Gene Selection Using Gene Expression Data in a Virtual Environment

Kunihiro Nishimura[1]
`kuni@cyber.rcast.u-tokyo.ac.jp`

Shumpei Ishikawa[2]   **Shuichi Tsutsumi[2]**
`{shumpei, shuichi}@genome.rcast.u-tokyo.ac.jp`

Hiroyuki Aburatani[2]
`haburata-tky@umin.ac.jp`

Koichi Hirota[2]   **Michitaka Hirose[2]**
`{hirota, hirose}@cyber.rcast.u-tokyo.ac.jp`

[1] Graduate School of Information Science and Technology, The University of Tokyo,
    7-3-1, Hongo Bunkyo-ku, Tokyo 113-0033, Japan
[2]   Reseach Center for Advanced Science and Technology, The University of Tokyo,
    4-6-1, Komaba Meguro-ku, Tokyo 153-8904, Japan

1   Gene Selection for Drug Discovery

The gene expression data of many normal and diseased tissues enables the selection of genes that satisfy the conditions for finding a new drug. In the process of selecting genes is required to determine many filtering parameters. The system of the selection of genes from all genes is required to have the following features: 1) visualization of the data to provide the users to grasp both holistic and detail views of the data, 2) interactivity to determine many filtering parameters, 3) provides spatial cue to understand intuitively because there are many filtering parameters. In order to visualize various information of large number of genes, the system also needs a wide field of view. Virtual reality technology satisfies these requirements above.

2   Virtual Reality

Virtual reality technology can help improve genome data analysis by enabling interactive visualization and providing a three-dimensional virtual environment. It is discussed that visualization and interactivity is effective in the gene expression analysis [1]. Immersive projection technology that can create three-dimensional virtual environments and provide high interactivity is a key to huge data mining [3]. In this paper, we developed the virtual workspace for gene selection of drug discovery using gene expression data that is satisfied the requirements above. The system is implemented to the immersive projection display, CABIN (Fig. 1). CABIN has five screens and generates as a 270-degree field of view [2]. It can provide a three-dimensional stereoscopic view and enables the user to manipulate the data using a spatial interface device.

3   Method

The gene expression data used in this system is obtained from Affymetrix Gene Chip. We used the gene expression data of 35 normal samples that are available to the public through the web [4] and the data of one lung cancerous sample.
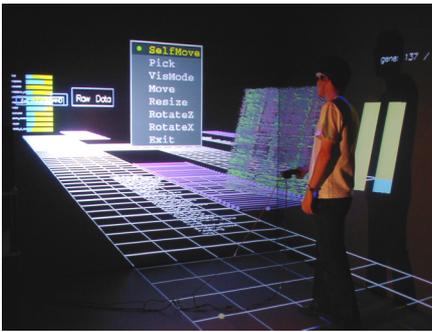

Figure 1: CABIN

Figure 2: Virtual environment for gene selection

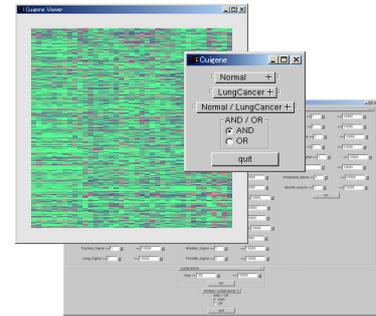Figure 3: Operation on slide bars to change filtering parameters

Figure 4: GUI version of gene selection system

In the virtual workspace, gene expression data and their filtering process are visualized as Fig. 2. There is a road and slide bars that are visualized the filtering parameters. A user can select genes that meet the conditions of drug discovery from among about 30,000 genes by setting the filtering parameters. The user can set these parameters interactively by using a three-dimensional input device shown in Fig.3. Spatial cues that are used in the presentation of parameters are thought to be effective in aiding the user's understanding when visualizing dozens of parameters in the virtual environment. The series of work processes are visualized as the metaphorical road, and the user selects work choices represented by branches in the road.

The system provides two kinds of filtering processes. The first one deals with the thresholds of raw expression level of each normal and cancerous tissue. The second one deals with the thresholds of gene expression ratio between normal and lung cancerous tissue. After the setting of the filtering thresholds, the filtering is carried out by multiplying "and" or "or" functions between the two kinds of processes. The selected genes' expression levels are visualized as a matrix and the selected genes' list is saved as a text file. The List contains gene name, UniGene ID, UniGene annotations, Locus Link, Gene Ontology, Conserved Domains, and the predicted gene localizations of a cell by PSORT.

3   Evaluations and Results

We compare the CABIN system and the system for PC that provides the same function using subjective evaluation. We developed the PC system using graphical user interface (GUI) shown in Fig. 4. The results indicate that the CABIN system is more effective than the PC system from the viewpoint of understanding the gene selection process. However, the working efficiency in the CABIN system is poorer than that in the PC system. The reason of this result derives from the difficulty of the pointing in the three-dimensional space.

Acknowledgements

References

[1] J. Seo, and B. Shneiderman, Interactively Exploring Hierarchical Clustering Results, IEEE Computer, 35(7):80-86, 2002.

[2] M. Hirose, T. Ogi, S. Ishiwata, T. Yamada, Development and Evaluation of the Imersive Multiscreen Display CABIN, Systems and Computers in Japan, 30(1):13-22, 1999.

[3] P. Call, Picture this: Drowning in data? New visualization techniques could help, Nature, 417:11-13, 2002.

[4] http://www2.genome.rcast.u-tokyo.ac.jp/database/