BioNetLab: A National Resource for Integrative Information Sciences in Support of Research and Education in Spatial Genomics

Michael Doyle, Ph.D., Eolas Technologies Inc./American Registry for Pathology
Mark Holterman, M.D., Ph.D., University of Illinois at Chicago
William Lennon, Ph.D., LLNL and University of California at Davis
Reagan Moore, Ph.D., San Diego Supercomputer Center
Adrianne Noe, Ph.D., National Museum of Health and Medicine
Charles Paidas, M.D., Johns Hopkins Medical Institution
Maurice Pescitelli, Ph.D., University of Illinois at Chicago/Eolas Technologies Inc.
Mark Pullen, D.Sc., George Mason University
Kent Thornburg, Ph.D., Oregon Health and Sciences University

Summary

There is a critical national need for an approach to genomics that studies the detailed activity of genomes within the morphological contexts of the actual biological structures in which they operate, providing answers to the fundamental questions of the nature and dynamics of organism development without which the promise of genomics for health care cannot be realized. This new subdiscipline is called "spatial genomics." We herewith propose a major new effort that will support this emerging field.

We propose to create a national resource that combines a large-scale knowledgebase with advanced analytical tools in an innovative online collaborative environment to support and advance the science of spatial genomics. In this proposal we will expand our existing project, which established the capability to utilize and expand the knowledge contained in a biomedically important very-large-scale data collection in human embryology. Information about the project can be found at http://netlab.gmu.edu/visembryo (please note particularly the article from *Science*). The resulting applications have produced direct and immediate benefits in the areas of biomedical research, clinical management and education. This expanded national-resource project has the potential to greatly improve understanding of the dynamics of human development and also will serve as a model for the optimal use of other massive data sets in the biomedical sciences.

Introduction

Although existing technologies have enabled numerous advances in biomedical science and industry, there are several long-recognized but unsolved needs for tools and methodologies to facilitate the understanding of the phenotypic expression of genomes within their physiological and morphological contexts.

One urgent need is to gather gene expression data in a manner that supports the types of exploratory research that can take advantage of broad-spectrum types of biologic activity analysis, such as that enabled by today's microarray tools. Further, there is a serious need for methods to visualize, characterize and utilize information on the spatial distribution of the biologic activity of a wide range of genes, across a wide array of species and tissue types. There is a great need for technology to allow the collection of large volumes of these types of data, to enable exploratory investigations into patterns of biologic activity that may provide insights into both normal and abnormal biologic states. And there is certainly a need to correlate gene expression data with morphological structure information in a useful and easy to understand manner.

Each of these needs is evident across all species and ages, however there is a particular need for these problems to be solved in order to enable researchers to make significant progress in the study of early development. Many breakthroughs in biomedical science will only occur through study of organism growth and development. Deciphering the delicate interplay between the spatial expression patterns of various genes and the timings of these biological events is among the most difficult of biomedical research questions. In order to solve such problems, tools are needed to allow the collection of larger volumes of expression data across a wider spectrum of gene types than ever before.

A new field of study, called "spatial genomics," is emerging to address these needs. Workers in this new field use an integrative approach to draw upon the capabilities of a variety of information science subdisciplines, such as multidimensional biomedical imaging, scientific visualization, knowledge management, and online collaboration, to reconstruct and study the distribution and activity of gene expression within the spatial context of precise reconstructions of organism anatomy.

Since spatial genomic approaches capitalize upon such a wide range of tools and information sources, it is necessary to "prime the pump" for such an effort by first building a base of knowledge and computational capabilities that can form a platform upon which to build an integrated view of the spatial mappings of gene activity within any particular organism.

A team from our eight organizations currently is completing this initialization/priming phase for the presently-proposed project through the creation of a digital library based on the Carnegie Collection of embryos at the National Museum of Health and Medicine, featuring digitized two- and three-dimensional cell-level image reconstructions of early human development. Our existing project, the Digital Library for Human Embryology (often called the "Visible Embryo Project") is now in its fourth and final year of support by the National Library of Medicine as a Next Generation Internet (NGI) demonstration of collaborative visualization supporting gene expression research, distributed embryology education, and clinical activities. Our effort has achieved significant results in three broad areas:

- In **Collaboration Technology,** we have built software for synchronous and asynchronous annotation of embryos, group collaboration, and delivery of embryology (and other) education to distant students.
- In the **Visible Embryo Repository**, we have made accessible digitized embryos from the Carnegie Collection amounting to nearly a Terabyte of cell-level two- and three-dimensional imagery. A total of twenty embryos, seven of them fully annotated, will be available in high-resolution form to the biomedical research community and as webpages to high school and college biology classes.
- In our **Medical Applications** we have demonstrated the potential of the NGI to support a range from medical activities, from collaborative examination and annotation of high-resolution photo-microscopy, to three-dimensional embryo animations for medical education (25 produced), to comparison with ultrasound imagery for medical consultation and patient information.

