UiT THE ARCTIC UNIVERSITY OF NORWAY

Faculty of Science and Technology
Department of Computer Science

Virtual Reality for the visualization of high-dimensional relationships in bioinformatics

Subtitle

—

Álvaro Martínez Fernández

INF-3990 Master's thesis in Computer Science November 2020

Contents

List of Figures

List of Tables

1

Introduction

1.1 Background

Technological advancement has revolutionized the field of genomics, which has led to a cost-effective generation of big amounts of sequence data. The sequencing of the first human genome (2002) took around 13 years and cost over $3 million to complete. Nowadays we can resequence a human genome for $1000 and can generate more than 320 genomes per week[7]. This technological innovation leads to the accumulation of vast quantities of genomic data, posing a tremendous challenge to scientists for effective mining of data to explain a phenomenon of interest. New ways of analysing the produced data have been therefore necessary in order to discover interesting patterns and make the most out of it. No matter how much resources we use into extracting the data if we don't get anything interesting out of it[14].

Some of the main problems that researchers face when analysing genomic data are information overload, data interconnectivity and high dimensionality. Visualization is one way of facing this problems. For this reason it is very important to implement efficient visualization technologies that can lead to find new patterns and the extraction of good conclusions of the data. In the field of system biology there are usually network representations where the nodes or bioentities are connected to each other, where these edges represent associations. Because of the improvements in technology, these networks can increase dramatically in size and complexity. We need therefore better visualization systems and more efficient alogirithms for the analysis of the data.

In 1.1 we can see a representation of the evolution for visualization of networks in system biology. From simple graphs in 2 dimensions, to 3D representations and nowadays also visualizations in virtual reality where we can interact directly with the data itself.

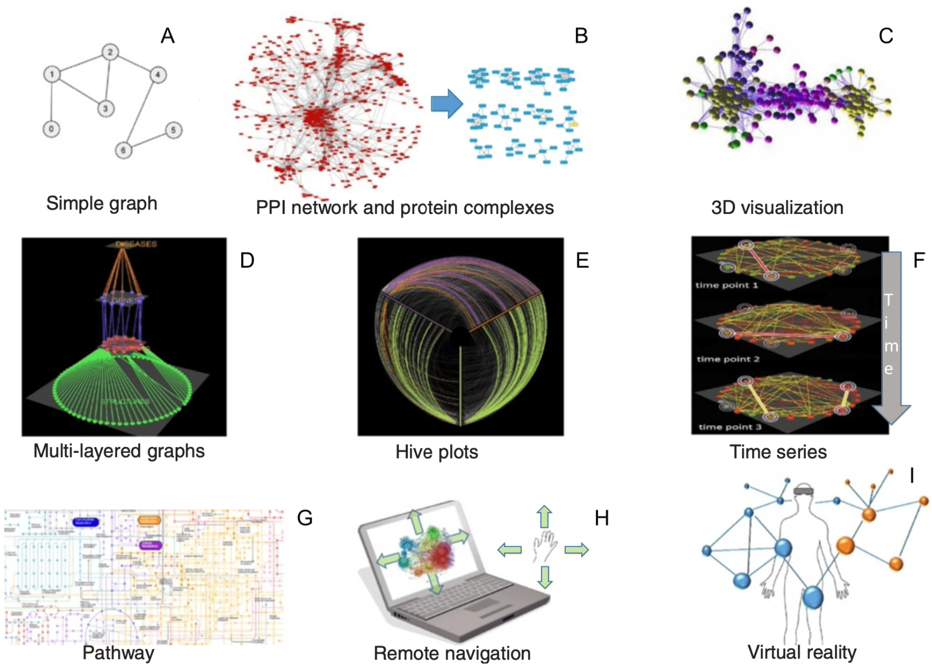


Figure 1.1: Visualization for network biology. a A simple drawing of an undirected unweighted graph. b A 2D representation of a yeast protein-protein interaction network visualized in Cytoscape (left) and potential protein complexes 3D identified by the MCL algorithm from that network (right). c A 3D view of a protein-protein interaction network visualized by BiolayoutExpress. d A multilayered network integrating different types of data visualized by Arena3D. e A hive plot view of a network in which nodes are mapped to and positioned on radially distributed linear axes. f Visualization of network changes over time. g Part of lung cancer pathway visualized by iPath. i Remote navigation and control of networks by hand gestures. h Integration and control of 3D networks using VR devices. Figure adapted[10].

