Bachelor Thesis

Multi-Omics Network Visualisation in Virtual Reality

Author:

Jip de Kok

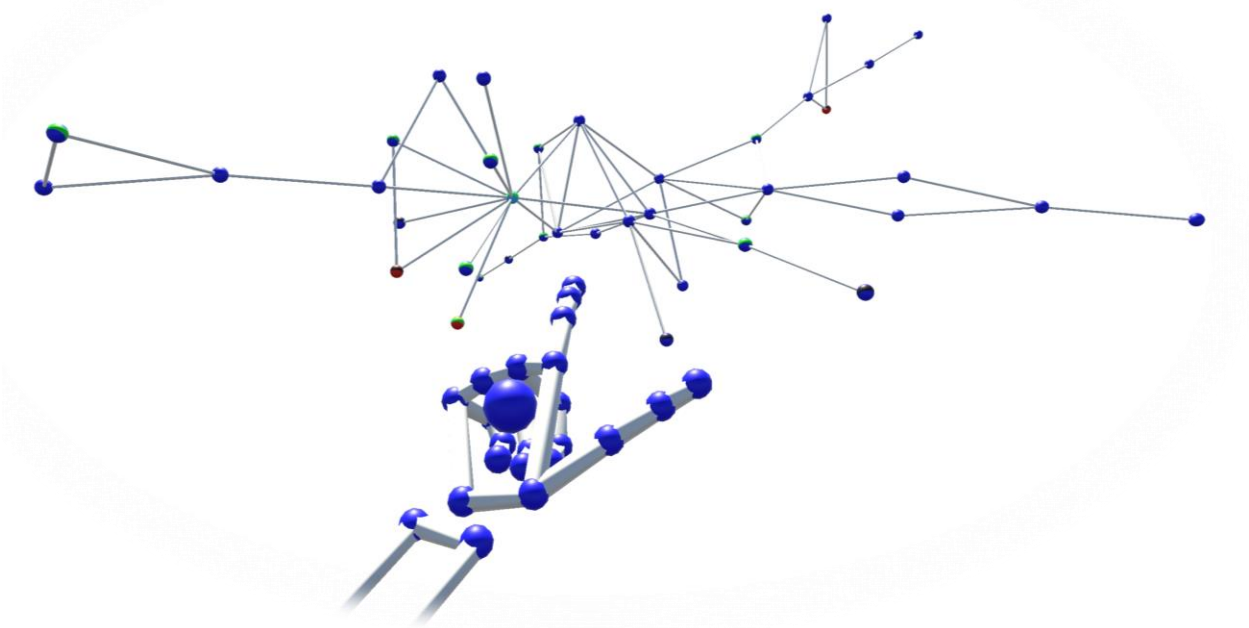
Jipdekok@hotmail.com

Internal Supervisor:

Christopher Pawley

External Supervisor:

Pieter Geelen



Abstract

This thesis aimed at the improvement of large multi-omics network visualisations by utilisation of virtual reality. Due to the increasing size of biological datasets, it is getting more challenging to properly visualise such datasets in 2D. Therefore, different approaches should be considered, like virtual reality. In this thesis, an interactive virtual reality network of multi-omics data was developed and compared to a pre-existing 2D network from the same dataset. Finally, the virtual reality visualisation was compared to the 2D visualisation through conducting a cross-over design study where participants had to test the visualisations and, subsequently, evaluate them with a Likert scale survey. Essentially, no significant differences were found between the two types of visualisations from the survey data. However, the feedback was generally positive indicating that virtual reality visualisation has the potential to serve as an appropriate way of visualising multi-omics data but should still be further developed and tested.

Keywords:

Network visualisation, Stereoscopic 3D (S3D), virtual reality (VR), multi-omics, Valproic Acid (VPA), Force-Directed placement, Fruchterman-Reingold algorithm, Likert scale, crossover design.

Contents

1 Introduction

In recent years, technology has been rapidly increasing. This has allowed for the collection of large amounts of data containing vital pieces of information for understanding a wide variety of different concepts. However, all this data will have to be stored and properly analysed. In the field of molecular biology, we see that information about molecular interactions in the human body is collected from previously conducted research and stored in online databases. An example of such a database is ConsensusPathDB (Kamburov, Stelzl, Lehrach, & Herwig, 2013), which combines interaction data from various other databases. Currently, it stores more than 603,000 unique interactions and is still increasing. However, without the proper tools to analyse it, this abundance of data is useless. One popular way of making sense of complex data, is visualising the data because it can simplify it in such a way that researchers can more easily understand it.

There are various aspects that influence the quality of the visualisation. The clarity is important because it should clearly illustrate the data in an easy to understand fashion. Also, aesthetics can improve the quality of a visualisation where aspects of the style of the visualisation like colours and contrast can play a big role in the visibility of certain aspects of the visualisation. At last, it should be accurate, such that no false conclusions are made from it. Currently, most visualisations are made in 2D, which is generally an appropriate way of visualising data. However, with the increasing amount of data it is getting harder to properly visualise these datasets, which raises the question how big data could be better visualised. In this thesis, VR is explored to see whether it can potentially improve the visualisation of biological networks.

This paper discusses the differences between 2D and 3D/VR visualisations, focussing on what aspects of VR could make it preferable over 2D, and when it might not have added value. Additionally, it is described how a VR visualisation was constructed for a multi-omics network and how the effectiveness of this visualisation, compared to a 2D visualisation, was tested. Based on the findings from these tests, it is finally discussed how the visualisations performed, if there is a future for multi-omics network visualisation in VR and how it could be further improved and implemented.

2 Literature Review

In this part of the paper, the literature upon which this thesis is based will be discussed. In the first section, the data that was used for the development of the visualisation is elaborated upon. Additionally, the algorithm that was used for constructing the visualisation is discussed, regarding its functionality, strengths and weaknesses. Also, another visualisation toolset is introduced, focussing on the functionality that makes it stand out and how this should be reflected in this thesis' visualisation. Finally, it is explained why a 3-dimensional virtual environment is used in this thesis for visualising biological networks. Subsequently, in section 2.2, it is described what methodologies were used for product testing and why these were ideal for this specific case. At last, in section 2.3, the statistical analysis of the data from the product testing is explained in more depth.

2.1 Visualisation

The first step of developing a good visualisation is getting acquainted with the data. There are countless ways of visualising data yet there is no such thing as 'the best way to visualise data' as this is very specific for the type of data that you are working with. In this thesis, data from a genomic- and epigenomic study on the repetitive exposure of cultured human liver cells to the drug valproic acid (VPA) was used to design a virtual reality visualisation (van Breda et al., 2018). The data consists of expression- and methylation values of 45 genes, making this a multi-omics research. These 45 genes are a subset of the total number of genes that had a significantly different expression under VPA treatment and were visualised in a network graph (see Figure 1), which essentially is a collection of nodes that are specifically interconnected with other nodes, indicating links between various entities. In this case, the links between the entities, which are the differentially expressed genes, represent gene- or protein interactions. The use of a network graph to visualise expression data is ideal as it can illustrate how these genes are interconnected and can also indicate the change in expression- and methylation values for these individual genes, which in this case was done by colour coding the nodes. The changes in gene expression can be useful to predict the effect it will have on a larger scale as it can be checked what biological pathways these genes are involved in, which could possibly explain certain effects of the treatment. A relatively novel aspect of this specific study is the analysis of gene methylation, which is often described as a gene silencing mechanism. However, this not always the case (P. A. Jones, 2012). In the dataset from

(van Breda et al., 2018) for example it is apparent that an increase of gene methylation is not always correlated with a decrease in gene expression. Visualising multi-omics datasets like these could give new insights in the functional properties of specific gene methylation and how these relate to expression patterns in the cell. Due to this data's biological complexity and relatively small size, it is ideal for the design of a new type of visualisation. Its biological complexity allows for comprehensive quality testing of the visualisation and the small size keeps the visualisation well observable during the development.

