IDC School of Design
अभिकल्प विद्यालय

Project 2

Data Visualisation of Disease-Carrying Genes along the Human Genome

Submitted by:

Niharika Mohile
18U130021

Guided by:

Prof. Venkatesh Rajamanickam
IDC School of Design, IIT Bombay

Declaration

I declare that this written document represents my ideas & interpretation of the data derived from literature, in my own words. I have tried (to the best of my capabilities) to adequately cite and reference the original sources wherever others' ideas or words have been included.

I also declare that I have adhered to all principles of academic honesty and integrity and have not fabricated any idea/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from those who have not granted the appropriate permissions.

Niharika Mohile
18U130021
IDC School of Design, IIT Bombay
8th July 2022

Acknowledgements

I express my sincere gratitude to Prof. Venkatesh Rajamanickam, whose guidance at various stages of this project was invaluable, and helped the project reach completion. I would also like to thank Profs. Swati Pal and Vivek Kant, for their feedback at regular presentations, which forced me to think further and identified the shortcomings I had to address.

I am thankful to my mother, without whom I really cannot do anything in life, and also especially thankful to my high school biology teacher, Ms Kumpal Walia, because of whom I knew what I did, on this topic.

I am also ever grateful to my peers for their continued support and motivation.

Contents

Aim

To create a data visualisation that draws interest into the field of genetics, while also being something fun to play around with, for those that already know the topic and would learn from engaging.

Introduction

DNA is the backbone of inheritance. It is DNA and RNA that carry all information possible into an organism since its conception, even before its birth. How to walk, breathe, grow, how to exist – is all programmed into an organism by its DNA. An explanation would be thus: the DNA of an organism contains several genes, each coding for a particular compound in the body, like an enzyme, hormone, metabolite or a structural protein – all of which affect how the body functions. This DNA is given to every organism by its parent(s), given to them by their parents. In our DNA, we carry millennia of history; relics of ancestors long gone.

I knew my legs were slightly bow-shaped because so are my aunt's, and her aunt's. I knew it runs in the family. But it was in classes eleventh and twelfth, where I learnt what 'running in the family' actually meant, and understood the mechanism behind it. I always loved the concepts of evolution and heredity, and genetics was the concretisation of these concepts. Within genetics, learning about the existence of pedigree charts, and the simplicity of their functioning, remains a highlight.

Genetics is a field where everything physical is microscopic, and the rest theoretical. Thus, data visualisation in this field has always used abstraction as a tool, to represent to others what is otherwise invisible. This project seeks to combine the current conventions of representing genetic data, with my own additions, in order to come up with a visualisation that is easy to understand while also being informatic.

Background

DNA – its Structure & Function

DNA is the medium of storage for all genetic data. Every bit of genetic information passed down from parents to offsprings is stored as the DNA of that organism. At the most basic level, DNA is a polymer of deoxyribonucleic acids. Each monomer consists of a deoxyribose sugar, attached to a phosphate group, as well as one of the four nucleotide bases: adenine, thymine, cytosine, and guanine (A, T, C, G). The sugar and phosphate group together form the backbone of a single DNA 'strand'. The nucleotide bases form hydrogen bonds with their complementary bases (A with T, C with G) in a reverse DNA strand. Together, the two strands form a ladder like structure, which is curved into a double helix structure, resulting in the 'twisted ladder' image of DNA that's most common.

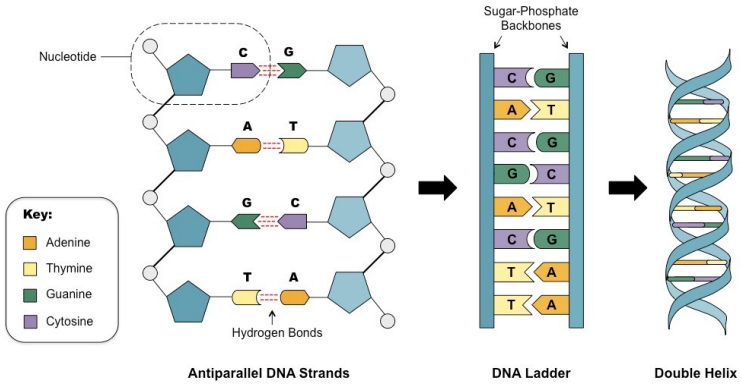

Fig 1. Structure of DNA [1]

This long strand of DNA is further coiled and looped, wrapped around proteins called histones, giving a 'pearls on a necklace' sort of appearance. This structure is called chromatin. Chromatins are further condensed, coiled into structures called chromosomes. Together, all the chromosomes are present in the nucleus of the cell.