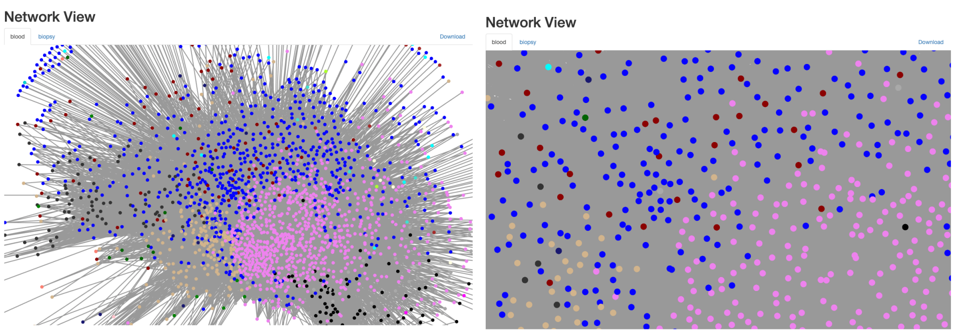
Virtual reality (VR) is still a field under exploration and that can be of great help in network analysis. VR can be very powerful because it takes advantage of the way the human being perceives and analysis things. We as human beings have a great ability to discover patterns, however we are biologically optimized to see the world and the patterns in 3 dimensions. VR is one of the best ways then for better discovery in spatial dimensions. It has been demonstrated that VR help scientists work more effectively in fields like medicine [6][12][3], biology[13][11] and neuroscience[1][8], to cite some examples.

1.2 Challenges and research problem

This project focus mainly on solving the problem of visualization of high dimenasional data from the MIxT project by using virtual reality. Furthermore the application allows the user to interact with the network created from the data in the virtual environment. It also allows the user compare the blood and biopsy networks at the same time in order to finde relationship, which wasn't possible in the MIxT web application as this only allows the user to visualize one network at a time.

MIxT[5] is a web application for bioinformaticians. Among other tools, it offers a network visualization of genes which are represented as nodes in the network and where the edges represent statistically significant correlation in expression between two nodes. This tool was used in a study[4] that identifies genes and pathways in the primary tumor that are tightly linked to genes and pathways in the systemic response of a patient with breast cancer. When exploring a network in MIxT, it can be hard to understand the data and its relationships because there is too much data. This problem is easy to occur when there are too many node and edges. In figure 1.2 we can see an example of the network visualization from MIxT. As we can see in Figure 1.2a, there are many nodes and relationships among them and when we zoom in in the network, it becomes very difficult to understand the data and the relationships as shown in in Figure 1.2b.

The network is also in 2-dimensions and what we propose in this project is to use a virtual reality 3d visualization in order to cope better with this problem.



(a) Network with several modules. **(b)** Zoom in the network.

Figure 1.2: Network view of the MIxT application where nodes repsent genes and the modules are repsented by colors. Relationships are represented by grey lines that connect a gene with another one.

1.3 Proposed solution

1.4 Significance and contribution

This project contributes in the exploration of the possibilities that Virtual Reality offers for visualization of big data in bioinformatics.

1.5 Outline

2

BigNet VR

BigNet VR is a virtual reality application for the interactive visualization of large-scale networks in a 3D space. The networks are represented using nodes and connections between them. In order to explore the data, the user can walk around, scale the network, move it around, filter the nodes using a 2D interface and also interact with the nodes to get more information about them. In Figure 2.1 we can see an example of the application running.

BigNet VR works with a dataset that contains the information of the nodes and relationships of the network. This dataset needs to be obtained from an external source or application. Once we have our dataset we can load it to the application and then run BigNet VR using a Head Mounted Display or HMD for VR. Finally we can explore the network and interact with it to visualize the dataset in a VR experience.