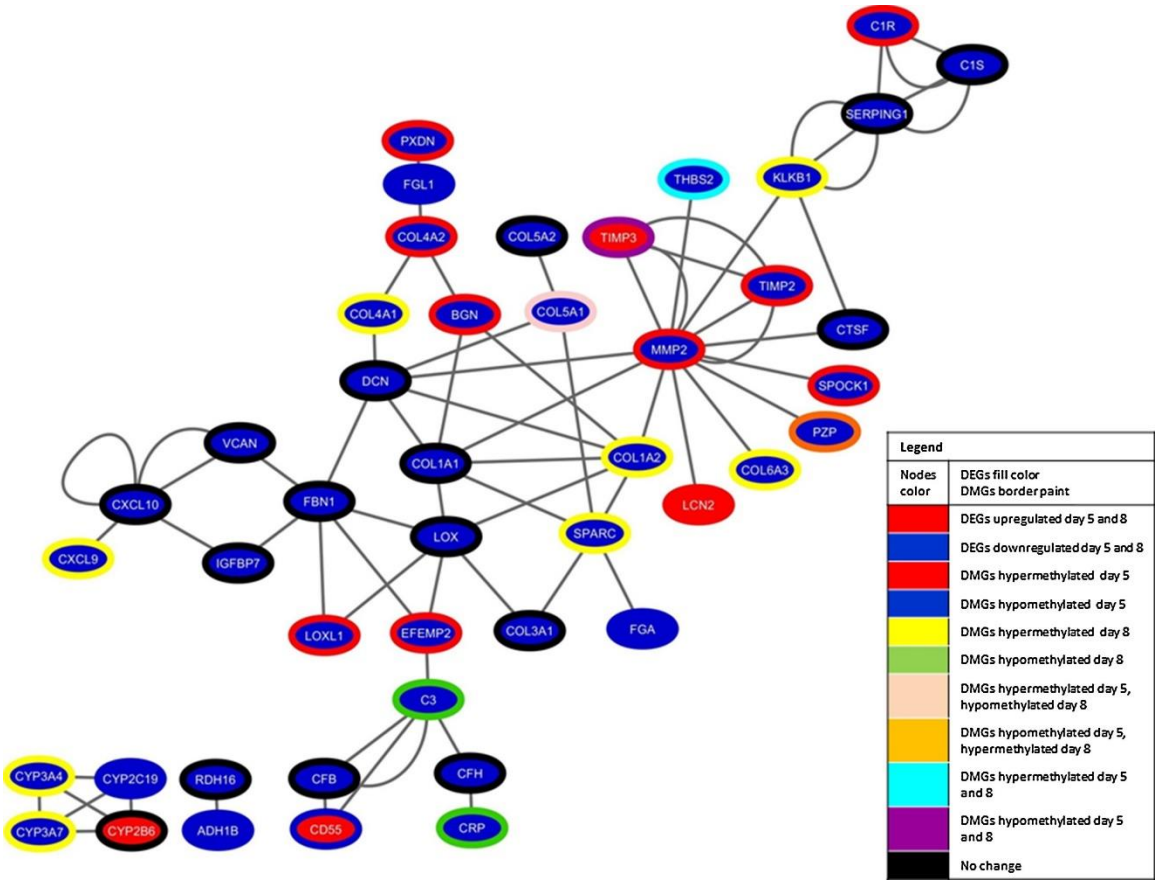


Figure 1: A network graph of 45 differentially expressed genes under VPA treatment with their interactions indicated by the edges. The colours are bound to two time points, day 5 which is after 5 days of treatment and day 8 which is after 5 days of treatment and a 3 day-long washout period. The fill colour indicates the change in expression. Red corresponds to an increase in expression and blue indicates a decrease in expression. The border colour defines the change in methylation, where black means no change in methylation. Red: increase in methylation after day 5; blue: decrease in methylation after day 5; Yellow; hypermethylation after day 8; Green: hypomethylation after day 8; Pink: hypermethylation after day 5 and hypomethylation after day 8; Orange: hypomethylation after day 5 and hypermethylation after day 8; Turquoise: hypermethylation after day 5 & 8; Purple: hypomethylation after day 5& 8.

The VR visualisation that was developed in this thesis is based on Force-Directed Placement, which is a method for constructing a network graph, for an example see Figure 2. Force-Directed Placement is based on spring-like forces between nodes, meaning that every individual node in the graph will repel all other nodes as if there was a contracted spring in between every pair of nodes that pushes the nodes away from each other. However, since it is a network graph, all nodes will be connected to at least one other node. These connected nodes will attract each other, as if there was an extended spring in between the two nodes that pulls them towards one another. This principle dates back to 1963 (Tutte, 1963) and was later refined by (Fruchterman & Reingold, 1991) who designed a force-directed placement algorithm for visualising undirected graphs with straight edges. They calculated the attractive force between every connected pair of nodes using the following formula:

$$F_{\mathrm{a}} = \frac{d^2}{k}$$

In this equation, the attractive force F_{a} is calculated by dividing the squared distance between the two connected nodes d by the ideal distance k. The repulsive force on the other hand, is calculated for every possible combination of nodes and is formulated as follows:

$$F_r = \frac{-k^2}{d}$$

In this formula, the repulsive force F_r is calculated by dividing the negative squared ideal distance k by the distance between the two nodes d. For both repellent and attractive forces, the ideal distance k is calculated using the following equation:

$$k = C\sqrt{\frac{area}{number\ of\ vertices}}$$

C is a constant that is experimentally determined. The ideal distance k is calculated like this to uniformly distribute the nodes based on the unit of area per number of nodes.

This algorithm was designed to be fast and simple. Moreover, it can evenly distribute vertices, results in uniform edge lengths and reflects symmetry (Fruchterman & Reingold, 1991), see Figure 2. However, the Fruchterman-Reingold algorithm is not flawless. One issue is the time complexity, specifically for the repulsive forces as it has an exponential time complexity of $\Theta(|V^2|)$, where V is the number of vertices, meaning that the required time to calculate the positions of the nodes exponentially increases for every

additional node. However, the attractive force calculations only have a time complexity of Θ(|E|), where E is the number of edges, which means that the time complexity increases linearly instead of exponentially. Various countermeasures are available for this issue. One option would be to minimise the number of nodes that exercise a force on each other. This could be achieved by only generating a repellent force between nodes that are within is specified distance from each other. Another approach would be to improve the hardware capabilities that perform the calculations for the force-directed algorithm. A novel variant of the Fruchterman-Reingold algorithm has been developed which is optimised for utilising parallel GPU architectures. Essentially, it makes use of multiple graphical processing units simultaneously, which are computer components with great processing power to speed up the algorithm, ultimately, giving it the capability of processing bigger datasets without losing any quality of the visualisation (Gajdoš, Ježowicz, Uher, & Dohnálek, 2016). These options should be considered when working with big data. For this research however, these countermeasures were unnecessary as big data was not yet fully explored.

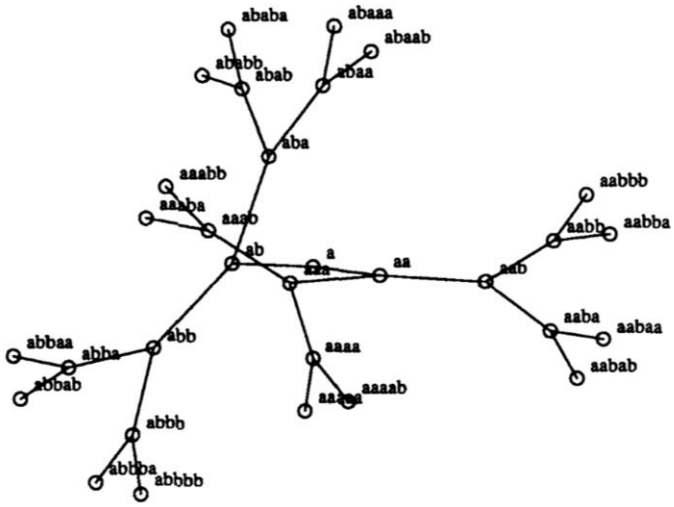


Figure 2: an example of a Force-Directed Graph, built with the Fruchterman Reingold algorithm(Fruchterman & Reingold, 1991).

There is a variety of software available for visualising biological data, one of which is Cytoscape(Shannon et al., 2003). Cytoscape is an open source toolset that can visualise many types of biological data and, interestingly, can also query the network to acquire additional information from large databases like ConsensusPathDB (Kamburov et al., 2013) using a plugin. This allows the user to incorporate pre-existing data, e.g. protein interactions, into their dataset. This enables the user to add interactions to their biological data and turn it into a network that can illustrate how the biological entities are connected. This thesis will incorporate the same interactions to illustrate the inter-relationship between genes.

Data visualisation in 3D has its strengths and weaknesses. The most straightforward positive aspect of 3D, when compared to 2D, is the increase in space availability, allowing for more possible shapes that the visualisation can take. Furthermore, various depth cues can effectively improve the understanding of structures. Occlusion is a very important depth cue, as from two partly overlapping objects, it is very easily determined which object is closer by. On the other hand, occlusion can also hide important information. However, with the ability to move in 3D space you get the phenomenon called motion parallax, where movement allows for a better understanding of a 3D structure as the occlusion will change and, thereby, can also reveal previously hidden information. From this, it is possible to get an understanding of the relative distance between objects and the structure they form in a 3D environment. Opposed from 3D rendering on a 2D display, it can also be projected on stereoscopic displays. These are two individual displays, one for each eye, as used in VR. The slight difference between the image projected on the left- and the right eye, also allows for improved depth perception (Munzner, 2014). These are some of the most important depth cues that could theoretically improve someone's understanding of familiar, but also abstract structures. However, it is no longer just theoretical as it has been shown that stereoscopic 3D (S3D) visualisations can increase analytical performance over 2D visualisations. Network path-tracing tests have been conducted, comparing 2D, S3D and self-motion parallax visualisations (van Beurden, Ijsselsteijn, & de Kort, 2011)(Sollenberger & Milgram, 1993). Self-motion parallax means that the movement some individual makes, is reflected on the image. Meaning, that it is simulated as if you were physically present in the virtual environment. From these path-tracing studies, a better performance was observed for both S3D and motion parallax visualisations, yet, specifically the combination of the two, resulted in a better performance. This is especially true for more complex networks. It has even been shown that S3D with incorporated self-motion parallax, allowed for an understanding of abstract informative graphs with a size of up to three times larger, compared to a 2D visualisation (Ware & Franck, 1996). Suggesting, that virtual reality through means of a head mounted display is the most effective way of visualising 3D structures. Additionally, the importance of detail has been studied. This was done by using high resolution S3D displays (3840x2400 pixels per display) and changing edges from 2D lines into 3D tubes (see Figure 3). From this, it was concluded that the level of detail does indeed make a difference, indicating the importance of the hardware that is used, but also the amount of detail in the visualisation itself to achieve the best performance (Ware & Mitchell, 2005). Furthermore, they found an even greater increase in possible dataset size that can be properly analysed than (Ware & Franck, 1996) as they found similar error rates for 'reading' 1000 nodes with self-motion

parallax S3D, and reading 33 nodes with 2D. In this context, 'reading' means identifying short paths between nodes.

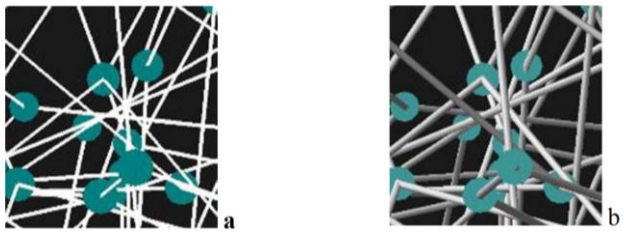


Figure 3: This figure illustrates the difference between 2D and 3D edges. a) Here the nodes in 3D space are connected by 2D lines, which makes it challenging to perceive the depth of the edges. b) Here the 2D edges are replaced with 3D cylinders, making the edges better observable.

However, S3D is not always preferable over 2D, some studies found no increase in performance when using an S3D visualisation. This could be explained by the type of data used in their visualisation, as S3D only seems beneficial when an understanding of the structure helps in analysing the data. Since, it is important when analysing biological networks to understand the structure of the network, where, being able to trace edges to different nodes can give an understanding of the underlying meaning, it is expected that S3D is an ideal platform for biological network visualisations. Furthermore, S3D also goes paired with some issues regarding discomfort for the user. Cyber sickness can occur when people use virtual reality. Some possible symptoms are headaches, disorientation, nausea, eye strain and sweating. Various theories on the cause of cyber sickness are described in (LaViola, 2000). There are many factors that can trigger cyber sickness, one of which is latency, thus, latency should be minimised in order to reduce cyber sickness (Hale & Stanney, 2002). Additionally, spatial orientation can induce cyber sickness as well. This is caused by observing self-motion without physically experiencing movement, which is referred to as motion sickness. A possible solution for this is the use of teleportation or 'jumping', instead of normal movement as we experience in real life. This means that you either instantly appear at the place where you wanted to go, or you successively appear on certain intervals between you and your destination (Lackner, 2014). Ideally, the user would not move at all within the visualisation, assuming the user is seated and thus not moving in the real world. Minimising the presence of cyber sickness is essential as it does not just make the product less enjoyable, it can also cause people to prematurely stop using the VR tool or stop using it at all. In order to achieve this in the VR visualisation, the viewpoint should be fixated to the position of the user's head and the network itself should be movable instead. Thereby, proper hardware should be used to minimise latency and other factors that could influence cyber sickness.

