Lecture Title and Date

An Individual's Perspective on Personal Genomes
02/05/2025

Objectives of the Lecture

By the end of this lecture, students should be able to:

1. Walk through Carl Zimmer's experience of sequencing his own genome / making sense of his sequencing results
2. Understand the impacts of genome sequencing on the individual, the field of medicine, and society at large, including ethical considerations

Key Concepts and Definitions

- Genome sequencing is falling rapidly in cost and may become a standard part of medical care
- Personal genome companies have been trying to break into the genomic medicine field
- Carl ZImmer's personal experience shows that there is a great deal of nuance and complexity (including with ethics) on knowing or not knowing what kind of variants you have
- Neanderthal DNA and its association with risk of severe Covid-19
- Greater understanding of the genome leads to information on complex traits (i.e. height)
- Metrics like polygenic risk scores can quantify the impact of having certain genetic variants on your health
- Genome editing is quickly approaching, but is fraught with ethical challenges

Main Content/Topics

Personal genome sequencing companies like Ancestry and 23andMe have been rising in popularity, with Ancestry gaining over 20 million people in its network in 2021 and expanding its market to 80 countries in 2022. The kind of data Zimmer received from Ancestry is a map indicating geographical regions his DNA came from. 23andMe gives results of whether or not the customer has disease-associated variants of specific genes (this does not mean it is impossible for that person to get that disease through environmental or other causes, just that they don't have those specifically studied variants). So, there are limitations to how useful results from these personal sequencing services can be.

Zimmer then gives some background on Gregor Mendel elucidating simple patterns of inheritance (recessive/dominant alleles) and how this does have relevance to certain traits/diseases such as Huntington's disease (fatal neurological disorder). This is due to a variation in the HTT gene, where if you have a certain number of CAG repeats in this gene, then you get Huntington's disease. The mechanism by which this variant HTT protein causes Huntington's disease is unknown, and there is also no treatment. There is a test that can tell you if you have this variant, but it's difficult for people to decide whether they want to get tested,

since there is no cure. Although much of what we know about Huntington's came from manually studying family trees, now we can learn a lot more from human genome sequencing. There have been major advances in decreasing the cost of sequencing a human genome, from $100 million in 2001 to $1000 in 2022 (thanks to Moore's Law).

While writing his book *She Has Her Mother's Laugh*, Zimmer had the opportunity to have his own genome sequenced by Illumina. He received a clinical report that stated he had no pathogenic variants found in the 1,691 genes evaluated, and some vague information about how likely he is to experience baldness, alcohol flush reaction, etc. He wanted to get the raw data and have his scientist friends help him analyze it.

With this data, Zimmer found out that he had a beneficial variant on IL23R, which is an immune gene. His variant was rs11209026, which reduces his odds of getting certain immune disorders. These types of variants are used for medical treatment, where people with certain autoimmune disorders can take a drug to mimic what his body does naturally. This applies to a range of autoimmune disorders, not only Zimmer's specific variant. Zimmer had also delved into ancestry with his brother and found they share identical runs of DNA, showing close relation.

Zimmer then shifts the topic to ancestry of the human population, with a particular focus on Neanderthals. All humans share a little bit of Neanderthal DNA, with African populations having a very low percentage and other populations having a range of 2-3%. This is because our species evolved in Africa and then expanded outwards, where they encountered Neanderthals. Neanderthal DNA in Africa is due to a large migration of people back into Africa.

Severe Covid-19 has many risk factors, and one such risk factor is inherited from Neanderthals. Zimmer proposed an idea to the class that this variant may have been effective against the types of viruses Neanderthals were encountering in Europe and Asia that may not be good when dealing with Covid-19. However, Neanderthal DNA is also shown to reduce risk of severe Covid-19, one specific example being on gene OAS3. Zimmer's DNA showed no alleles for increased risk on chromosome 13 and two heterozygous SNPs for decreased risk on chromosome 12.

Zimmer then discusses completion of the human genome and height. Only in 2021 were scientists able to finish sequencing the whole human genome, and there is ongoing work on a telomere-to-telomere consortium. Understanding of the human genome will likely give rise to complex information, with height being a good example. The first variant associated with height was discovered in 2007, but this number grew over time, ballooning to 12,111 variants in 2022. A paper aimed to predict heights based on these variants used a polygenic score, and they were only able to do a decent job. However, the performance dropped when individuals were not of European descent, a result of genomic data being heavily biased towards people of European ancestry.

It is also important to note that the data is not static. The example presented in class shows women from Canada being taller than women from Barbados in 1896, but data from 1996 shows the opposite. The heights of women in 1996 were also taller than women in 1896. Over that time, diet and healthcare have changed, and these changes are reflected in women's height.

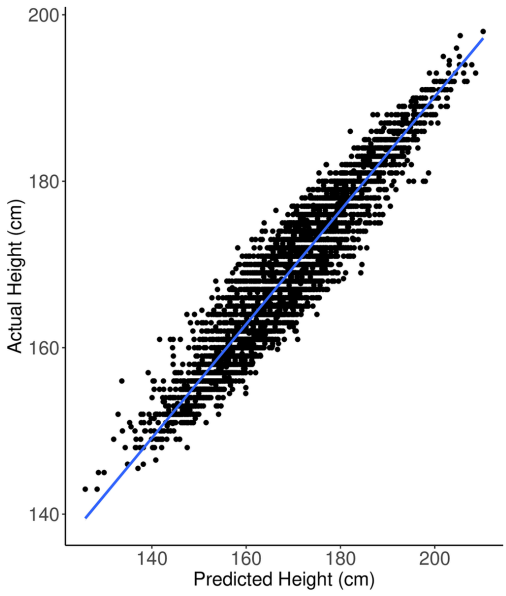
One way of quantifying the impact of genetics on health is by way of a **polygenic risk score**, which is a numerical value that estimates an individual's genetic predisposition to developing a complex disease by summing effects from one or many different SNPs calculated

from an individual's genotype and relevant genome-wide association study (GWAS) data; therefore, polygenic risk scores explain the relative risk of disease. They do not provide a baseline or timeframe for disease progression. They only show correlations, not causations.

