ARIZONA STATE UNIVERSITY

Computational Biosciences • Bioinformatics • Genomics

Dr. Rosemary Renaut, renaut@asu.edu
Director, Computational Biosciences
www.asu.edu/compbiosci

ARIZONA STATE UNIVERSITY

Computational Biosciences • Bioinformatics • Genomics

PSM PROFESSIONAL SCIENCE MASTER'S

A Professional Masters Program

The School of Life Sciences

Mathematics and Statistics

Computer Science Engineering

W. P Carey Schoool Of Business

Program Overview

<table>
<tr><td>Semester 1</td><td>Introduction to Structural and Molecular Biology
4C.H.</td><td colspan="2">Multivariate Statistical Analysis
3C.H.</td><td>Mathematical Modeling and Introduction to Computational Biology
4C.H.</td></tr>
<tr><td>Semester 2</td><td>Business Issues and Ethics I
(CBS 510 or MGT 591 or MCB 590)
3C.H.</td><td colspan="2">Applications & Complex Problem Solving in Computational Biology
4C.H.</td><td>One of:
Biophysical Chemistry
Computational Genomics
Mathematical Model of Ecology
Mathematical Cell Physiology</td></tr>
<tr><td>Summer 1</td><td colspan="4">Continuation of Consulting Activity of Summer Course (Optional)</td></tr>
<tr><td rowspan="2">Semester 3</td><td>Comp. Mol. Biology Track I (9C.H.)
Mathematical Aspects of Biotechnology
3C.H.
Two of:
Advanced Cell Biology
Biomembranes
Cell Biotechnology Laboratory</td><td>Bioinformatics-Genomics Track II (9-10C.H.)
Data Warehousing and Data Mining
3C.H.
Two of:
Molecular Genetics
Functional Genomics
Adv. Statistical Genomics
Statistical Methods in Molecular Evolutionary Genomics</td><td>Quantitative Ecology Track III (9-10C.H.)
Mathematical Modeling in Biology & Environment
Two of:
Population & Community Ecology
Ecosystems
Populations:
Evololutionary Ecology</td><td>Physiological Modeling Track IV (9C.H.)
Neural Modeling
Two of:
Neurophysiology
Cellular Physiology
Bioelectric Phenomena</td></tr>
<tr><td colspan="4">Business Issues and Ethics II
(CBS 515 or MGT 591 or MCB 590) 3C.H.</td></tr>
<tr><td>Semester 4</td><td colspan="2">Internship
6C.H.</td><td colspan="2">Experimental Design
3C.H.</td></tr>
</table>

(blue)	Required CBS Course
(green)	Optional Course
(white)	Required Existing Course

Core Requirements

- *Scientific Computing for Biosciences(4)*
- *Case Studies/ Projects in Biosciences(4)*
- *Structural and Molecular Biology(4)*
- *Statistics and Experimental Design(6)*
- *Business Practice and Ethics(6)*
- *Internship and Applied Project(6)*

Electives

–Genomics/Proteomics

–Data Mining Data Bases,

–Medical Imaging

–Molecular/Functional Genomics

–Microarray Analysis

–Individualized to meet student needs

Enforced Prerequisites

- –Calculus and Differential Equations
- –Basic Statistics (junior)
- –Discrete Algorithms and Data Structures
- –Programming skills(C++/Java)
- –Cell biology, genetics(junior level)
- –Organic and Bio Chemistry (junior)
  - –Motivation, creativity, determination!

Features/Strengths

- –Interdisciplinary Training/Team Work
- –Internship/Applied Project Report
- –Business, Management and Ethics
- –(Health Services Administration MBA)
- –Small Groups/Close Faculty Involvement
- –Computer Laboratory
  - –**Extensive Project work/Consulting**

DATA

- Year 3: total 43 students
- Graduates: 10 + 5 in Fall 04
- Internships: NIH, ASU, Tgen, AZ Game and Fish, Water conservation lab, AZ biodesign
- Jobs: Tgen, ASU, Independent company, 3 pregnancies, Medical record keeping
- Future Doctoral programs

  many are interested!