Within the nucleus, there is an entire complex mechanism through which data from DNA is used. In short, parts of the chromosomes are opened up and uncoiled into simple DNA segments. This DNA is 'unzipped' (ie, part of it is made single-stranded, and the hydrogen bonds between the nucleotide bases are broken) with the help of enzymes. This DNA is then translated into proteins. The DNA itself is a series of ATCG. Each consecutive triplet in this series codes for one particular amino acid, and is called a codon. For example, the codon AAA codes for the amino acid Lysine. In this way, the DNA produces a string of amino acids – the building blocks for proteins. These bond together to form the complete proteins. Each string of DNA that codes for a particular protein is called a gene.

Genomes and the Human Genome Project

A genome refers to all the genetic information stored in an organism. For most eukaryotic organisms, this is the sum total of all the DNA in the organism - both nuclear DNA as well as mitochondrial DNA. This genetic information

defines everything about the organism - its physical structure, its biological processes, and its information passage to its progeny. The genome is akin to an instruction manual, defining the rules for each cell, on how to grow, live, and reproduce.

The Human Genome Project (HGP) is an international scientific research project, started in 1990, which aimed at identifying the sequence of base pairs of the entire human genome. It was declared complete at 85% identification in 2003, and 99.9% complete more recently in 2021. Obviously, no two humans have the exact same genome. The HGP hence took samples from a number of individuals to collate and combine into their resultant genome. The project managed to identify the sequence of over 3 billion nucleotide bases that make up one haploid genome set, and also identified the sequence, function, position etc. of the ~22,500 genes the genome contains. It has now resulted in a huge database of information, of all the different genes that we have, their location, size, function, and much more.

Inheriting Mutations – and thus Diseases

DNA is the blueprint for the building and functioning of every part of our body. This information is carried forth by genes – the sequence of bases in the gene determines the sequence of amino acids in the subsequent protein.

At times, however, defects creep into DNA. This could happen during DNA replication before cell division, due to certain viral infections, or because of harmful chemicals or radiation. These defects are called mutations. A mutation could refer to deletion of one or more nucleotide bases from the DNA (for example, the sequence ATTCGAT becomes ATTGAT), or could refer to the substitution of one or more bases with different bases (for example, TGGCGCA becomes TAACGCA). The change in DNA sequence leads to a change in the amino acids it produces, and hence a change in the protein. At times this can be harmless. Other times, the mutations lead to diseases, like cancer. If these mutations are present in the germ cells (the sperm or the egg), then they can get passed down to the offspring.

Several diseases are hence hereditary. While some diseases are caused by bacteria, like Typhoid, hereditary diseases are caused by defects in DNA. Actually, several diseases have multiple causes. For instance, male pattern baldness is an issue widely known to be genetic [2] – it is common to hear among young men that they would go bald young because their father did. But, it is actually also affected by several other factors in the person's life, like their diet, stress levels, medications, etc [3]. Most genetic diseases have a certain percentage of cases which are caused by genetics. For instance, 5-10% of all melanoma cases show familial heredity, caused due to mutations in the PTCH1, PTCH2, or CDKN2A genes [4].

Basics of Human Inheritance

Human beings have 23 pairs of chromosomes. While most cells in the human body have the full 23 pairs of chromosomes, the germ cells are haploid – they only have 23 chromosomes each. The egg has 22 autosomal chromosomes, and one sex chromosome, the X chromosome. From the sperm too, the zygote gets 22 autosomal chromosomes, and another sex chromosome. This sex chromosome, which one gets from the father, could be either X or Y, and determines what sex the baby is assigned at birth (XX for female and XY for male).