As we near completion of a basic online repository with a range of embryos spanning the stages of development and a related products in research, education, and clinical activity, we have come to recognize the extremely high potential of this work for understanding techniques for preventing, diagnosing and treating not only prenatal and childhood diseases but also genetically-linked adult diseases such as cardiovascular disorders. As a result, we are seeking future avenues to pursue the rich possibilities of the repository and the collaboration technology we developed to construct and apply it. In this new phase, bioinformatics rather than networking/collaboration technology is at the forefront of our efforts; we therefore envision that leadership of the project would pass from George Mason University to the National Museum of Health and Medicine/American Registry of Pathology, where Dr. Michael Doyle would have a special appointment to serve as Principal Investigator.

Some of the most promising possibilities we see for the Visible Embryo repository are:

- research in embryo gene expression that can identify mechanisms offering treatment potential for aging organ systems, for example the cardiovascular
- patient and physician information in maternal fetal medicine that can increase doctor effectiveness and expectant mother confidence
- increased effectiveness in medical school embryology, a cornerstone course, by distributed education using cross-institutional teams; the same capability can support continuing medical education
- knowledge management and 3-D visualization activities for clinical, research, and education applications
- growth of the Visible Embryo digital library as needed to support these processes and many more that can be expected to be stimulated by its availability

Developing the ability to do these things is closely connected with advanced Internet, Web and Grid information technologies. Our eight organizations, drawn from both biomedical and information technology research groups across the United States, have worked together very effectively on the Visible Embryo project, sponsored by the Next Generation Internet program at the National Library of Medicine. This teamwork will continue to play an important role as we create the BioNetLab national resource for spatial genomics. As in the current project, the named organizations will lead efforts that incorporate other team members as needed to support biomedical and technical requirements of the work. The full scope of work described below will require an estimated 5 years to complete at a cost of roughly $1.5 million per year. All of our results and products will be available to the research and education communities through a centralized online resource center. While specific groups are identified with the various research areas described below, we also rely on the high level of synergy among the team members, with each group contributing in some measure to every research area. This unusually effective collaboration has served us very well in the current project and will continue to distinguish our team in the future.

The study of spatial genomics is an inherently integrative activity. In order to fully exploit the synergies enabled by this approach and to establish a national-resource center, an ongoing effort will be required to undertake a number of subprojects that will augment each other in order to form a larger whole.

For example, the morphological context which lies at the philosophical core of the discipline requires, by definition, a large base of data from which that context can be easily reconstructed. Such a data collection effort must be ongoing, and needs to draw from unique existing collections (see AFIP below) and new experimental laboratory techniques (see OHSU and Eolas below). In order to make emerging knowledge immediately relevant, approaches need to be developed to allow correlation with in-vivo data from living patients (see JHMI below). Canonical models of multidimensional morphology need to be created for each organism to be studied, so that disparate data sources can be correlated through mappings to homologies in those models. In order to make this possible, new nomenclature systems need to be developed to facilitate those mappings (see OHSU below).

All of these activities draw from a remarkably wide range of fields of knowledge. An unavoidable consequence of the multidisciplinary expertise necessary to work in the developing field of spatial genomics is that there is a distinct shortage of an existing body of experts who can collect, characterize, correlate and promulgate the various kinds of information dealt with by spatial genomics research. Collaboration and education, therefore must both be key components of the national resource center proposed here. Innovative technologies need to be exploited in order to tap into the small number of existing geographically-distributed experts and to allow them to work together effectively. Further, new educational approaches must be developed and widely deployed to allow the training of new populations of researchers and practitioners for the field. The BioNetLab project will provide an environment for biomedical informatics activities operating at a scale never before attempted. In order to make all of these

activities possible, the state-of-the art in data management, visualization, data analysis and communications will need to be exploited and expanded upon. Within the BioNetLab, SDSC specializes in technology for storing, indexing, and visualizing very large datasets, GMU specializes in networking and collaboration technologies, and Eolas specializes in the marriage of information technology with biomedical applications, knitting together the SDSC and GMU contributions into a seamless whole that meet the needs of spatial genomics.

These activities form key pieces of a larger whole that, when fully operational, will create a resource of knowledge, expertise, tools and capabilities that will act as a catalyst for the nation to rapidly take the lead in spatial genomics. An examination of the subprojects described in further detail below will show that each of these efforts will act as an important and integral component that will be essential in allowing the BioNetLab project to set new standards in the synthesis of knowledge to further an understanding of the development of biological form and function.

