Interactive Visual Analysis on Gene Regulatory Pathways

Dong Hyun Jeong, William Ribarsky, Alireza Darvish, Kayvan Najarian, and Jing Yang

College of Information Technology, University of North Carolina at Charlotte, USA

Abstract

Dynamic regulatory pathways, estimated using DNA microarray time-series information, provide predictions of dynamic interactions among genes. Even though several analysis methods have been studied, most of them fall short in analyzing or observing these interactions in detail. In contrast to other methods, we have developed a visual approach based on the gene regulatory pathway prediction method [DHN04]. The prediction method applies two different techniques, clustering and an Auto Regressive (AR) model. Even though this approach provides a successful prediction of the dynamic pathways involved in biological processes, it is not easy to analyze the gene interactions. To counteract this, in this paper we designed a visual analysis of gene regulatory pathways. This framework allows the user to easily find the dynamic interactions among genes, highlighting interesting gene information, and showing the detailed annotations of the selected genes. In order to support greater analysis capability, it also provides a simple graph display, a pixel-based gene visualization technique, and a relation-displaying technique among gene expressions and gene regulatory pathways.

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [User Interfaces]: Graphical user interfaces (GUI) / I.3.6 [Methodology and Techniques]: Interaction techniques / J.3 [Life and Medical Sciences]: Biology and genetics

1. Introduction

The human body and other organisms are dauntingly complex. In order to significantly improve medicine and human health, we need a more detailed understanding of the vast networks of molecules whose interactions and regulatory behavior affect drugs and disease.

Precise knowledge of gene regulatory pathways can provide an understanding of the time-dependent enhancement and suppression of gene activity and drug effectiveness. We can learn, for example, how the effects of drugs are “turned on or off” and what combinations of drugs may be effective. But since biological pathways represent complex interactions among the genes, visual representations are necessary to facilitate the exploratory analysis and understanding of these interactions.

In this paper we present a new framework for supporting visual analysis among complex gene pathways. The framework uses a new gene regulatory pathway prediction method [DHN04] to predict the gene expressions over time. The method employs clustering to group the gene pool into a number of biologically-meaningful clusters as a pre-processing step. Then an Auto-Regressive (AR) model is applied to relate the expression levels of each of the prototypes. The visualization framework maps the clustered pathway information onto 3D space. In 3D space, further exploration of interactions among clusters is performed using simple mouse clicks. The visual environment provides an overview and a detail-view in a rich, dynamic context as shown in Figure 1.

The major advantage of using a 3D representation technique is the increased possibility of integrating additional visual information into the representation [TSWS05]. But there are some drawbacks to understanding visual information in overcrowded and cluttered displays, and the display environment in Figure 1 is designed to overcome these drawbacks. The user can move easily over time and between the overview and the detail view in order to maximize efficiency and create understanding. We developed the design based on our knowledge of bioinformatics needs and processes. (Two of us are working bioinformaticists and the rest of us have been working with bioinformatics professionals on interactive visualization for some time.)

In the following sections, several design properties and techniques used in the visual analysis framework will be described further:

- Biological pathway prediction techniques using clustering and an AR model,
- A new 3D information visualization technique to display clustered pathway information over time in 3D space,
- A novel overview and detail-view approach with which the user can employ interactive analysis to create new understandings of dynamic pathways, gene properties, and relationships,
- Design schemes used in the detail view to continuously reveal the patterns and relations using simple graph drawing and pixel-based visualization techniques,
- An integrated analysis tool that can be launched and whose results can then be compared.

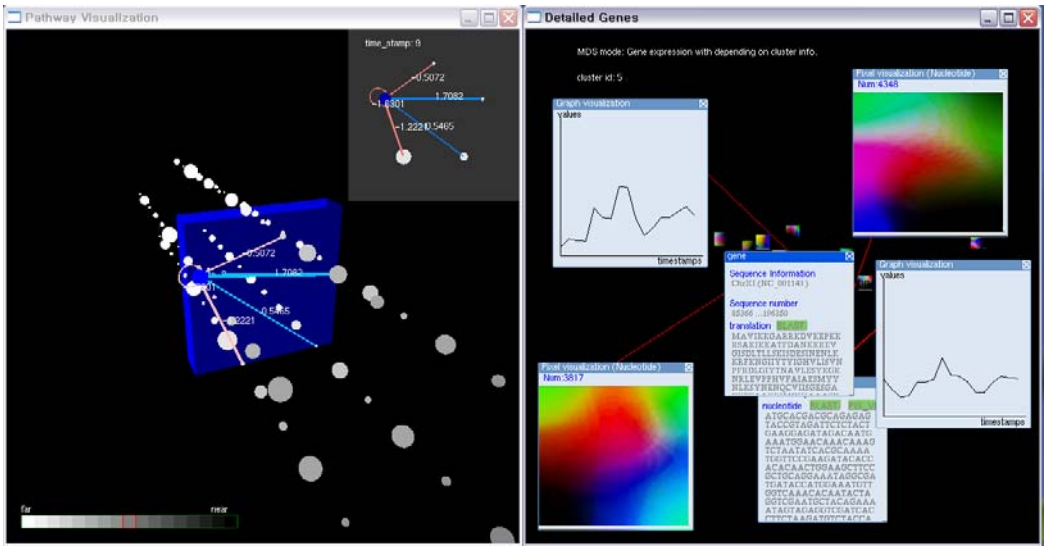


Figure 1: *The overall layout of the visual analysis system, which consists of two main windows: one is for showing regulatory pathway information depending on time (left), and the other is for providing interactive analysis of gene information by displaying as pixel-based and line-graph visualizations and as textual annotations (right).*