The implementation of BigNet VR is done in Unity, a cross-platform game engine. This engine is used for a wide range of applications, especially for the development of videogames in 3D and 2D, VR applications and engineering solutions. The programming language used to develop the application inside Unity is C#. We also used VRTK, a VR toolkit to build VR solutions in Unity. As for the VR hardware, we used an Oculus Quest headset. This type of headset is an oll-in-one HMD, which means that it doesn't need to be connected to a PC to run an application, it can be run inside the hardware of the headset itself. However during the development process, the headset needs to be connected to the PC and the application can be run directly from Unity.

For this chapter we will use dataset examples from MIxT to illustrate the concepts in this chapter. MIxT is a web application that is used for the visualization of bioinformatic data[5][4] and the datasets used here contain genetic information about a woman with breast cancer. There are in total 2 datasets, the first one is from a blood sample and the second one is from the tumor sample. In Figure 2.1 we can see an example of the application running using the blood dataset from MIxT.

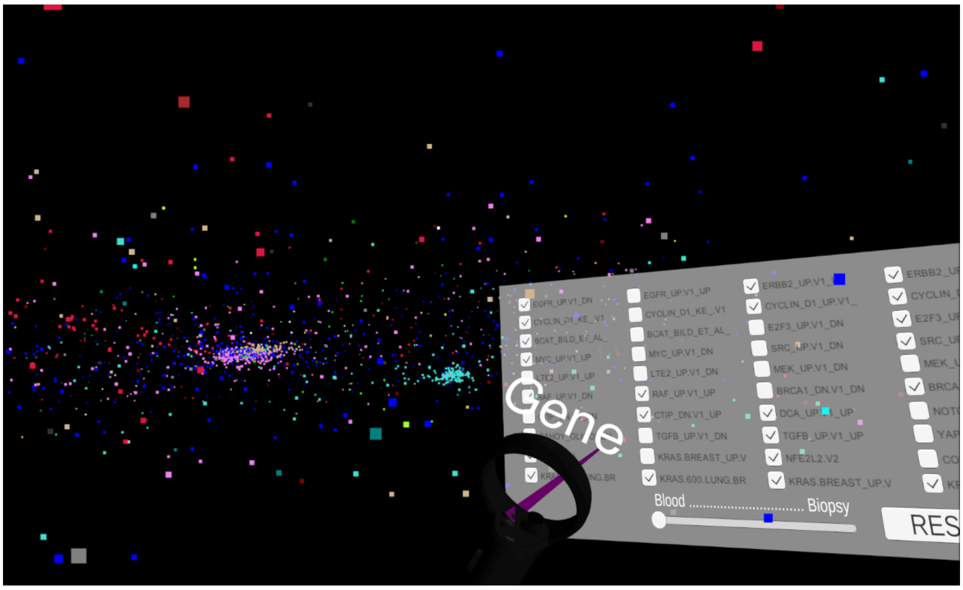


Figure 2.1: BigNet VR. Example of the application running on a Oculus Quest.

2.1 Visualization and interaction with the network

Virtual reality headsets offer a rich inmmersive experience. It's not only about inmersing the user into a 3D environment, but also giving the user the possibility to interact with the environment itself. This makes it possible to build complex VR applications where the user can do almost anything in a virtual world. Some examples of what it is possible to do in a VR application are moving around by teleporting to a different spot, grabbing objects, interact with the environment, 2D interfaces and menus and interact with them, etc. In this section we will explain the techniques that we have implemented to visualize and interact with the network. These techniques correspond to how the user can move around in the network and how the user can manipulate it.

Oculus Quest is an all-in-one VR headset and so it doesn't need a PC nor wires to run the applications. Apart from the headset it comes with 2 controllers, one for each hand. These controlers have inputs as buttons, thumbsticks and triggers that can be used to activate actions in the VR application. We have used some of these inputs available in the controllers in BigNet VR and mapped them to different actions that allow the user interact with the network and the environment. In Figure 2.2 we can see which action correspond each input from the controllers.