Anatomical Mapping of Gene Expression

Oregon Health Sciences University

Gene expression mapping is the key enabling technology in spatial genomics, and will be at the core of the effort to create this national resource. It is one of the most information-intensive biomedical research activities and also requires the collaborative expertise of a team of specialists. Our goal in this application area is to develop tools for generating anatomically accurate 3-D graphical models capable of displaying multi-gene expression datasets, using a format compatible with existing Visible Embryo voxel visualization technologies. We will focus on one anatomical area, the heart, to demonstrate the potential of our approach. The approach we propose has two key aspects:

(1) Create anatomical maps of gene expression patterns in the embryo heart: Using a collaborative, networked environment, we will generate high-resolution serial cross-sectional images from whole-embryo specimens in the chick and mouse with *in-situ* hybridization for cardiac-related proteins. Voxel datasets will be produced from these images. Dr. Scott Fraser of Caltech will be a key participant in this activity. Magnetic resonance microscopy (MRM) image data from Caltech will provide *in vivo* chick embryo heart morphology information based on magnetic staining (GFP) of vascular endothelium and cardiac myocytes. This will provide a means for obtaining accurate physical dimensions of the developing heart needed as canonical references and will serve as a prototypical data acquisition process for regional gene-expression data in embryos of other species.

(2) Symbolic model development for embryological anatomical nomenclature: Integrate symbolic model anatomical transformational abstraction (ATA) ontology into information resources developed under the Visible Embryo project. Dr. Rosse of the University of Washington will be a key participant in this activity, which includes review and revision of existing embryo annotations and providing domain expertise to Rosse when needed.

We view the ideal 3D embryo heart visualization as having these features:

- single-cell resolution (<10 um)of embryonic anatomical structures;
- voxel formats for subsurface, arbitrary plane visualization;
- full 3D morphing capability;
- micro-adjustable XY planes for registration correction;
- RGB and gray-scale capabilities;
- single/multiple voxel referencing for user control of color, brightness, and opacity;
- scalable ontology describing structural, temporal, systemic, and semantic data; and
- links to online gene, protein databases (NCBI, Genbank, etc).

Voxel imaging technologies created for the Visible Embryo project will serve as a foundation upon which gene expression visualization tools can be built. Our work in this project will use *in-situ* hybridization technology for labeling the genes of interest, with data imported into a voxel format for 3D viewing. More specifically, we aim to integrate voxelized embryo datasets with image data from our current efforts of labeling embryos with mouse anti-myosin followed by fluorescent conjugated anti-mouse antibodies to create a 3D description of localized cardiac-specific proteins.

Spatial Analysis of Genomic Activity

Eolas Technologies Inc.

Eolas' Spatial Analysis of Genomic Activity (SAGA) work is focused on providing new robotically-controlled systems capable of automatically generating spatial morphological maps of gene expression from any given biological tissue specimen.

A key feature of this system is that it will make it possible for a biological tissue specimen to be imaged in multiple dimensions to allow both morphological and functional reconstruction. The tissue specimen is physically sampled in a regular raster array, so that tissue samples are taken in a regular multidimensional matrix pattern across each of the dimensions of the tissue specimen. Each sample is isolated and coded so that it can later be correlated to the specific multidimensional raster array coordinates, thereby providing a correlation with the sample's original pre-sampling morphological location in the tissue specimen. Each tissue sample isolate is then analyzed with broad-spectrum biological activity methods, using tools such as microarrays, providing information on a multitude of biologic functional characteristics for that sample. The resultant raster-based biological characteristic data may then be spatially mapped onto the original multidimensional morphological matrix of image data. Various types of analysis may then be performed on the resultant correlated multidimensional spatial datasets.

Just as Celera's "shotgun sequencing" technology made the laborious process of gene sequencing scalable enough to rapidly generate huge amounts of sequence data in a short period of time, the SAGA process will enable the creation of massive amounts of spatial genomic data through the deployment of many SAGA systems that can run in parallel to reconstruct and mapping many three-dimensional specimens simultaneously. This will allow the project to rapidly populate the BioNetLab knowledgebase with information that integrates both form and function in the embryological context.

Collection Asset Management and Web-Based Access Tools

National Museum of Health and Medicine

The BioNetLab/Visible Embryo Project coalesces around the Carnegie Human Embryology Collection at the Human Development Anatomy Center (HDAC), which is a component of the National Museum of Health and Medicine (NMHM). The influence of the Carnegie Collection, HDAC, and NMHM to continue to exert a strong positive effect on the networked community we will build. The Human Developmental Anatomy Center will continue to maintain and manage the collection and related materials, as well as function as an access point for both the database and for animations and images form other sources by acquiring and housing the images and animations and/or by providing descriptions, thumbnails and links.

The HDAC website and NMHM are well known. They are a logical place for educators, students and researchers to seek data. The HDAC website will become more user-friendly for this purpose. We will fully integrate the Visible Embryo repository into the local search engine so researchers, when searching the database, will be searching across the whole site accessing all the collateral information available for a particular specimen or stage. This will include incorporation of partners' tools as they are developed, particularly knowledge management systems. We also will provide increased historical information contained in the collection on the website, including details as to which specimens have had prior study

and what those studies were. Tailored access levels will enable rapid, secured, and selective access to the database.

