On the way to building an integrated computational environment for the study of developmental patterns and genetic diseases

Andrei L Turinsky
Christoph W Sensen

Sun Center of Excellence for Visual Genomics, University of Calgary, Faculty of Medicine, Department of Biochemistry and Molecular Biology, Calgary, AB, Canada

Abstract: Genetic diseases and developmental patterns should be studied on several levels: from macroscale (organs and tissues) to nanoscale (cells, genes, proteins). Due to the overwhelming complexity of the life science data, it is common that disparate data pieces are meticulously stored but never fully analyzed or correlated. We have begun to develop a novel methodology based on virtual reality techniques for the study of these phenomena. Our key approach to knowledge integration is a top-down mapping of data onto visual contexts. For each organism that we want to study, a structural model is created and used as a visual "wireframe" onto which other data types are superimposed in a top-down assembly. Data analysis tools, visual controls, and queries are enabled so that users can interactively explore data. Our visualization technology gives users an opportunity to map disparate data onto a common model, and search visually for hitherto unknown patterns and correlations contained within the data. It is our goal to eventually transform genomics research from measuring various data pieces analytically into a fully interactive exploration of combined data in a 4D immersive visual environment that best matches a researcher's intuition.

Keywords: 4D bioinformatics, visualization, virtual reality, complex genetic diseases, CAVE automated virtual reality environment

Why 4D bioinformatics?

The diverse sets of large-scale data currently produced by genome research efforts require new data integration approaches to reveal their full potential. Meaningful integration of knowledge about life processes is frequently described as a main challenge for the bioinformatics community (Chicurel 2002; Stein 2002). In nanomedicine, the difficulties are often exacerbated by the inherent disparity between the nanoscale of the applied techniques and the macroscale of the eventual health effects. This formidable incongruity often requires the researchers to piece together available data across different scales of analysis, while learning to tolerate knowledge gaps on levels where data are unavailable. Bioinformatics technologies for gap-tolerant data integration will therefore greatly improve both the accuracy and the efficiency of nanomedical applications.

The focus of our work is to build the next generation of bioinformatics tools for the comprehensive multilevel studies of genetic diseases and developmental patterns. This technology is based on advanced 4D (spatial and temporal) visualization. Our approach to the study of genetic diseases is to create fully integrated computer models of data available for characterization of diseases over time. This includes generic medical information, advanced imaging data, test results such as biopsies, gene expression data, and proteomics information. Using Java 3D^{TM}-based computer

Correspondence: Christoph W Sensen
Sun Center of Excellence for Visual Genomics, 3330 Hospital Drive NW, Calgary, AB, Canada, T2N 4N1
Tel +1 403 220 4301
Fax +1 403 210 9538
Email csensen@ucalgary.ca

models of the commonly studied organisms, such as human or mouse, the methods for 4D data visualization, data querying, and data mining will help to elucidate the nature of genetic diseases.

Learning from other environments, such as the study of airline disasters, it became clear that most successful reconstructions of very complex phenomena were done top-down, ie, the knowledge base was organized around a framework that outlined the context. For instance, in the case of a downed plane, a wireframe outlining the shape of the plane would be constructed initially, onto which all recovered pieces of the plane would be mounted in order to understand how the accident happened.

Unlike in the example above, we discovered that most of the gene expression studies and proteomics efforts (see Gavin et al 2002; Ho et al 2002) were working bottom-up, basically reconstructing the biology one gene or protein at a time. This also appears to be the predominant direction of research in the emerging area of systems biology. Most of the recent initiatives in systems biology are aimed at reverse-engineering complex biological processes from individual protein interactions, to pathways, to information networks, to cells, etc (Ideker et al 2001). Kitano (2002) likens systems biology to an airplane reconstruction, but the analogy proceeds to searching for possible local pairwise matches between all the listed pieces instead of mapping them onto a global top-down context (wireframe). While bottom-up approaches seem to be easier to initiate, eg, using abundant microarray gene expression data, they inherently do not work very well with knowledge gaps and suffer from massive combinatorial complexity. Several projects that use data mapping onto anatomical context have been initiated, such the Edinburgh Mouse Atlas Project (Davidson et al 1997) or the CYTOMER project (Wingender 2004). Our approach differs from these efforts in both its explicit adherence to the top-down visual data mapping and in its emphasis on the state-of-the-art 4D visualization.

