An open book

Authors: Sebastián Vishnopolska

Submitted: 18. July 2020
Published: 21. September 2020
Volume: 7
Issue: 5
Affiliation: El Gato y La Caja Journal. Buenos Aires, Argentina.
Languages: English
Keywords: A libro abierto, El Gato y La Caja, journal project.
Categories: Life Sciences, Medicine
DOI: 10.17160/josha.7.5.699

Abstract:

In the 1990s, scientists started to discuss how to sequence the human genome, the complete list of genes encoded in human DNA. Two research groups, one directed by James Watson and funded by NIH grants, and another one supervised by Craig Venter and financed by private funding, started a race to be the first one to obtain the complete sequence. In 2000, the US President Bill Clinton, together with Venter and Francis Collins (Watson's successor), announced that the first draft of the human genome was available. In the last years, cost reduction for sequencing the whole human DNA open several research fields, such as cancer genomics, pharmacogenomics, diagnosis of rare diseases, and genomics of pregnancy, among many others. In the present article, originally published in El Gato y La Caja, the history and impact of this discovery is presented.

JOSHA
Journal of Science, Humanities and Arts
July 2020 Volume 7, Issue 4

An open book
– a libro abierto

By Sebastián Vishnopolska

Date of previous publication in *El Gato y la Caja*: 17/10/2018
Link to the original article: https://elgatoylacaja.com.ar/a-libro-abierto/
Keywords: What is the story of the discovery of the human genome?
What impact does (and will) knowing our genes have?

Abstract:

In the 1990s, scientists started to discuss how to sequence the human genome, the complete list of genes encoded in human DNA. Two research groups, one directed by James Watson and funded by NIH grants, and another one supervised by Craig Venter and financed by private funding, started a race to be the first one to obtain the complete sequence. In 2000, the US President Bill Clinton, together with Venter and Francis Collins (Watson's successor), announced that the first draft of the human genome was available. In the last years, cost reduction for sequencing the whole human DNA open several research fields, such as cancer genomics, pharmacogenomics, diagnosis of rare diseases, and genomics of pregnancy, among many others. In the present article, originally published in El Gato y La Caja, the history and impact of this discovery is presented.

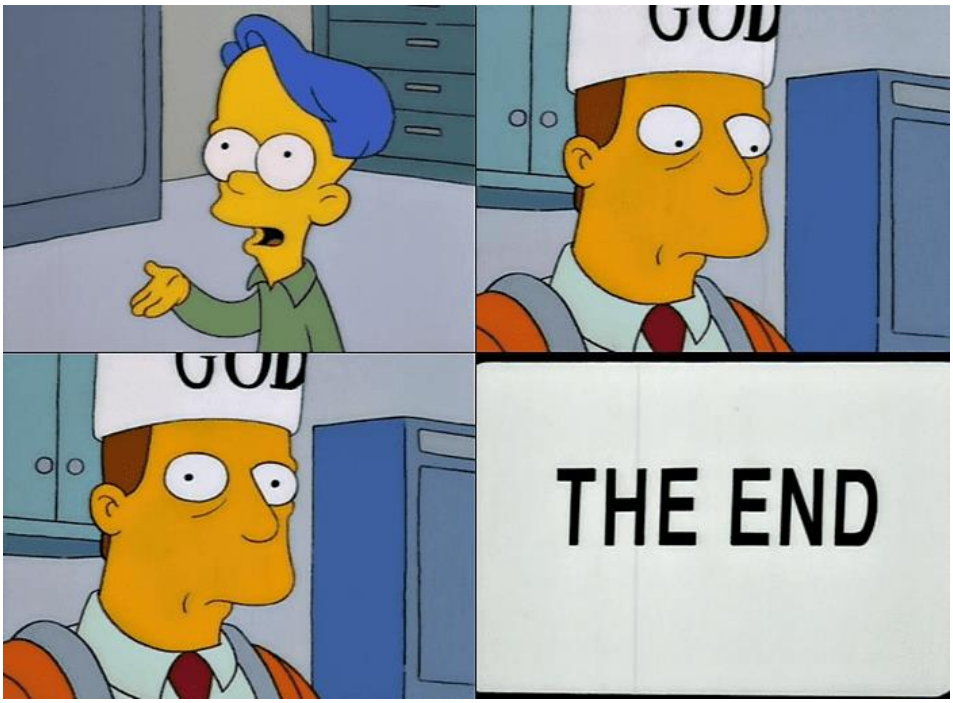
-Look, he's got daddy's eyes.

-Yes, and mommy's curls.

-Although the nose got it from the nonna.

