RESEARCH SUMMARY
of
DAVID WILLIAM KNOWLES

Life Sciences Division,
Lawrence Berkeley National Laboratory
University of California
84R0171, 1 Cyclotron Road, Berkeley, CA 94720
510-486-6082
DWKnowles@lbl.gov
http://dwknowles.lbl.gov/

Research Interests:

My laboratory develops statistical, image analysis and visualization methods for unraveling basic mechanisms of biology. We are particularly interested in multicellular systems and assembling quantitative system-wide information derived at cellular and subcellular resolution. We are currently developing image-based techniques to study nuclear organization, gene expression, cellular interactions, epigenetics and morphology, and we are applying these to human mammary epithelial tissues as well as *Drosophila* embryos.

Our goal is to build computational, multi-feature atlases to provide quantitative morphological and macromolecular information at cellular resolution, and to use these atlases to reveal new biology about the multicellular system. Multidisciplinary collaborations are essential to our work and we believe there is an enormous and exciting future, in fact a mandate, for bioimaging research to increase basic understanding of biological processes and provide new knowledge and tools for the greater good.

Current Projects:

Two major projects in my lab are independently funded by the National Institutes of Health.

1) Image-based screening of mammary tumors:

The goal of this project is to develop methods for turning high resolution fluorescence images of human mammary epithelial tissue into tissue-maps which report the probablities of nonneoplastic, premalignant and malignant phenotypes at cellular resolution. The hypothesis is that malignant transformation of cells is accompanied by changes in gene expression which alter the nuclear organization of chromatin-related proteins. Such alterations will permit an epigenetic, imaging-based screening of nuclei, and provide macromolecular information about a tissue and enable the recognition of subtle differences in tissue morphology and behavior, and better detection of benign and malignant lesions.

Currently, histological classification of biopsied mammary tissue plays a determining role in the treatment decision. Unfortunately, the risk associated with benign or pre-invasive disease, the appropriate treatment or adjuvant treatment, and the risk of reoccurrence remain poorly understood. As a consequence, treatment decisions are based on epidemiological findings, which fail to address the complexity of the disease and thus the needs of individual patients. Our goal is to aid the treatment decision process of breast cancer patients by providing pathologists with new tools, capable of objectively and quantitatively determining premalignant and malignant lesions, to support current subjective histological classification of biopsied breast tissue.

To do this, we are developing methods where by epithelial tissue can be imaged in three dimensions, the nuclear organization of specific nuclear proteins analyzed on a cell-by-cell basis, and the results mapped into the tissue-context. For example, we have shown that the radial nuclear distribution of Nuclear Mitotic Apparatus Protein (NuMA) is a biomarker capable of distinguishing nonneoplastic and malignant cultured epithelial tissue models [1, 2]. Figure 1 demonstrates our ability to combine quantitative macromolecular information, on a per cell basis, with the gross tissue context.

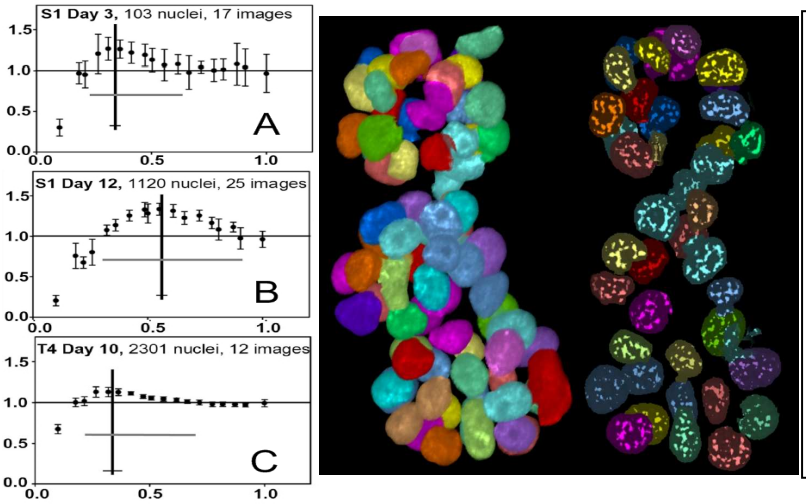

Figure 1. **Combining tissue morphology and measured nuclear distributions into a single 3D map**. We have developed techniques for measuring the radial nuclear distrbution of specific proteins on a per-cell basis. Panels A, B and C show clear differences in NuMA organization of normal early-proliferating, normal differienciated and malignant human mammary epithelial cells, respectively [1]. These results can be mapped back onto a model of the tissue to provide a novel multiscale visualized of macromolecular information and the tissue morphology in 3D, right panel.