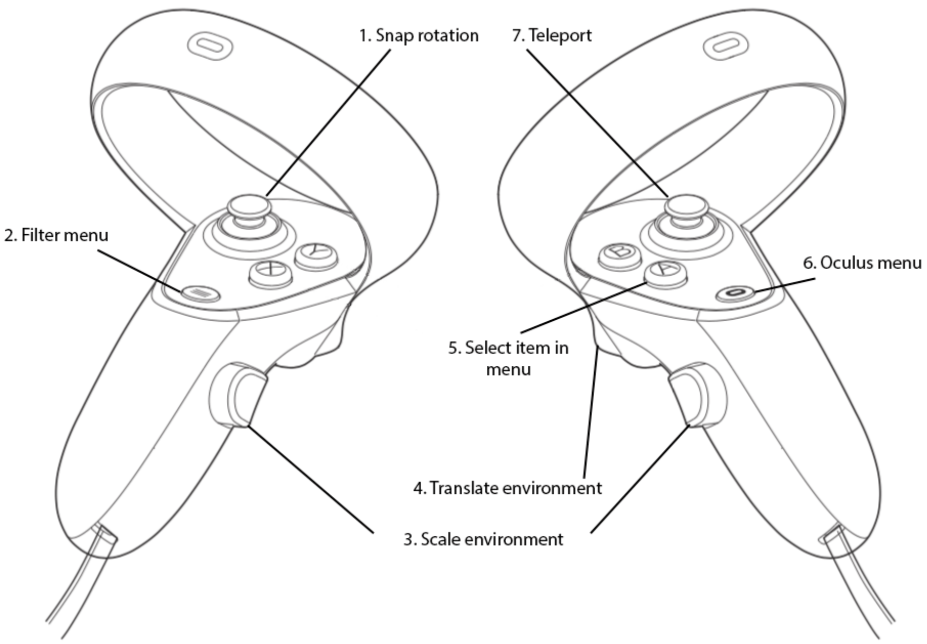


Figure 2.2: Mapping of the Oculus Quest controllers for the different actions implemented in BigNet VR. 1. Snap rotation. 2. Filter menu. 3. Scale environment. 4. Translate environment. 5. Select item in menu. 6. Oculus menu. 7. Teleport. Adapted figure from Oculus developer's page[9].

2.1.1 Locomotion

Locomotion is one of the most important ways of interaction in virtual reality experiences. It can be defined as a self-proppelled movement in virtual environments. Even though moving around is not the main goal in most of VR applications, it is an important aspect for the user's perspective in order to move the user's viewpoint in the virtual world.

Locomotion can have an strong influence in the user's experience. A poorly designed locomotion technique can reduce the user's immersion and even introduce motion sickness, which is related to the movement that the locomotion technique produces. HMDs like Oculus Quest allow the users to control the position and the orientation of the viewpoint by moving their heads and walking. However large virtual environments such as BigNet VR needs a big physical tracked area which cannot be covered by just walking around. It is for this reason that we need to use a locomotion technique that makes it possible to move around without the having to walk[2].

The locomotion technique that we use in BigNet VR is called teleportation. It consists in choosing a spot on the floor were we want to teleport to. To do this the user has to move forward the joystick from the right controller (see input 7. Teleport from Figure 2.2). In addition it is possible to choose which direction the user will face once the teleportation is done. To do this we just need to rotate the same joystick to the desired direction. Once the user releases the joystick, a black flash will show followed by the new position. This black flash is very important when implementing some techniques because it prevents from producing motion sickness and disorientation.

In Figure 2.3 we can see an example of how the teleportation technique is used in BigNet VR. As we can see, a parabolic arc is shown as if we are throwing an object to the spot where we want to teleport. A round green circle is also shown on the floor on the place where we will teleport. This includes also an arrow, indicating the direction that we will face once we are teleported.

