profile of a candidate

before offering him/her a job[16]). Combined with other forms of retrieving one´s data, such as profiles on social media, credit card purchases or public video recordings, the disclosure of genetic information may accelerate the process of population surveillance. As Caulfield (Park, 2019, p. 617) stresses "[i]ncreasingly, these databases are valuable sources of information that can be used for everything from marketing to biomedical research to public health surveillance".

Concerning the possibility of discrimination based on genetic information (ii), Schwab (Park, 2019, p. 616) highlights, in the context of the USA laws, that "[…] one key point that everyone should be aware of is that genetic information can be used in determining whether to offer and to set the premiums for long-term disability and life insurance". Unfortunately, in Europe and Brazil, for instance, the laws regarding genetic and personal information are not well defined, requiring further developments in order to guarantee citizen´s rights. Moreover, cyber-attacks configure a real threaten to one´s data.

A growing number of studies aim to investigate the psychological consequences of genetic testing (iii). Baudhuin (Park, 2019, p. 613) emphasizes that "[m]isinterpretation may cause not only medical mismanagement but also psychological distress and/or mishandling of their own healthcare". In an interview about the psychological risks of taking prophylactic treatment on breast cancer, Patenaude points out that "[d]irect-to-consumer [genetic] testing further complicates the scene because typically there is no counselling and patients often misinterpret the results" (Hamilton, 2013). The lack of counselling may not just increase consumers anxiety, but also overload healthcare systems once the results may motivate people to search for medical advice after taking those tests.

The fourth (iv) implication concerning genetic privacy comprises the need for a broader notion of privacy. Borrowing the idea of an "extended mind" (Clark et al., 1998), I propose the discussion of the notion of "extended privacy". An individual decision to share one´s genetic information also implies sharing the information of one´s relatives. In other words, taking a genetic test and sharing its results may compromise the autonomy of other family members that do not consent to have their genetic information disclosed or shared in a public platform. The concept of extended privacy means that one´s genetic information is not only personal, but also belongs to his/her family. The idea of an extended privacy enables the discussion of applying responsibility to a moral agent that shares his/her genetic information while not requesting his/her relatives informed consent. According to the concept of extended

[16] Although it would be considered illegal in the USA, for example, such possibility cannot be dismissed.

privacy, policy makers should regard disclosing one´s personal genetic information as harnessing one´s relatives' autonomy. Research about informed consent may provide solid grounds to deal with such challenge.

A scarier consequence about (the lack of) informed consent is reported by the New York Times report (Bala, 2020) which inquires: "Why are you publicly sharing your child's DNA information?". Bala (2020) explains that "[t]he problem with these [genetic] tests is twofold. First, parents are testing their children in ways that could have serious implications as they grow older — and they are not old enough to consent. Second, by sharing their children's genetic information on public websites, parents are forever exposing their personal health data". By testing their children´s DNA before they are mature enough to consent, parents condemn the children´s right to not know a given information about their health-related risks, ancestry, or cilantro aversion. This example challenges the "right to be forgotten" (Floridi, 2015) once companies´ personal data policies may pose barriers for an individual to claim the deletion of his/her data.

As an answer to those challenges, in 2018, the companies 23andMe, Ancestry, Helix, MyHeritage, and Habit signed a term stating "Best practices to protect privacy of consumers of genetic testing" (Martinez et al., 2018). Among these practices, they included: (1) transparency about the consumer's genetic information, how it is used, collected and shared; (2) allowance for the consumer to choose, through informed consent, to participate in research as well as to destroy his/her genetic sample; (3) improve consumers protection regarding their genetic information according to the established laws (Martinez et al., 2018). However, with the expansion of genetic databases, the internationalization of data storage, and the lack of investment in proper education, it is hard to believe that such measures will soften privacy problems. Moreover, studies show that the re-identification of a person from his/her genetic data is fairly easy, what compromises the promise of anonymization. Caulfield warns that

> The technologies that are making genomic research more sophisticated—such as Big Data and artificial intelligence—are also intensifying the privacy issues. It is becoming easier to reconnect anonymous data to identifiable individuals. […] As information technology continues to evolve, it seems likely the privacy issues will intensify […] it will certainly become more difficult to make promises about data protection and anonymity (Park, 2019, p. 615).

The development of Genomics, Big Data analytics, and Artificial Intelligence, together with the growing commercialization of genetic testing that offers attractive promises for a

better future, imposes urgent measures and the need for better policies. Furthermore, it becomes necessary to disseminate scientific discoveries about the mechanisms behind the genome to the general public. Better education of key-concepts in biology and epistemology could soften misinterpretation of genetic results as well as it may preclude the general public to believe in dubious promises that the attractive and compelling advertisements of genetic testing companies attempt to offer.