To expand access to the existing Visible Embryo Project, and to evolve the project towards the BioNetLab national resource for spatial genomics, we will develop additional web-based tools. They will facilitate tours of the embryology information on line, guided by a human "digital docent" using educational and conferencing tools developed by the project. Groups of students will be able to tour NMHM and HDAC and receive knowledgeable answers to questions without leaving their classrooms. Our tour will complement the federal and state curricula for science literacy. We will make available materials for pre-visit classroom discussion that will facilitate enhanced-value educational experiences. Our intention is to demonstrate the viability of the approach, which has great potential.

Patient and Physician Information in Maternal Fetal Medicine

Johns Hopkins Medical Institution

Currently, first trimester prenatal diagnosis is limited to chromosomal studies and biochemical analyses. Throughout the first trimester, clinicians obtain measurements (greatest length, crown rump length) but there are no morphological correlations with the developmental stage of the embryo. Our aim is to establish and validate a morphological database utilizing multimodal imaging techniques (magnetic resonance microscopy (MRM) and serial microscopy reconstructions) generated from the virtual collaboratory, to evaluate embryo ultrasonography. We have had excellent success in the current Visible Embryo project by demonstrating unequivocal comparisons of first Trimester images from weeks 6, 8 and 9 post fertilization with annotated Carnegie specimens. The MRM and serial histology images are supplied by AFIP and transmitted to SDSC as a repository of large digital files of images. This collaborative step is an essential primary step for generating multiple specimen comparisons with a series of embryo ultrasound images.

In addition to the AFIP and SDSC collaboration, AFIP will coordinate the collection of new high resolution MRM images of specimens from the Carnegie Collection by Dr. Tom Mareci at the Advanced Magnetic Resonance Imaging and Spectroscopy (AMRIS) facility of the University of Florida and others. By taking advantage of SDSC's abilities to morph and map multiple types of visual information into single coherent visualizations, a series of annotated ultrasound landmarks will be generated using these 3 image modalities - MRM and serial microscopy reconstructions, and ultrasound- through voxel-level visualization. Some of the landmarks are readily identified in the First Trimester and these include rhombencephalon, mandible, heart, stomach, cord insertion, limbs and bladder. Each of these tissue landmarks will be visualized, annotated at the voxel level on the 3-D visualizations, and validated using the non-mapped images in the database of the virtual collaboratory. The goal is to identify specific features in the ultrasound embryo images that are unique and specific for a particular Carnegie stage. These morphological landmarks will then be used as reference landmarks to assess the first trimester data because they are well known in the maternal fetal medicine literature.

Secured first trimester embryo ultrasonography data will be obtained from a timed sequence of ultrasounds performed by the Maternal Fetal Medicine Unit of The Johns Hopkins Hospital. Woman cared for by the In Vitro Fertilization Unit, Department of Obstetrics and Gynecology, at The Johns Hopkins Hospital will be selected as possible candidates for the First Trimester Ultrasound study; they are appropriate for inclusion because the precise day of conception must be established to manage these patients. A total of twelve subjects will be followed every other day through the eight post fertilization weeks when organogenesis is complete. Scanning will be performed using standard transvaginal techniques with high resolution probes. Each pregnancy will be followed to completion to ensure no untoward side effects of the scanning process, as well as to obtain knowledge about the status of the newborn. This portion of the application will be submitted to the Internal Review Board of the Johns Hopkins Hospital for approval. The American Society of Ultrasonography has sanctioned the use of first trimester ultrasonography. Dr. Paidas currently is certified by the Human Use Protocol Committee of the Johns Hopkins Medical Institution to submit and perform these studies.

The morphological database will be collated into a clinically relevant document incorporating the SDSC 3-D multi modal visualization techniques in the secure virtual collaboratory environment. It will provide valuable information for spatial genomics, to be integrated into the knowledgebase accessed by the BioNetLab. The multimedia formats and visualization tools will be usable over a wide range of network capacities, including consultation capability from hospital to home and hospital to hospital.

In addition to the clinical capabilities, the annotated visualizations of ultrasound landmarks will be used for medical education and embryological research. In coordination with UIC, an on-line course and CD-ROM will be assembled from the images and used as an educational tool for embryology education at both the basic science and clinical level. Another CD-ROM and on-line course accessible through the HDAC website, will be created for educating both technicians and medical staff, on the techniques of first trimester ultrasonography.