ARIZONA STATE UNIVERSITY

Computational Biosciences • Bioinformatics • Genomics

PSM PROFESSIONAL SCIENCE MASTER'S

Future Developments/in progress

–Undergraduate: NIH MARC:

– Quantitative Skills (sophomore)spring 05

– Modeling Comp Bio (Junior) spring 05

–Doctoral Program Computational Biosciences

– Life Sciences

– Mathematics

Example Projects 2004

–Database Construction/Mining of Pathology Specimens (Tgen)

– Gegenbauer high resolution reconstruction for MRI, ASU

–TLS-SVM for Feature Extraction of Microarray Data, ASU

–Automated video analysis for cell behavior. Tgen

–EST DB for Marine Dinoflagellate *Crypthecodinium cohnii*, *ASU*

–Data mining for microsatellites in ESTS from *arabidopsis thaliana and brassica* species (US Water Conservation Laboratory)

–The Genome Assembler- Tgen

–A user interface to support navigation for scientific discovery ASU

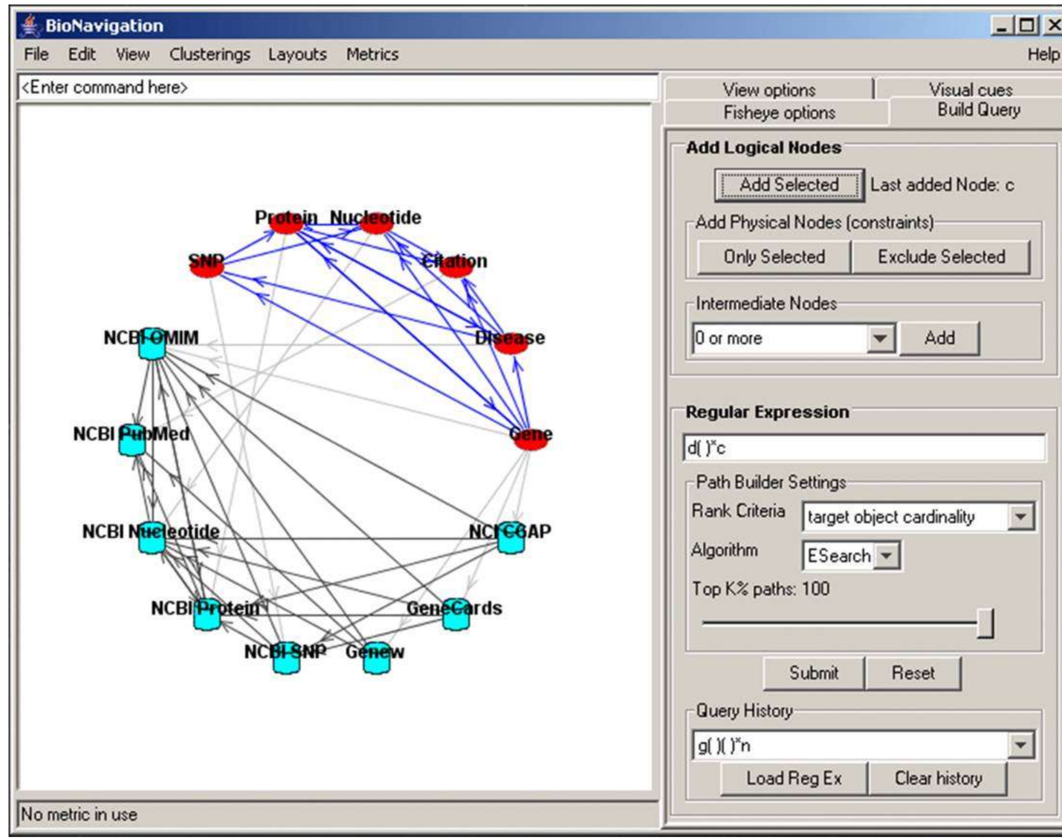
–Cell Migration Software Tool Tgen

BioNavigation: Selecting Resources to Evaluate Scientific Queries

Kaushal Parekh

Advisor: Dr. Zoé Lacroix, Arizona State University

Answering biological queries involves the navigation of numerous **richly interconnected scientific data sources**. The BioNavigation system supports the scientist in exploring these sources and paths. Scientific queries can be posed at the conceptual level rather than being restricted to particular data sources. The BioNavigation interface provides a scientist with information about the available data sources and the scientific classes they represent. The user can **graphically create a navigational query**. BioNavigation will evaluate the query and will return a list of suitable paths through the data sources identified by the ESearch algorithm. ESearch searches a graph for paths satisfying a regular expression and ranks them using benefit and cost metrics. BioNavigation thus complements the traditional mediation approach and **provides scientists with much needed guidance in selecting data sources** and navigational paths.