Collaborators on this project include: 1) Associate Professor Sophie Lelièvre, Cancer Center Investigator, Oncological Sciences Center investigator, Basic Medical Sciences, Purdue University. 2) Associate Professor Sunil Badve, Department of Pathology, Indiana University School of Medicine. 3) James Sethian, Professor, Department of Mathematics, University California Berkeley, 4) Bernd Hamann, Professor, Computer Science and Visualization, University California Davis.

Dr. Knowles is a member of the Breast Oncology Program at the UCSF Comprehensive Cancer Center.

2) Morphology and Gene Expression Atlas of *Drosophila*:

The goal of this project is to create a computational, quantitative, cellular resolution atlas of morphology and gene expression for one of the most studied model animals: *Drosophila melanogaster*. The hypothesis is that cellular resolution, whole animal morphological maps will provide the most comprehensive repository for creating atlases of data including gene expression, signaling pathway, and subcellular localization. Such atlases will foster and allow unprecedented interrogation of biological events that range in complexity from the analysis of cell types and processes to the analysis of entire systems like the transcriptional network.

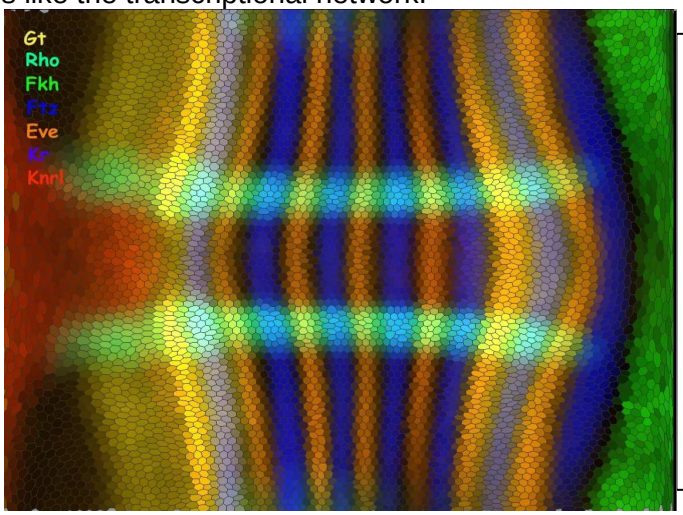

Figure 2. **Drosophila Blastoderm Atlas.** Our Blastoderm atlas is the first to provide 3D spatio-temporal morphology and gene expression data at cellular resolution, for an entire animal [3,4]. This figure shows, in cylindrical projection, early acting expression patterns of the genes *gt*, *rho*, *fkh*, *ftz*, *kr* and *kni*, which pattern the surface of the embryo. By analyzing these patterns we have uncovered previously unkonwn biology of the system, recovered known transcriptional regulatory interactions, and predicted hundreds of new ones [5].

In our current work, which is part of the Berkeley *Drosophila* Transcription Network Project (BDTNP.lbl.gov), we have created the first morphology and gene expression atlas of the early stage of the *Drosophila* embryo. Using high resolution three dimensional imaging and image and statistical analysis we have created spatiotemporal atlases for multiple *Drosophila* species and have collected expression pattern data for 24 of the principal regulators and over 80 putative target genes [3].

Figure 2 shows, in cylindrical projection, 7 genes that pattern the embryo surface. These are quantitative maps of gene expression and cell position for every cell in the embryo. By analyzing these data we have revealed previously unknown features about the biology of the system. We are the first to show that anterior/posterior patterning genes have complex dorsal/ventral dependencies and that these are lost in dorsalizing mutants, which implies that the anterior/posterior and dorsal/ventral patterning systems are not independent as was previously thought [3]. We have shown that the local nuclear

density on the embryo surface, the canvas on which the gene patterns are expressed, is neither constant nor static [4]. This implies that the spatiotemporal dynamics of both gene expression and morphology are needed for accurately modeling the transcriptional network. Using a simple linear model of this transcriptional network, we are able to recover known regulatory interactions from our spatio-temporal expression atlas, and predict hundreds of new interactions [5]. Comparisions between species has revealed significient expression pattern differences between every gene we have compared. These finding pose many important questions regarding the functional importance of expression pattern differences between species, and impact our understanding of speciation.

