DANCER: a program for digital anatomical reconstruction of gene expression data

Matti Kankainen and Garry Wong*

Laboratory of Functional Genomics and Bioinformatics, Department of Neurobiology, A.I. Virtanen Institute for Molecular Sciences, University of Kuopio, PO Box 1627, FIN-70211, Kuopio, Finland

Received July 22, 2003; Revised and Accepted September 11, 2003

ABSTRACT

A digital anatomy construction (DANCER) program was developed for gene expression data. DANCER can be used to reconstruct anatomical images from in situ hybridization images, microarray or other gene expression data. The program fills regions of a drawn figure with the corresponding values from a gene expression data set. The output of the program presents the expression levels of a particular gene in a particular region relative to other regions. The program was tested with values from experimental in situ hybridization autoradiographs and from a microarray experiment. Reconstruction of in situ hybridization data from adult rat brain made by DANCER corresponded well with the original autoradiograph. Reconstruction of microarray data from adult mouse brains provided images that reflect actual expression levels. This program should help to provide visualization and interpretation of data derived from gene expression experiments. DANCER may be freely downloaded.

INTRODUCTION

Current high throughput functional genomic techniques allow for the production of massive amounts of gene expression data (1,2). These include large gene expression screens using in situ hybridization, in situ synthesized oligo-arrays and two-color microarray techniques. Studies using these techniques are aimed at a variety of biological questions including expression profiling to compare healthy versus disease tissues, understanding drug effects, comparative genomics and transcriptome analysis.

Transcriptome analysis seeks to catalog the qualitative and quantitative amounts of genes within a cell type or tissue. The goals of these studies are to provide a list of expressed genes and their expression levels which can then be analyzed to investigate mechanisms of cell- or tissue-specific regulation or to discover novel disease gene markers (3).

A traditional method to investigate neuroanatomical gene expression patterns has been by in situ hybridization. While resolution is high and anatomical structures remain intact, the method is labor intensive (4). Microarrays have also been used to investigate neuroanatomical gene expression patterns (3). While throughput with respect to the number of genes is high, the anatomical resolution is low and, presently, there is no method to reconstruct an anatomical image from the data.

In order to merge the advantageous features from these two gene expression analysis techniques, localization and visualization from in situ hybridization, and massive gene numbers from microarrays, we have created a tool called DANCER to digitally reconstruct anatomical images from microarray data. Since anatomical templates can be drawn by the user, the tool is flexible for virtually any type of anatomical reconstruction. Since data is entered in rows and columns as numerical values, gene expression or other types of quantitative data, such as protein expression, can also be entered.

MATERIALS AND METHODS

Principles of the program

The task of the program is to reconstruct anatomical images from values provided for each anatomical region. The selected areas which correspond to anatomical regions must be enclosed and contours must be drawn in black for correct filling by the program. After recognizing areas from the figure, the program fills them with colors based on numerical values for each region. The user can determine the coloring scheme. Fill colors are calculated automatically either from selected gene minimum and maximum expression values or from user-determined minimum and maximum values. The user can also perform global normalization by selecting these values from the data set or from selected gene group minimum, maximum or average expression values that are determined automatically by DANCER.

The fill colors used to color the regions are scaled relative to minimum and maximum values. If gene expression values are 0, 2.5, 5, 7.5 and 10 and minimum and maximum values are set to be 0 and 10, the chosen sections will be colored as follows: minimum (0 value), a color between the minimum and middle (2.5), middle (5), a color between the middle and maximum (7.5) and maximum (10). If gene expression values are 0, 5, 10, 15 and 20 and the minimum and maximum values are set to be 0 and 20, the user will have exactly the same output as in the first case. Therefore, to make outputs relative, the user has to set the minimum and the maximum values to be

*To whom correspondence should be addressed. Tel: +358 17 162108; Fax: +358 17 163030; Email: garry.wong@uku.fi

Present address:
Matti Kankainen, Structural Genomics Group, Institute of Biotechnology, University of Helsinki, PO Box 56 (Viikinkaari 5), FIN-00014, Helsinki, Finland
0 and 20 in both sets. The program also does not treat nor transform values from the data set. If the user wants to perform log or any other type of transformation, it must be done prior to data entry. The DANCER user interface is shown in Figure 1.

DANCER accepts, as a data input, a file which contains the gene expression data set in a tab-delimited, Excel or Unicode format. The second input file is a picture file which contains the anatomical figure to be filled. DANCER is able to use picture inputs, which are in bmp, gif or jpg format. To color the regions of the figure, the user chooses the regions which will be filled by clicking onto the region. This is matched with a column on the data set from which the value for filling will be taken. The user then selects the gene from a row in the data set. As a result or output, DANCER creates an image which shows the relative expression levels of the selected gene in marked regions. Created pictures can be saved afterwards.

**Test values**

The sensitivity and functionality of DANCER were tested by using different data sets and by comparing the DANCER outputs to original outputs, if it was possible to do so. The first and second data sets presented expression levels of genes or expressed sequence tags (ESTs) from experimental animal brains. The source of the first data set was rats and the source of the second data set was mice. The first data set contained 146 ESTs quantified in situ hybridization autoradiograms (5) (original data can be found at http://www.uku.fi/aivi/synexpre.htm). The levels of mRNA expression of ESTs were quantified by a video-based computer image analysis system (MCID/M4; Imaging Research, St Catharines, Ontario, Canada). The second data set contained array-based gene expression data sets (GeneChip; Affymetrix, Santa Clara, CA). The total number of probe sets was 13 069, which corresponded to more than 10 000 genes and ESTs (3) (Mu11KsubA/B) (data can be found at ftp://ftp.gnf.org/pub/papers/brainstrain/).