We all hear a dialogue like that sometime, probably because many people know that physical traits are inherited, sometimes more clearly, sometimes less, always through the environment in which the organism develops. So, some of the things (not all) that make us who we are, are also printed in our DNA, that famous double helix, if there are any famous helixes.

We have long known two fundamental things about DNA: that it works like an instruction manual, that is, a jumble of information that allows cells (or viruses) to be as they are. And also that it passes from parent to parent; it is inherited, making the teacher say 'you have your daddy's eyes'. No, sir, excuse me for getting technical but I don't have his eyes, I'm rather expressing genes that came in a daddy's sperm that was more successful than the rest in finding another half of the genome (and a bunch of gene expression regulators that have more to do with mommy than with him), and multiplying to form these beautiful eyes.



Mr. McClure, what does DNA mean?

DNA was one of the superstar molecules for much of the last two centuries. In that time we went from postulating the existence of a particle that was passed down from generation to generation to understanding some of its functions, and also to knowing a lot about its structure. So much so that over time we matured the idea that this instruction manual was written in a code. And that changed everything. If we know that there are diseases that pass from parents to children, why not thinking that they are in the DNA? If only there was some way to break that code, to read it and understand what it means.

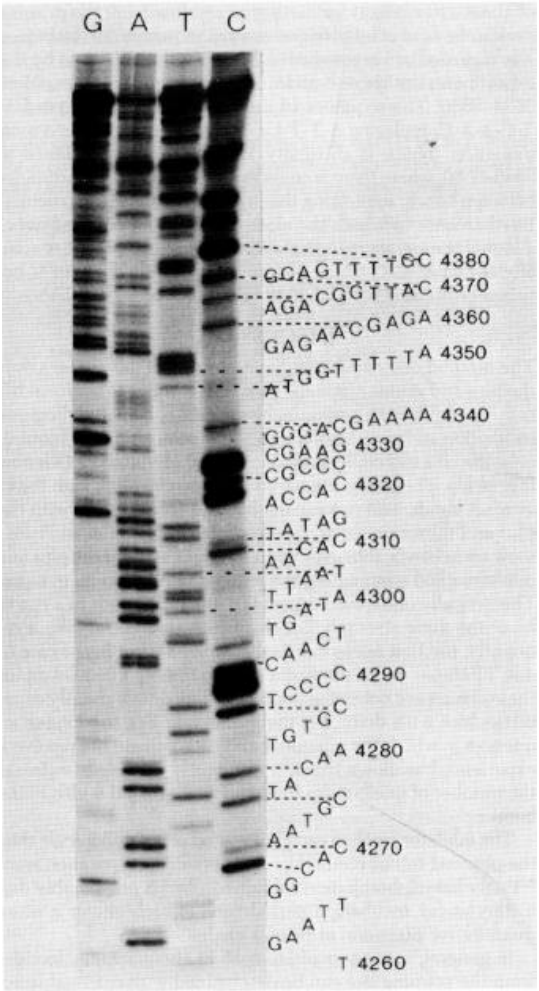
In 1990, a group of scientists working in the U.S. began to wonder what the human genome would look like, that is, what and how is the complete (or almost complete) list of genes (and other information that is not necessarily transcribed or translated, but super important) present in human DNA. They wanted to see what information in our cells is what makes us who we are and not just any other living thing.

They started with what all scientists do when they start a project: ~~devise a work plan~~...look for grants. After a little convincing, they got the NIH (National Institutes of Health) and the US Department of Energy to give them a little money - about $3 billion - and then put together a plan to read and interpret this instruction manual that we all carry inside. What happened next was fascinating. Intrigues, resignations, competitions. It turns out that...

Hold it! First, a little bit of history

DNA is a huge chain formed by a sequence of similar units (called nucleotides) that may differ in one of its parts. Thus, each piece of the DNA chain may contain a different nitrogenous base that we symbolize with the letters A, C, G and T (there are many more but, in principle, we start with those). Then we can say that a sequence is formed in an alphabet of only 4 letters. And since it is a chain, the order of these letters defines what is 'written' in that sequence and which gene it is.

The truth is that in the middle of the 20th century it was already known that this sequence was a key component in all living beings and viruses, but the problem was reading it. A little later, during the 1970s, a scientist called Frederick Sanger, who had already come up with the idea of solving riddles related to other 'letters', those of proteins (which is why he won his first Nobel Prize), developed a method to be able to sequence DNA and know which letters (nucleotides) and in what order they were part of a DNA molecule. And with this technique in 1977 he sequenced the genome of a virus, the first totally sequenced nucleic acid in history, with a length of 5370 'letters' or nucleotides.