Furthermore, when someone uses the tool behind their desk, which is a preferable workplace for many people, it is favourable if the required tools are as minimalistic as possible (Zielasko et al., 2017). Therefore, it was chosen for this VR visualisation not to use any kind of physical controller but the Leap Motion. The Leap Motion is an infrared sensor that can be attached to a head mounted display and can track hands in real time, enabling the user to interact with the visualisation using their bare hands.

2.2 Product Evaluation Setup

To check whether the VR visualisation is well functioning, and preferable over 2D visualisations for analysing biological networks, the product had to be tested. In this section, some testing methodologies will be discussed that were used for evaluating the VR visualisation.

To capture the test subject's opinion about the visualisations, a Likert scale was used, which is a psychometric scale that is often used in surveys to quantify an individual's opinion about a specific matter. It consists of statements, which are referred to as Likert items, and the test subjects have to answer by stating their level of agreement with that statement. It is often used as a 5- or 7-point scale ranging from "Strongly agree" to "Strongly disagree". It is typically an uneven numbered scale such that the middle of the scale can be "Neither agree nor disagree" (see Figure 4). However, sometimes it can be preferable to take an even scale without the undecided option, forcing participants to state their opinion even if they are undecided. Thereby, people might choose the middle option because they think their opinion is socially unacceptable. However, if it is not present this might cause them to take the option they think is socially preferable opposed to stating their true opinion (Garland, 1991). The Likert scale's simplicity and reliability make it an ideal method for conducting a survey. However, there are some negative aspects that should be kept in mind. For example, people tend not to choose extreme answers like "totally agree" or "totally disagree". The middle answer tends to be unpopular as well, since it usually does not properly depict an opinion. Typically, this results in two peaks, one for answer 2 & 3, and one for answer 5 & 6, on a 7-point Likert scale. Furthermore, social aspects can influence the honesty of the participant's answer. Acquiescence bias could be one reason, which is the tendency of people to agree with someone else's opinion/statement. One reason could be that the participant wants to favour the person conducting the experiment. Alternatively, they could try to portray themselves in a way they think is socially preferable (Salkind, 2010). Unfortunately, these limitations are difficult to tackle as you can never be sure on what ground the participant's answers are based. Properly introducing the survey,

clarifying the importance of honesty and ensuring their anonymity, is of great importance to minimise the impact of social bias on the reliability of the data. Every question can be analysed individually, or related items can be summed up together to create a score for a selection of Likert items. Individual questions can be analysed using the Mann-Whitney U test. Alternatively, if multiple questions are summed up, they can be analysed using ANOVA testing (Bertram, 2007), these tests are elaborated upon in section 2.3.

Figure 4: An example of a Likert scale.

When you want to compare two tests that could have an influence on each other. Meaning, that having done one test affects your performance in the other test. It can be challenging to properly compare the results of these two tests. One option would be to split the sample in two and let both groups only do one test. However, this is not very efficient as you will only use half of your sample size. An alternative to this is a crossover design (Sibbald & Roberts, 1998). A crossover design is an experimental setup where participants get a specific sequence of experiments. The participants are split up in groups that will each have a different experimental sequence. The groups can be constructed using randomized stratification. Essentially, this means that the total population is split up and put in so called 'strata', which are subgroups based on a certain category like age, sex or profession. From these strata, random individuals are placed in either sample group such that each group will end up with the same number of subjects from each stratum, maintaining the same proportions of a specified category in the samples as it was in the total population. A crossover design can be preferable when resource availability or sample size is limited, because it allows for usage of the entire sample without taking the risk that one treatment could influence the results of the other treatment(s) for all subjects. This made it a preferable experimental setup for this thesis. However, a disadvantage is the carryover effect. This entails, that one experiment might influence the participant's performance in a subsequent experiment. In the case of testing two types of visualisations based on the same dataset, this would mean that familiarity with the dataset can affect how well the participant can analyse the dataset in the other format, and how this influences his opinion on this visualisation method. Assuming, the carryover effect is present, experiments held later in the sequence, cannot be properly statistically compared (Salkind, 2010). An often-used solution to this, is

introducing a washout period, where a time interval is set in between trials such that the carryover effect has disappeared by the time they conduct the subsequent test. Unfortunately, this was an unviable option for this research due to time restrictions and possible external learning factors during the washout period. However, the carryover effect can also be used for knowledge gain, as it can give information on whether one method serves as a better training model for the other. In this thesis, that would mean that it could be analysed whether it is beneficial to combine the two types of visualisations, opposed to picking only one of the two, and if so, in what order they should be used to get the best performance. Conducting a crossover design study with two different tests and two groups is called a 2x2 design (B. Jones & Kenward, 2003), which is what should be used for analysing the two types of visualisations as was done in this thesis.

2.3 Data Analysis

For the analysis of the individual Likert items, the Mann-Whitney U test was used, which is a non-parametric test, meaning it is not exclusively based on probabilistic parameters like the mean, variance and so forth. Moreover, it requires independent sample groups and ordinal observations yet does not depend on a specific distribution. Having independent sample groups means that the results from one sample have no influence on the results from another group. Ordinal scales have variables that are ordered yet are unevenly distributed. Likert scales are ordinal, as for example the difference between "somewhat agree" and "agree" and the difference between "agree" and "strongly agree" is not identical. The fact that it does not require a normal distribution was ideal for this research due to its small sample size. The Mann-Whitney U test can be used to test the null-hypothesis, essentially, checking whether two samples have a significantly different outcome. The null hypothesis will be rejected if the Mann-Whitney U test shows that the median of one sample is significantly different from the other. Since, the Likert scale is an ordinal non-parametric test, the Mann-Whitney U test is an appropriate way of analysing the individual Likert items from the survey conducted in this thesis (Salkind, 2010) (Bertram, 2007).

Analysis of variance (ANOVA) is a statistical analysis method that can compare the averages of two or more groups. This allows you to check whether there is a significant difference between these groups. With the Likert data from the conducted tests, ANOVA can compare summed Likert items to check whether there is a difference between these bundles of questions. For this, a fixed-effects one-way ANOVA approach was used as this method compares two or more treatments to test whether the

population means are statistically different, which is precisely what is required to compare two visualisations (Salkind, 2010). It is often suggested that parametric tests, like the ANOVA test, are unsuitable for ordinal scales, like the Likert scale. Especially, when sample size is small or if the data is not normally distributed. Although, sample size and distribution can affect statistical significance, it does not render ANOVA inappropriate for the data (Norman, 2010).

3 Methodology

3.1 Visualisation Development

For the development of the biological network visualisation in virtual reality using a force-directed placement algorithm, the Unity software was used. Unity (Creighton, 2010) is a game engine which is software that is typically used for both 2D & 3D game-design. Moreover, it supports Rigidbodies, which are components that can be attached to objects, giving them physical properties like mass, drag, collision detection and more. This feature can be used to apply the force, calculated with the force-directed placement algorithm (explained in section 2.1), to the individual nodes to construct the 3D network. Additionally, it has native VR support, making it an ideal application to develop the VR visualization with. All scripts for Unity were written in the C# programming language.

The dataset used for the development of the VR visualisation, contains gene expression- and methylation data. Essentially, it consisted of two data files, one with some general information about the individual genes present in the network like their symbol, change in expression-/methylation values and a brief description of the gene. The other data file was an edge list, meaning it contains all known connections between the genes from the first dataset, which will form the edges between the nodes in the network (van Breda et al., 2018).

First, to initialise the nodes of the network, which represent genes in this case, the dataset had to be loaded into unity. This was done by the first two scripts which parse the two data files into two-dimensional string arrays which can be processed by Unity. From the first data file, the third script instantiates a set of Gameobjects, each representing a specific gene. Subsequently Rigidbodies components are attached to these Gameobjects to enable their physical properties (for Rigidbody properties see appendix Figure 11). Additionally, the colour of the nodes is based on their according expression- and methylation values. The second data file is used to create a 'Connectivity Matrix' that stores the indices of all the connected nodes from the array they were put in. Using this Connectivity Matrix, edges, which are cylinders, are created in between the nodes such that they connect all the nodes that have an interaction with each other. Now, the scene is initialised and the FixedUpdate method runs the methods that calculate the attractive and repulsive forces, also, it updates the position of the nodes and the edges between the nodes based on these forces every frame.

The attractive forces were calculated only for the connected nodes. An ideal edge length can be set by the user, such that the attractive force will become negative (repellent) when the nodes are too close together. The formula that was used to calculate the attractive force is:

$$F_{\mathrm{a}} = log_{10}\frac{d}{k} * P$$

In this formula, attractive force F_{a} is calculated by taking log_{10} of the ratio between the distance d, which is the distance between the two nodes, and ideal distance k scaled by the pull speed P. Here, k & P can be changed by the user. This formula diverts from the Fruchterman-Reingold algorithm as they calculate it as the squared distance over the ideal distance. This alternative formula was used as it gave a better result.

The repellent forces were calculated for every possible combination of nodes. The algorithm used for calculating the repellent force is:

$$F_r = -\frac{k^2}{d} * R$$

Here the repulsive force F_r is calculated by multiplying the negative ratio of the squared ideal distance k and the distance d by the repulsive speed R, just like the Fruchterman Reingold algorithm (Fruchterman & Reingold, 1991).

Since, a relatively small network was used during the development of the VR visualisation, there were no performance issues. However, an option was implemented with which a maximum repulsion distance can be set. Essentially, when this feature is enabled, only the repulsive forces between nodes whose distance to each other is smaller than the maximum repulsion distance will be calculated, potentially, increasing the performance when working with a big network.