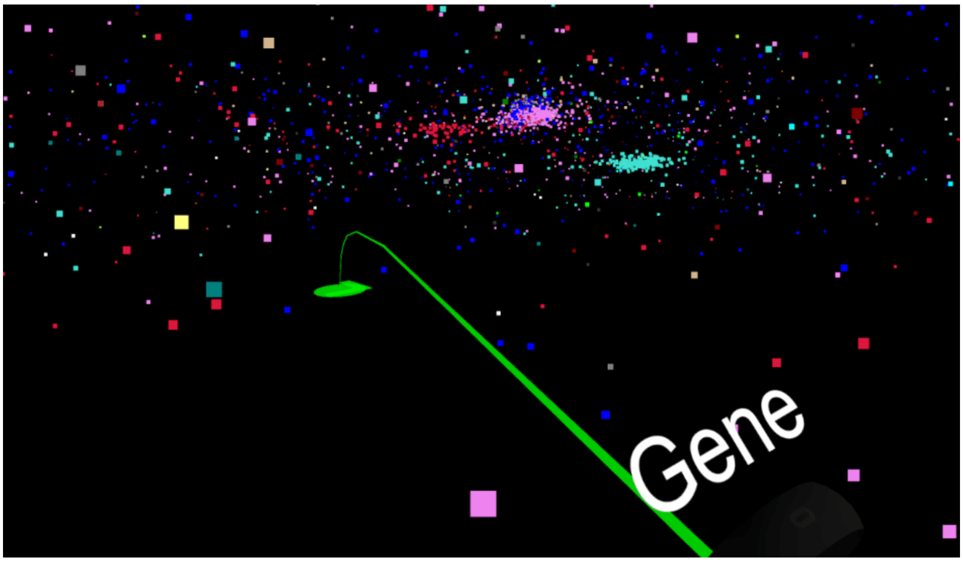


Figure 2.3: Teleportation technique. The user can use the jystick from the right controller to teleport to a different spot. To choose the spot a parabolic arc will appear.

2.1.2 Translation of the network

By teleporting to different places in the environment we allow the user see the network from different perspectives. However it is also interesting being able to move the network and specially move in a precise way so that the user has more control over it. The user might for instance be able to see the network or a specific node or cluster from above or also from below. To do this we have implemented a functionality to translate the network to wherever we want.

In order to translate the network in BigNet VR, the user just needs to press on the 4. Translate environment button from the right controller as shown in 2.2. Then the user needs to keep holding this button down and move the hand to the direction to which the user wants the network to move to. This is a very intuitive approach because it feels like we are just pulling from a rope that is connected to the network so we just need to move it to the direction we want without involving a lot of learning for this process.

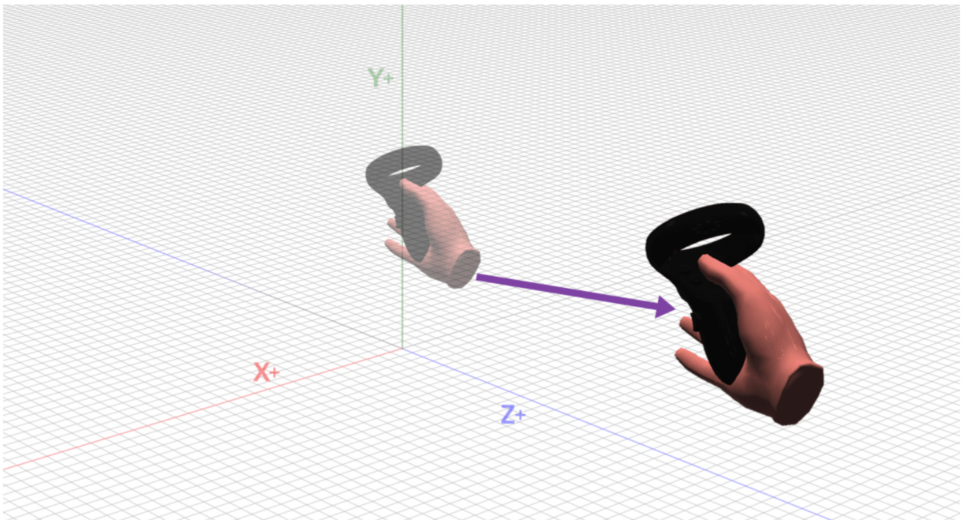


Figure 2.4: Translation of the network functionality. The user holds the translation button on the Oculus controller and moves the hand to the direction where he or she wants the network to translate.

2.1.3 Zooming in the network

When exploring a big network with hundreds of nodes, sometimes the information can be too crowded. In our example dataset that we use in BigNet VR, there are some clusters of nodes that have too many nodes close to each other and it gets very hard to visualize them properly. A way to cope with this problem is for instance by "zooming" in the part of the network that we want to explore better. We implement then a scaling functionality that makes the network bigger or smaller and so simulating a zooming action.

