

Imaging Bioinformatics for Mapping Multidimensional Responses

B. Parvin, G. Cong, G. Fontenay, J. Han, H. Chang, and M.H. Barcellos-Hoff
Lawrence Berkeley National Laboratory
Berkeley, CA 94720
parvin@media.lbl.gov

The response of tissues and biological material to ionizing radiation is often heterogeneous and requires a significant amount of data for detailed characterization. These responses are often multidimensional in both space and time, can be imaged using digital microscopy, and have a spatio-temporal choreography that represents a number of challenging issues for gaining insight and knowledge representation. These challenges are considered along two dimensions of novel algorithms for quantitative analysis and imaging bioinformatics. Quantitative analysis enables detailed information on protein localization as a function of cell morphology and tissue architecture at each end point. The imaging bioinformatics provides the necessary schema for capturing experimental factors (e.g., model system, treatment), plate design, and sample preparation in terms of controlled vocabularies. However, informatics and quantitative analysis are tightly coupled together so that pertinent informations can cross-mapped and queried effectively.

With respect to detailed quantitative analysis, a number of algorithms have been developed for quantifying the DNA repair mechanism following ionizing radiation exposure. These techniques operate on model systems that are either grown in 2D or 3D (e.g., mammosphere). An important step in quantitative analysis is segmentation, and sample results are shown in Figure 1. With respect to the informatics system, its functional design is shown in Figure 2, and a staging system used for training and developmental research is located at <http://biosig2.lbl.gov:8443/biosigstruts> for reviewers' evaluation. Access to the staging system is unsecured, and clients are allowed to annotate, upload data, and visualize pertinent information. The backend database is open source PostGreSQL, which can easily be ported to any other platform.

The data model has evolved from its previous version to a new design that leverages emerging new standards in microscopy and experimental design. In this context, the data model is influenced by the Open Microscopy Environment (OME) and the MAGE model for managing microarray data. OME provides a syntax for tracking samples, instrument configuration (complete description of the optical light path), the type of analysis that has been performed on data, and, more importantly, a homogenized representation of image data in five-dimensional format. This representation has been extended with in-house software to read a variety of data types (e.g., tiff, stk) in multiple formats (8, 16, and 32 bits). A secure Web-based interface has been built to share image-based information, and to upload and download images and their feature-based representation. OME is an extensible data model providing a "semantic data type" structure with four levels of granularity: Global, Dataset, Image, and Features. Within this framework, a subset of MAGE (experimental designs and specifics of the assay development) is embedded into the Global semantic type. The MAGE model provides a concise definition of the experimental factors (e.g., cell-line, treatments) and protocols (e.g., plating, incubation time with a reagent at a specific concentration, number of washes, fixation, etc.). Furthermore, the use of controlled vocabularies from the NCICB database, the MAGE model, and in-house-specific terminologies facilitates uniform annotation of

database content. The in-house-specific terminologies are being annotated with Life Sciences Identifier (LSID) from IBM through a joint collaboration. Two views of the BioSig presentation manager are shown in Figure 3. The system consists of four modules:

1. *Resource Manager (RM)*: RM maintains a list of cells (their attributes and growth conditions), small molecules (e.g., siRNA), and imaging agents. Some of the resources are also coupled with external databases such as Hugo and Ensemble. These resources can be coupled with their location in a specific container and their PubMed indices.
2. *Experimental Design Manager (EDM)*: EDM provides the user with the necessary interfaces to define experimental variables and is tightly coupled with RM for enforcing controlled annotations. For an imaging experiment, the end result is a physical bioassay that can be laid out in a variety of plate format. Furthermore, while the plate layout maintains the static content of each well (e.g., cell line and treatment), the temporal activities of each plate (e.g., plating time, incubation time, number of treatments, harvest time) is specified through a protocol definition.
3. *Data loader Manager (LDM)*: LDM enables legacy data of different formats to be uploaded into the system from any workstation at any location. The system maintains two copies of uploaded data: (i) their original format, and (ii) a homogenized five-dimensional format that is supported by an extended OME image server.
4. *Visualization Manager (VM)*: VM facilitates image- and feature- based viewing. The image-based viewing allows raw and processed data to be visually examined for each set of experimental factors. The feature-based viewing allows plots and scatter diagrams that correspond to quantitative representation of images. This capability enables the end user to query the system for repair kinetics following ionizing radiation.

As a case study, the entire system is used to quantify kinetics of foci formed by phosphorylation of histone γ H2AX following ionizing radiation as a function of the number of experimental factors and computed profile. A live demonstration will conclude the presentation. A more detailed demonstration of quantitative tools and the informatics system will be continued during the poster session.