One thing to note is be that each biological parent has pairs of each chromosome, as mentioned several times above. Out of these, they only pass on one of each chromosome number to the offspring. If only one parental chromosome carries the diseased gene, then there is a 50% chance of the child getting the diseased gene.
It is possible that your parent carries a gene, but it remained dormant in them and never presented, and hence they did not get the disease. They could still be 'carriers' and pass on such disease-carrying genes to you.

While genes come only from parents, doctors still take the entire family's history - because diseases that your grandparents or uncles and aunts have are indicative of the genes that your parents have. So if your dadi had breast cancer, it is possible that your dad has the mutated gene, and passes it on to you.

Here comes in handy the notation system used to understand diseases across your family tree, one that is also used in my final prototype. These are called pedigree charts. They use shapes to represent people and lines to establish relationships between them. A square represents a male, and a circle stands for a female. A horizontal line between two shapes indicates marriage or mating, and a vertical line going down from a marriage line leads to offsprings. Filled shapes are used to show those that were affected by a particular disease.

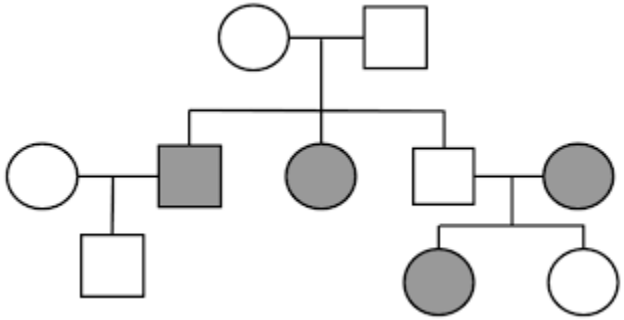

Fig 2. A Pedigree chart

Existing Visualisation Techniques

There are multiple ways of visualising genes currently in use within the scientific community. I started my secondary research by attempting to understand these visualisations, hoping to find one I would want to recreate. These visualisations, I found, differed mainly upon their spatial orientation.

Circular Plots

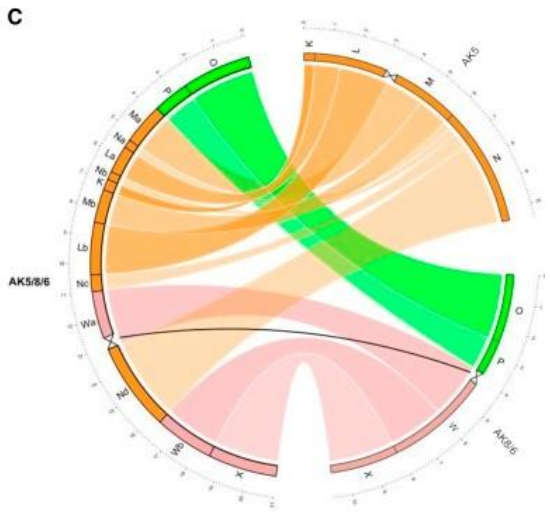

Fig 3. A circular plot showing collinearity between homoeologs in Cardamine amara and Cardamine rivularis and a chromosome in the Cardamine pratensis and in C. × schulzii [5]

These were the most commonly used plots, and were the inspiration behind most of my initial ideation. These visualisations used two main methods of depicting data. In the first, different fields of information about the genes were represented along the circumferences of concentric circles. The other method was using curves of varying thicknesses to connect one part of the circumference to another, to establish a relationship between them. Some plots used both of these simultaneously, using the second method for the innermost circle among multiple concentric circles.

Simply Linear Plots

Fig 4. A snippet from the genome browser Ensembl, for the gene HSD17B14

These plots are most commonly seen in genome browsers. They represent different fields of information about the gene in different parallel lines, using colour to differentiate between the segments of the gene. They are the most intuitive to understand, and the quickest for a layperson to grasp an understanding of. However, for representing large quantities of data, these are not comfortable because one

has to either scroll through data or squint at minute details.