For the implementation of interaction in the visualisation, the leap motion, which is an infrared hand tracking device, was used. The Leap Motion Orion SDK, core version 4.3.4 was used together with the Leap Motion Interaction Engine (1.2.0), which allows to interact with 3D objects and user interfaces. For the user interface (UI) itself the Hovercast VR Interface was utilised. This asset allows the use of an arc-shaped interface on the top of your fingertips (see Figure 5). This UI has been setup in such a way that it incorporates all functionality of the visualisation. The top UI item is the appearance menu, here, the

colour of the nodes, the ideal edge length, attractive- and repulsive forces can be changed. Additionally, there is an option that makes the force-directed graph spherical, in the sense that all nodes have an additional force towards the centre of the graph, creating a sphere-like structured network around the VR headset. The second UI item is the transform menu which has a submenu that allows the user to scale, move and rotate the network. The third UI item is a toggle which can pause/start the real-time simulation of the network. This can be convenient when the frame rate is too low as pausing the simulation increases the performance of the system. The last UI item is the selection mode toggle, which can be used to pinpoint nodes with a visible raycast from the right index fingertip. A raycast is a line which reaches from a specific starting point A to another point B, if it collides with anything in between these points it will return what object was hit. Here, point A would be the fingertip and point B a specific arbitrary distance away from point A in the pointing direction of the finger. If a node is hit, its material is illuminated, indicating which node is selected. Moreover, a floating graphical user interface (GUI) will show the name of the selected gene. This GUI can be moved when the 'Move UI' toggle from the appearance settings menu is enabled, by pinching your right thumb and index finger together and dragging it to the desired location. Furthermore, an interaction behaviour component was added to the nodes on initialisation such that the users can move and grab them with their hands if they are in reach. The VR headset that was used for the development and testing of the visualisation is the HTC vive pro, which has a resolution of 1440 x 1600 per display.

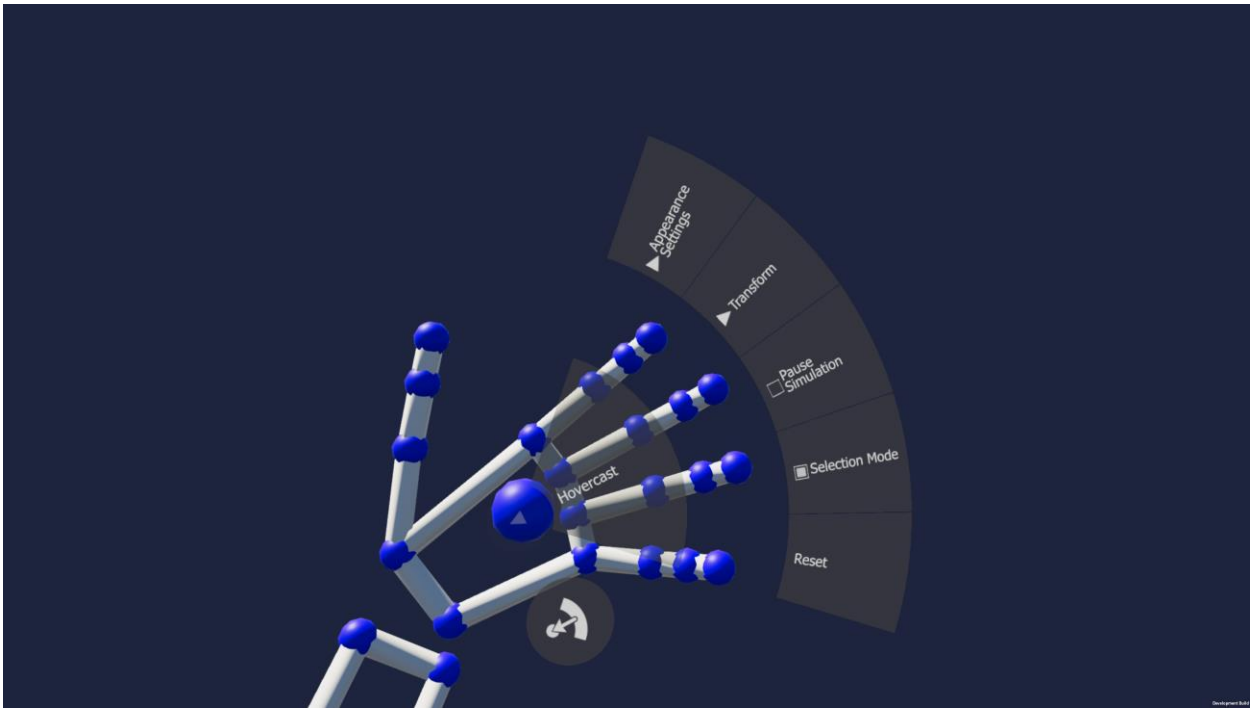


Figure 5: A picture of the left Leap Motion hand with the Hovercast menu. Here the root menu is illustrated which has the two submenus 'Appearance settings' and 'Transform'. Additionally, it has the 'pause simulation' and 'selection mode' toggle and a reset button. These menu items can be selected with the right Leap Motion hand which is not present in this image.

3.2 visualisation Evaluation

To assess the quality and usability of the developed VR visualisation, a comparative study was conducted. Tasks were designed which could be completed using either the 2D- (see appendix Figure 12) or the VR visualisation. The total testing population of 11 people was split into two groups of equal size. The groups were constructed using randomized stratification based on age and occupation such that both groups had similar experience in analysing visualisations. A crossover design was used, meaning that both groups started with a different visualisation as indicated by table 1. Prior to conducting the tests, the subjects were introduced into the topic. A brief presentation was given, explaining the multi-omics data type that was visualised and how network graphs work. Subsequently, a 2D visualisation was shown that had an identical style as the 2D visualisation that was later used in the test, the network itself however was random, this was done to ensure their understanding of the visualisation. Additionally, a random network was demonstrated in VR with the exact same visual style and interactivity as the original VR visualisation. During this demonstration the entire interactivity with the network was explained and illustrated. Finally,

when the subjects were clearly informed of the test and had no more questions, they were individually assigned to a visualisation according to the group they were in. For the VR visualisation the screen was recorded such that their interactivity with the tool could be later analysed. Also, they were lead through the process by dictating what to do and asking them questions which they had to answer based on the visualisation. For the tasks that they were given when evaluating the 2D and the VR visualisation see appendix Figure 13.

Table 1: The two test groups A & B with their according testing sequence.

	1st test	2nd test
Group A	2D visualisation	VR visualisation
Group B	VR visualisation	2D visualisation

After each test, the subjects had to fill in a 7-point Likert scale survey containing statements about the quality and usability of the visualisation. For the VR visualisation there were some additional statements regarding its interactivity and presence of cyber sickness. For the survey see appendix Table 6. Subsequently, the surveys from the 2D and VR tests were compared using the Mann-Whitney U test. This was done for both parts of the sequence, meaning that the VR test results from the participants that first tested the VR visualisation were compared the 2D results from the participants that tested the 2D visualisation first. The visualisations that were second in the sequence were similarly compared. Also, the tests were compared regardless of the sequence they were in. For these three scenarios, the U value was calculated for each Likert item and compared to the critical value of U for $p < 0.05$ based on the sample sizes. Additionally, ANOVA testing was used to compare the summed Likert items. Here, all the Likert items were summed up together for each of the participants and compared using a one-way ANOVA test. This was done for the same three scenarios as was done with the Mann-Whitney U test.

4 Results

The multi-omics VR visualisation was successfully developed. The 3D network took shape as would be expected when using a force-directed placement approach as described by (Fruchterman & Reingold, 1991). Also, all the functionalities of the Leap Motion interaction menu worked properly, allowing the users to fully test the VR visualisation. A snapshot of the 3D network in the VR environment is shown in Figure 6.

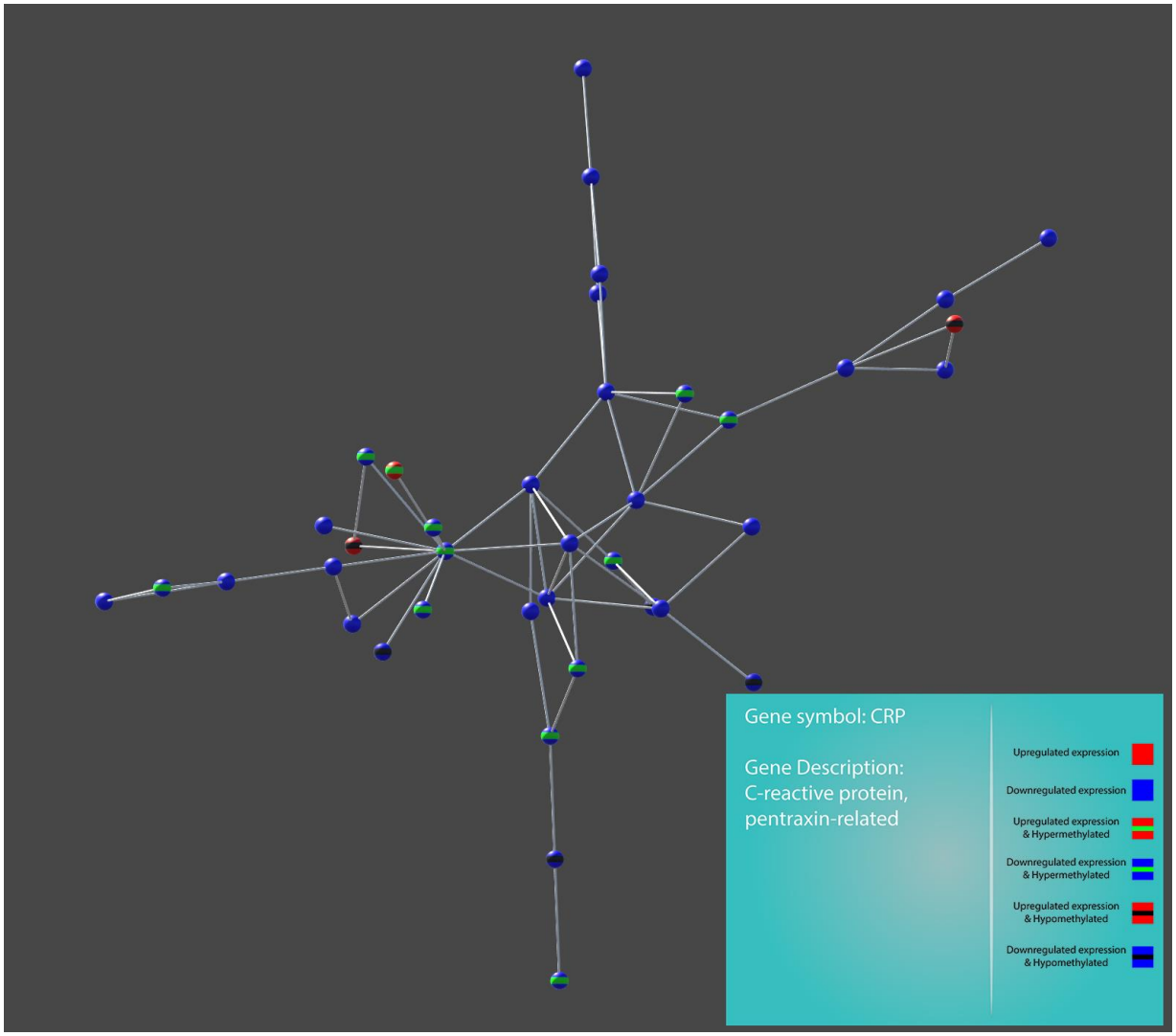


Figure 6: A picture of the 3D force-directed network that is used in the VR visualization. The nodes represent genes and the cylinders that connect them are interactions. The blue canvas on the right bottom is the GUI that users can also see in the virtual environment. On the right of the GUI, all the different colour combinations are explained. On the left of the GUI, the selected gene will be described by stating the symbol and a brief description, here an example is shown for the gene CRP.

