TranScape VR

Transcriptomic land**scape** visualisation tool in Virtual Reality

Dr. Annika Kreikenbohm[1], Anda Iosip[2]

[1]Institute of Physics, University of Wuerzburg, Germany, [2]Center for Computational and Theoretical Biology, University of Wuerzburg, Germany

INTRODUCTION

Nowadays RNA-seq is the norm for investigating transcriptome-wide gene expression. These data cover the expression of thousands of genes over many different tissues and conditions. Finding the biological story hidden in this massive amount of data is challenging and standard data analysis methods using 2D, non-interactive visualization are limiting and daunting. **To upgrade to state-of-the-art technology, we developed TranScape VR, a new tool for visualizing transcriptomic landscapes in an immersive, explorative, holistic and entertaining Virtual Reality (VR) space.**

METHODS

Soft & Hardware: Unity (version 2019.3), C# Scripts, Steam VR 2.0 package, Valve Index VR glasses.

Data: we use publicly available data of seven different organs and tissues of the Venus flytrap (*Dionaea muscipula*), a deadly carnivorous plant [1].

Concept: In search for special plant carnivory-related genes, we represent transcriptomic data as an interactive network, mapping genes to each tissue core-node, forming gene clusters. Biological parameters (e.g. gene expression level, tissue specificity) are mapped to visual parameters such as colour, size, distance, and gene position within the 3D space. The visualization offers different modes (*see panels 1 – 3*) which depict different properties of the data set. All modes can be interactively explored in order to form and test hypothesis in real-time (*see panels A – D*).

GENE MAPPING

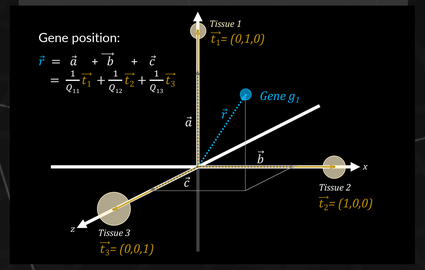
Genes are represented by points in 3D space. Positions are defined by the sum of each tissue vector that hosts the gene, weighted by the inverse of the specificity. The tissue specificity of a gene is determined using the Shannon Entropy method [1,2]. It is measured by the Q-value. A very small Q-value indicates a very high specificity level and vice versa.

The position of a gene k found in tissues i is:

$$\vec{r}_k = \sum_i \frac{1}{Q_{ki}} \vec{t}_i,$$

where $\vec{t}_i$ is the position vector of tissue T_i and Q_{ki} is the specificity of gene k to tissue i.

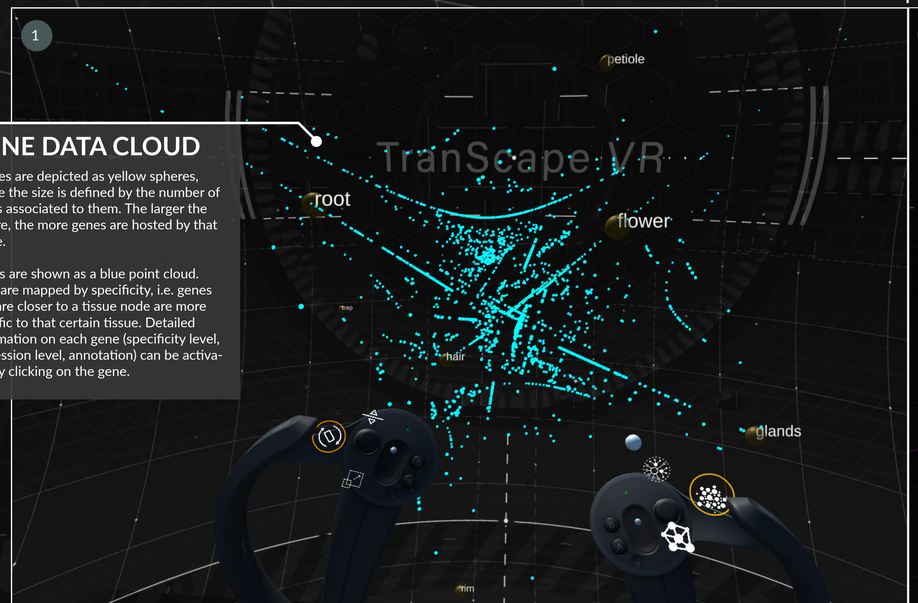
An example is given in the sketch below. Gene g, is found in tissues T1, T2, and T3 with specificities Q_{11} = 1.4, Q_{12} = 1.5 and Q_{13} = 1.25, respectively. Hence, the tissue vector are scale by a factor of 0.71, 0.66 and 0.8 and summed up to yield the gene positon (0.66, 0.71, 0.8), depicted as blue point.