Linear Orthogonal Plots

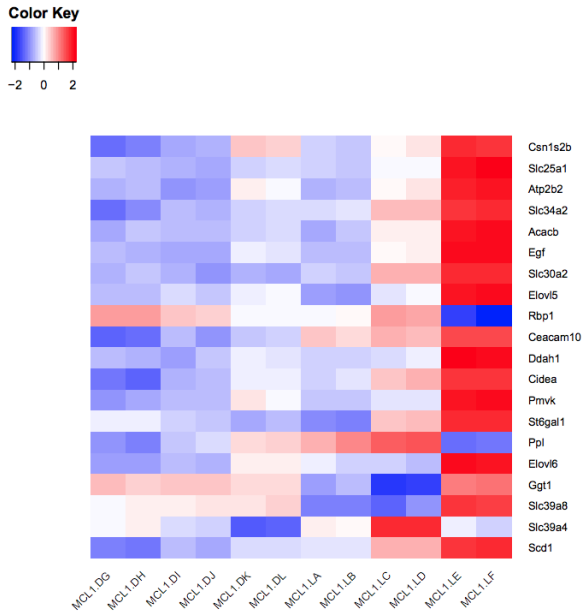

Fig 5. A heatmap comparing expression of genes in lactating mice vs pregnant mice [6]

These plots represent two different categories of information perpendicular to each other, along the x and y axes. They are also fairly intuitive to read. A shortcoming of orthogonal plots is that they are limited in the amount of data they carry.

While I took certain elements of representation from such existing visualisations, most of my final outcome strays from conventional methods of representation. I sought to combine that which is intuitive from the existing visualisations, with the narrative sort of visualisation that I wanted to build.

Project Definition

Design brief

To create a simple, engaging visualisation that helps better understand the roles of genes and inheritance in our lives

Objectives

- Create interest in children who find the topic boring
- Be a tool to help understand, for children who find it hard
- Be fun and informatic enough for even the general public to play with

Approach

- Keep content and representation simple enough to understand for even those with brief knowledge
- Make it engaging while maintaining an attitude befitting the seriousness of the content
- Use simple interactivity to increase engagement, while making sure it is not complex to use
- Make it personal to make it more relatable and interesting, and bring more real-life value into it

Initial Ideation

Following were some of the ideas I had for this project:

a) **A circular plot comparing the different chromosomes of the human being**
It would visibly compare features of different chromosomes like their lengths, centromere positions, number of genes they carry, etc. The positives of this idea were that the data was easily available and tools like Circa are available to make such plots. Drawbacks were that for one, the idea was not novel, and had been done multiple times before. I also had no particular push towards this, and was not sure of what I wanted to show through it. Finally, because I wasn't sure of what it could show, it didn't seem particularly useful or educational, or seem like one would gain anything from it.

b) Comparing the human genome to that of the dog, or neanderthal, or banana
Another idea I had had was about comparing the genome of human beings to that of the neanderthal, or something like the banana. While this was to be the content, I had not ideated on the form of the visualisation. This visualisation was mainly an idea because of the 'wow factor' it gives people when they see how much similar our DNA is to that of what is seemingly unrelated, like the banana. This idea too, however, was not novel, and one did not gain much by it.

c) X chromosome versus the Y chromosome
The X and Y chromosomes, also called the sex chromosomes, are what determine the sex of the individual. Visually, they are already very different, with the Y chromosome being much shorter than all the others, including the X chromosome. This project had potential across multiple avenues. Heredity along sex chromosomes is quite interesting, with diseases like colour blindness, which are sex-linked. The project could also discuss modern-day politics around gender, and could also talk possibly about those who have chromosomal abnormalities due to an extra X chromosome etc. This topic quite interested me, and also seemed to be of value. It did, however, seem to stray a little from my original discussions of visualisations of DNA sequences, and worries of it not being scientific enough led me to not choose this for the final outcome.