The rest of the paper is organized as follows. In Section 2, prior works concerning biological pathways visualization tools and existing visualization techniques are introduced. Section 3 gives the details of applied prediction model. In Section 4, several interactive visual analyzing techniques are briefly described. Section 5 provides conclusions and future work.

2. Prior Art

Several different applications supporting the capabilities of understanding of the functions of genes have been designed. Several different applications supporting these capabilities have been designed. Interactive exploration techniques have been proposed to interact with hierarchical clustered datasets [SS02]. To model and analyze biological pathways based on microarray data, KnowledgeEditor [TK03] was designed. With this tool, biological experts can model biomolecular network graphs and find molecular interactions. One of the broadly used applications for analyzing microarray data is GenMAPP [DSV*02]. It visualizes gene expression data on maps representing biological pathways and groupings of genes. Even though the majority of the applications have good features for analyzing pathway values, almost all such models and tools are developed for "static models" that cannot be used for the emerging field of microarray time-series analysis. In addition, due to the lack of suitable logical connections between the analytical tools used for analysis and the devised visualization tools, the use of these tools has not proved to be as biologically insightful as it could be. However, evaluation has been applied to provide the basic requirements for pathway visualization systems [SND05].

In information visualization, several interaction techniques have been designed for 2D or 3D display spaces. In our application, we have used several techniques such as Overview+detail [TSWS05], Focus+context [PCW01], and Pad++ [BH94]. We have previously found that those techniques are useful for visualizing genomic information in a richly annotated, multiscale display space [HJS*05].

3. Predicting Biological Pathways

We have applied the dynamic gene pathway prediction method we have developed [DHN04]. The prediction method is based on a unified pattern classification and Auto Regressive (AR) modeling. In order to analyze pathways with a very small number of genes, one can apply the AR model directly to all genes expression variables and therefore predict the expression value of each gene at time *t* based on expression values of all genes at time *t*-1. However, knowing that almost all biological pathways are composed of a large number of genes, this approach is not computationally feasible. In other words, the number of genes in a biological process is often so large that it is impossible to develop reliable AR models for the typically short time-series microarray data while directly incorporating all genes as individual AR variables. This is due to the fact that when the number of the variables (i.e. genes) increases the number of model parameters also increases, and therefore the estimation of many parameters with few time-series examples becomes unreliable and even infeasible in many cases. In addition, a blind application of AR models to molecular biology problems would not clearl y represent the insightful clustering effect of genes involved in a biological process, i.e. the model would fail to insightf ully display the massive grouping and parallelism in the ge netic network. To address these issues, we exploit the fact t hat many genes behave very similarly in the biological sens e and therefore the role and effects of these genes can be co mbined by a suitable clustering technique before dynamic modeling.

To implement this idea, we employ a preprocessing step that uses K-means clustering to groups the gene pool into a number of biologically meaningful clusters. Each of these gene clusters is represented by a prototype that reflects the

overall time trends of the cluster. After the preprocessing step, the AR model is applied on the prototypes of clusters as opposed to individual genes. Since the number of the clusters is small enough, the AR model can be reliably developed for the prototypes. The model relates the future expression level of the prototypes of gene clusters to the values of the prototypes in past time step(s). The model also considers uncertainty inherent to the model by considering a noise factor in the equations. In its most general form, the model is a linear system of difference equationsThe details of how to estimate and optimize the AR model have been provided in [DHN04].

To illustrate the method, we use a dataset containing the time-series expression values of almost 200 genes involved in the cell cycle of the budding yeast *S. Cerevisiae* [CCW*98]. The gene expression values were collected in 17 time points. It is known that there are five major phases in cell cycle development: Early G1phase, late G1 phase, S phase, G2 phase and M phase. In each phase only genes whose biological functions correspond to the changes occurring in that phase are active. Based on the known biologically-distinctive functions of these five phases, it is reasonable to cluster the genes into five clusters. The results in [DHN04] show that the resulting model can very accurately predict the expression values of almost all genes in the future steps.

The main application of this dynamic modeling method is twofold. First a dynamic regulatory network governing the quantitative interactions among the prototypes of the main biological trends can be obtained, i.e. the model discovers the effects of each gene group on itself and on other groups in time. Secondly, by using the resulting dynamic network, the expression level of each gene at time *t* can be predicted based on its expression level and expression level of other genes.