The results from the Likert scale survey, show that the participants generally appreciated the VR visualisation. They stated that it was an appropriate way of visualising the data in a clear and easy to understand manner. Also, the tasks were supposedly easy to answer. However, the opinion seemed to be a little bit more divided as to whether they would use this type of visualisation for analysing similar data, see Figure 7. For all the individual responses on the survey, see Table 7 & Table 8 of the appendix.

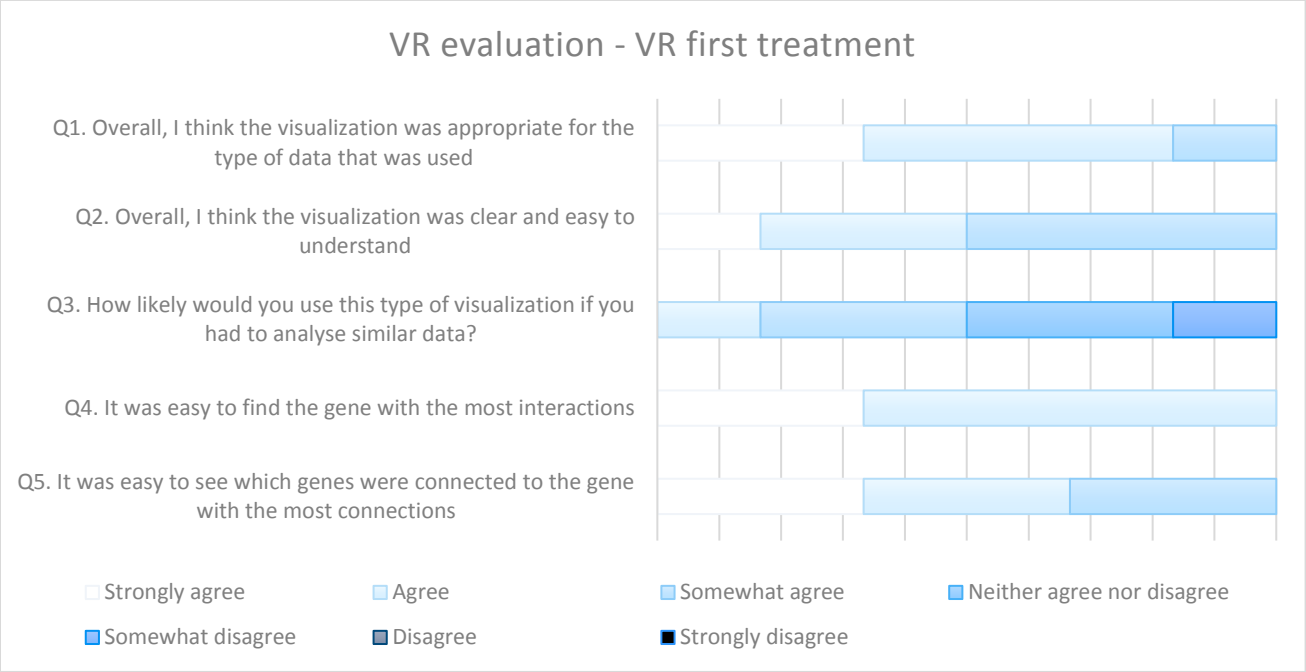


Figure 7: The Likert scale survey responses on the VR visualisation. Only data from participants that evaluated the VR visualisation before the 2D visualisation was taken. Their level of agreement with the statements is indicated by the blue colour scale as indicated in the legend.

The 2D visualisation also got positive feedback. The visualisation properly illustrated the data and they stated it to be likely for them to use a similar type of visualisation if they had to visualise this specific type data. They executed the tasks well, although, some said they found it a bit more challenging to see which genes were connected to the gene with the most interactions. These findings are illustrated in Figure 8.

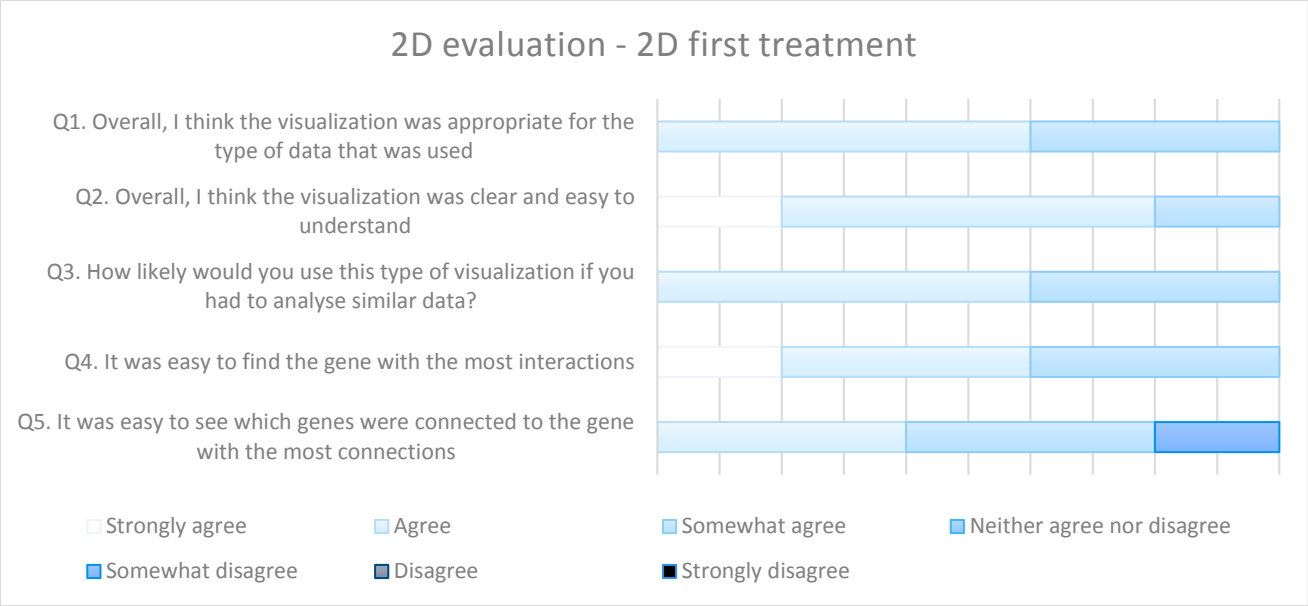


Figure 8: The Likert scale survey responses on the 2D visualisation. Only data from participants that evaluated the 2D visualisation before the VR visualisation was taken. Their level of agreement with the statements is indicated by the blue colour scale as indicated in the legend.

Using the Mann-Whitney U test, the survey data from the VR- and the 2D visualisations was compared. Essentially, no significant differences were found in the data. For the comparison of the VR- and 2D visualisations as the first tests in the sequence (Table 2), and the comparison of the VR- and 2D visualisations as the second tests in the sequence (Table 3), the critical U value was 3 for $p < 0.05$. This means that the calculated value for U must be smaller than 3 for the median of the two populations to be significantly different. This was not the case for any of the Likert items, the smallest value for U that was found was 5.5 for question 3 when the first sequenced visualisations were compared and for question 2 when the second visualisations in the sequence were compared.

Table 2: The results from the Mann-Whitney U test on the comparison of the VR evaluation responses from the subjects that first tested the VR visualisation and the responses from the subjects on the 2D visualisation which tested the 2D visualisation first. The columns indicate the different questions. The first row 'U value' shows the calculated value for U, the second and third row hold the median values of the VR and 2D survey responses respectively. The individual cell on the right bottom of the table indicates the critical value for U.

VR vs 2D 1st	Q1	Q2	Q3	Q4	Q5	
U value	14	11	5.5	9	8	
median VR	2.5	2.5	3.5	2	2	
median 2D	2	2	2	2	3	U < 3

Table 3: The results from the Mann-Whitney U test on the comparison of the VR evaluation responses from the subjects that first tested the 2D visualisation and the responses from the subjects on the 2D visualisation which tested the VR visualisation first. The columns indicate the different questions. The first row 'U value' shows the calculated value for U, the second and third row hold the median values of the VR and 2D survey responses respectively. The individual cell on the right bottom of the table indicates the critical value for U.

VR vs 2D 2nd	Q1	Q2	Q3	Q4	Q5	
U value	14	5.5	8	14.5	13	
median VR	2	2	2	2	3	
median 2D	2.5	4	4	2	2.5	U < 3

Additionally, the Likert items were compared regardless of their test sequence (Table 4). Here, the critical value for U was 30 for $p < 0.05$ and the smallest value for U was 31 for question 1. Thus, also here there was no significant difference between the two visualisations.

Table 4: The results from the Mann-Whitney U test on the comparison of the VR evaluation responses and the 2D response regardless of the sequence in which the tests were conducted. The columns indicate the different questions. The first row 'U value' shows the calculated value for U, the second and third row hold the median values of the VR and 2D survey responses respectively. The individual cell on the right bottom of the table indicates the critical value for U.