Medical School Embryology Distributed Education

University of Illinois, Chicago (UIC)

The efforts to create the BioNetLab large-scale knowledgebase on spatial genomics described above assume the existence of a body of medical professionals with a solid grasp of embryology, a fundamental course in the preparation of physicians. Comprehending embryology prepares the medical student for the complexities of anatomy, physiology and pathology. The developmental process is also a real life dynamic showcase of the interplay between biochemistry, genetics and cell biology. Unfortunately, many medical students do not receive an adequate preparation in embryology during medical school. Thus, many practicing physicians even in relevant fields such as pediatric cardiology do not understand concepts that would assist them in understanding the diseases they treat. With recent discoveries in fetal programming of adult disease and in gene expression patterns that will profile adult disease, knowledge of embryology is becoming increasingly important to the medical curriculum. The major reasons for the inadequate preparation of medical students in the principles of developmental biology are (1) shortage of allocated time in the medical school curriculum due to competition from larger courses; (2) the inherent difficulty of understanding embryology; and (3) the worldwide shortage of qualified embryology educators. We can improve these factors significantly: using networks, we will combine educational expertise and achieve a synergism in teaching effort; using animated computer graphics, we will improve student comprehension.

The developmental process begins with the fertilized ovum and ends with the fully formed fetus at the end of the first trimester. The journey from a single cell to the complex form of the human body at birth occurs rapidly and involves many organs and tissue types growing simultaneously. Visualizing these three dimensional changes over the fourth dimension of time is very difficult for the typical medical student to accomplish in the limited amount of time that he or she can devote to embryology. We have experienced considerable success in attacking this problem by developing a series of teaching tools using computer graphics animations and models that are designed to transmit the salient teaching points to the students. We developed the tools in a networked, collaborative environment and so came to understand the high potential of collaboration for addressing the shortage of skilled teachers of embryology. In this regard, embryology can form a model for many under-taught areas in medical education. Our project will show powerful ways to use networked computer technology that can be applied in many such areas, over a range of different network performance levels.

During the Visible Embryo Project we have developed a technique that allows us to prepare accurate surface models and animations of human embryology. This process involves the conversion of photomicrographs of sectioned human embryos from the database into registered digital images. Using the collaboratory annotation tools developed by the Visible Embryo Project, the annotations of structures within these micrographs are stored in the database. By using the multi-modal mapping capabilities of SDSC we can combine these annotations and images with other 3-D visualization data to create three

dimensional models of the organs and structures in perfect alignment with each other and within the context of the external shell of the embryo. The 3-D representations of the embryos and their components from each embryologic stage can then be arranged in time sequence and serve as templates for the animation process. Visualization experts from SDSC, artists from UIC, and embryologists from around the country will work together to accurately depict the developmental process. The resulting animations will provide a clean representation of the salient teaching points that previously were extremely difficult to visualize. Interactive 3-D fly-throughs of the models and visualizations will add a unique dimension to the final product.

This process results in compelling depictions of embryology. We have concentrated our efforts on animating changes in the thoracic region of embryo stages 12 through 18 and on the cardiac, digestive and respiratory systems. The resulting animations have been used in a limited number of medical schools and have met with enthusiasm by the educators and students alike. The most common request we received was to provide additional images. In the proposed new work we will complete fifty animations, approximately twice as many animations as have been completed in the current project, and will combine them with the detailed heart annotations from OHSU, the clinical ultrasound images from Johns Hopkins, and new annotated MRM volume visualizations to create truly unique set of educational tools. Selection of particular subjects will be based on priorities of the educational process.

We propose to combine the power of the network with that of computer graphics in a teaching and learning environment for embryology education that will illuminate the path to improving a wide range of topics in medical education. We have recruited experienced embryology educators at multiple institutions to create an Internet distance-learning embryology course that will be available to medical students across ten participating institutions. Each school will be responsible for the on-site mentoring of their respective students. Each educator will prepare one or two lectures, using the body of annotated embryos from the database, animations produced by UIC and incorporating 3-D fly through visualizations of genomics material from SDSC. These lectures will be provided over the Internet to participating students. Lectures will be recorded for delayed delivery to accommodate conflicting class schedules and for student review. The ability of each educator to concentrate on the development of one or two embryology lectures will allow them to concentrate on optimizing the educational content of their lectures.

After the medical school distributed courses are in place, we will prepare and administer a Continuing Medical Education (CME) Program with these lectures. Working with Johns Hopkins Medical Institutions we will create a series of lectures for ultrasound clinicians and technicians on the use and identification of Carnegie Landmarks in ultrasound examinations, in addition to providing lectures in embryology, This will be available over the Internet or via CDROM as a self-study-teaching module for which physicians can earn additional CME credits for their re-certification requirements

We also will publicize our results with the intention of continuing the project using the medical schools' internal resources and expanding to other institutions who provide their own support for participation. By building the collaborative distance education effort without providing financial support for participation, we will have demonstrated the economic viability of the distributed education approach.

Repository Technology

San Diego Supercomputer Center (SDSC)

The Visible Embryo Project image collection is managed as a digital library built upon a distributed data handling system. The digital library provides image management, data discovery, and image display services. SDSC currently provides data grid technology based upon the Storage Resource Broker (SRB) to support image ingestion and collection management. As we expand this system to support the BioNetLab national resource for spatial genomics, we propose extending this environment to support knowledge-based discovery on spatially annotated images, with advanced three-dimensional display technology for manipulating the images.