d) Comparison of disease-carrying gene to 'regular' gene
This idea started with an explanation of genes, how they are sequences of triplets of the nucleotides, and how each triplet (codon) codes for a particular amino acid. In this way, each gene codes for a particular protein, each protein being a polymer of

amino acids. The project would then take a particular gene as an example, and show how mutations in this gene lead to changes in the protein formed, leading to diseases. This idea seemed fairly informative, and also feasible given my skill level and timeframe. One of its drawbacks was that it felt more like a textbook explanation, and there was nothing particularly novel about it.

e) **Comparison of different isoforms of a gene**
This visualisation would begin same as the last one, with an introduction to the structure and functionality of genes. However, it would then move on to try to explain the more complex concept of isoforms of a gene, with examples. This topic had the same shortcomings as the last, of being more of a textbook explanation than a creative visualisation.

f) **Find a gene that I have and visualise it**
I tried to change my approach to ideation, make it less of a textbook explanation of scientific concepts, and more of something personalised. I thought for this I could identify a gene I have, like that for brown eyes, or bad eyesight, and then visualise it. In the visualisation, I planned on using Prof. Venkatesh's process of replicating an existing visualisation, finding flaws, and then redesigning it. I later abandoned this idea because the visualisation itself did not seem to convey much, and would only be understood by those already working in the field of biotechnology or genetics.

g) **A collage of different genes I have**
Inspired in part from the previous idea, and in part from idea (d), this one combined the idea of doing something personal while also explaining a scientific concept. Through this idea, I thought I could explain a bit of the structure of the genome, with different sized chromosomes etc, and show particular genes on the chromosomes. These genes would be genes that I myself have, and would make the visualisation a sort of self-portrait. Finding raw data for this project would be harder, since I do not actually know what genes I have. But this idea seemed genuinely interesting, and also novel.

Idea Selection

From the different concepts I had gathered in the initial ideation, I had to choose one to go forward with. I did this by evaluating them on the following factors:

1. **Novelty-** required the idea to be unique, and also not simply be a copy of what is taught from textbooks
2. **Gainful Impact-** needed to be something that boosted one's concepts, or was educational
3. **Topic Relevance-** topic needed to be related to genetics and heredity, and not stray too far from what is taught
4. **Engagement-** evaluated how engaging users would find the tool, based on preliminary feedback from friends, and how interesting the idea seemed to me

	Novelty	Gainful Impact	Topic Relevance	Engagement	Total
(a) A circular plot comparing the different chromosomes of the human being	1	3	3	1	8
(b) Comparing the human genome to that of the dog, or neanderthal, or banana	1	1	1	3	6
(c) X chromosome versus the Y chromosome	3	9	3	3	**18**
(d) Comparison of disease-carrying gene to 'regular' gene	3	3	3	3	12
(e) Comparison of different isoforms of a gene	3	3	1	1	8
(f) Find a gene that I have and visualise it	3	1	3	1	8
(g) A collage of different genes I have	9	3	3	3	**18**

To force a decision, I used a scoring system of 1-3-9. The two ideas that scored the most through this decision matrix were the X chromosome versus Y chromosome one, and the collage of the different genes I have. Between these two, I picked the collage because it was more personalised, and because I thought it was relatively more engaging.

Final Visualisation

The prototype for the final visualisation is available at: https://www.figma.com/proto/pkrFcTR5XccZcm4kYDqq77/P2?page-id=205%3A2&node-id=205%3A621&viewport=-169%2C383%2C0.6&scaling=contain&starting-point-node-id=205%3A621

Process

I started work on the final visualisation by collecting family history on the diseases that have been present, and whether they have a genetic component. Upon collecting this information, I made a database with these diseases, the genes that cause them, and all the other data that I have represented in the visualisation.