5. Final considerations

In the present paper, I have discussed epistemological and ethical implications of direct-to-consumer genetic testing. I have shown that Genomics, conceived as a data-intensive science, is developing together with other technologies such as Big Data analytics and Artificial Intelligence. A challenge brought by the growing automatization of Genomics lies in the theoretical commitments involved, for example, by the concept of gene. Moreover, by exploring multiple aspects of the concept of gene, one can comprehend the difficulties of consumers in understanding their genetic tests' results without any professional assistance.

This article also discussed potential privacy risks brought by the disclosure of genetic information to private companies and open-access platforms. Even though privacy laws are being developed and companies have been discussing better practices to protect consumers' privacy, there are real privacy threats. Cyber-attacks, leaks, changes in the companies' privacy policy, and pressure from the government to have access to their database, configure some of the risks to the privacy rights of citizens that live in data-rich societies.

By selling the promise to become "better than well", private companies are building huge databases which, in a data-rich society, means increasing their power over the population. Kass (2003, p. 16) warns that "[w]e are right to worry that the self-selected non-therapeutic uses of the new powers, especially where they become widespread, will be put in the service of the most common human desires, moving us toward still greater homogenization of human society". So, what can be done to soften such prognosis? The immediate, feasible, answer to this question is education.

According to Scott (2013), there are three genetic-related concepts that should be well understood by society: evolution, adaptation, and phenotype. Evolution is the idea that every living being has a common ancestor and, together, forms a tree of life that connect us all. Adaptation is the idea that, by natural selection, the traits and the genes that affect the

organisms will prevail over time through the process of the co-evolution between organism and environment. Lastly, Scott (2013, p. 1) proposes the understanding of phenotypes as "[…] the result of the effect of the genes plus the environment plus their interaction". By these definitions, both genes and environment play an important role in the process of evolution by adaptation through natural selection. I believe another two key-concepts should be added to Scott´s list in order to better inform the general public interested in direct-to-consumer genetic testing: circular causality and extended privacy.

In addition to these five key-concepts, I argue that society should become aware that science is a dynamic, undergoing practice of representing and intervening in different aspects of a given studied object (Hacking, 1983). When dealing with complex systems, the practice of investigating a given system may, by circular causality, change its properties. The role of researchers should not be only to understand and control nature, but to reflect upon the possible consequences of their research to science and society. Scientists should be able to disseminate their results in order to open a sincere dialogue about the scope and limits of scientific knowledge with society.

References

Allen, A. G.; Chung, C.; Atkins, A.; Dampier, W.; Khalili, K.; Nonnemarcher, M.; Whigdahl, B. et al. [2018]. Gene Editing of HIV-1 Co-receptors to Prevent and/or Cure Virus Infection. *Frontiers in Microbiology*, 9, 1-14, 2018.

Bala, N. [2020]. Why are you public Sharing your child´s DNA information? *New York Times*, January 2nd 2020. Available at < https://www.nytimes.com/2020/01/02/opinion/dna-test-privacy-children.html>. Accessed on January 2020.

Charney, E.; English, W. [2012]. Candidate genes and Political Behavior. *American political science review*, p. 1-34, 2012.

Clark, A.; Chalmers, D. [1998]. The extended mind. *Analysis*, 58, 7-19, 1998.

Crick, F. [1958]. On protein synthesis. *Symposia of the Society for experimental biology*, 12, 138-165, 1958.

El-Hani, C.; Meyer, D. [2013]. The Concept of Gene in the Twenty-First Century: What Are the Open Avenues?. *Contrastes*, 18, 55–68, 2013.

Ellen, B. [2017]. What I learned from home DNA testing. *The Guardian*, July 23rd, 2017. Available at < https://www.theguardian.com/science/2017/jul/23/what-i-learned-from-home-dna-test-kits-are-they-accurate-or-worthwhile>. Accessed on February 2020.

Falcon, A.; Cuevas, M. T.; Rodriguez-Frandsen, A.; Reynes, N.; Pozo, F.; Moreno, S.; Ledesma, J.; Martinez-Alarcon, J.; Nieto, A.; **Casas, I.** [2015]. CCR5 deficiency predisposes to fatal outcome in influenza virus infection. *Journal of general virology*, 96, issue 8, 2074-8, 2015.

Floridi, L. [2012]. Big Data and Their Epistemological Challenge. *Philosophy & Technology*, 25, Issue 4, p. 435-7, 2012.

Floridi, L. [2015]. "The right to be forgotten": a philosophical view. *The annual Review of Laws and Ethics*, p. 1-18, 2015.

Griffths, P.; Stotz, K. [2013]. *Genetics and Philosophy:* An introduction. Cambridge: Cambridge University Press, 2013.

Hacking, I. [1983]. *Representing and Intervening, Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press, 1983.

Haken, H. [1999]. Synergetics and some applications to psychology. In: Tschacher, W.; Daualder, P.-P. Daualder (Eds.). In: *Dynamics, synergetics, autonomous agents*. London: World Scientific, 1999, p. 3-12.