Sanger's puzzle.

His method consists of literally putting together a puzzle piece by piece. But this technique had a problem: it allowed up to a thousand and some letters to be read at a time, while the

human genome has about 6 billion. Tiny bit of a limitation. The experimental part was very slow, expensive and a bit difficult to do. But anyway, this first version was very useful to start reading short gene sequences, formed by a few letters.

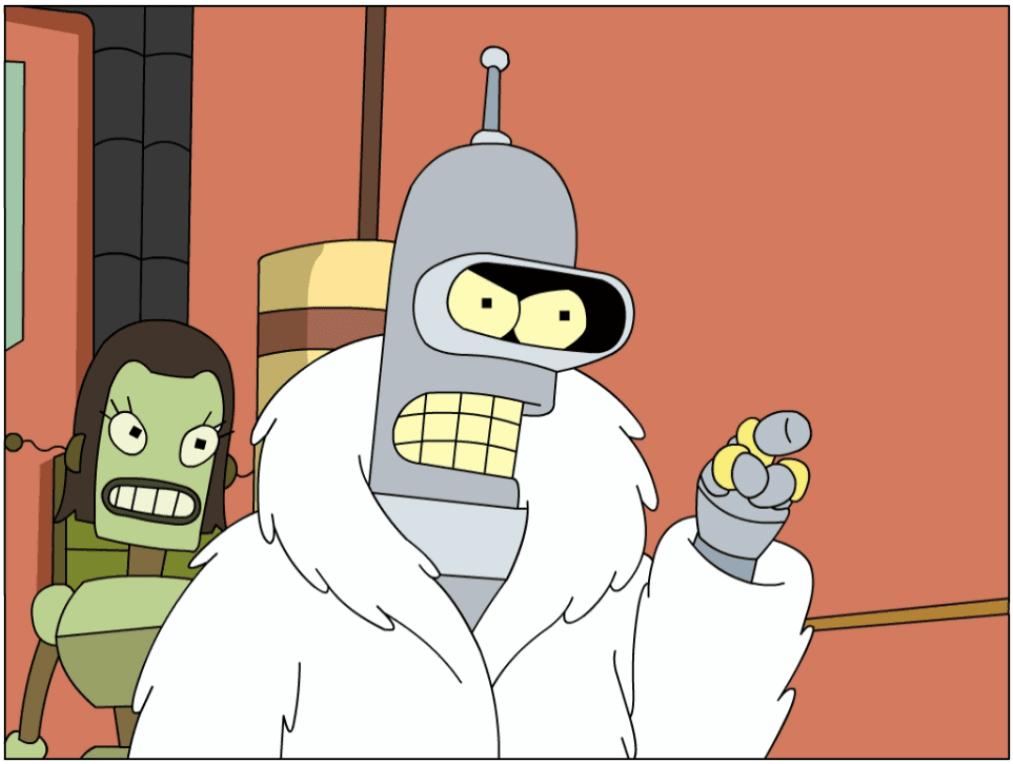
With each small advance, curiosity grew. **What would we discover if we could know the entire sequence of the human genome? Would we know exactly what makes us human? Do we have more genes than other organisms? Can we know the cause of genetic diseases?** All these questions and more were asked by a group of people in the late 1980s while they were getting these grants, which, by the way, did not come very selflessly: the NIH wanted to find out how our genome could help advance medicine, while the Department of Energy wanted to find out what effects the use of radiation and atomic energy could have on human DNA.

As a first approach, they started with genomes that were known to be smaller: in 1995 that of some bacteria or in 1996 that of beer yeast. But eventually the big step had to be taken, and as with other organisms, the methodology for sequencing the human genome was simple but laborious. It involved extracting DNA from anonymous donor cells and fragmenting it into smaller pieces. Each institution involved in the project received several of these ordered fragments that it had to fragment even more (and randomly) to obtain 'readable regions', or approximately one thousand 'letters', because this was the only way they could be read by the technology Sanger had developed. Thus, by sequencing many, many overlapping fragments, the original molecule could be reconstructed. In other words: the machinery that existed at that time only allowed to sequence a few fragments in a couple of days, **so the only way to sequence millions of fragments was to have many parallel machines working continuously and then put everything together.**

Like a cybercafe but better

At that time, **Craig Venter**, an NIH researcher working on bacteria, had been developing a more efficient method for sequencing DNA, known as 'shotgun sequencing', which proposed reading thousands of fragments simultaneously in no particular order. While this may sound strange at first glance, in 1995 Craig succeeded in sequencing the genome of a bacterium in record time. This caused him to step up the sequencing of the human genome and began a struggle of egos with the director of the Human Genome Project: **James Watson**, who won the Nobel Prize for discovering the structure of DNA (in 1953) and was always characterized by being difficult to treat.