I did my own research on which of the multiple genes that cause the disease fits my case the best, and is likely to be the gene that caused it in my family member. For some, it was easy – mutation in BRCA2 is the leading cause of genetic breast cancer, and is easily most likely to be the gene in my family as well. On the other hand, for Parkinson's, there are generally PARK 1-8 mutations – 8 loci on the genome that cause it (they are known with other names as well as PARK#). While the one that is most common is not PARK8 (another name for LRRK2), PARK8 is the one that is most responsible for late-onset Parkinson's, the kind my great grandmother had. Others like PARK1 are responsible for a higher number of genetic Parkinson's cases, but are responsible for early-onset Parkinson's disease. Thus, reading up about the different genes that could be responsible, I came up with a database of the genes I would code in my visualisation.

Within the database, I collected multiple points of information on each gene – its exact position, size, mode of action, and so on. Based on professor Kant's feedback, of how I may not necessarily have the genes I have represented, I also found for each disease the number of cases that are likely to be genetic. This, along with the aforementioned in-depth research done before choosing each gene, I think addressed the issue. In the database, I also included some personal points of data, like how much that particular disease has impacted my life, at what age the disease expressed itself (or generally expresses itself), etc.

After this, I ideate on ways of encoding this data into a visualisation. The first prototype I made was entirely different from the current one – it was more artistic, less abstract, and encoded much more information. However, upon discussing with Prof Venkatesh, I realised the multiple flaws it had. Though it encoded more data, it was encoded in a way that was neither conventional, nor

intuitive, making the visualisation impossible to understand without explanation. For instance: it encoded, for size of the gene, two separate measures of size – the conventional base pair count, and also the number of exons the gene has. In the visualisation, genes were, for lack of a better word, scribbles, on the screen. The length and breadth of the binding box of a scribble were representative of the exon count and base pair count of the respective gene. This is not something a layperson would think of or understand, unless it were written down or orally explained to them. A data visualization that needs exhaustive explanation did not fit the requirements I had set for myself, so I ideated further.

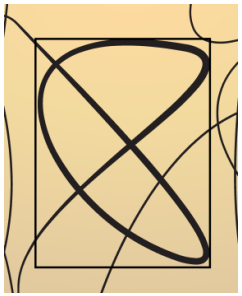

Fig 6. Example of a scribble representing the gene SLC6A4; the height of the box represents 41683 base pairs, the width represents 15 exons

I initially tried to fix the existing visualisation, addressing the points Prof Venkatesh had listed. For the issue mentioned above, for instance, I tried to encode the size of gene as what appeared to be the visual size of the scribble – a combination of its length and thickness. A toggle interaction helped change between base pair count and exon count as the measure of size. However, while working on this, I did feel that some measures of information I have may not necessarily be relevant, I was coding the information merely because I had found it – a baseless reason. Moreover, with Prof Venkatesh I had discussed the level of abstraction I was using. He had explained how my current visualisation was at an awkward level of abstraction, and suggested I play around with it to find one that made sense. In fact, it was his advice to completely stretch out the entire genome, which is what I did in my final visualisation. Playing around with pen and paper, on encoding methods and abstraction levels for the final visualisation, I arrived at what is now my final visualisation.

Features

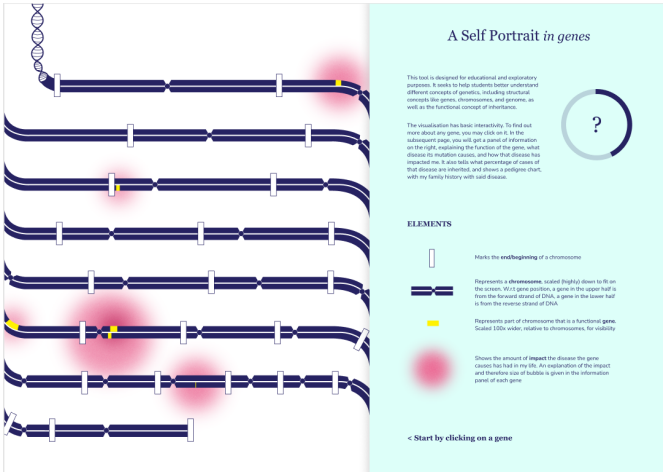

Fig 7. Screenshot of final visualisation, main page

The human genome is shown as a series of dark blue parallel bands which flow from left to right. The left and right ends have curved dips up and down respectively, to indicate a continuous flow between the bands, such that they together constitute one whole genome.