**Preparation of template figures**

Templates for the brain and the embryo that come with the program package are based on figures taken either from a rat brain atlas (6) or from the mouse development atlas (7). Figures were first drawn and then scanned. In the next step, figures were managed by an image processing program (Adobe Photoshop; ADOBE Inc., San Jose, CA). These steps contained amplification and black coloring of contours by using tools which recognize and strengthen the contours and edges from the figure. Pictures sizes were also shrunk, so that they would fit into the program’s picture form. Figures were also treated to be more distinct by removing some brain sections that were too small to be quantified.

**Availability and running the program**

DANCER is written in Microsoft Visual Basic language and runs in a Windows operating system environment. DANCER has been tested with Windows 2000 and Windows XP operating systems, and in those systems it has worked without problems. The source code and the program package can be downloaded from the website (http://www.uku.fi/aivi/dancer.htm). The program package is stored in a compressed file which includes the executable program, example templates, example data sets and a user manual.

**RESULTS**

To determine whether DANCER could reproduce an in situ autoradiograph image, a picture file was produced from a rat brain atlas and data from an autoradiograph were quantitated (Fig. 2A). The original expression level of the gene in each brain region is presented as a histogram (Fig. 2B). The DANCER output files are shown in Figure 2C and D. The first output figure (Fig. 2C) displays brain sections that were filled with monochrome colors, where white represents the lowest value and black represents the highest value. The second output figure (Fig. 2D) displays brain sections that were filled...
with red, cyan and green colors, where red represents the lowest value and green represents the highest value. The second output figure shows that it is also possible to choose other color schemes within the program. When comparing the original in situ hybridization figure with DANCER figures, the correspondence is good.

DANCER was also used to perform anatomical reconstruction on array-based gene expression data sets. In this experiment, the template figure was made from an atlas in a manner similar to Figure 2. As a result, a figure was created by DANCER which contains relative gene expression values (Fig. 3A). The actual values from the microarray data set are shown below it for comparison (Fig. 3B). Adjustment of the offset (minimum value) allows for better contrast in the DANCER-generated image compared with the histogram.

To demonstrate the sensitivity and the resolution of DANCER, a mouse embryo template was constructed (Fig. 3C). The template picture was produced from a mouse development atlas (7) at 10.25–10.5 d.p.c. At least 34 distinct regions can be filled in this template.

**DISCUSSION**

In this work, we created a novel computer program that can be used to perform digital reconstruction of anatomical images from gene expression and microarray data. This is a novel approach, and to our knowledge a program to perform this work does not yet exist. The principle idea was to create a computer program that would aid in the easy visualization of gene expression data in an anatomical manner.

Since gene expression experiments are designed to provide a framework to understand drug sensitivity, diseases and the interplay between the genes that give rise to complex behaviors and unique brain functions (3), it is important to have tools which give quick and easy ways to handle data sets generated from these projects. DANCER offers a novel technique to create anatomical in silico hybridizations. It is also recognized that accessing the anatomical distribution of a single gene at a time may be limiting. The DANCER program has a player that can run through the list of genes at a speed determined by the user. This feature, combined with sorting and finding tools, may make it useful to read in a rapid fashion the results of a microarray experiment where genes are clustered into functional pathways. It is also possible to store shrunken images into the figure collection window, where users can compare their constructed image results side by side.

In contrast with existing techniques which provide a way to visualize the microarray data, DANCER gives the output as an anatomical figure. Most of the microarray tools that are accessible for academic use, such as AVA (8) or Arrayplot (9), visualize data into clusters, patterns, dot-plots or some other mathematical set. Nonetheless, these programs seem to

---

**Figure 2.** Gene expression values in different brain regions; the accession number of the gene is AI100885. (A) An in situ autoradiogram of the gene AI100885 (5). (B) Quantified expression values of the autoradiogram. (C) The template from adult rat brain (bregma 0.48 mm) that has been colored with DANCER with monochrome colors. (D) The template from adult rat brain (bregma 0.48 mm) that has been colored with DANCER with red, cyan and green colors. The abbreviations for the regions are: Cpu, caudate putamen; Cg, cingulate cortex; Fr, frontal cortex; Par, parietal cortex; Den, dorsal endopiriform nu; Tu, olfactory tubercle; Lsd, lateral septal nu, dorsal; Lsi, lateral septal nu, intermediate; Lsv, lateral septal nu, ventral; aca, anterior commissure, anterior; VDB, nu vertical limb diagonal band; cc, corpus callosum; I, insular cortex; Vp, ventral pallidum.
dismiss the anatomical idea of genes: the notion that genes are expressed in tissues and regions. One DANCER application is to render anatomy from a microarray experiment, and thus obviate the need to perform in situ or northern blot experiments once the microarray experiments have been done.

When computational power becomes greater, it might be possible to perform 3D anatomical reconstructions as some groups have shown using different techniques (10). Before this can take place, some problems, such as whether gene expression from a dissected region can be generalized to the entire 3D tissue region, need to be addressed. There is also a need to use transparency colors that enable visualization of the entire tissue.

Since users can draw their own anatomical templates, DANCER has no limitation for tissue, species or developmental age. Since anatomical templates may be common for many users, sharing of template files could increase the usefulness of the program while decreasing unnecessary labor. Since DANCER accepts data sets which contain numerical values, DANCER can accept protein as well as gene expression data. In summary, users should be able to reconstruct any kind of anatomical figure using any kind of numerical data.

ACKNOWLEDGEMENTS

We wish to thank Markéta Marvanová, Markus Storvik and Petri Törönen for helpful discussions and providing data. This work was supported by a grant from the Academy of Finland (G.W.)

REFERENCES