Software Development for Alzheimer's Disease Diagnosis and Research

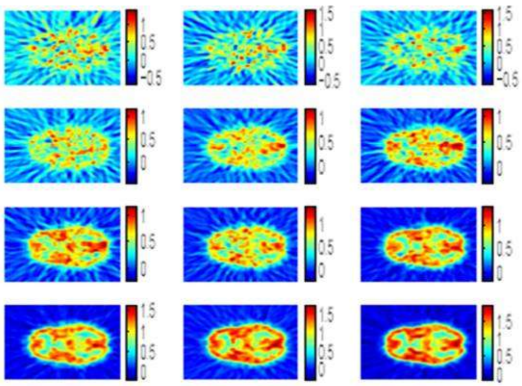
Guadalupe Ayala

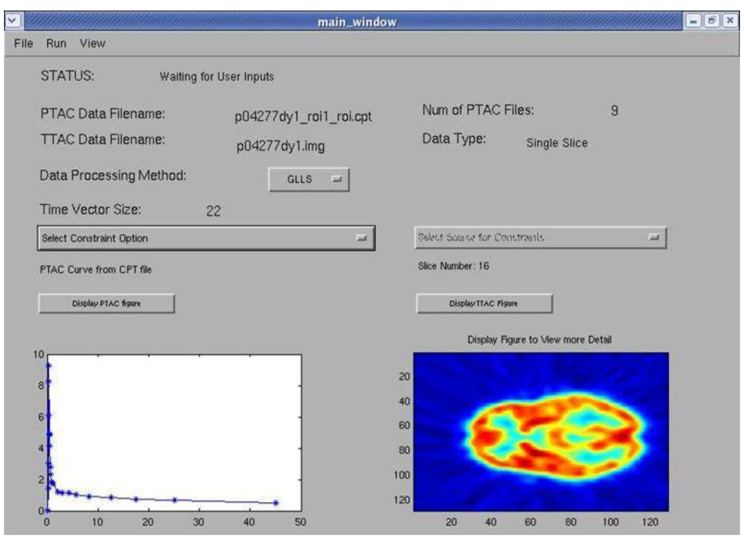
Advisor: Rosemary Renaut, Arizona State University

Figures:

(top) sample input images

(bottom) user interface





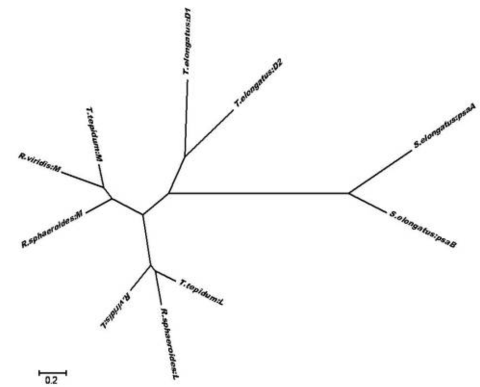
We are interested in using PET to image brain activity in patients with Alzheimer’s disease (AD). In AD studies, one way to measure disease progression is by **measuring Flouro-Deoxy-Glucose** (FDG), which is an analog of glucose uptake in the brain. Studies which determine a **local cerebral metabolic rate** (LCMR) of FDG uptake in a region of interest have proved successful in understanding AD progression. More specific information may be obtained by estimating the individual kinetic parameters which describe FDG metabolism. In particular, it is believed that the individual **FDG kinetic parameters may be used for early detection of AD**. We had developed an application that is used to estimate the kinetic parameters in order to be able to focus towards understanding the spatial distribution of kinetic parameters in AD, and as well as towards developing a precise measure for utilization in the early detection of AD. It uses dynamic PET data obtained from one-dimensional, two-dimensional or three-dimensional measurements. It also allows the user to compare results with respect to the computational and estimation methods, filters, constraints, and input sources chosen by the user. Comparing the results could help find out what are the optimal estimation methods, what are the constraints or what is the best filtering technique that provides optimal results. Results could be compared with expected results according to theoretical information, and an educated decision can be made on what are the optimal computational methods to use for every situation.

Evolution of Reaction Center in Photosynthetic Organisms: Conserved sequences in Photosystems