We will provide intelligent access and display of information within the Visible Embryo digital library for users having a wider range of network performance. The interactive network load imposed by a large user community can be much greater than the bandwidth needed to support the collection formation. To enable wider use of the image collection, advanced knowledge management systems are needed to facilitate discovery and guide the visualization process. This challenge has three major components: application of a knowledge management system for the spatially annotated and segmented images, integration of a knowledge discovery environment on top of the existing information and data management system, and integration of spatial manipulations of the data with 3-D display systems. These components are under development at SDSC through various federally funded projects; we will apply them to the Visible Embryo repository.

The SDSC 3D Visualization Toolkit can display 3D images of arbitrarily large data sets. The system has been integrated with the SRB for remote data access, supports paging of data from disk, and optimizes visualization operations on the data to drive 3D color fly-throughs. The system supports scalable display, from laptop-sized panel to large power walls. The software runs on single workstations as well as large clusters. The integration of the Model-based Mediation system with the 3D Visualization Toolkit will enable the display of segmented volumes and the integration of annotated surface representations with 3D data. SDSC will apply this approach to registered slices that have been segmented and annotated in the Virtual Collaboratory project. The combined system will be applied as the successor to visualization methods developed for the Visible Embryo project. It will support discovery by embryology concepts, access to the data through the SRB data grid, selection of a particular image segment, and 3D visualization of the segment for remote researchers with various levels of Internet access. Ultimately it will be possible to stream data from the repository through a client that runs in end-user workstations to produce and control the visualization.

Collaboration and Distributed Education Technology

George Mason University and Eolas Technologies Inc.

As the final annual demonstration in the Virtual Embryo Project, we demonstrated a “master class” in embryology, taught by eminent embryologists. That class required the resources of the NGI to deliver instruction to Portland, Oregon and Bethesda, Maryland from presenters in Chicago. However, GMU has shown that a very effective level of education can be distributed over much smaller network capacity, even that of a dialup modem. We propose to support the biomedical collaboration and teaching described above by bringing together a leading-edge suite of adaptive, scalable end-to-end networking technologies that can provide highly effective distributed education over available Internet-compatible networks, be they NGI/Access Grid, commodity Internet, dialup service, or wireless. The bioinformatics expertise of Eolas complements the multimedia networking skills of GMU, supporting development of a rich suite of easy-to-use tools for networked teaching and collaboration.

Our method will follow that of the open-source software community: obtain the best freely-available working code, enhance it as needed to support the required function, and then post it back on the website

for others to improve further. Already we have audiographic, video, and web browser tools working together seamlessly. Under this project, we will integrate a range of visualization and other applications for collaborative work in a framework that provides for adaptive use of whatever network capacity is available. We will seek always to maintain the simplest possible graphic user interface (GUI) so the result is accessible to users with lower levels of information technology sophistication (including medical doctors). Fine examples of this are the Visible Embryo annotator developed by Eolas and the Network EducationWare system developed by GMU to teach its own distributed classes and made available to the Visible Embryo project (and the rest of the world) as open source software. The interface to this software, which incorporates in one simple graphic the ability to test computer sound, adjust input and output levels, start and stop transmitting, and see the level of sound being transmitted. Working together, GMU and Eolas will continue to produce software that underpins the BioNetLab national resource for spatial genomics with highly effective tools for networked collaboration and education.

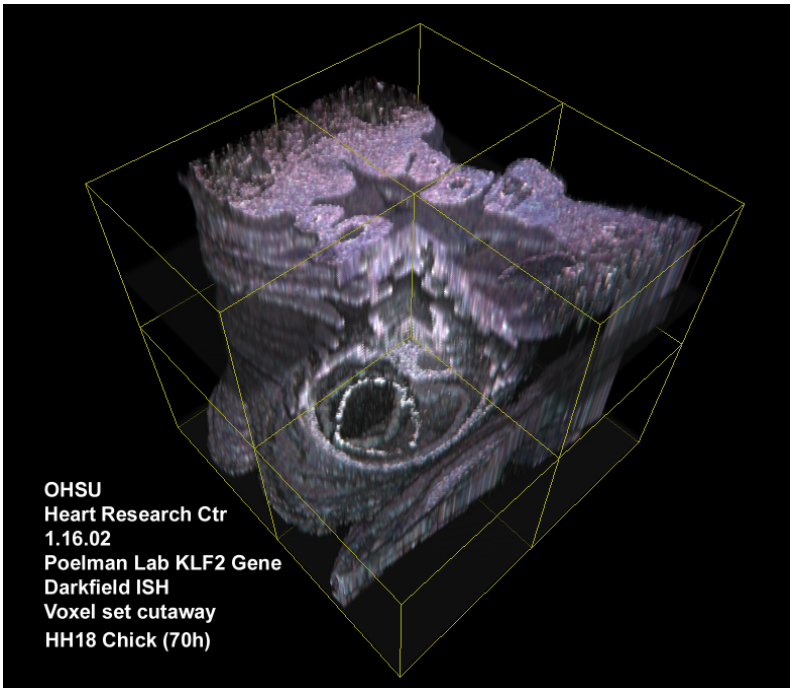
Resource Availability and Reliability Technology
Lawrence Livermore National Laboratory, GMU, Eolas Technologies and SDSC