REFERENCES:

[1] Iosip AL, Böhm J, Scherzer S, Al-Rasheid KAS, Dreyer I, Schultz J, et al. (2020): PLoS Biol 18(12) (doi:10.1371/journal.pbio.3000964)

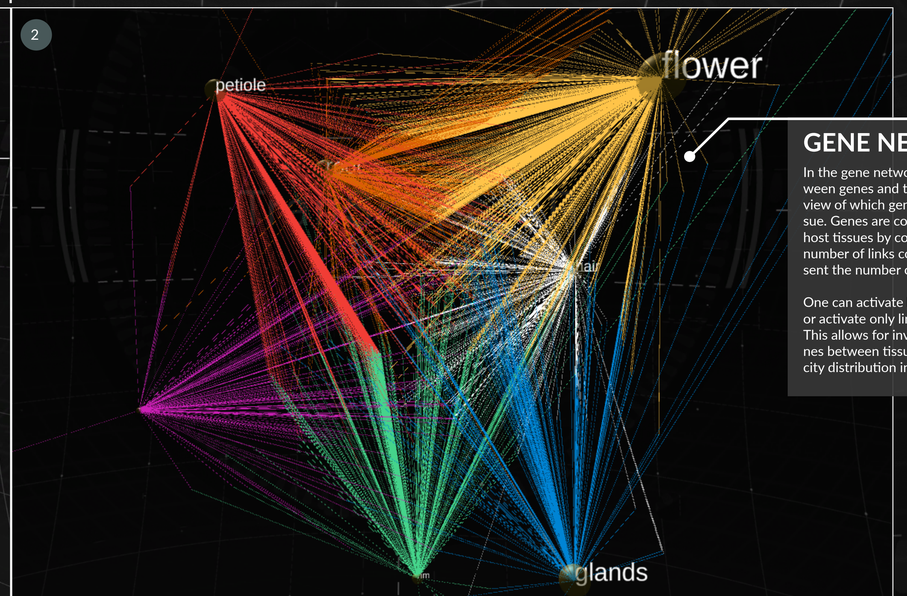
[2] Schug J, Schuller WP, Kappen C, Salbaum JM, Bucan M, Stoeckert Jr, CJ, (2005): Genome Biology 2005, 6:R33 (doi:10.1186/gb-2005-6-4-r33)



GENE DATA CLOUD

Tissues are depicted as yellow spheres, where the size is defined by the number of genes associated to them. The larger the sphere, the more genes are hosted by that tissue.

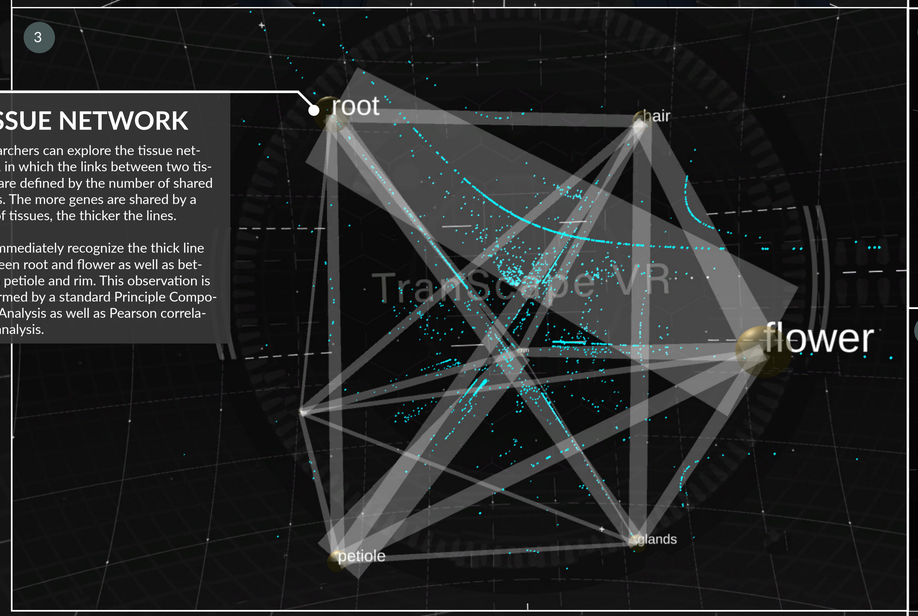
Genes are shown as a blue point cloud. They are mapped by specificity, i.e. genes that are closer to a tissue node are more specific to that certain tissue. Detailed information on each gene (specificity level, expression level, annotation) can be activated by clicking on the gene.