Hamilton, A. [2013]. Addressing Psychological Impacts of Genetic Testing on Patients, Families: Six questions for psychologist Andrea Farkas Patenaude, PhD. Available at <https://www.apa.org/news/press/releases/2013/05/genetic-testing> Accessed on February 2020.

Haselager, F. G.; Gonzalez, M. E. Q. [2002]. Causalidade Circular: uma saída para a oposição internalismo versus externalismo?. *Manuscrito*, 25, p. 217-38, 2002.

Hoffmeyer, J. [2008]. Biosemiotics: an examination into the Signs of Life and the Life of Signs. Translated by Hoffmeyer J. and Favareau, D. Edited by Favareau, D. University of Scranton Press: Scranton and London, 2008.

Ideker, T.; Winslow, L. R.; Lauffenburger, D. A. [2006]. Bioengineering and systems biology. *Annals of Biomedical Engineering*, 34, p. 1226-33, 2006.

Kass, L. R. [2003]. Ageless bodies, Happy souls: Biotechnology and the pursuit of perfection. *The New Atlantis*, p. 9-28, 2003.

Keller, E. F. [2000]. *The Century of the Gene*. Cambrige, Massachussets: Harvard University Press, 2000.

Keller, E. F. [2014]. From gene action to reactive genomes. *The journal of physiology*, 592, p. 2423-9, 2014.

Koepsell, D.; Covarrubias, V. G. [2016]. The rise of big data and genetic privacy. *Ethics, Medicine and public health*, 2, p. 348-55, 2016.

Kuska, B. [1998]. Beer, Bethesda, and Biology: How "Genomics" came into being. *Journal of National Cancer institute*, 90, p. 93, 1998.

Leonelli, S. [2014]. What difference does quantity make? On the Epistemology of Big Data in Biology. *Big data & society*, junho, 1(1), p. 1-16, 2014.

Leonelli, S. [2015]. What counts as scientific data? A relational framework. *Philosophy of Science*, 82, p. 810-21, 2015.

Leonelli, S. [2019]. Philosophy of Biology: the challenges of big data biology. *Elife*. Available at https://elifesciences.org/articles/47381. Accessed on February 2020.

Lewis, R. [1992]. *Complexity:* Life at the edge of chaos. Nova York: Collier Books/Macmillan Publishing Company, 1992.

Lim, J.; Murphy, P. [2011]. Chemokine control of West Nile virus infection. *Experimental Cell Research.* março, 317:5, p. 569-74, 2011.

Mahdawi, A. [2018]. Why did I risk my privacy with home DNA testing? I blame my Neanderthal heritage. *The Guardian*, March 28th 2018. Available at https://www.theguardian.com/commentisfree/2018/mar/28/what-did-i-learn-from-my-diy-dna-test-how-foolish-i-was-to-sign-my-life-away. Accessed on February 2020.

Martinez, C.; Bates, M. Future of Privacy Forum and Leading Genetic Testing Companies Announce Best Practices to Protect Privacy of Consumer Genetic Data. *Future of Privacy Forum*. Available at <https://fpf.org/2018/07/31/future-of-privacy-forum-and-leading-genetic-testing-companies-announce-best-practices-to-protect-privacy-of-consumer-genetic-data/>. Accessed on February 2020.

Mola, R. [2020]. Why DNA testing are suddenly unpopular? *Vox*, February 13th 2020. Available at <https://www.vox.com/recode/2020/2/13/21129177/consumer-dna-tests-23andme-ancestry-sales-decline>. Accessed on February 2020.

More, M. [2013]. The Philosophy of Transhumanism. In: *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future*. p. 3-17, 2013.

Moss, L. [2003]. *What genes can´t do*. Cambridge: The MIT Press, 2003.

Navarro, F. C. P.; Mohsen, H.; Yan, C.; Li, S.; Gu, M.; Meyerson, W.; Gerstein, M. [2019]. Genomics and data science: an application within an umbrella. *Genome Biology*, 20, p. 1-11, 2019.

Park, J. Y. [2019]. Privacy in Direct-to-consumer genetic testing. *Clinical Chemistry*, 65, p. 612-7, 2019.

Santos, B. S. [1998]. Um discurso sobre as ciências na transição para uma ciência pós-moderna. *Estudos avançados*, 2, p. 46-71, 1998.

Scott, E. C. [2013]. This I believe: we need to understand evolution, adaptation, and phenotype. *Frontiers in Genetics*, 3, p. 1-2, 2013.

Su, P. [2013]. Direct-to-consumer genetic testing: a comprehensive view. *Yale journal of biology and medicine*, 86, p. 359-65, 2013.

Watson, J. D.; Crick, F. H. C. [1953]. A structure for deoxyribose nucleic acid. *Nature*, 171, p. 737-8, 1953.