They are typically calculated as a weighted sum of trait-associated alleles and only account for genetics, ignoring environmental factors. For instance, a *single-gene disease* like cystic fibrosis is caused by variants in the CFTR gene on chromosome 7. Zimmer's example in the lecture is on coronary artery disease, which as a *complex disease* is linked with many genetic variants.

Example: Consider two people with high polygenic risk scores for having coronary artery disease. The first is 22 years old, while the latter is 98. Although they have the same polygenic risk score, they will have different lifetime risks for the disease. This is in contrast to absolute risk, which shows the likelihood of a disease occurring. For instance, women who carry a BRCA1 mutation have a 60-80% risk of breast cancer, which would be true even without any comparison to any groups of people.

They are widely used in animal breeding and plant breeding due to their efficacy in improving livestock breeding and crops. The below example illustrates a plot of predicted versus actual height in humans through the use of a polygenic risk score (source). Papers and companies are now being written and established on the concept of predictive medicine, which naturally leads to the question of whether we should try and change these pre-determined outcomes by way of genome editing.



But should we change our genomes? Proponents argue that we can drastically reduce the rate of debilitating diseases such as Alzheimer's, but opponents like Visscher et al. argue that supposedly positive consequences of editing embryo and germline genomes can have serious unintended consequences and that we must observe certain ethical counter-considerations. Most seriously, it could renew interest in eugenics and deepen health inequalities; we must respect individual liberty and individual values. In addition, one gene can have myriad targets in the body; altering one gene can have cascading unintended side effects.

Discussion/Comments

- In regards to human height, the number of variants associated have been increasing and have not yet plateaued. If ample data on individuals that are not of European descent is collected, how much larger would the number of variants grow? In addition to this, since prediction performance is worse in non-Europeans, how would these variants be weighted across different races when determining polygenic scores?
- Another possible concern with genome editing, both in humans and other species, would be that it could negatively affect biodiversity and lead to subsequent ecological damage.

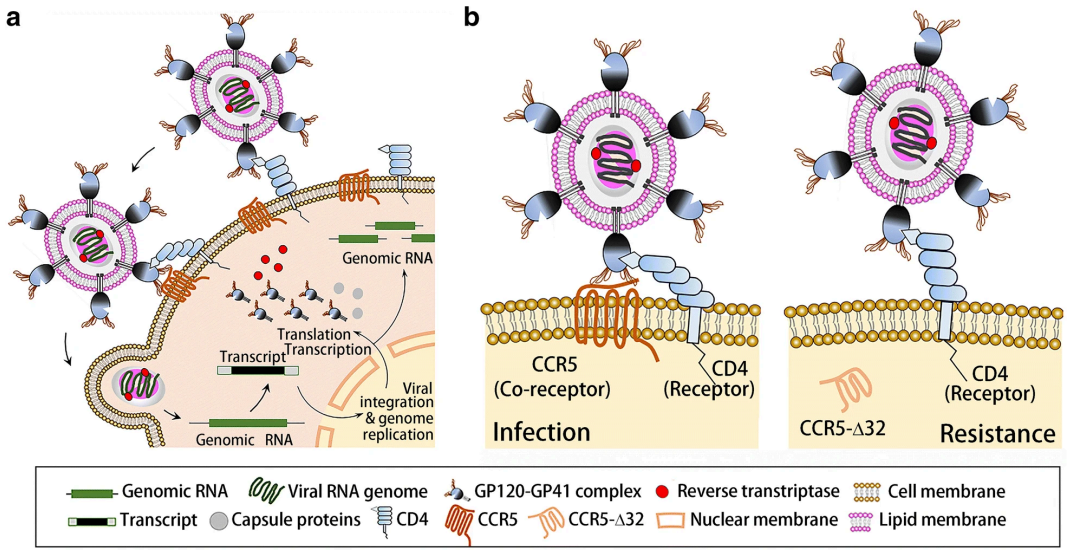
Suggested readings:

Zimmer's recommendations:

- She Has Her Mother's Laugh: The Powers, Perversions, and Potential of Heredity (Carl Zimmer, 2018)
    - Book about heredity, how science has informed us about heredity, how ideas about heredity have been used to justify oppression/genocide/etc.
- Air-Borne: The Hidden History of the Life We Breathe (Carl Zimmer, 2025)
    - History of science regarding airborne diseases, aerobiology, and airborne biological weapons during the Cold War.
    -

Other suggested references for many of the key concepts

The case of Dr. Jiankui He, a Chinese scientist who in an infamous 2018 affair claimed to have created the world's first genome-edited babies (he said he altered the CCR5 gene in a pair of twin girls, supposedly conferring onto them HIV resistance), offers a cautionary tale of what can happen with inadequate guardrails surrounding genome editing. Dr. He received widespread international criticism for his experiments and was fined and sentenced to prison for three years in China. (Though he's back out now and is now the director of the Institute of Genomic Medicine at Wuchang Technical College in Wuhan.)



Mechanism of CCR 5-related HIV resistance (source)