An operative assumption that underlies our entire proposed effort is that the National Resource that we create will be both available and trustworthy: The developed tools must work reliably. Privacy must be understood and maintained. Information must be accessible only to authorized users with authenticated identities. Retrieved information must be accurate, properly logged, preserved and unaltered. The information must safe and cannot be lost. Finally, these characteristics must remain true as the Resource grows to meet new needs. In short, the system must be scalable in all respects.

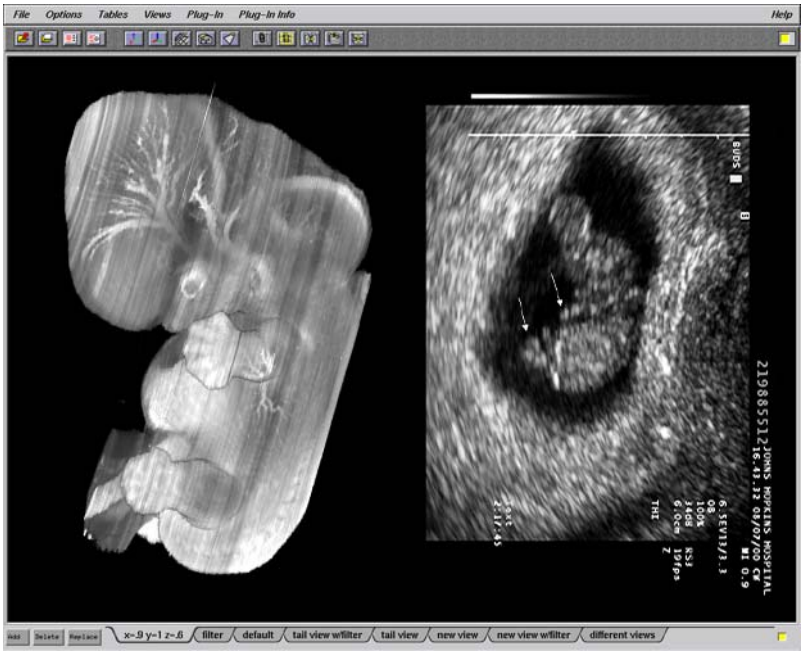
Considerable effort is required to satisfy these requirements. Procedures must be established early on to guarantee that the requisite capabilities are designed into all of the tools and operational procedures throughout the project. At present there are no agreed-upon metrics in the bio-informatics community to characterize the elements of and to predict the overall trust level of such a National Resource. We will develop innovative new metrics to fulfill these needs, and will adopt those that prove most effective. As we proceed, we will also track such developments by others and adopt those that prove most suitable to the task. There are, for example, well known existing techniques for assuring data integrity and security. These include emerging strategies for user-friendly authentication and authorization, proven mechanisms for data-integrity certification, and promising approaches for providing 24/7 availability of distributed databases and computation resources. We will both investigate and evaluate such existing methods, and develop new methodologies as needed.

When this project is complete, we need to be confident that the tools and resources will remain available with a high degree of reliability. In short, we will invest every effort to insure that the National Resource just works. When we have accomplished that goal, the system will then truly be equipped to meet its full potential to catalyze rapid advances in the field of spatial genomics, and to thereby address pressing and evolving needs for the nationwide biomedical research community.