This genome is divided into 23 chromosomes using rectangular ‘dividers’ at the beginning and end of each chromosome. These not only divide the bands into 23 distinct segments, but they also enable one to see and compare the sizes of the different chromosomes.

Each chromosome is further divided into two vertical halves by a single white line, along with a circle somewhere along the length of the chromosome. The dividing line and circle serve dual functions. First, they give each ‘chromosome’ the look of an actual chromosome, dividing it into the classic X shape of chromosomes. Further, the circle represents the centromere of the actual chromosome, and each circle is placed according to the centromere’s position on the chromosome. The line also divides the chromosome into 2 parts, which are then used as the forward and reverse DNA strands. Genes present on the forward DNA strand are shown on the upper half of the blue band of chromosome, while genes present on the lower half of the blue band are actually found in the reverse strand of the DNA of that chromosome.

The genes mentioned above are shown as yellow portions on the blue bands of chromosomes. Due to their significantly smaller size as compared to the size of the chromosome, the genes would not actually be visible, if shown to scale. Hence, these genes are shown at a 100x magnification in length. They are still present at the correct position on the chromosome, aligned centrally.

Finally, each gene is responsible for a particular disease present in my familial history. How much this disease (and thus, indirectly, the gene) has impacted my life is shown using pink fading circles behind the gene. The size of the circle is representative of the amount of impact. 3 sizes of circles are seen, to show high, medium, or low impact. The largest circle (on the 13th chromosome) shows the impact of the gene for Major Depressive Disorder, a disease that has had a large impact on my life. On the other hand, though I am likely to have breast cancer later in life, it has had minimal impact on my life so far, and hence the gene BRCA2 has the smallest circle behind it.

Each gene (yellow segment) on the main page is clickable, and leads to a page that gives more information about that gene and its disease. Firstly, on the bands, there are more genes highlighted. Mutations in all of these genes lead to the same disease as the clicked gene. While this sounds a little confusing and counter-intuitive, it is important to recognise that there are multiple loci across the genome, where mutations for each disease are found – not just one. Some diseases can be caused by mutations in just one,

while others need a combination of mutated genes to be present, in order for the disease to be expressed. The mechanics of these are quite complex and beyond the scope of this project.

Upon clicking, the right-hand side panel gives the following information about the gene:

1. Full name(s) of the gene, its size in base pairs, which chromosome it is on and where
2. What disease it is responsible for
3. Mode of action, that is, how the mutated gene causes the disease
4. Some personal excerpts of my history (or lack thereof) with the disease
5. Percentage of cases of that disease that are genetic
6. Pedigree chart showing familial history with respect to that particular disease

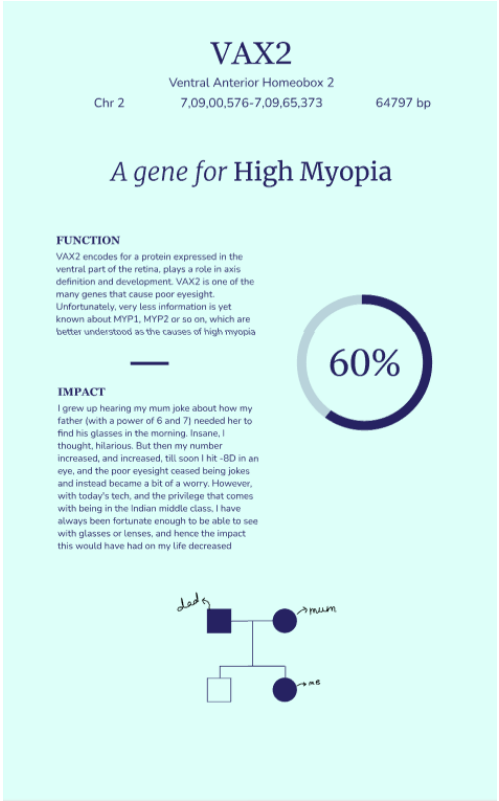

Fig 8. Screenshot of right-handside panel for gene VAX2

Evaluation

Evaluation was conducted with a single user, of the target demographic. R, 18F, is in her first year of medical school and has recently completed 12th class biology.