In this article we review our experience in this exciting research area, and outline the strategic directions of our research. We have made an extensive use of the advanced computational infrastructure at the Sun Center of Excellence for Visual Genomics (University of Calgary, Canada), which includes the world's first Java 3D™-enabled Cave Automatic Virtual Reality Environment (CAVE®). Besides the unprecedented resolution and convenience for the users, the CAVE has been an indispensable testbed tool for the developers of 4D data analysis solutions. Whereas users can in principle be satisfied with a flat-screen display of the final 3D models, the developers must first ensure the validity of the biomedical models, from which the validity of 2D and semi-3D projections would follow. For example, flat-screen projections of 3D vascular structures suffer from the "spaghetti bowl" effect, where individual blood vessels are very hard to distinguish. However, validation of their accuracy can be done within seconds in a 3D CAVE. Therefore, regardless of subsequent usage of the 4D bioinformatics solutions, including those for nanomedical applications, their initial design and implementation will likely require an extensive utilization of virtual reality techniques. This article will reveal the key processes and methods that our Center uses for the 4D bioinformatics development.

Case study in 2D

The origins of many 4D visual techniques lie in 2D visualization. 2D bioinformatics is a well established area and a number of 2D approaches can be extended into the more advanced 3D and 4D settings. Our team has worked extensively on the automated analysis and annotation of genomic information in the context of 2D visualization. Our ongoing 2D projects often serve as important test cases in the development of novel bioinformatics visualization techniques extendable into 4D. We describe some of the relevant experiences below.

Over the last 10 years, we have built tools that can start with the DNA or protein sequence in plain text format, attach functional assignments to the respective regions of the sequence in question, and present the results in graphical and tabular formats. Our initial system, called MAGPIE (Multipurpose Automated Genome Project Investigation Environment) (Gaasterland and Sensen 1996; Turinsky et al 2005), was a static system that pre-computed all results, stored them on a file system, and served static hypertext-based webpages to the end-users (http://magpie.ucalgary.ca). It became clear over time that this system was not sufficient to fully explore the knowledge base. Therefore, we began to build a 2D data exploration system that connects to the MAGPIE knowledge base as well as other sources of genomic information in eXtensible Markup Language format (XML) and can display the information dynamically (Gordon and Sensen 2000; Gordon et al 2004). This system, which is called Bluejay, is now in use in more than 1000 laboratories around the world (http://bluejay.ucalgary.ca).

As a case study of top-down data integration, we recently developed a system of visual mapping between the Bluejay genomic browser and the Multi-expression Viewer (MeV)

toolkit for gene expression data analysis developed at The Institute for Genomic Research (TIGR; http://www.tigr.org/software/tm4/mev.html). This visual mapping system enables gene expression data from external sources to be visualised directly within the image of the DNA of an organism. For example, the user can examine groups of similarly expressed genes and their locations on the chromosome or watch a "movie" of gene activity over time, directly on the genome. The underlying mechanism manages hierarchical XML genomics data structures, and inserts new data as additional XML branches into the hierarchy. We are now extending this mechanism into 4D bioinformatics, with two major differences: (1) instead of the XML data, the system should manipulate Java 3D scene graphs, inserting visual nodes into the scene; and (2) instead of an organism's chromosome, the visual context is a 3D anatomical model, so the matching of the data pieces to their intended locations is much more complex.

The two systems mentioned above were major steps forward in the integration of knowledge related to genome research, but they fall short of enabling the understanding of more complex phenomena, such as gene expression and regulation, protein–protein interactions or protein modification in the cell. What all of these problems have in common is that they have spatial and temporal components, which cannot be resolved in a 2D approach. Two-dimensional systems typically also have very low resolution; in the common PowerPoint presentations often thousands of genes are represented in a 640 by 480 pixel image. Thus the Sun Center of Excellence for Visual Genomics began in 2001 to study what the capture of these data spaces would entail.