The way we implemented the zooming functionality in BigNet VR is by using the side buttons with the name "3. Scale environment" shown in Figure 2.2. In the first place the user needs to press and hold this button from both right and left controllers and in the second place the user needs to expand or contract the arms, as if we were streching out or contracting the network itself. This is also an intuitive acction to do since the user might think that he or she is actually stretching the network with the hands.

In Figure 2.5 we an see a visual example of how the zooming works using the Oculus controllers. In this example the user is stretching the hands out in order to make the network bigger. The user starts in a intial position, then holds the zooming buttons from both controllers and then moves the hands out. If we wanted to make the network smaller we would have to do the opposite action, by contracting the hands to the inside.

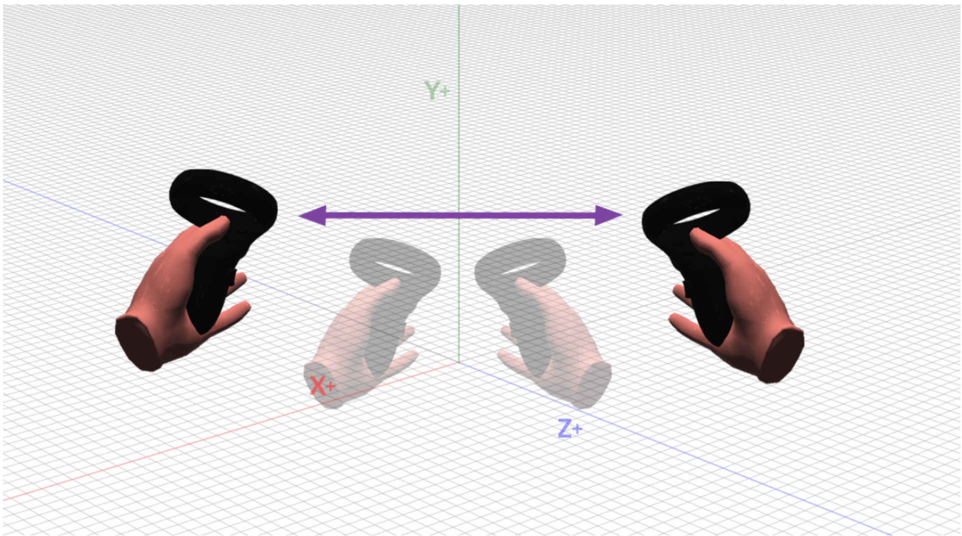


Figure 2.5: Zooming in the network functionality. The user can hold the scaling buttons on the Oculus controller to make the network bigger or smaller. In this example if we strech our hands outside, the network will expand.

category	genes
brown	ARHGAP30 FERMT3 ARHGAP25 CD53 PLEK IRF8 DOCK2
cyan	SAFB MOB3A RAB35 ABR ASCC2 CDC37 ANKFY1 GLTSCR1
darkgrey	RAB40C ZNF213 ZNF263 PIGQ RHBDF1 RAB11FIP3
darkorange	TCEB1 MRPL13 ENY2 MTERF3 UBE2W WDYHV1

Table 2.1: Fragment of the dataset with the categories and the genes belonging to each category from the biopsy sample.

source	target	weight	id
AAMP	ARGLU1	0.102486209330144	AAMP-ARGLU1
ACADM	FOXN2	0.107506881676173	ACADM-FOXN2
ACADM	MBNL1	0.12269622045714	ACADM-MBNL1
ACADM	PPM1B	0.103496640767895	ACADM-PPM1B

Table 2.2: Fragment of the dataset used to build the network relationships of the blood sample.

2.1.4 Interaction with the nodes

2.1.5 Showing node relationships

2.2 Creation of the network in a 3D space

2.3 Filtering information in the network

How the network is filtered.

3

Related work

I will focuse in this chapter on VR applications found in the literature for the visualization of bioinformatic data.