The relationship between two such intense characters did not end well and, in 1999, Venter ended up saying goodbye to the NIH and founded Celera Genomics, his own company that began its own human genome project but with private funding.



"I'll form my own 'Human Genome Project' with private funds."

Lax and unofficial interpretation of what happened.

In those years, many researchers who discovered genes patented them; that made them the only ones able to continue researching those genes, sweeping away the competition. And this opened the door to intense debate in science. Could someone 'own' the genome of our species? **Can the human genome be patented?** Venter's move to a company and the issue of patenting **started a race between public and private projects on who would be the first to have the complete sequence of the human genome.**

The dispute ended around the year 2000, when U.S. President Bill Clinton decided to get in the middle and reach an agreement. In June of that year he held a press conference with Venter and Francis Collins (Watson's successor in the public project) **announcing the publication of the first draft of the free, public human genome, accessible to everyone.**

"So, one wanted to sequence everything in an orderly fashion, and the other wanted to sequence everything at random, and I told them, hey, get together, and bam bam, here we are, all smiling and with the thing sequenced" (unofficial statements, likely made-up ones). Craig Venter, Bill Clinton and Francis Collins at the White House on June 26, 2000. Photo: Ron Sachs.

The following year, both projects were published in the two prestigious scientific journals Nature (public project) and Science (private project).

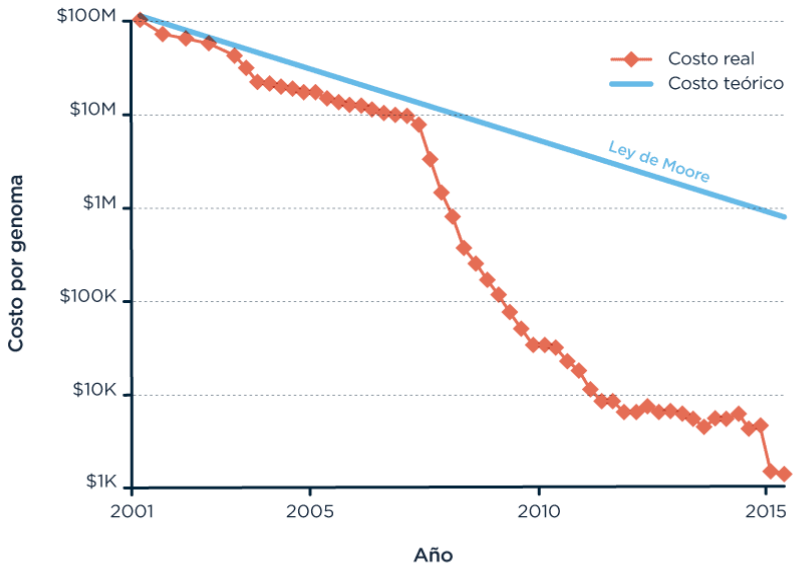
The impact

This opened a door of possibilities to the understanding of ourselves but also of many other species. For instance, thanks to the sequencing of the human genome (Homo sapiens) and the genome of the Neanderthal (Homo neanderthalensis), we learned **that our species separated about half a million years ago, and then, when they crossed back up, they made it so that now almost all humans have a bit of Neanderthal DNA in our genome.**

But perhaps the area where it can have the greatest impact is the application of genomic studies on patients. Until now much of medicine has been handled with concepts derived

from statistics: 'most people with a certain condition are served by such a drug, so every new patient who appears with the same symptoms is prescribed that same drug or treatment'. The problem is that there are a lot of cases where that doesn't work because, of course, statistics. But if the disease was caused by something genetic only and we knew the relevant bits of the genome, it would be possible to have a clear and precise marker that would allow us to make better decisions about what treatments to prescribe. No longer in statistical terms, but personalized and directed.

It has been a long time since the sequencing of the first human genome in 2001. **New technologies have made it possible to reduce the costs of sequencing a genome considerably.** The first genome cost many, many hours of work and several billion dollars, but today a genome can be sequenced for a few thousand dollars and the trend is for the price to continue to fall. So much so that, if this continues, it **may be that in the near future sequencing a genome will become as easy a protocol as any routine check-up.**



A comparison of the theoretical cost projection with the actual cost of sequencing a genome. Moore's law, badly and soon, is a projection of the speed of technological advance. Where linearity is broken, there we find a technological disruption, as were the second (early 2008) and third generation (3 or 4 years ago) sequencing techniques.