Towards 4D capabilities

Our progress on the path towards 4D capabilities can be subdivided into three general phases. In the first phase, we built the hardware environment, including the world's first Java 3D™-enabled CAVE automated virtual reality environment (Sensen 2002). This system can be programmed using the (almost) universal programming language Java, which is freely available as an open-source software package from Sun Microsystems (http://java.sun.com). Through the built-in *ConfiguredUniverse* package of utilities, the viewing environment and the program executable are separated, allowing for complete portability of the Java code, regardless of the viewing environment that was initially used for development. The CAVE was opened in February 2002 and has been used in numerous projects. Our system features three walls and one floor display, thus creating a cubed viewing environment measured in billions of voxels (three-dimensional pixels), a much higher resolution than the usual 2D displays (1–2 million pixels).

In the second phase, we have begun to create a middleware layer that allows us to view and explore datasets that were not initially meant for use in the CAVE environment. This system, which is called JABIRU (Java 3D Application Behavior Immersive Virtual Reality Utilities; Stromer et al 2005), allows the users to interact with the visual objects in a seamless manner, using tools such as the magic wand in the CAVE, or keyboard and mouse in a personal computer. Any Java 3D program for which the source code is available can be adopted for use with JABIRU, allowing us to utilize such systems as PDB file-viewing tools for the study of molecule structures in the CAVE without access to a keyboard or mouse.

In the third phase, we have now begun to build the equivalent of the "wireframe" described above for some of the models we are working with. Being located at the Faculty of Medicine, our main focus is the study of human diseases. Therefore, we have begun to build a Java 3D model of the human body. Initially, we tried to extract the anatomical features using the Visible Human dataset, as distributed by the National Library of Medicine at the National Institute of Health (NIH; http://www.nlm.nih.gov/research/visible/visible_human.html), but we found the results of automated

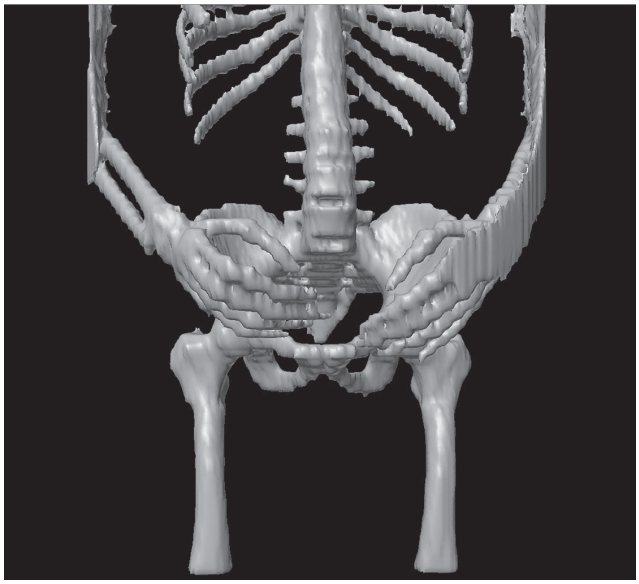

Figure 1 Attempt to automatically extract bone information from the Visible Human dataset (NIH) using the equidensity approach. It is apparent that the extraction process results in a very crude image, which would need a lot of manual interference to become useful.

extraction tools too crude to work with and therefore unusable for our purpose (Figure 1).

Learning from this, we are now building an anatomical model in collaboration with Kasterstener Inc from Red Deer, AB, Canada (Lajeunesse et al 2003) using images constructed from multiple sources, including anatomy text books and work in cadaver laboratories (Figure 2). The work is object oriented: each object (anatomical part) is rendered in extremely high resolution and with several image maps (coloration, bump map, shades). We expect to eventually obtain similar models for the major model organisms for medical research such as mice and rats.

Using the building blocks mentioned above as the basis, we want to expand our efforts to achieve the integration of all data relevant to a particular phenomenon (ie, a disease pattern or data related to development) into a coherent model. It is our goal to bring together the specialists from a number of disciplines and have them work together on the same model, which is visualized in virtual reality and which allows "natural" interactions in order to better understand the underlying biological and medical mechanisms of disease and development. We will now discuss the main research problems involved in the 4D research, using an example from liver tumor studies to illustrate the strategic importance of these problems.