Methodology

The user was not given any overview of the project. They were first given the prototype to explore freely, and a concurrent think-aloud protocol was applied. The user was asked to speak aloud whatever she was thinking, questioning, or understood.
After this, a semi-structured interview was conducted, to find out shortcomings of the current prototype.

Observations

1. While reading the description of the project on the main page, it took the user a while to understand that the genes are my own.
2. When first encountered gene names like MYP1, was not sure what it means.
3. Subheadings which declared the size and position on chromosome of the gene were vague, user did not understand what the numbers or the abbreviation ‘bp’ meant
4. User said that the language used in text blocks is sometimes vague and hard to understand
5. User related the red circles along the genome with the red circles they use to mark pain points on the body, and appreciated thee similarities
6. When asked particularly about the visual elements, user said she recognised the chromosomes instantly because of the X-shape, and did not need the legend
7. Even when probed towards an answer, user was unable to articulate ‘genome’ as what the string of chromosomes represents
8. User mentioned enjoying the pedigree charts - “It was fun to try to figure them out myself, and then see if I was correct by rolling my mouse over it”
9. User seemed to easily grasp the overall concept, that genes can cause diseases, and these are the genes that I think I may have

Insights

1. More information needs to be given on the main page, regarding subheadings, percentage chart, pedigree chart, etc. that is mentioned in the information panel.
2. Assumptions that students would recognise the abbreviation ‘bp’, or alphanumeric strings like ‘MYP1’ as gene names have crept in, and need to be amended
3. While the overall concept seemed to be understood, smaller concepts like the blue bands representing the genome were missed out on.

Challenges and Learnings

1. Had I planned my timeline better, I would have liked to do a proper evaluation of the visualisation and made it better before submission
2. I did not meet my guide enough during the semester, and so my project suffered from lack of valuable feedback, which could have made the visualisation better and the final outcome less text-heavy
3. An irrational fear of coding led me to stay staunchly away from a lot of projects I thought were interesting, a mistake I will try my best to avoid in the future
4. In hindsight, my ideation was narrow and limited. Too late into the semester, while explaining my project to friends, did I realise that a visualisation I would have enjoyed more and been better at would be something that explains basic theory, rather than unnecessarily building upon advanced topics

References

1. Cornell, B. "DNA Structure | Bioninja". Ib.Bioninja.Com.Au, 2016, http://ib.bioninja.com.au/standard-level/topic-2-molecular-biology/26-structure-of-dna-and-rna/dna-structure.html.

2. Hagenaars, Saskia P et al. “Genetic prediction of male pattern baldness.” PLoS genetics vol. 13,2 e1006594. 14 Feb. 2017, doi:10.1371/journal.pgen.1006594

3. Cranwell W, Sinclair R. “Male Androgenetic Alopecia.” Feingold KR, Anawalt B, Boyce A, et al., editors. Endotext [Internet]. South Dartmouth (MA): MDText.com, https://www.ncbi.nlm.nih.gov/books/NBK278957/

4. Bethesda, MD. "PDQ Genetics of Skin Cancer." PDQ® Cancer Genetics Editorial Board. National Cancer Institute. Updated 05/04/2022. Available at: https://www.cancer.gov/types/skin/hp/skin-genetics-pdq.

5. Mandáková T, Kovarík A, Zozomová-Lihová J, et al. “The more the merrier: recent hybridization and polyploidy in cardamine”. Plant Cell. 2013;25(9):3280-3295. doi:10.1105/tpc.113.114405

6. Doyle, Maria. “Visualization of RNA-Seq results with heatmap2 (Galaxy Training Materials)”. 2021. https://training.galaxyproject.org/training-material/topics/transcriptomics/tutorials/rna-seq-viz-with-heatmap2/tutorial.html