Even though the prediction method is designed to predict the signal patterns and dynamic interactions, understanding and analyzing the information predicted by the model may not be straightforward. In particular, concepts such as cluster prototypes, time steps, and excitatory/inhibitory interactions need to be effectively presented to users. In the next sections, we describe visual analysis techniques we have developed to address these needs.

4. Interactive Visual Analysis of Genes

As described above, biological pathways are clustered and the AR model is applied in order to predict the future expression levels. One approach is to then display the clustered information as objects in 3D space along with the interactions among the clusters [TSWS05]. However, this approach does not take time dependence into account and has drawbacks such as overlapped information, which can become highly cluttered in display space. To meet the need for detailed, specific information while avoiding clutter, our visualization framework has been designed with two additional views, a detail-view and an annotated view (Figure 1). The annotated view has a feature to show detailed information of selected clusters. The detail-view has several visual interaction techniques to support visual analysis, as discussed further in Section 4.2.

4.1. Visualization of Clustered Gene Expressions

Our approach is to use a 3D information visualization concept, as it provides enriched possibilities of integrating additional visual information into the representation space [TSWS05]. The clustered gene expressions are laid out in 3D space at successive time steps. (In the case of the *S. Cerevisia example above*, 17 temporal points are used.) Each cluster is represented as a 3D sphere. The perspective view permits the user to maintain an overall context in time while exploring dynamic interactions in detail. As shown in Figure 2, this layout is accompanied by a temporal selecting bar at the bottom left and an annotation window in the top right corner.

The clusters are colored along a gray scale from white (17^{th} time step) to dark gray (1^{st} time step). This provides a natural mapping between time and saturation that allows the user to identify positions in the temporal space while permitting color to be used for other information display. The radius of the cluster (3D sphere) represents the normalized mean value of the clustered signal pattern. Thus a bigger sphere represents stronger overall contributions of the cluster to the regulatory network at that time step.

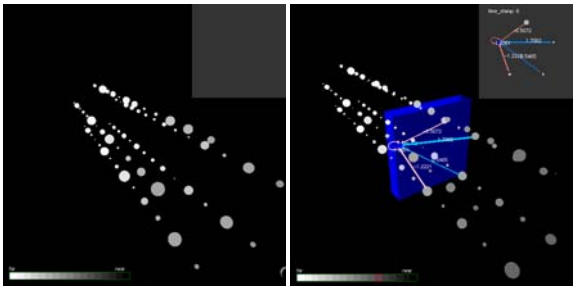


Figure 2: *Cluster representations in an initial layout (left) and a selected time step showing several interaction features such as transitional arrows, a different color coding technique and detailed annotations (right)*

Each cluster makes use of a color coding [SWTS05] to enhance visualization of cluster values. When selected, a cluster sphere changes color to blue. Arrows colored blue indicate positive transitional pathways between cluster spheres, and arrows colored red indicate negative transitions (Figure 2). That is, the arrows show control of one cluster over another with blue indicating enhancement of gene expression and red indicating suppression. Loops indicate control of the cluster on itself. With this simple palette, users can quickly discern important pathway differences and the state of the regulatory network over time. The user can also select individual arrows for quantitative information.

The temporal selecting bar is designed to support dynamic animation of cluster pathways over time. The selecting bar ranges over all time steps in the dataset. When the user makes a selection on the bar, a thin blue semi-transparent box appears at the selected time step (Figure 2). If the user drags the mouse along the selecting bar, the box moves accordingly along the time steps. Alternatively, the user can grab and drag the box directly. In either case an

animation showing changes over time is produced. In addition, selections can be made within the box (e.g., a cluster is selected and its arrows displayed). These selections are then maintained at later or earlier time steps as the box is moved around. By employing this animation capability, the user can quickly analyze and compare behavior in time with minimal effort. [RCM93, BMB00].

An 'Overview + detail' concept [HJS*05] is used for interactively dealing with the clutter than can result in the 3D view. The currently selected time step is shown in the 2D overview annotation window (with additional annotations).

4.2. Visualization of Individual Gene Expressions

The above method gives the user a quick overview of the dynamics of the gene regulatory network, but significantly more detail is needed on individual gene expression and function within the cluster. To demonstrate how this detail can be represented in a useful way, we develop and compare several methods such as Pixel-based visualization to represent gene expressions [WWFT03], Multi-dimensional Scaling (MDS) for positioning gene information [ARL01], and 'Focus + Context' for displaying the detail information [PCW01, HJS*05]. We then add interaction techniques to support interactive analysis and incorporate with sequence analysis tools, in this case BLAST [AGM*90], to provide a much more powerful correlative analysis capability.

4.3. Pixel-based gene representation

DNA sequences range in length from a few thousand to 20 million. In contrast to DNA sequences, gene (protein) sequences range about a few thousand in length. To manage data for a single protein sequence and provide contextual pattern information in a useful way, a pixel-based visualization technique [WWFT03] has been used. Through this technique, color glyphs that represent protein data are generated.