VR vs 2D both	Q1	Q2	Q3	Q4	Q5	
U value	31	46	60	40	40	
median VR	2	2	3	2	2	
median 2D	2	2	3	2	3	U < 30

Furthermore, the summed Likert items were compared with one-way ANOVA testing, see Table 5. Just like the Mann Whitney U test this resulted in no significant differences between the two visualisations. The lowest value for p was obtained when comparing the entire sequences resulting in $p \approx 0.066$ which is much lower opposed to analysing the individual sequences which resulted in $p \approx 0.6$.

Table 5: Results from the ANOVA test, comparing the 2D and VR visualisation summed Likert items. The columns represent the different testing sequences, the first column compares the results from the VR- and 2D tests from the groups that took those according tests first. The second column represents the data from the groups that took the corresponding tests second and the third column represents the comparison of the two testing sequences combined. The means of both visualisation tests are indicated by row 2 & 3. The last row shows the p-value that was obtained by comparing the specific visualisation test results using one-way ANOVA.

	First in sequence	Second in sequence	Entire sequences
VR test mean	11.3333333	15.83333	11.63636
2D test mean	12	20.2	14.09091
p-value	0.635481	0.667232	0.066162

Some specific Likert items were made for the VR visualisation, regarding its usability and comfort. In Figure 9 it is illustrated that most of the participants found it easy to interact with the visualisation and that only a few experienced some discomfort during the utilisation of the VR visualisation. For the raw data from the survey see Table 9 in the appendix.

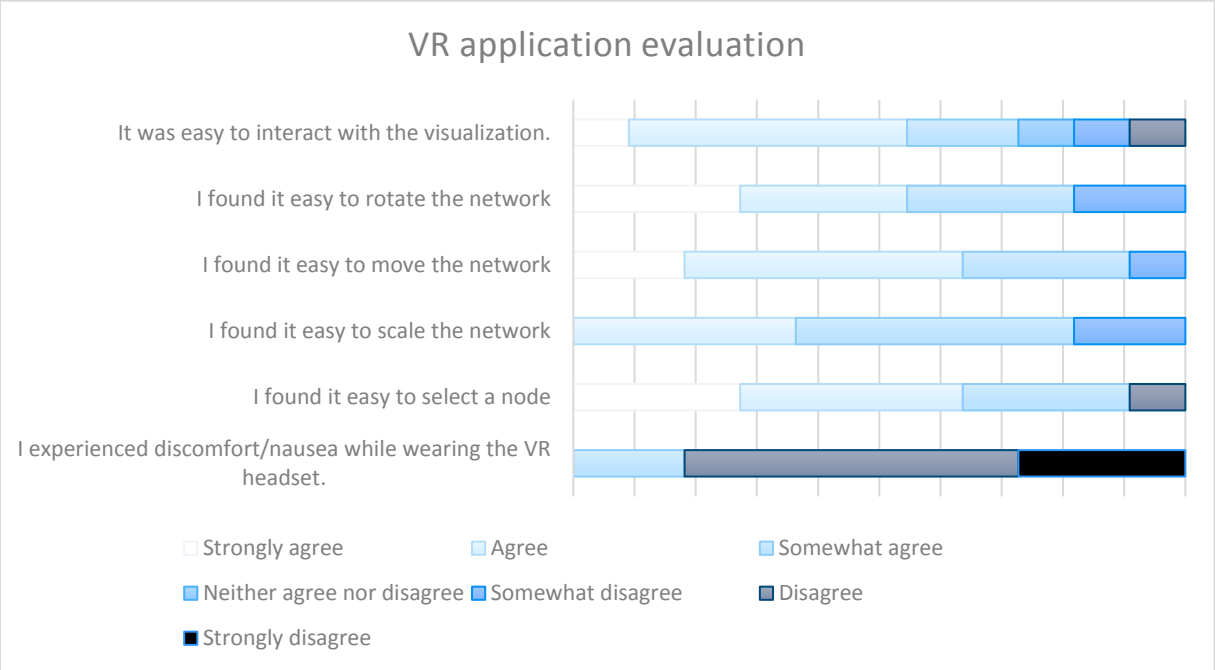


Figure 9: Likert scale survey responses from the VR visualisation evaluation regarding the interactivity and comfort of the virtual reality application. Their level of agreement with the statements is indicated by the blue colour scale as indicated in the legend.

Also, one question was included in the survey where the participants were asked what they think should be improved about the visualisation. The responses to this question for the VR visualisation were quite diverse, some pointed out that the interactivity should be optimised, and others made remarks on the lack of quality of the VR headset. Furthermore, some comments were made on the possibility of adding information to the user interface, turning it into a more comprehensive analysis toolset. Moreover, the absence of the time points at day 7 was said to clarify the visualisation but also to decrease the amount of information one can deduct from the visualisation. For the 2D visualisation, the main concern turned out to be the complexity of the colour coding and the visibility of some of these colours. The comments on the visualisations can be found in Table 10 & Table 11 of the appendix.

5 Discussion

The goals of this thesis was to research whether multi-omics networks could be better visualised in VR than in 2D. This is interesting to investigate since increasing amounts of data make it challenging, if not impossible, to properly visualise such data in 2D. However, when visualising big data, more aspects should be considered than just virtual reality. Since, developing a virtual reality force-directed graph with which a user can interact using their bare hands is already an exhaustive task, it was decided not yet to optimise it for big data. This tool aimed at illustrating the potential of multi-omics VR visualisations and its flaws. Not only would designing it for big data have resulted in a worse quality product, it would also have been more difficult to properly test it. First, since big data is typically not visualised as a network like the one in this thesis, the VR visualisation would have to be compared to either a different type of visualisation or a similar network visualisation which is not applied for this type of data in real life. Second, a novel way of interacting with the visualisation was designed for the VR visualisation, which is already challenging to get acquainted with using a small network. Testing the VR visualisation with a big network would have been more time consuming and challenging for the participants, potentially, decreasing the sample size even further. All in all, the small network was ideal for both the development and testing of the VR visualisation. No quantitative measurements like time and error rates were measured because these measurements, do not depict the true quality of the visualisation. If any of such measurements got better results in one visualisation, this could be caused by various aspects besides the quality of the visualisation itself. For example, the type of interaction with the VR visualisation, which is not required in the 2D visualisation, could influence the performance of the participants. Also, you would need a large sample size to get useful data from such tests. Since, it was anticipated that a big sample size might not be obtained, rendering quantitative measurements unsuitable as presumably no significant results would have been found, the quantitative measurements would have been useless. Instead, a Likert scale survey was used, as even without a big sample size or significant results, surveys can still serve as a proper evaluation of a product. Additionally, it gives a better indication of the true quality of the visualisations.

Overall, the VR multi-omics visualisation ended up as an appropriate tool to visualise multi-omics network data in an interactive virtual environment. The tests indicated that it was usable by experienced professionals, but also by amateurs. As to whether it was better than its predecessor, the 2D network visualisation, this cannot be proven nor rejected. Both the mean and median of the survey responses for the VR visualisation were generally lower opposed to those of the 2D visualisation (see Table 2, Table

3,Table 4 & Table 5), indicating a more positive feedback for the VR visualisation. However, they were not significantly different which is presumably caused by the small sample size. If a bigger sample size was used, more significant results could have been obtained, although, it is not safe to say that this would have meant that the VR visualisation would have gotten more positive feedback than the 2D visualisation. Survey data can be tricky to analyse as it can be greatly influenced by the individuals. However, it does give an indication of the quality of the product and can reveal aspects that should be improved.

In the survey, one question was included where the participants were asked what they think should be improved about the visualisation. Some pointed out that the VR headset was not very comfortable to wear and the image was slightly blurry, which should be improved over time as VR is still at an early phase of development. The blurriness is presumably caused by the resolution of HTC vive pro, which is already relatively high (1440 x 1600 per display) but not as high as the displays used in the research of (Ware & Mitchell, 2005) where they show that high resolution displays (3840x2400 pixels per display) can improve line tracing performance in a network. Thus, an even higher resolution than the one of the HTC vive pro might be preferable for VR data visualisation. Also, it was suggested that additional information could be included in the user interface, ranging from analytical tools giving quantitative feedback on the visualisation like how many genes are upregulated or how strongly genes are upregulated. However, also qualitative information could be included like descriptions of the connections such that it is clear how two genes interact with each other. This is something that was also pursued in this thesis but was not yet achieved. Finally, some remarks were made on the presentation of the change in methylation, where, time point 7 was eliminated from the VR visualisation. This was done to clarify the visualisation as the abundance of different colours makes it harder to distinguish specific genes. However, to make up for the information loss this caused, a slider was meant to be incorporated with which the time point could be changed, allowing the user to change between time point 5 and 7 within the VR visualisation. This feature should have been added to the visualisation before the tests to maintain similarity between the two types of visualisation yet was not. The feedback on the 2D visualisation was mainly focussed on the colour coding since there were many different combinations of colours possible, making the visualisation more challenging to understand with respect to the different types of changes in methylation. Moreover, some colour combinations were said to be hard to see. This indicates that time point 7 should indeed be removed from the VR visualisation, although, the data should still be accessible by including a feature to observe different time points.

The interactivity of the VR visualisation worked properly as all the test subjects were able to complete the tasks and utilise all types of interaction. Figure 9 illustrates that most of the participants could easily interact with the VR visualisation, specifically, moving the network was not an issue. Rotating and scaling the visualisation seemed to be a bit more challenging. Although there are some improvements to be made on the interactivity, it served well for this prototype.

6 Conclusion

In this thesis, an interactive virtual reality visualisation for multi-omics network data was successfully developed. From the tests, some valuable feedback was obtained. The visualisation clearly illustrated the data and the interactivity works properly but should be refined such that it is easier to use. Also, various features should be included which could not only improve the visualisation but could also add additional information to the network itself, this could be statistical information on the network or additional content like specific interactions. However, no statistical evidence was found as to whether the VR visualisation performed better or worse at specific elements of data analysis than the traditional 2D network. Therefore, future research should be done with an improved version of the visualisation which should be tested with a bigger sample size. All in all, Virtual Reality visualisations of multi-omics data is a promising way of novel data analysis that could be used to explore big biological datasets when it is further developed.