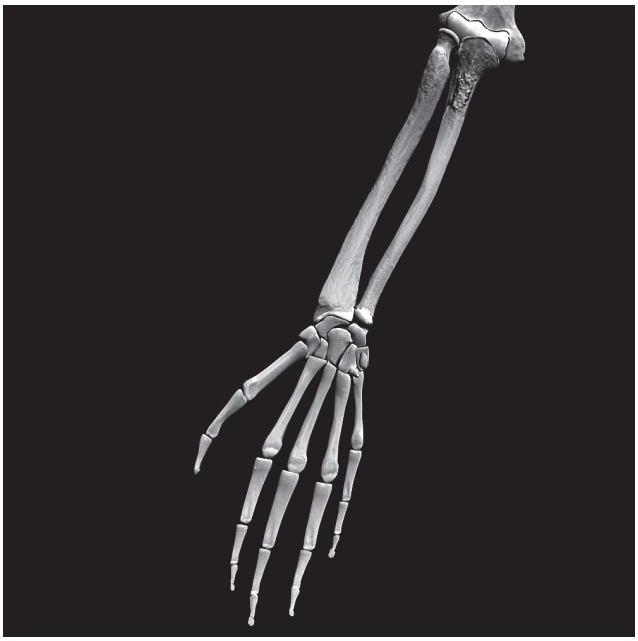

Figure 2 Hand and arm information from the Kasterstener virtual human body. The resolution of the objects shown is extremely high. In addition, each component shown is already a separate object, which is advantageous for the construction of the final model. The data are also anatomically correct, which allows the animation of the body parts in the future. The Kasterstener body data can be used directly as the input for the Java 3D™-enabled model.

Problem 1: Data integration

We have begun to build the computational environment that allows us to bring together diverse datasets such as information derived from advanced imaging experiments, gene expression studies, and medical information such as patient data records and clinical studies. Beginning with an anatomical model of the organism in question, we are building tools that can superimpose spatiotemporal information from diverse data sources.

The amount of data accumulated in both academia and industry, in genomics and other areas of biology, is growing very rapidly. The sheer size and complexity of the resulting data collections overwhelm domain experts, leading to what is known in data mining as "write-only data", or "data tombs" (Fayyad et al 2001). In the presence of overwhelming complexity, the key difficulty is the inability of humans to organize information in a meaningful way. From the technical standpoint, our task is to enable the loading of relevant data onto a combined 4D (spatiotemporal) visual model of the organism, which shows genetic and physiological changes over time. This will allow users to monitor the 4D image while being immersed in the model.

Our goal is to visualize genetic and physiological processes together, in a medium that naturally reflects the way life scientists think of living organisms. We are building an environment where anatomical and medical information is merged with data derived from advanced medical imaging techniques, such as functional magnetic resonance imaging (fMRI), computer tomography (CT and μCT data), and genomic information (gene expression, proteomics, and metabolomics data). Using the CAVE environment, researchers can interact with the data in an unprecedented spatial and temporal resolution, and interrogate it using complex queries.

As an illustration, the researcher will no longer have to comprehend disparate tables of numerical measurements that describe the genetic signature of liver cancer in a patient. Instead, scientists can simply "touch" the virtual liver model. The system should automatically access the required data and convert it into either static images of the gene expression in the biopsied 3D liver locales, or a "movie" of the gene expression changes over time.

Problem 2: Data reduction

We are currently using a mainframe computer with 20 CPUs and 80 gigabyte of main memory. Even with this very generous computational environment, it is not possible to load the entire Java 3D-enabled human body in full

resolution, let alone superimpose other information at the same time. We have begun to build smart technologies based on skeletonization into our tools (Cooper et al 2003) to reduce the amount of data handled at any time, while still maintaining as much information as possible on the objects directly visible to the user.

In our previous work, we have handled very large datasets without loss of performance. For example, we have skeletonized 3D volumetric images of bone samples into graph descriptions of the image so that image quantification could be performed on a much lighter data object (Cooper et al 2003). We generally base our approach on the development of a semantic level-of-detail (LOD) management to allow the handling of massive amounts of bioinformatics data. We are building semantic zoom capability into our tools (Loraine and Helt 2002), which automatically balances the breadth of the region of interest, or current image tile, with the depth of the level of details. For instance, in Bluejay, a tiled semantic zoom shows high-resolution details only when the user zooms into a small region of interest on the genome. The scalability of our method is based on the exploitation of hierarchical data structures in the visual model. Hierarchical structures, such as the ones built into the nested XML markup or Java 3D scene graph, are commonly used in bioinformatics data management. XML has enjoyed a wide acceptance as a de facto data standard (Achard et al 2001; Wang et al 2002; Orchard et al 2004). The underlying mechanism reduces very large data structures into much lighter versions of themselves, by dynamically pruning and restoring the lower, more detailed branches of the hierarchy.