The pixel-based visualization technique we use here consists of several processes of arranging the pixel information and revealing the features of gene sequences. To map from gene sequences to pixels, space-filling methods are quite useful because they produce spatial patterns that have consistent locality, even for long gene sequences. Several methods of this type have been designed. Here we use a Hilbert curve ordering method to arrange sequence information mapping with color information, since it has the advantage of providing continuous curves while maintaining good locality of sequence information [WWFT03]. Figure 3 shows the basic Hilbert curve ordering. Whenever the order number (n) changes, the overall size of the mapped matrix will be different.

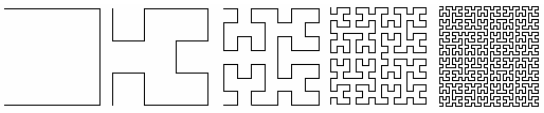
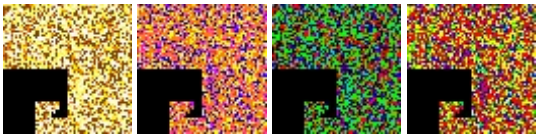


Figure 3. Order dependent Hilbert curve ordering which is always self-similar at 2^n x 2^n (n=1, 2, 3, …).

For mapping with gene sequences, we set the Hilbert curve order is set to 12 which cover 2^{12}x2^{12} sizes of gene sequences having the same sequential patterns. Color coding is then used to represent the sequence information.

DNA is a linear polymer made up of sequences of four nucleotides: adenine, guanine, cytosine, and thymine – designated A, G, C, and T. As discussed above, gene regulatory pathways can involve hundreds or more genes from which different proteins can be expressed. Hence, two different color mapping approaches have been made, one for the gene sequence and the other for the expressed protein.

For determining the color codes, four commonly used color maps proposed by other researchers are used. All images in Figure 4 are generated using the Hilbert curve ordering method. These show that it is often difficult to clearly represent patterns and features in the sequence. Therefore, a pixel enhancement technique and digital image-processing filters are applied [WWFT03]. First, a Gaussian filter is used to smooth the high-frequency values. And then Histogram equalization is applied to modify the dynamic range and contrast of an image depending on color channels (for example, R, G, and B channels). Finally, saturation values are increased using extrapolation as a saturation adjustment technique.



a b c d

Figure 4: *Pixel-based gene (SWItching deficient (SWI4)) visualization having several different color codings; (a) A (white), C (yellow), G (orange), T (dark brown) [WWFT03], (b) A (orange), C (blue), G (purple), T (yellow) [SGM*05], (c) A (green), C (blue), G (black), T (red) [ECR97],and (d) A (red), C (blue), G (green), T (yellow) [RJSS00]. Black space located in left bottom of each image represents the empty part of the space-filling curve.*

Even though all pixel-enhanced images have similar results (Figure 5) with respect to the patterns in terms of R, G, B channels, the pixel representations in the 3rd column of Table 1, using color codes (A (green), C (blue), G (black), T (red)), show consistent patterns in all 3 channels. Furthermore, merging all channels and applying saturation adjustment (Figures 5e and 5f) show in clearest detail which area of the image has denser information for adenine, guanine, cytosine, or thymine, and what its shape is. We thus apply the result of Figure 5f in the following. To arrange the gene images in 2D space, we must next apply Multi-dimensional Scaling, as discussed next.

It is important to note that this is one, rather simple, way to construct gene images. More sophisticated methods, which might take into account sequence structure and meaning, could be applied. However, the pixel mapping would be the same.

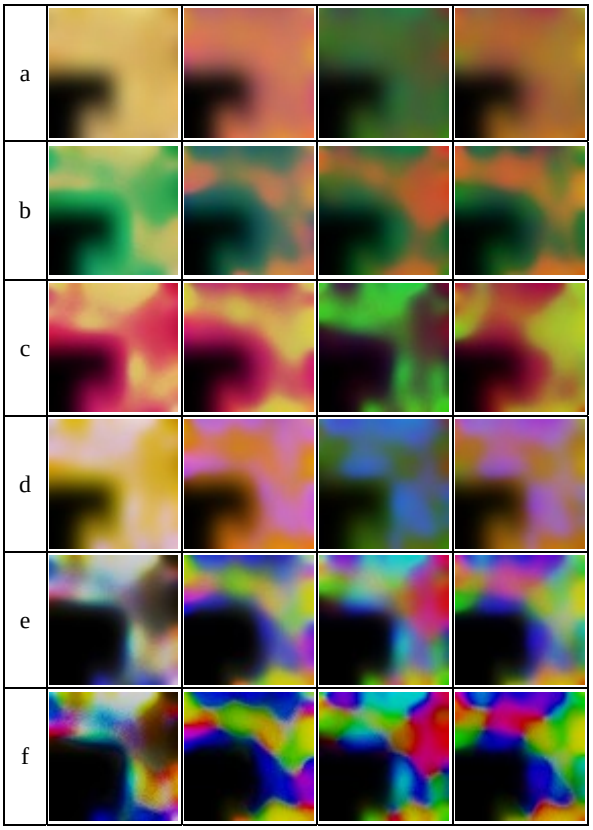


Figure 5: *Rows represent several image processing filters and a pixel enhancement method are applied to pixel maps shown in Figure 4 (represented by columns). (a) Gaussian filtering, (b) Histogram equalization on red channel, (c) Histogram equalization on green channel, (d) Histogram equalization on blue channel, (e) Merged all color channels, (f) Saturation adjustment is applied with the saturation value of 2.5*