3.1 BioVR

3.2 CellexaVR

3.3 BigTop

4

Evaluation and discussion

Questions to answer in the evaluation chapter:

1. How big can the graph be so that it is comfortable visualizing the network?
   What is comfortable? Number of FPS? How can we scale the graph? By adding nodes and spread them around, by adding more interconnexions? Should the experiment split in several parts? Scaling, filtering, moving around, etc. What is the performance by using Oculus Link and the performance using just the Quest hardware? -We can use the Unity GPU Profiler for Oculus Quest and Go in order to see the performance.
   See: Getting Started w/ The Unity GPU Profiler for Oculus Quest and Go

2. How is this way of visualizing the graph better by using VR?
   We are researchging the technology and the test with actual users is for future work.

3. In what way can the application and the visualization of the graph be improved?
   Argue in the discussion part.

5

Conclusion and future work

Bibliography

[1] Corey J. Bohil, Bradly Alicea, and Frank A. Biocca. "Virtual reality in neuroscience research and therapy." In: *Nature Reviews Neuroscience* 12.12 (2011), 752–762. DOI: 10.1038/nrn3122.

[2] Evren Bozgeyikli et al. "Point & Teleport Locomotion Technique for Virtual Reality." In: *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play - CHI PLAY 16* (2016). DOI: 10.1145/2967934.2968105.

[3] "Commentary on Rose, F.D., Brooks, B.M., & Rizzo, A.A., Virtual Reality in Brain Damage Rehabilitation: Review." In: *CyberPsychology & Behavior* 8.3 (2005), 263–271. DOI: 10.1089/cpb.2005.8.263.

[4] Vanessa Dumeaux et al. "Interactions between the tumor and the blood systemic response of breast cancer patients." In: *PLOS Computational Biology* 13.9 (2017). DOI: 10.1371/journal.pcbi.1005680.

[5] Bjørn Fjukstad et al. "Building Applications for Interactive Data Exploration in Systems Biology." In: *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology,and Health Informatics - ACM-BCB 17* (2017). DOI: 10.1145/3107411.3107481.

[6] KE. Laver et al. "Virtual reality for stroke rehabilitation." In: *Cochrane Database of Systematic Reviews* 11 (2017). ISSN: 1465-1858. DOI: 10.1002/14651858.CD008349.pub4. URL: https://doi.org//10.1002/14651858.CD008349.pub4.

[7] Mike May. "LIFE SCIENCE TECHNOLOGIES: Big biological impacts from big data." In: *Science* 344.6189 (2014), pp. 1298–1300. ISSN: 0036-8075. DOI: 10.1126/science.344.6189.1298. eprint: https://science.sciencemag.org/content. URL: https://science.sciencemag.org/content/344/6189/1298.

[8] Matthias Minderer et al. "Virtual reality explored." In: *Nature* 533.7603 (2016), 324–325. DOI: 10.1038/nature17899.

[9] *OVRInput*. URL: https://developer.oculus.com/documentation/unity/unity-ovrinput/.

[10] Georgios A. Pavlopoulos et al. "Visualizing genome and systems biology: technologies, tools, implementation techniques and trends, past, present and future." In: *GigaScience* 4.1 (2015). DOI: 10.1186/s13742-015-0077-2.

[11] David A. Thorley-Lawson, Karen A. Duca, and Michael Shapiro. "Epstein-Barr virus: a paradigm for persistent infection – for real and in virtual reality." In: *Trends in Immunology* 29.4 (2008), 195–201. DOI: 10.1016/j.it.2008.01.006.

[12] James Xia et al. "Three-dimensional virtual-reality surgical planning and soft-tissue prediction for orthognathic surgery." In: *IEEE Transactions on Information Technology in Biomedicine* 5.2 (2001), 97–107. DOI: 10.1109/4233.924800.

[13] Yuting Yang et al. "Integration of metabolic networks and gene expression in virtual reality." In: *Bioinformatics* 21.18 (July 2005), pp. 3645–3650. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/bti581. eprint: http://oup.prod.sis.lan/bioinformatics/article-pdf/21/18/3645/520673/bti581.pdf. URL: https://doi.org/10.1093/bioinformatics/bti581.

[14] Jimmy F. Zhang et al. "BioVR: a platform for virtual reality assisted biological data integration and visualization." In: *BMC Bioinformatics* 20.1 (2019). DOI: 10.1186/s12859-019-2666-z.