To continue the liver cancer example, the researcher should be able to correctly perceive the general shape of the liver, the relative position of its tissue types, and the locations of biopsies performed. However, the full complex network of the minor vascular structures that permeates the liver, especially those away from the biopsies, should be left out of the initial view, and should become visible only on demand. The liver data should therefore be semantically organized based on their locations and levels of detail.

Problem 3: User interaction

One of the main problems in handling complex datasets is the usually steep learning curve for new users. It has long been our goal to provide the user with graphical environments that express the knowledge through imaging already familiar to the user from other experiences. Examples include the graphical display of genome maps, or the mapping of additional information (for example, the tRNA genes with introns) onto the universal codon usage table. Similarly, we expect the "top-down" effort, when based on anatomical models, to resonate with the user community, especially with medical doctors and biologists.

Given that many biologists have limited knowledge of computer technology, ease-of-use is one of the most important practical considerations in the design of bioinformatics tools. In the context of bioinformatics visualization, we have identified two key usability factors: fully visual data exploration and software portability. Fully visual exploration means that no configuration, command line parameters, or scripting are required by the end-user. Full portability means that Java 3D bioinformatics applications, commonly developed for 2D desktops, can be reconfigured at runtime for a required physical environment, such as an immersive virtual reality environment, using JABIRU (Stromer et al 2005).

In the liver example, the users should be able to map their genomics data onto the same data-ready 3D liver model from within any computational environment: either an immersive CAVE with a 3D joystick device, or a mouse-and-keyboard UNIX client as part of the Laboratory Information Management System (LIMS), or even on a personal laptop at home. In particular, scientists can use and evaluate the same 4D visualization tools on common mouse-and-keyboard desktops, without the need to visit the CAVE. This offers a tremendous improvement over other existing approaches to virtual reality development, in which tools are commonly developed for only one device setup and require a recompilation and/or code adjustment for another setup, hindering desired exchange of methods and ideas between researchers.

Problem 4: Modeling

It is our goal to create true models, not just high-resolution graphical representations of static knowledge. Although the rendering of large 4D biomedical images in real time is a daunting task, it is just a small part of visual analysis. The main potential of the 4D visualization technologies lies in the ability to model biomedical phenomena. This involves (a) the creation of the structural image model per se, and (b) the enhancement of the model with additional functionalities for data analysis and human–computer interaction. The implications in the liver example are rather obvious: the anatomical model should contain explicit structure to be easily manipulated at several levels of detail and to be responsive to data processing queries via 3D interface.

The modeling relies heavily on computationally intensive data pre-processing. We are pursuing the 4D visual modeling in relation to both surface models and volumetric models. The original volumetric data and the extracted structural model complement each other, presenting two different ends of the cognitive spectrum to the user. On one hand, the original volumetric image represents real biomedical data and is therefore indispensable for biomedical quantitative studies, but it is often too imprecise, fuzzy, and noisy. On the other hand, the structural surface model represents processed data and is thus prone to processing errors, but it helps the user perceive the conceptual boundaries and mutual interrelations between the image parts.

The main challenge is that the original imaging data typically contains no explicit structure, such as 3D MRI or CT scans, which are essentially collections of 3D pixels (also called voxels) without any internal structure. The structure therefore must be extracted using image processing algorithms. It is achieved typically by a combination of such methods as image segmentation, thresholding, and skeletonization (Jain and Farrokhina 1991; Golland et al 1999; Selle et al 2001). All these techniques are computationally intensive, and their accuracy depends considerably on the algorithm used and requires expert knowledge. We are therefore beginning to use the high-resolution surface-based atlas approach (Toga and Thompson 2001). When the new fuzzy volumetric images are presented, the data will be registered and warped onto the atlas, instead of being individually processed. This approach requires investment of resources into the atlas creation, but is less error prone and allows for better standardization of tasks.