4.4. Data positioning in 2D space

Multi-dimensional scaling (MDS) is one of the most broadly used methods to attack high dimensional data and represent them in a lower dimensional space. We apply MDS to display the collection of gene sequences in 2D space. Since the time series data are organized in clusters, we can apply MDS to these data and the gene sequences with or without clusters. In each case, the first step is generating a distance matrix depending on dimensional information. We opted for a simple approach that references the numbers of nucleotides when generating the distance matrix. First, gene sequences are counted in terms of adenine, guanine, cytosine, and thymine. Based on the counts, one of the most commonly used distance measurements, the Euclidean distance function, is used to measure gene similarity. Finally all gene sequences are mapped into 4 dimensional distance matrixes. After that MDS is applied to map the generated dimensions into 2D display space. More detailed measures of similarity could be applied. However, this approach is quick and gives a rough idea of contextual similarity. The user can then follow up with the interactive tools described below to get more details.

In Figure 6, two different glyph forms are used, the pixel-based representation of gene sequences and a simple line graph representing the DNA microarray time-series information. As described above, our visual analysis method is closely tied to the dynamic pathway prediction model. The line graph display for each gene, where the gene expression level is plotted along the vertical axis and the time steps along the horizontal axis, provides an option that emphasizes each gene's time history. Distances between genes are then measured according to the similarity of their line graphs. Hence, there are 4 possible ways to construct gene expression patterns, as indicated in Figure 6. It is seen that the pixel-based gene representation does not provide major differences depending on whether clustering is considered or not. But the graph-based representation shows that there are major differences between considering clustering or not, with the latter case producing highly clumped results. It is thus clear that clustering should be applied in this case because it will provide more diverse patterns.

This combination of approaches provides the user with a rich set of capabilities for exploring correlations from multiple perspectives over time. The interactive analysis methods discussed next enrich these capabilities.

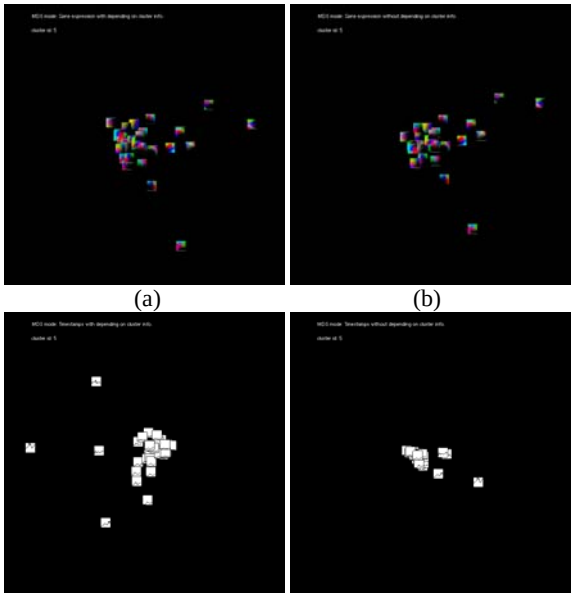


Figure 6: *Glyph representations of (a, b) pixel-based gene sequences and (c, d) graph-based DNA microarray time-series information with (left column) and without (right column) including the cluster information located in the fifth group.*

4.5. Interactive analysis

Interaction is a very useful exploration and discovery tool for large scale or complex data. In our visual analysis framework, we apply several interaction methods such as animation (described earlier), modified excentric labeling,

navigation (panning and zooming), scaling, arrangement, selection, and comparison.

To heighten the knowledge content of the glyph representations in Figure 6, we need to have a way of showing brief annotations for specific genes such as scientific name, etc. To display these annotations efficiently, an excentric labeling [FP99] option is used. In this method annotations are shown as the cursor is passed over glyphs. Although several different excentric labeling techniques have been proposed, there are drawbacks with respect to clutter. Therefore, we have designed a modified excentric labeling technique that includes a position shifting operation [CRMK95]. Manual position shifting is quite useful for dense information visualizations. For our visual framework, we designed an automatic shifting method. This reduces the user attention and effort needed in manual shifting and permits the high value annotation knowledge to flow unimpeded to the user. Our approach measures the size of the overlapped region between labels. With a strength based on this size, each label has a repulsive force against any overlapping label, so that the labels are pushed apart.

The excentric labels can contain a variety of information. Here the labels contain gene names and, in the case of the line graph representation, arrows indicating the point in the line graph corresponding to the current time step (Figure 7). This feature is quite useful for analyzing the microarray time-series details and relations.