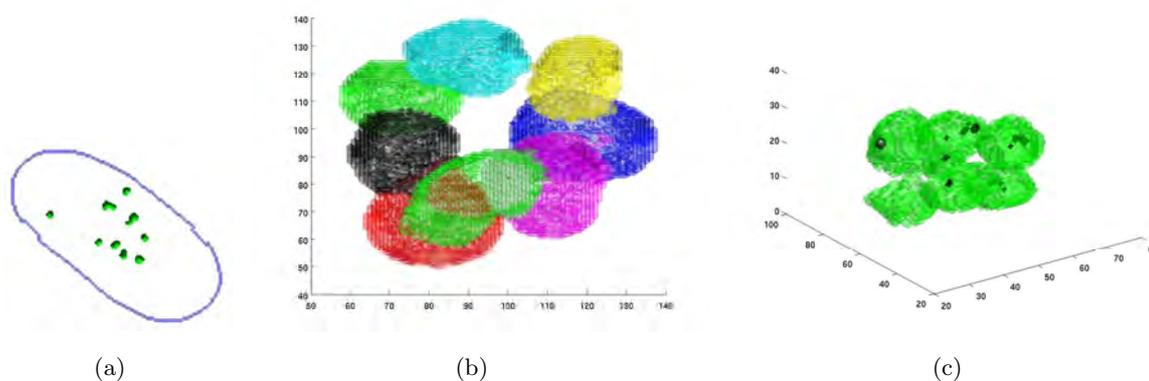


Figure 1: Visualization of foci in two model systems: (a) foci visualization in a monolayer; (b) delineation of each nuclear region in a mammosphere; and (c) foci formation in the mammosphere.

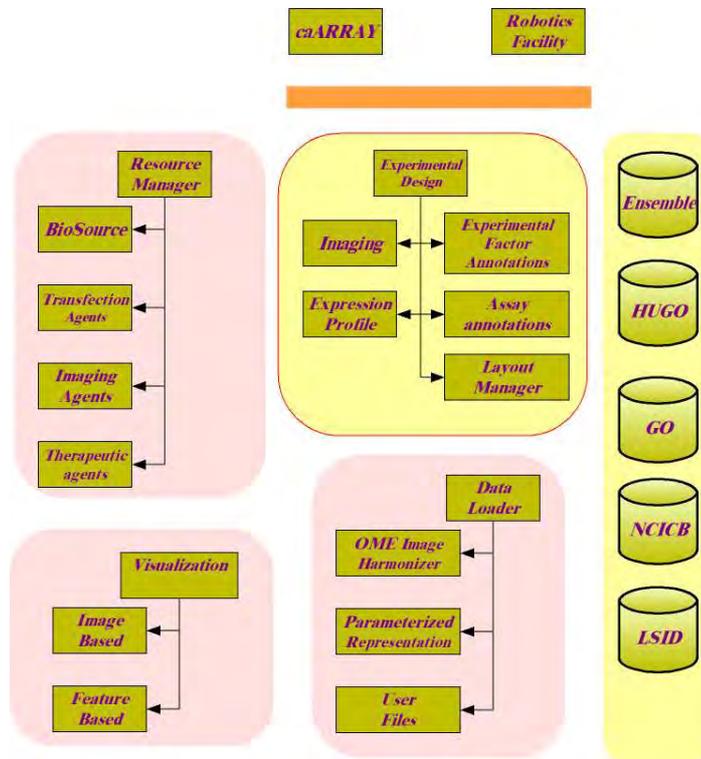
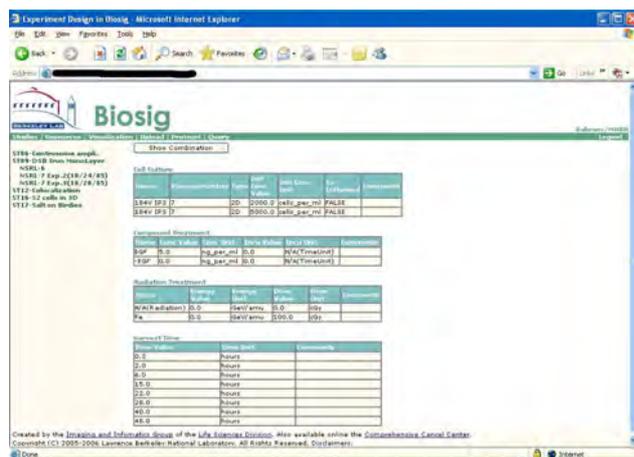
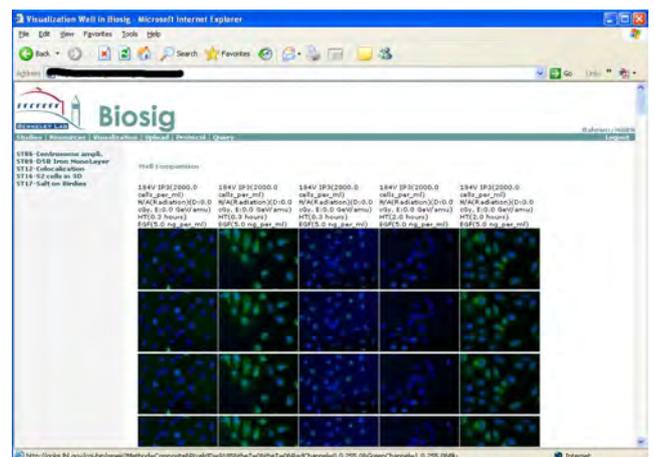


Figure 2: Software architecture for BioSig Imaging Bioinformatics Framework.



(a)



(b)

Figure 3: Two views of presentation layer: (a) a view of experimental design identifying biosource, compound treatments, type of exogenous treatment, and harvest time; (b) a view of images collected for this experiment. Each image is a thumbnail to a larger high-quality image that is also coupled to its quantitative representation.