Collaborators on this project include 1) Mark Biggin, Scientist, Genomics Division, Lawrence Berkeley National Lab, 2) Jitendra Malik, Professor, EECS, University of California Berkeley, 3) Michael Eisen, Associate Professor, Genetics, Genomics and Development, Department of Molecular and Cell Biology, University of California Berkeley, 4) Bernd Hamann, Professor, Computer Science and Visualization, University California Davis,

Future Goals:

This is an exciting time to be coordinating collaborations in bioimaging, biophysics, electrical engineering, computer science and biology. The creation of comprehensive, multi-feature atlases that quantitatively describe biological systems at cellular resolution is at the frontier of bioimaging research. There is tremendous need for post-genome based systems analysis, and these atlases will provide the essential framework for such analysis, allowing us to better understand a vast range of complex biological systems, such as early onset of malignancy and developmental programs in animals.

For the breast cancer project we are developing methods that can extract new macromolecular features that can be targeted at multiple nuclear proteins with the goal of defining a set of quantitative biomarkers that can be used to objectively classify different histological classes of tumors. We are interested to know if there are epigenetic traits of low-grade tumors that can help identify patients with non-metastatic disease who have been histologically classified at high risk. For example, infiltrating ductal carcinomas constitute approximately 70% of invasive mammary cancers but span the entire histological grading range. On the other hand, tubular carcinomas, which account for about 5% of invasive tumors, are low grade, usually do not enter the lymph system and almost never metastasize to distant sites. The question is what specific macromolecular epigenetic traits of tubular carcinomas can be used to identify infiltrating ductal carcinomas of low grade that do not metastasize. By combining macromolecular information at cellular resolution with the tissue context, we hope to be able to answer these types of long standing problems.

The next step for the *Drosophila* project is to extend our atlas through embryogenesis. There are some 6000 nuclei at the Blastoderm stage, but approximately 40,000 cells of 70 different type by the end of embyogenesis. To tackle this vast increase in complexity we are developing new image acquisition methods and have demonstrated that the majority of the nuclei through embryogenesis are computationally distinguished. This lays the groundwork for creating static morphological frameworks to describe *Drosophila* anatomy at various stages. Challenges lie ahead in defining the extent of the cells and classifing and anotating the tissue and cell types, but we predict that this will be achievable based on embryo morphology and tissue-specific gene expression markers. The goal is to produce a computational quantitative cellular resolution atlas of *Drosophila*. The atlas will be an enormous resource to the *Drosophila* community, will produce novel techniques in image analysis, vision research and visualization, and the techniques developed will server as a general strategy for creating anotated atlases in biology.

Selected References:

1. **David W. Knowles**, Damir Sudar, Carol Bator-Kelly, Mina J. Bissell, and Sophie A. Lelièvre **2006** *Automated local bright feature image analysis of nuclear protein distribution identifies changes in tissue phenotype* **Proc. Natl. Acad. Sci. USA *103*, 4445-4450**

2. Fuhui Long, Hanchuan Peng, Damir Sudar, Sophie A. Lelièvre, and **David W. Knowles 2007** *Phenotype Clustering of Breast Epithelial Cells in Confocal Images based on Nuclear Protein Distribution Analysis* **BMC Cell Biology 2007, *8*(Suppl 1):S3**

3. Cris L Luengo Hendriks, Soile VE Keranen, Charless C Fowlkes, Lisa Simirenko, Gunther H Weber, Angela H DePace, Clara Henriquez, David W Kaszuba, Bernd Hamann, Michael B Eisen, Jitendra Malik, Damir Sudar, Mark D Biggin and **David W Knowles 2006** *3D morphology and gene expression in the Drosophila blastoderm at cellular resolution I: data acquisition pipeline* **Genome Biology 2006, *7*:R123**

4. Soile VE Keranen, Charless C Fowlkes, Cris L Luengo Hendriks, Damir Sudar, **David W Knowles**, Jitendra Malik and Mark D Biggin **2006** *3D morphology and gene expression in the Drosophila blastoderm at cellular resolution II: dynamics* **Genome Biology 2006, *7*:R124**

5. Charless C. Fowlkes, Cris L. Luengo Hendriks, Soile V. E. Keränen, Gunther H. Weber, Oliver Rubel, Min-Yu Huang, Clara Henriquez, Lisa Simirenko, Mike B. Eisen, Bernd Hamann, **David W. Knowles**, Mark D. Biggin, Jitendra Malik **2008** *A quantitative spatiotemporal atlas of gene expression in the Drosophila blastoderm* **Cell *133*:364-374 April 18, 2008**