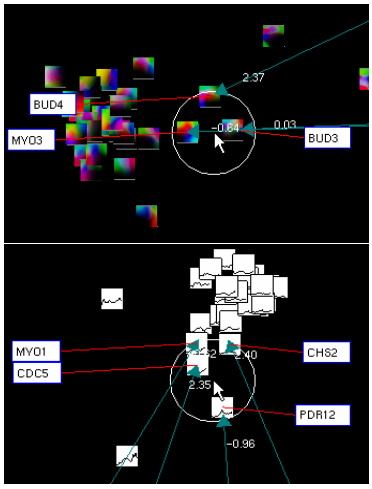


Figure 7: *A modified excentric labeling provides labels in a focus region. The focus region dynamically changes its location while the cursor moves over the display. Labels of the objects located in the focus region are updated smoothly and dynamically depending on the focus region.*

The modified excentric labeling also supports changing the size of the focus region centered on the cursor. This can be useful for showing labels when the scale of the glyphs is modified or when the user is interested in a larger or smaller region. The effective rearrangement and smooth transition caused by the force fields help the user maintain context and keep track of new annotations even under continuous cursor movement.

All these capabilities are incorporated in the detail view window (Figure 8). Zoom and pan navigation within the detail view window is designed based on the "Pad" [PF93] metaphor and its extension, Pad++ [BH94, FB95]. In this metaphor, the visual space is considered as an infinite 2D plane (called Pad), which can be stretched by orders of magnitude at any point to investigate details. The technique has been already applied in our genomic visualization system, GVis, [HJS*05] and was found to be an important capability for finding details and relations at all scales within a context of thousands of genomes.

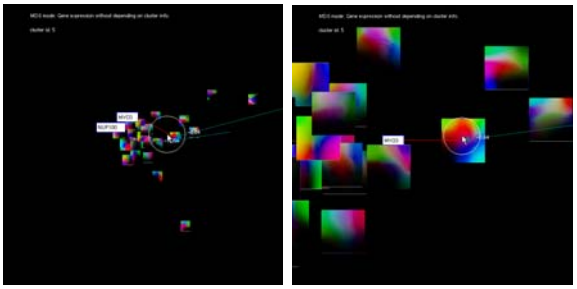


Figure 8: *Navigation in the zoomable space provides the capability for smoothly finding and examining interesting data objects in overview (left) or close-up (right).*

By using the Pad++ metaphor, a user can easily compare objects' patterns and find their differences over the whole object space and at multiple scales through successive pans and zooms. Figure 8 shows zoomed-out and zoomed-in navigation states. When the glyphs are in the zoomed-out state, we can find the overall arrangement of the gene expressions with respect to each other. Upon zooming in, we can take a close look at glyphs to find complete details in the gene expressions.

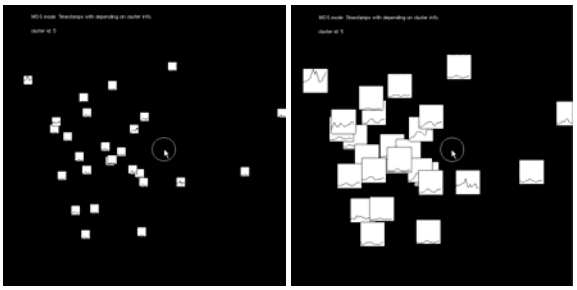


Figure 9. *Scaling makes it possible to see glyph details without disrupting their location information.*

In addition to navigation, scaling is also necessary since it shows the glyphs at larger or smaller scale without changing their positions or the viewpoint. Scaling is a broadly used technique to avoid or reduce overlaps, or to increase glyph sizes to show details. Figure 9 shows the smaller or larger glyphs without losing the overall pattern and positioning.

Although navigation and scaling often work well in avoiding overlaps in the display space, they are not perfect and additional tools are needed. To overcome this problem, we permit the reordering of overlapped glyphs in display space. Selected glyphs can be moved to the front or back with respect to the user's viewing perspective. In this way a clear view of the details of a glyph can always be obtained.

In addition, we use selection to show additional annotations about gene expressions in a "details on

demand" mode. A similar approach has been applied previously in our GVis system [HJS*05]. Since genomic data can sometimes contain many pieces of knowledge on sequence, active sites, expressed proteins, etc., we have found a modified selection technique to be quite useful.

Figure 10: *When selection is made, a pop-up window displays additional information about the selected glyph.*

Upon selection a small window will pop up with the text record of the item, including the standard name, the original DNA reference name, the sequence start and end, the gene's function, the proteins it creates, and so on (Figure 10). By panning or scrolling within the pop-up window, text of any size is accessible to the user. The windows align themselves dynamically in the 2D space, near the glyphs they represent, so that overlap is minimized as several boxes are opened. Each window can be positioned into another location by dragging it. It also can be scaled up or down to show the text information more clearly or to avoid overlaps.

To support analysis, the pop-up window has buttons for pixel-based visualization, line-graph visualization, and BLAST. The BLAST button is designed for launching the sequence analyzing tool called BLAST (the Basic Local Alignment Search Tool). When one of the buttons is clicked by the user, another pop-up window appears showing either the pixel-based display, line-graph display, or BLAST results (described further below).