GENE NETWORK

In the gene network mode, the links between genes and tissues provide a clear view of which genes belong to which tissue. Genes are connected to each of their host tissues by color-coded lines. Thus, the number of links connected to a gene represent the number of tissues that host it.

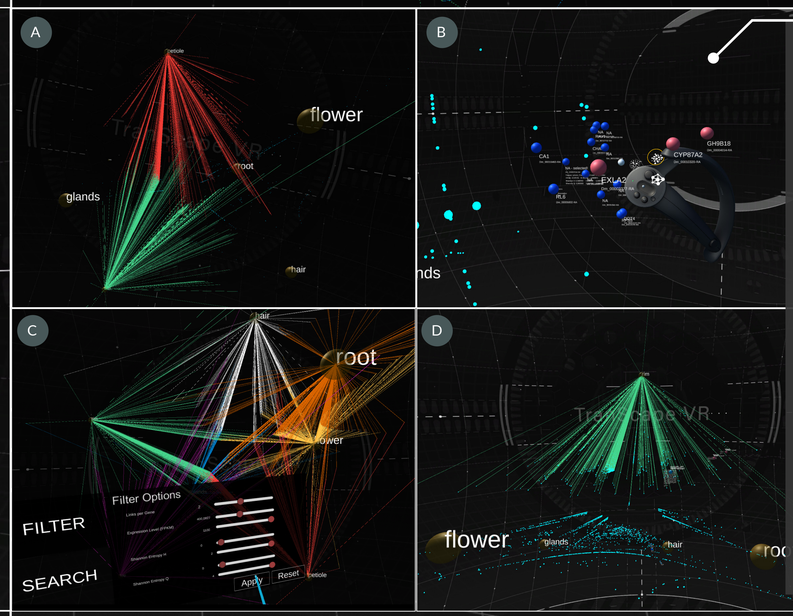
One can activate links to all tissues at once or activate only links to selected tissues. This allows for investigation of shared genes between tissues as well as the specificity distribution in more detail.



TISSUE NETWORK

Researchers can explore the tissue network, in which the links between two tissues are defined by the number of shared genes. The more genes are shared by a pair of tissues, the thicker the lines.

We immediately recognize the thick line between root and flower as well as between petiole and rim. This observation is confirmed by a standard Principle Component Analysis as well as Pearson correlation analysis.



VISUAL EXPLORATION

The user can visually explore and analyse the multidimensional dataset by re-scaling, interacting, retrieving genes' description as well as filtering or highlighting genes based on keywords.

- **Scale, rotate and move** the gene cluster for different (over) views
- **Rearrange tissue nodes** to investigate different mapping constellations in real-time (*e.g., panel D*)
- **Activate tissue links** to find assosiated genes and **activate gene details** for quantitative data and description (*e.g., panels A, B*)
- **Filter parameters** (e.g., expression level, specificity level) to define which genes are displayed and **search for (groups of) genes** using keywords (*e.g, panel C*)

Through visual exploration we can observe patterns which reflect and are confirmed by standard analysis methods, e.g.:

- **Correlation Analysis,**
- **Principal Component Analysis,**
- **Tissue Specificity Analysis**

SUMMARY

- The visualization reflects multiple standard analysis in one shot through visual exploration
- The depth and rescaling features allows users to get an overview and provides details on demand
- Large and interconnected data sets can be interacted with and handled in an intuitive and fun way
- Different mapping strategies as well as types of parameters can be implemented, making it highly custumizable

The project is publicly available at
https://github.com/annok23/TranscapeVR

For more information contact:
annika.kreikenbohm@uni-wuerzburg.de
anda-larisa.iosip@uni-wuerzburg.de