We are also equipping our models with data-query behaviors. Up to now, 4D human–computer interactions were often custom-designed for specific applications. We have previously implemented some of these functionalities in a generic way in (Quon et al 2003), where 3D visual slider objects were used to select the range of RNA sequences for phylogenetic analysis, and also to implement the 3D zooming mechanism. We are also building fully portable interactive behaviors in our JABIRU project. JABIRU also allows users to enhance pre-existing Java 3D applications at runtime with additional behaviors. Further enhancements will include visual query interface for interactive image editing and annotation; object retrieval by image features, semantic name, physical measurements, or spatiotemporal arrangement; and interactive classification (Gupta and Jain 1997).

Problem 5: Automated discovery

While we expect that many discoveries can be made through the interaction of the users with the integrated models, it should be clear that much will remain undiscovered unless data mining approaches are applied to unearth the hidden connectivity among the diverse datasets (Jain et al 2000). We will build automated agents that can mine the spatiotemporal knowledge base and present the user with suggestions for additional relationships within the dataset. Users will be able to verify the suggestions and capture the newly discovered facts. This process will be iterative, as many of the discoveries will depend on previous knowledge.

We advocate an evolutionary approach to mining the spatiotemporal models: once sufficient understanding of the common research queries is gained, new automated features are gradually introduced. This is done by implementing automated steps to optimize common tasks, with human experts, not the software itself, remaining the ultimate data explorers.

As an example, let us consider the task of spatiotemporal data exploration in the context of liver tumor research. Once the relative position of the tumor is viewed and understood, the user explores the implications of the shape and location of the tumor to determine its type and severity. This subtask may be automated by a spatiotemporal image mining module that performs classification based on tumor location and shape dynamics (Koperski et al 1996; Ester et al 1997). Next, the user may wish to examine the genetic signature of the disease by exploring the gene expression data from biopsies. Given the complexity of genetics data, much of this exploration will be impossible without a readily available automated data mining module. Data mining tools such as clustering, support vector machine classification, or principal component analysis (Eisen et al 1998; Quackenbush 2001) will assist the user with the computationally intensive data-mining tasks. The expert then examines the suggested evidence from all aspects, and either accepts, adjusts, or rejects the automatically detected patterns.

Conclusion

We are using a completely integrated approach, combining advanced imaging with Genome Research information and other data, such as patient information, to better understand

the mechanisms of genetic diseases and development. Our main goal is to merge a large number of diverse data types into coherent disease models on a range of levels, from macroscale to nanoscale. Using the CAVE automated virtual reality environment, we will allow users to interact with 4D representations of the models in a natural way. This includes finding the techniques for runtime data mapping; developing a visual query mechanism that is fully portable across operating systems, visualization environments, and devices; developing level-of-details management and semantic zoom for the virtual reality systems; and enabling a range of advance pattern discovery tools.

Our focus at present is on the development of the technical capabilities. The 4D visualization technology allows the users to visualize data and models related to developmental patterns and complex genetic diseases, such as cancer, Alzheimer disease, or diabetes; observe interesting patterns; formulate new ideas and hypotheses, and test them with advanced data analysis tools. Much research needs to be done to achieve the goal, but the final outcome will be useful for the study of all complex genetic diseases and developmental patterns, and we expect new and exciting results. We expect to eventually create a generic system for the integration of life sciences data, which can be used in an even broader context, allowing the study of such phenomena as the growth and development of plants or the interactions in microbial communities. It will be exciting to use these new tools for the exploration of medical and biological questions.

Acknowledgments

This work has been supported by: Genome Canada through Genome Prairie's Integrated and Distributed Bioinformatics Platform Project; the Government of Canada and the Government of Alberta through the Western Economic Partnership Agreement; iCORE/Sun Microsystems Industrial Research Chair program; Alberta Science and Research Authority; Western Economic Diversification; Alberta Network for Proteomics Innovation; and the Canada Foundation for Innovation.

References

Achard F, Vaysseix G, Barillot E. 2001. XML, bioinformatics and data integration. *Bioinformatics*, 17:115–25.

Chicurel M. 2002. Bioinformatics: bringing it all together. *Nature*, 419:751, 753, 755.

Cooper DML, Turinsky AL, Sensen CW, et al. 2003. Quantitative 3D analysis of the canal network in cortical bone by micro-computed tomography. *Anat Rec*, 274:169–79.

Davidson D, Bard J, Brune R, et al. 1997. The mouse atlas and graphical gene-expression database. *Semin Cell Dev Biol*, 8:509–17.

Eisen MB, Spellman PT, Brown PO, et al. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*, 95:14863–8.