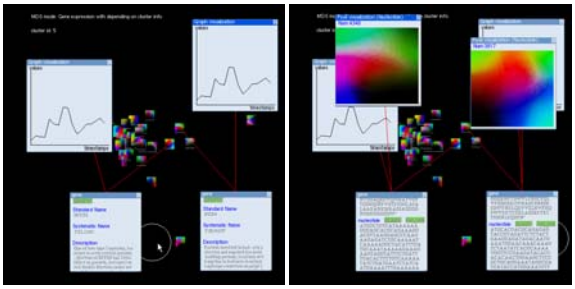


Figure 11: *Pixel-based gene representations and line-graph visualization are added into a second inner window.*

This set of buttons provides a fast and comprehensive view of the available gene sequence information such that, for example, if the gene collection is represented using the pixel-based glyphs, the detailed line-graph view is always available for any selected glyph. As shown in Figure 11, a richly detailed visualization results that permits close comparison of two or more gene sequences.

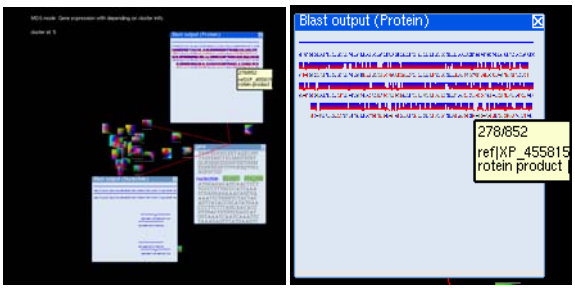


Figure 12: *The BLAST outputs are shown with protein and nucleotide sequences (left); a close-up of BLAST output shows similar sequences aligned for comparison (right).*

As mentioned above, our application incorporates a broadly used sequence matching tool, BLAST, which detects relationships among sequences that share only isolated regions of similarity [AGM*90]. The National Center for Biotechnology Information (NCBI) provides BLAST utilities such as network-based BLAST, stand-alone BLAST, etc. When the BLAST button is clicked, network-based BLAST is launched and searches for similar alignments in the NCBI database. Depending on the option selected, results are provided either in terms of alignments with the selected nucleotide sequence or its expressed proteins, which are then displayed in the pop-up window (Figure 12). The outputs are aligned in a graphical representation that permits the user to get overviews or comparing in detail by zooming out or in.

To complete the interactive analysis framework, we have incorporated the GVis system [HJS*05]. GVis supports in-depth study of the structure and function of the genomes, genes, and their expressed proteins involved in the dynamic regulatory pathways. This study is the follow-on to the fast exploratory analysis carried out with the above interactive tools. GVis provides this by automatically taking genes selected from the above analysis and allowing detailed comparison of their fully annotated genome environments with any other genome environment (which may be selected by iterative applications of BLAST or other comparative analysis tools). GVis is capable of permitting interactive exploration of tens of thousands (or more) of genomes from overviews down to the level of the annotated nucleotide sequences.

5. Conclusions and future works

In this paper, we designed an interactive analysis framework, with which the user can develop understanding of the dynamic regulatory pathways among genes by using visual analysis coupled with a prediction method. The framework implements a powerful integrated analysis that supports both understanding of the gene interactions over time and the understanding of gene function and structure itself (through comparative analysis). Several interactive analyzing features are provided such as time-based cluster visualization, pixel-based visualization, simple line-graph layouts, multi-layered navigation tools, and sequence analyzing features. With these features, the user can easily and quickly move among different perspectives to build an understanding of the time series structure, the gene interactions, their annotations, and their functions. This

integrated approach is quite important because it supports what the analyst must do anyway and, up to now, has had to do without an integrated set of tools.

Our future work is to extend the application to display correlative analyses of genomic data in a more comprehensive and effective way. We also plan to add another gene regulatory pathway prediction method [DNJR05] that uses nonlinear factor analysis methods to provide significantly improved identification of dynamic regulatory pathways. Finally, although our team includes bioinformaticists and the visual analysis framework has been developed, tested, and used with their help, we will offer the framework to a wider group of bioinformaticists for their use. As part of this, we are planning thorough evaluations and comparative testing.

References

[AGM*90] ALTSCHUL S. F., GISH W., MILLER W., MYERS E. W., LIPMAN D. J.: Basic local alignment search tool. J. Molecular Biology 215 (1990), 403-410.

[ARL01] AGRAFIOTIS D. K., RASSOKHIN D. N., LOBANOV V. S.: Multidimensional scaling and visualization of large molecular similarity tables. Journal of Computational Chemistry 22, 5 (2001), 488-500.

[BH94] BEDERSON B. B., HOLLAN J. D.: Pad++: A zooming graphical interface for exploring alternate interface physics. Proc. UIST ’94 (Nov. 1994), pp. 17-26.

[BMB00] BROWN J., MCGREGOR A., BRAUN H.-W.: Network performance visualization: insight through animation. In Proc. PAM2000 Passive and Active Measurement Workshop (Apr. 2000), pp. 33-41.