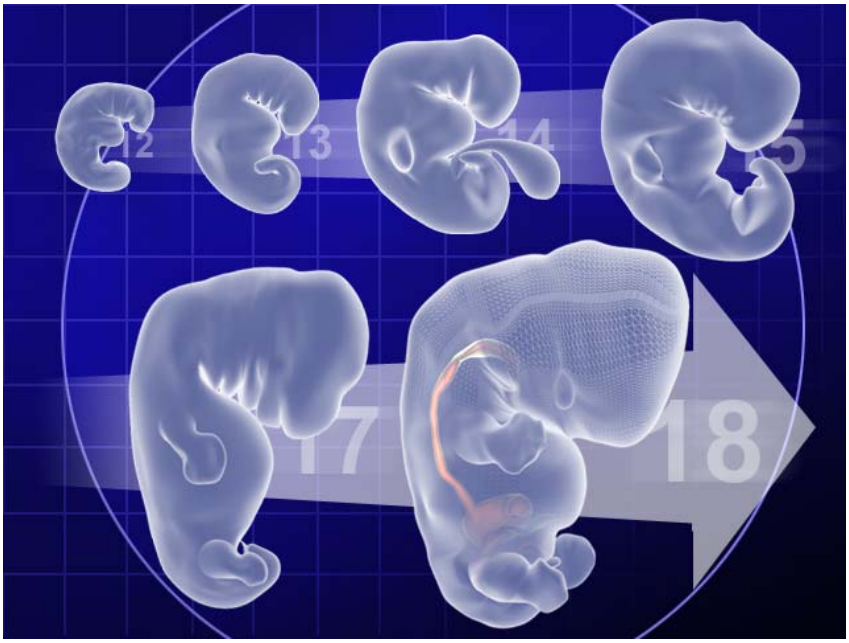
SOME PRODUCTS OF THE “VISIBLE EMBRYO” PROJECT



OHSU Gene Expression Visualization

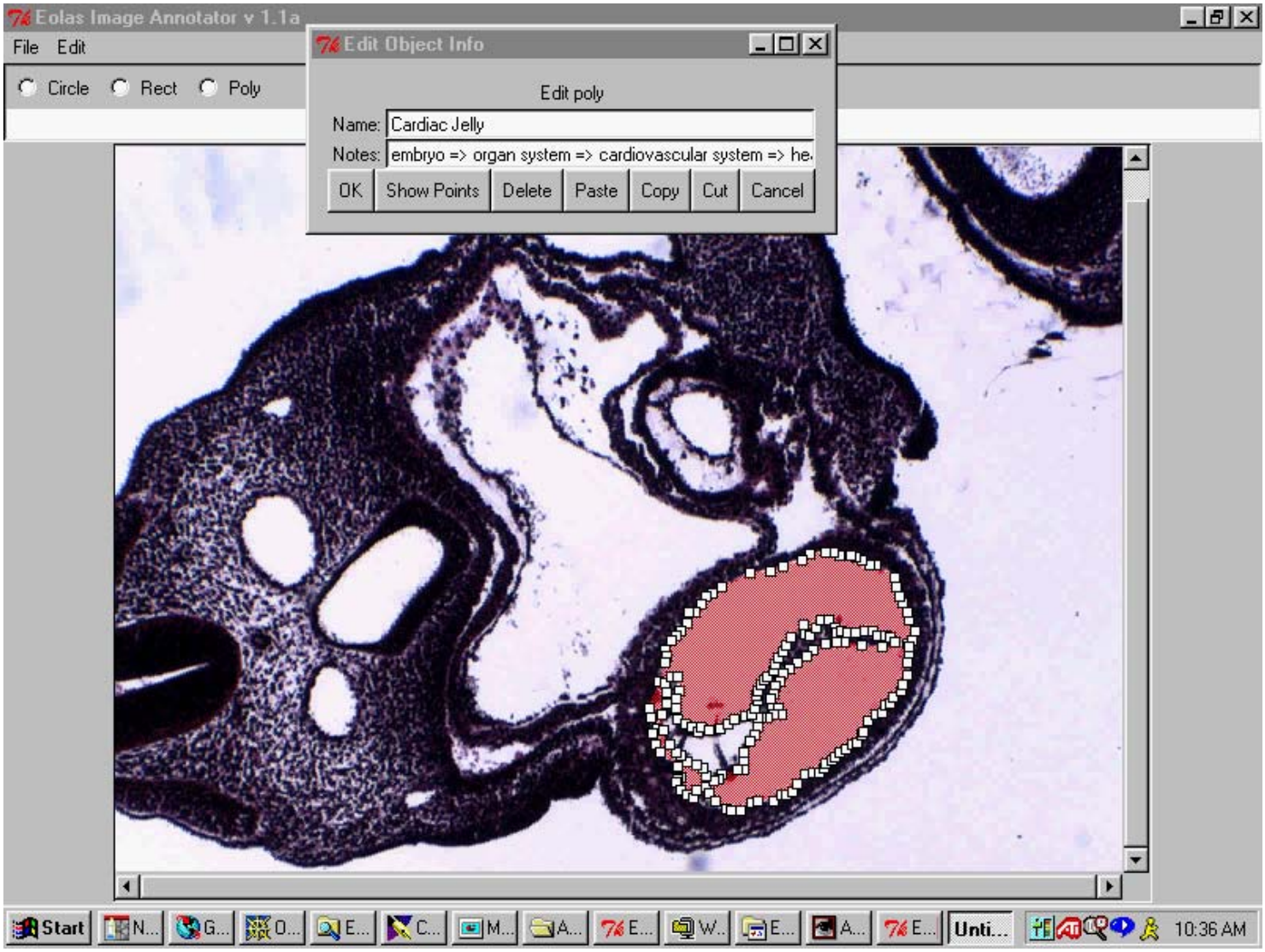


JHMI Carnegie and Ultrasound Rhombencephalon Saggital View Stage 18



UIC Embryo Animation Rendering

SDSC Embryo Fly-Through Rendering



Eolas Technologies Annotator