7 Future prospects

In this thesis a VR visualisation was developed which serves as a proof of concept. This novel type of data visualisation still needs to be perfected for it to be properly usable by the scientific community. The VR visualisation was designed for one specific multi-omics dataset which was relatively small. Ultimately, it should be compatible with a wide variety of data that can be visualised as a network graph and should be able to process bigger datasets as well. Due to the high time complexity of the force-directed algorithm which constructs the VR network, it can be challenging to use it for big data visualisations, however, optimisations could be made for this. To tackle this issue, the time complexity could be reduced, either by limiting the number of repulsive forces or making better use of hardware by for example taking a parallel computing approach as described in (Gajdoš et al., 2016). However, this not the biggest issue with big data as the network could always be pre-calculated and then used as a static network opposed from a real-time simulated network as this thesis' VR visualisation. The real issue is the observability of such a network due to its high interconnectivity. Big datasets can get very cluttered making it challenging to properly analyse such a network. Therefore, additional visualisation strategies should be considered such as filtering and clustering. There are many ways to filter or cluster a network and depending on the data and what is being looked for in the data, different methodologies might be preferable. One methodology that is useful in general for highly interconnected networks is edge bundling. This is an approach that curves edges in a way that they cluster together making it more easily observable (see Figure 10) (Holten & Van Wijk, 2009).

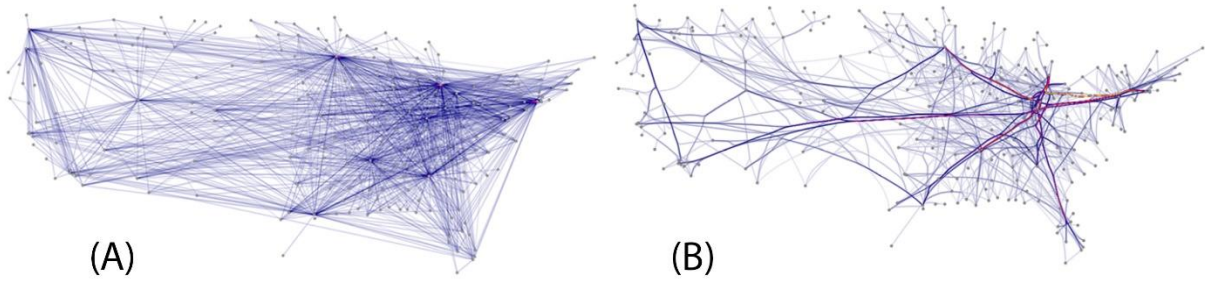


Figure 10: Example of a network-graph on which edge bundling is applied. (A) shows the original network without edge bundling (B) is the same network modified by force-directed edge bundling with an inverse-quadratic model. (Holten & Van Wijk, 2009)

On the other hand, adding new features to the toolset will also make it more complicated to use. Already, the VR interactivity seems challenging for people to get accustomed to, meaning it will require time and effort to get used to this novel type of interaction. Therefore, it might be useful to allow users to change settings in a regular 3D environment with mouse and keyboard, with which they are more experienced, so they can observe the visualisation in VR afterwards. This way, the toolset will remain relatively easy to use although it can hold many different features. Moreover, the way of interacting with the visualisation using the leap motion should have more gesture-based interactions such that you don't have to open a menu for every action.

From the biological point of view, there are also interesting features that could be included. One improvement that was pursued during this thesis was acquiring more specific interactions between genes. However, this was unfortunately not achieved since ConsensusPathDB was used which is ideal for getting many interactions but does not offer specific information on the interactions. Cross checking these interactions from other databases might allow for more specific information on some interactions which could be used to exclusively show certain interactions. This might give a better insight in why specific genes are changed in a certain way when exposed to a drug.

Finally, the hardware that was used was appropriate but far from perfect for usage on a regular basis. The VR headset was quite heavy and big, making it very inconvenient for storage in a small room and for transport. Additionally, the two sensors were also too large and some interference with mirrors was experienced reducing the accuracy of the headset tracing. Thereby, the resolution was sufficient but could be improved even further. Additionally, the Leap motion worked sufficiently but was slightly buggy at times, where hands would suddenly flip upside down or vibrate uncontrollably. These issues never persisted very long but could decrease productivity if the tool is used in the long term.

8 Critical reflection

Overall, the thesis went smoothly, apart from the rough start due to my late admission. This problem should have been tackled by me pushing harder to get the thesis sorted earlier, which would have made the start of the thesis easier for both me and my supervisors. At the start, my focus was to develop the visualisation as it was my highest priority to advance my programming skillset and to make sure I ended up with a good product. Although I am still glad I pursued and managed this, I should have been a bit more open to the suggestions of working more on other parts of the thesis as well, like literature review and planning the testing methodology. I did take such advice but still resisted it slightly which made the process unnecessarily hard and less enjoyable. Moreover, a bit more structured working order could have been useful, where I typically started working on one thing until I either completed it or could no longer properly work on it due to a loss of concentration and then think of something else to do on the spot. This works for me relatively well, yet, if I set more specific goals for myself per day, this could have motivated me to do even more and the structure might have helped me stay focussed when I haven't reached the specific goal I made yet.

9 Bibliography

Bertram, D. (2007). Likert Scales. *University of Calagary, Department of Computer Science*, pages.cpsc.ucalgary.ca/~saul/wiki/uploads/CPSC681/. https://doi.org/10.1002/9780470479216.corpsy0508

Creighton, R. H. (2010). *Unity 3D Game Development by Example*. *PACKT Publishing*. Retrieved from https://www.packtpub.com/unity-3d-game-development-by-example-beginners-guide/book

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, *21*(11), 1129–1164. https://doi.org/10.1002/spe.4380211102

Gajdoš, P., Ježowicz, T., Uher, V., & Dohnálek, P. (2016). A parallel Fruchterman-Reingold algorithm optimized for fast visualization of large graphs and swarms of data. *Swarm and Evolutionary Computation*, *26*, 56–63. https://doi.org/10.1016/j.swevo.2015.07.006

Garland, R. (1991). The mid-point on a rating scale: Is it desirable? *Marketing Bulletin*, *2*, 66–70. https://doi.org/citeulike-article-id:4775464

Hale, K. S., & Stanney, K. M. (2002). *Handbook of Virtual Environments: Design, Implementation, and Applications*. CRC press. Retrieved from https://www.crcpress.com/Handbook-of-Virtual-Environments-Design-Implementation-and-Applications/Hale-Stanney/p/book/9781138074637

Holten, D., & Van Wijk, J. J. (2009). Force-Directed edge bundling for graph visualization. *Computer Graphics Forum*, *28*(3), 983–990. https://doi.org/10.1111/j.1467-8659.2009.01450.x

Jones, B., & Kenward, M. G. (2003). Design and analysis of cross-over trials, 2nd edition.

Jones, P. A. (2012). Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, *13*(7), 484–492. https://doi.org/10.1038/nrg3230

Kamburov, A., Stelzl, U., Lehrach, H., & Herwig, R. (2013). The ConsensusPathDB interaction database: 2013 Update. *Nucleic Acids Research*, *41*(D1), 793–800. https://doi.org/10.1093/nar/gks1055

Lackner, J. R. (2014). Motion sickness: More than nausea and vomiting. *Experimental Brain Research*, *232*(8), 2493–2510. https://doi.org/10.1007/s00221-014-4008-8

LaViola, J. J. (2000). A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin*, *32*(1), 47–

56. https://doi.org/10.1145/333329.333344

Munzner, T. (2014). *Visualization Analysis and Design*. Retrieved from https://books.google.de/books?id=dznSBQAAQBAJ

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, *15*(5), 625–632. https://doi.org/10.1007/s10459-010-9222-y

Salkind, N. J. (2010). Encyclopedia of Research Design. *Thousand Oaks, Calif: SAGE Publications*. https://doi.org/10.4135/9781412961288

Shannon, P., Markiel, A., Owen Ozier, 2, Baliga, N. S., Wang, J. T., Ramage, D., … Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, (13), 2498–2504. https://doi.org/10.1101/gr.1239303.metabolite

Sibbald, B., & Roberts, C. (1998). Understanding controlled trials Crossover trials. *Bmj*, *316*(7146), 1719–1720. https://doi.org/10.1136/bmj.316.7146.1719

Sollenberger, R. L., & Milgram, P. (1993). Effects of stereoscopic and rotational displays in a three-dimensional path-tracing task. *Human Factors*, *35*(3), 483–499. https://doi.org/10.1177/001872089303500306

Tutte, W. T. (1963). How to Draw a Graph. *Proceedings of the London Mathematical Society*, *s3-13*(1), 743–767. https://doi.org/10.1112/plms/s3-13.1.743

van Beurden, M. H. P. H., Ijsselsteijn, W. a., & de Kort, Y. a. W. (2011). Evaluating stereoscopic displays: Both efficiency measures and perceived workload sensitive to manipulations in binocular disparity. *Proceedings of SPIE Electronic Imaging*, *7863*, 786316–786316. https://doi.org/10.1117/12.872566

van Breda, S. G. J., Claessen, S. M. H., van Herwijnen, M., Theunissen, D. H. J., Jennen, D. G. J., de Kok, T. M. C. M., & Kleinjans, J. C. S. (2018). Integrative omics data analyses of repeated dose toxicity of valproic acid in vitro reveal new mechanisms of steatosis induction. *Toxicology*, *393*, 160–170. https://doi.org/10.1016/j.tox.2017.11.013

Ware, C., & Franck, G. (1996). Evaluating stereo and motion cues for visualizing information nets in three dimensions. *ACM Transactions on Graphics*, *15*(2), 121–140. https://doi.org/10.1145/234972.234975

Ware, C., & Mitchell, P. (2005). Reevaluating stereo and motion cues for visualizing graphs in three

dimensions. *APGV '05 Proceedings of the 2nd Symposium on Applied Perception in Graphics and Visualization*, 51–58. https://doi.org/10.1145/1080402.1080411

Zielasko, D., Weyers, B., Bellgardt, M., Pick, S., Meibner, A., Vierjahn, T., & Kuhlen, T. W. (2017). Remain seated: Towards fully-immersive desktop VR. *2017 IEEE 3rd Workshop on Everyday Virtual Reality, WEVR 2017*. https://doi.org/10.1109/WEVR.2017.7957707

10 Appendix

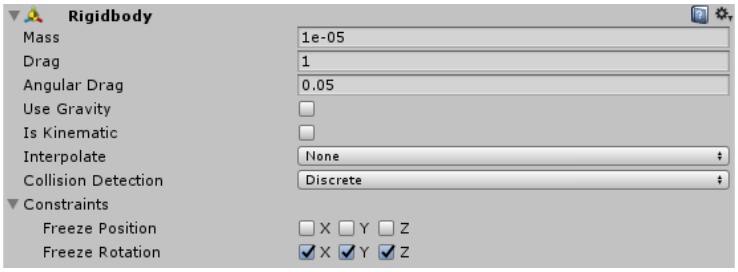


Figure 11: Rigidbody properties of the nodes in the network.