[CCW*98] CHO R. J., CAMPBELL M. J., WINZELER E. A., STEINMETZ L., CONWAY A., WODICKA L., WOLFSBERG T., GABRIELIAN A., LANDSMAN D., LOCKHART D., DAVIS R.: A genome-wide transcriptional analysis of the mitotic cell cycle. Molecular Cell 2, 1 (1998), 65-73.

[CRMK95] CHUAH M. C., ROTH S. F., MATTIS J., KOLOJEJCHICK J.: Sdm: malleable information graphics. In Proc. Information Visualization (INFOVIS ’95) (Oct. 1995), pp. 36-42.

[DHN04] DARVISH A., HAKIMZADEH R., NAJARIAN K.: Discovering dynamic regulatory pathway by applying an auto regressive model to time series dna microarray data. In Proc. 26th Annual International Conference of the IEEE/EMBS (2004) (Sept. 2004), pp. 2941-2944.

[DNJR05] DARVISH A., NAJARIAN K., JEONG D. H., RIBARSKY W.: System identification and nonlinear factor analysis for discovery and visualization of dynamic gene regulatory pathways. In IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (Nov. 2005), pp. 76-81.

[DSV*02] DAHLQUIST K. D., SALOMONIS N., VRANIZAN K., LAWLOR S. C., CONKLIN B. R.: Genmapp, a new tool for viewing and analyzing microarray data on biological pathways. Nature Genetics 31, 1 (2002), 19.20.

[ECR97] E. C. ROUCHKA R. MAZZARELLA D. J. S.: Computational detection of cpg islands in dna. In Technical Report (1997). Washington University, Department of Computer Science, WUCS-97-39.

[FB95] FURNAS G., BEDERSON B. B.: Scale space diagrams:

Understanding multiscale interfaces. In Proc. of Human Factors in Computing Systems (CHI’95) (May 1995), pp. 234-241.

[FP99] FEKETE J.-D., PLAISANT C.: Excentric labeling: Dynamic neighborhood labeling for data visualization. In Proc. of Human Factors in Computing Systems (CHI’99) (May 1999), pp. 512-519.

[HJS*05] HONG J., JEONG D. H., SHAW C. D., RIBARSKY W., BORODOVSKY M., SONG C.: Gvis: A scalable visualization framework for genomic data. In Eurographics / IEEE VGTC Symposium on Visualization (EuroVis 2005) (June 2005), pp. 191-198.

[PCW01] PIROLLI P., CARD S., WEGE M. V. D.: Visual information foraging in a focus + context visualization. In Proc. of Human Factors in Computing Systems (CHI 2001) (Apr. 2001), pp. 506-513.

[PF93] PERLIN K., FOX D.: Pad: An alternative approach to the computer interface. In Proc. of ACM SIGGRAPH ’93 (Aug. 1993), pp. 57-64.

[RCM93] ROBERTSON G., CARD S., MACKINLAY J. D.: Information visualization using 3d interactive animation. Comm. Of the ACM 36, 4 (1993), 57-71.

[RJSS00] REICHERT J., JABS A., SLICKERS P., SUHNEL J.: The imb jena image library of biological macromolecules. Nucleic Acids Research 28, 1 (2000), 246.249.

[Shn96] SHNEIDERMAN B.: The eyes have it: A task by data type taxonomy for information visualization. In Proc. of the IEEE Conference on Visual Languages (VL’96) (Sept. 1996), pp. 336-343.

[SND05] SARAIYA P., NORTH C., DUCA K.: Visualizing biological pathways: requirements analysis, systems evaluation and research agenda. Journal of Information Visualization 4, 3 (2005), 191-205.

[SPGM*05] SALES-PARDO M., GUIMERA R., MOREIRAA. A., WIDOM J., AMARAL L. A. N.: Mesoscopic modeling for nucleic acid chain dynamics. Physical Review E 71, 5 (2005). Art. Num. 051902.

[SS02] SEO J., SHNEIDERMAN B.: Interactively exploring hierarchical clustering results. IEEE Computer 35, 7 (2002), 80-86.

[SWTS05] SCHULZE-WOLLGAST P., TOMINSKI C., SCHUMANN H.: Enhancing visual exploration by appropriate color coding. In Proc. of International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG’05) (Jan. 2005), pp. 203-210.

[TK03] TOYODA T., KONAGAYA A.: Knowledgeeditor: a new tool for interactive modeling and analyzing biological pathways based on microarray data. Journal of Bioinformatics 19, 3 (2003), 433-434.

[TSWS05] TOMINSKI C., SCHULZE-WOLLGAST P., SCHUMANN H.: 3d information visualization for time dependent data on maps. In Proc. of the Ninth International Conference on Information Visualisation (IV’05) (July 2005), pp. 175-181.

[WWFT03] WONG P. C., WONG K. K., FOOTE H., THOMAS J.: Global visualization and alignments of whole bacterial genomes. IEEE Trans on Visualization and Computer Graphics 9, 3 (2003), 361.377.