Ester M, Kriegel H, Sander J. 1997. Spatial data mining: a database approach. *Proc 5th Intl Symp Spatial Databases*, 47–66.

Fayyad U, Grinstein GG, Wierse A. 2001. Information visualization in data mining and knowledge discovery. San Francisco: Morgan Kaufmann.

Gaasterland T, Sensen CW. 1996. Fully automated genome analysis that reflects user needs and preferences – a detailed introduction to the MAGPIE system architecture. *Biochimie*, 78:302–10.

Gavin AC, Bosche M, Krause R, et al. 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, 415:141–7.

Golland P, Grimson W, Kikinis R. 1999. Statistical shape analysis using fixed topology skeletons: Corpus callosum study. *Intl Conf Inf Process Med Imaging LNCS*, 1613:382–8.

Gordon P, Sensen CW. 2000. Bluejay: a browser for linear units in Java. In Pollard A, Mewhort DJK, Weaver DF (eds). High performance computing systems and applications. Kingston: Kluwer Academic Publ. p 183–94.

Gordon PMK, Stromer J, Turinsky AL, et al. 2004. Bluejay: a biologcial sequence browser featuring knowledge integration. *Proc Virt Conf Genomics Bioinformatics*, 3:4–11.

Gupta A, Jain R. 1997. Visual information retrieval. *CACM*, 40:70–9.

Ho Y, Gruhler A, Heilbut A, et al. 2002. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature*, 415:180–3.

Ideker T, Galitski T, Hood L. 2001. A new approach to decoding life: systems biology. *Annu Rev Genomics Hum Genet*, 2:343–72.

Jain AK, Duin P, Mao J. 2000. Statistical pattern recognition: a review. *IEEE Trans PAMI*, 22:4–37.

Jain AK, Farrokhina F. 1991. Unsupervised texture segmentation using Gabor filters. *Pattern Recognit*, 23:1167–86.

Kitano H. 2002. Systems biology: a brief overview. *Science*, 295:1662–4.

Koperski K, Adhikary J, Han J. 1996. Knowledge discovery in spatial databases: progress and challenges. In: Proceedings ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery; 1996 Jun; Montreal. p 55–70.

Lajeunesse D, Edwards C, Grosenick B. 2003. Realism: a study in human structural anatomy. Red Deer: Kasterstener Publ.

Loraine AE, Helt GA. 2002. Visualizing the genome: techniques for presenting human genome data and annotations. *BMC Bioinformatics*, 3(1):19.

Orchard S, Hermjakob H, Julian RK Jr, et al. 2004. Common interchange standards for proteomics data: Public availability of tools and schema. *Proteomics*, 4:490–1.

Quackenbush J. 2001. Computational analysis of microarray data. *Nat Rev*, 2:418–27.

Quon GT, Gordon P, Sensen CW. 2003. 4D Bioinformatics: a new look at the ribosome as an example. *IUBMB Life*, 55:279–83.

Selle D, Spindler W, Preim B, et al. 2001. Mathematical methods in medical imaging: analysis of vascular structures for liver surgery planning. In Engquist B, Schmid W (eds). Mathematics unlimited – 2001 and beyond. Berlin: Springer. p 1039–59.

Sensen CW. 2002. Using CAVE® technology for functional genomics studies. *Diabetes Technol Ther*, 4:867–71.

Stein L. 2002. Creating a bioinformatics nation. *Nature*, 417(6885): 119–20.

Stromer JN, Quon GT, Gordon PMK, et al. 2005. JABIRU: harnessing Java 3D behaviors for device and display portability. *IEEE Comput Graph Appl*, 25:70–80.

Toga AW, Thompson PM. 2001. Maps of the brain. *Anat Rec B New Anat*, 265:37–53.

Turinsky A, Gordon PMK, Xu E, et al. 2005. Genomic data representation through Images: The MAGPIE/Bluejay System. In Sensen CW (ed). Handbook of genome research. Weinheim: Wiley-VCH. p 383–414.

Wang L, Riethoven JJ, Robinson A. 2002. XEMBL: distributing EMBL data in ML format. *Bioinformatics*, 18:1147–8.

Wingender E. 2004. TRANSFAC, TRANSPATH and CYTOMER as starting points for an ontology of regulatory networks. *In Silico Bio*, 4:55–61.