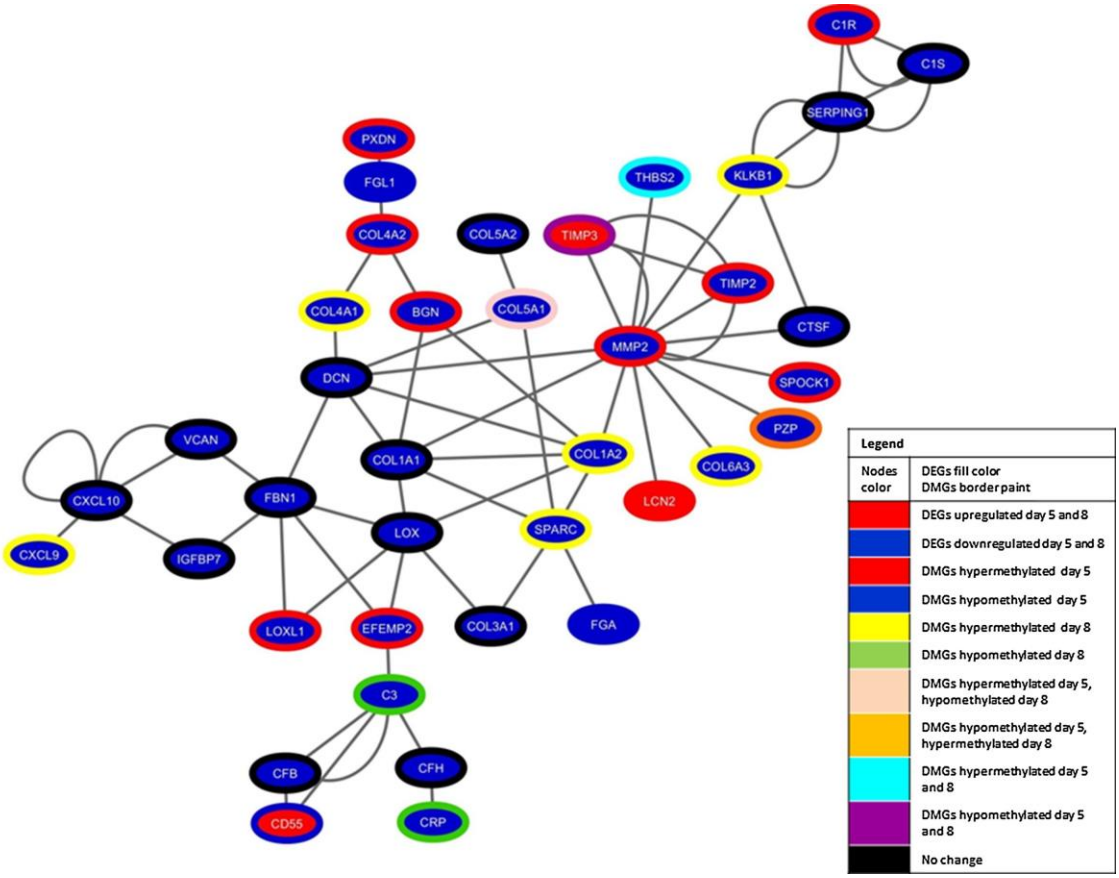


Figure 12: The 2D network graph that was used during the testing phase to compare the quality of the VR visualisation compared to this 2D visualisation.

0) Scale the visualization to make it the size you think is best.
1) Move it to the position where it is best observable
2) Rotate it to get a better impression of the position of the nodes.
3) Place the UI (the blue user interface on your left) where you think it should be.
4) Enter the selection mode, to select genes.

5) Find the gene with the most interactions with other genes.
    a. What is the name of this gene?

    b. How many interactions does it have with other genes?

6) Find the genes with an increase in expression value.
    a. How many genes are upregulated?

    b. Do any of the upregulated genes have a direct interaction with each other?

7) Find the genes that are downregulated and also hyper methylated.
    a. How many genes are downregulated and also hyper methylated

Figure 13: A picture of the tasks that had to be answered by the participants during the test. For the 2D test only questions 5 - 7 were used. For the VR all tasks were used.

Table 6: A table listing the Likert items for the survey. Each row is a different statement. For the VR survey all Likert items were used. For the 2D survey only Q1-Q3 & Q10-Q11 were used.

Q1	Overall, I think the visualization was appropriate for the type of data that was used
Q2	Overall, I think the visualization was clear and easy to understand
Q3	How likely would you use this type of visualization if you had to analyse
Q4	It was easy to interact with the visualisation.
Q5	I found it easy to rotate the network.
Q6	I found it easy to move the network.
Q7	I found it easy to scale the network.

Q8	I found it easy to select a node.
Q9	I experienced discomfort/nausea while wearing the VR headset.
Q10	It was easy to find the gene with the most interactions.
Q11	It was easy to see which genes were connected to the gene with the most connections.

Table 7: The Likert scale survey responses of the VR visualization evaluation.

The first visualisation type the participant tested	Overall, I think the visualization was appropriate for the type of data that was used	Overall, I think the visualization was clear and easy to understand	How likely would you use this type of visualization if you had to analyse similar data?	It was easy to find the gene with the most interactions	It was easy to see which genes were connected to the gene with the most connections	Participant number
VR	3	3	5	2	2	11
2D	1	2	2	1	2	36
VR	2	3	4	2	3	18
2D	1	2	2	2	2	74
VR	2	3	3	2	2	61
2D	2	2	5	3	2	22
VR	1	2	4	2	1	53
2D	2	2	2	2	3	69
VR	2	2	3	1	3	68
2D	2	2	2	1	2	84
VR	1	1	2	1	1	39

Table 8: The Likert scale survey responses from the 2D visualisation evaluation.

The first visualisation type the participant tested	Overall, I think the visualization was appropriate for the type of data that was used	Overall, I think the visualization was clear and easy to understand	How likely would you use this type of visualization if you had to analyse similar data?	It was easy to find the gene with the most interactions	It was easy to see which genes were connected to the gene with the most connections	Participant number
VR	6	7	6	3	2	11
2D	3	1	2	1	2	36
VR	3	5	5	2	2	18
2D	3	3	3	2	5	74
VR	2	2	3	3	3	61
2D	2	2	3	2	2	22
VR	2	6	5	2	6	53
2D	2	2	2	3	3	69
VR	3	3	2	2	3	68
2D	2	2	2	3	3	84
VR	1	2	2	1	1	39

Table 9: Likert scale survey responses from the VR visualisation evaluation regarding the interactivity and comfort of the virtual reality application.

It was easy to interact with the visualization	I found it easy to rotate the network	I found it easy to move the network	I found it easy to scale the network	I found it easy to select a node	I experienced discomfort/nausea while wearing the VR headset.	Participant number
5	5	3	5	3	6	11
1	1	2	3	2	7	36
2	3	2	2	1	7	74
4	3	3	3	2	6	18

3	2	2	3	6	3	22
2	2	1	2	3	6	61
2	1	3	2	3	6	53
2	3	2	3	1	6	69
2	2	2	3	2	6	68
3	1	1	2	2	7	84
6	5	5	5	1	3	39

Table 10: The answers to the question "How do you think the visualisation should be improved?" for the VR visualisation.

How do you think the visualisation should be improved?	Participant number
The good thing was that the time dependency of the methylation data had been simplified, so you didn't need to look for multiple colours to find the hyper methylated genes. On the other hand this means that you have lost information. The interface would take time to get used to. The question is whether this overhead is worth it for this type of visualisation, which will depend on how often you need to analyse data in this way. This also depends on whether you are wanting to make a visualisation for data exploration purposes or for showing particular features of your data to other people. If you regularly need to explore such data then the overhead of learning the interface may be worthwhile, but for a casual reader of a journal article or who just wants to look at the image and extract the main points as mentioned in the text, this is a different scenario.	11
Perhaps by setting it at the most convenient position for most people. For experienced users this is a very nice way to express data	18
a bit sharper view, especially for text	74
Rotation around the axis could be improved, VR-machine is warm and heavy, maybe some calculation possibilities. Rotation ball was a bit hard to use sometimes, but a nice option.	61
Display gene information on user interface Summary statistics of the network having number of upregulated and downregulated genes can be displayed on the user interface	22

It was hard to differentiate between the fully blue nodes and the black and blue nodes.	53
It could be optimised to be more fluid when the network has to be moved or rotated (reducing the rigidity of movement).	69
maybe less sensitive to hand movements. Function to zoom in into specific parts of the network. In addition to balls, also the ability to use pyramids, or cubes.	68
it needs a bit of practice to find out how to move, scale and zoom. that makes the visualization easier to use. if the network is less extended it is easier to get the overview to select hub genes. The panel sometimes comes in the way of the network. it would be good if you could get quicker info on the genes, and if there would be a second layer of more biological details. This is what contributes to understanding more complex responses.	84
your arms are getting tired after some time. But the visualization is perfect. Compliments!	39

Table 11: The answers to the question "How do you think the visualisation should be improved?" for the 2D visualisation.

How do you think the visualisation should be improved?	Participant number
The representation of time was the most problematic factor. The number of different colours with different meanings was confusing. As there were not too many timepoints you could either make a separate visualisation for each timepoint, or split the shape into parts which are coloured in a more consistent way according to their change at each timepoint (e.g. stripes for each timepoint). 3d and animation will also help with clarity giving extra dimensions to show things in.	11
Don't use a pink indicator when using a white background	36
explanation of colours could be better, maybe explain border colour and filling colour separately.	18
clearer connections, the round arrows are quite easy to miss.	74
Too many colours, no option to count edges, or to interact with the visualisation.	61
Could use colors apart from blue and red for border colors, I found it confusing with blue-red combination for both node and border colors	22
Use higher contrast for the colors.	53

it should be easier to identify genes with specific characteristics (direction of gene expression and methylation change), plus the connections between the genes.	69
The color code is rather complex; you really have to puzzle to answer relatively simple questions. It would be easier if the hyper methylated genes would have variants of the same colour (shades of green for example) and the hypo shades of purple.	68