This substantial reduction in costs allows for new lines of research and application of technologies in today's medicine. There are four major areas that are booming in the use of genomic information, in this new area known as Medical Genomics:

The first is **Cancer Genomics**, and it is worth remembering at this point that cancer is (also) a disease of the genome. A series of changes can transform a normal cell into a tumor cell, which grows and divides uncontrollably, creating tumor masses in the body. Knowing which mutations a tumor has could allow us to guide the treatment to fight it. It should be clarified that this would only work in certain types of tumors, generally the less aggressive ones; since the most aggressive tumours are usually composed of a non-negligible diversity of cell populations (which is born precisely from the high mutation rate of its genome).

The second is **pharmacogenomics**, because we are finding that the way in which medications are administered is largely inaccurate. People may respond differently to certain drugs and one of the causes may be because their genome is different. Thus, a drug that may be beneficial to one person may be ineffective to another. Today, 7% of the drugs approved by the FDA (the U.S. Food and Drug Administration) have information that can help establish the right dosage based on the patient's genome.

It also opens up the possibility for the **diagnosis of rare diseases**. A rare disease is considered to be one that affects less than 1 person in 2000. Knowing this, and that about 8000 different rare diseases are known, the result is that together they affect about 7% of the world's population. In Argentina alone, it is estimated that 3.2 million people suffer from one of these diseases during their lifetime. Many of these families with affected members spend years trying to get an accurate diagnosis to explain their affliction (on average up to 7 years), not counting the many health professionals they visit. Genomics can help to diagnose some of these patients earlier and in a shorter time, thus helping families in a targeted way for their treatment.

The last one, with very interesting ethical aspects, is **Genomics in Pregnancy**. In Argentina, neonatal studies are mandatory and try to identify those babies with any of the 5 rare congenital diseases (Phenylketonuria, Galactosemia, Biotinidase Deficiency, Primary Congenital Hypothyroidism and Congenital Adrenal Hyperplasia) that, if detected and treated

early, can be reversed. Prenatal genomic studies are currently being considered to detect numerous diseases in a single test and in a non-invasive way: just taking some blood from the mother, without touching the fetus at all. In addition, babies are born acutely ill all the time, and many times they die soon after because they do not have an early diagnosis (in Argentina, for example, in 2016, 3,027 newborns died during the first week of life). Sequencing and early detection of genetic diseases could help save some of these lives.

Although it seems like science fiction, it is something that is beginning to be implemented. In Argentina, the first personalized medicine projects began in patients with rare congenital diseases. In fact, there is a slightly cheaper way of doing this that has less to do with sequencing an entire genome but with seeing if certain bits of the genome are found in a person's cells. These new sequencing technologies could be brought to patients with genetic diseases who never had a clear diagnosis. They could be told 'what you have - or what your child has - is this disease'. Whether there is a treatment available or not, this is a great advance and a huge burden that is taken away from the patient and their families.

Not without a certain elegance, knowing more and more about genomes made us realize that there are also many things that cannot be explained by looking only at the DNA sequence that makes us what we are, reinforcing once again the idea that it is necessary to look at the context in which those genomes are found, but at the same time embracing the need not to deny the differences in that long chain that gives the individual his starting point. We are still investigating not only what function each part of our genome fulfils, but how each of these parts interacts with our environment. Genomes are not everything, but perhaps they are that first piece that guides us and allows us to put together this puzzle that we call life.

References

Lander, Eric S., et al. “Initial sequencing and analysis of the human genome.” Nature 409.6822 (2001): 860-921.

Venter, J. Craig, et al. “The sequence of the human genome.” science 291.5507 (2001): 1304-1351

Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., … & Smith, M. (1977). Nucleotide sequence of bacteriophage ϕX174 DNA. nature, 265(5596), 687-695

Ko, K. H. (2016). Hominin interbreeding and the evolution of human variation. Journal of Biological Research-Thessaloniki, 23(1), 17.

Wang, Liewei. “Pharmacogenomics: a systems approach.” Wiley Interdisciplinary Reviews: Systems Biology and Medicine 2.1 (2010): 3-22.

http://fadepof.org.ar/info_epof

https://www.argentina.gob.ar/salud/glosario/pesquisaneonatal

http://www.conicet.gov.ar/campana-de-diagnostico-para-personas-con-enfermedades-raras-o-poco-frecuentes/

About the author:

Half molecular biologist, a bit of bioinformatician, and a zip of illustrator. I work with genes, usually with more than one at the same time. // Medio biólogo molecular, algo de bioinformático, un poco menos pero igual de significativo de ilustrador. Trabajo con genes, casi siempre más de uno a la vez.