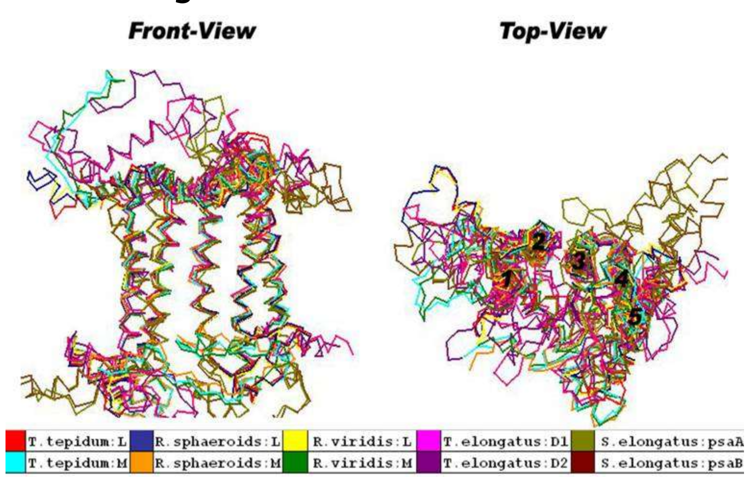
Sumedha Gholba

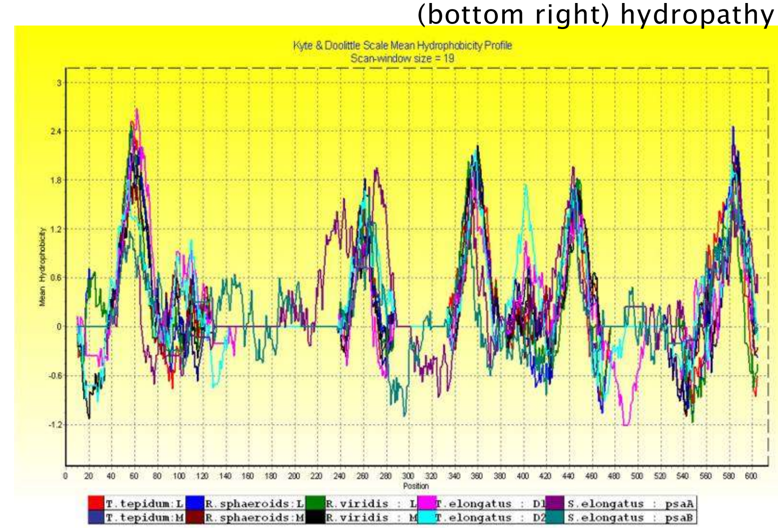
Advisor: Robert Blankenship, Arizona State University

The study of reaction center **proteins from both the photosystems and the primordial reaction centers** from bacteria reveal the conservation of certain amino acids. The multiple sequence alignment and phylogenetic trees created from the proteins show high degree of conserved regions in photosystem-II and bacterial reaction center-II, implying common genealogy. Also, the similarity between photosystem-I heterodimers and reaction center I homodimer proteins, indicate them having a single precursor. It is seen that even though L-M and D1-D2 show similar evolution with gene duplication, L-M proteins show step-by-step diversification whereas the other branch bifurcates into D1s and D2s just at the end. The reaction center I homodimer is placed nearly at the center between the photosystem I and II portions of the tree, suggesting it to be an ancestral type of reaction center. The structural alignment of these proteins depicts five well aligned α-carbon helices. Their **sequences show good amount of similarity in the hydrophobic domains forming the transmembrane helices, which are the main functional regions**.



Figures:
(top) phylogenetic tree
(bottom left) structural alignment
(bottom right) hydropathy plots



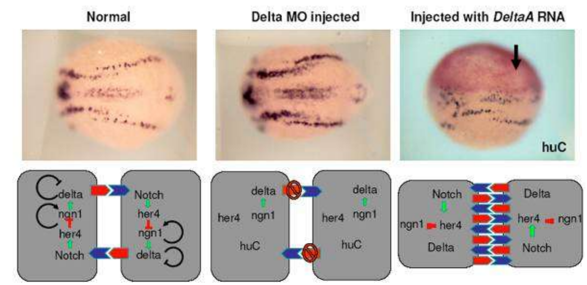


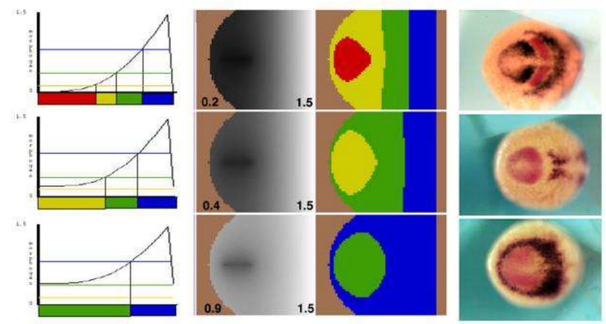
Simulating gene expression patterns during zebrafish embryo development

Ei-Ei Gaw

Advisor: Ajay Chitnis, National Institutes of Health, Maryland

During embryo development, it is essential that relatively homogenous **groups of cells undergo differentiation to form spatially different patterns** and eventually take on many different functions. Intercellular communication and morphogen gradients are two aspects that have been shown to play roles in determining cell fate. To understand better how these activities result in pattern formation, we utilized the NetLogo programming environment to simulate these processes. We were able to observe visually the possible **pattern formation of gene expression from activation and inhibition of genes, intercellular interactions** (top figure), **and the exposure to morphogen gradients** (middle and bottom figures). The three developmental phenomena of zebrafish embryos we studied were neurogenesis, somitogenesis, and morphogenesis during anterior/posterior patterning. The model for neurogenesis examines how autocatalysis and lateral inhibition (notch signaling) are required to form stable patterns. In addition, we investigated to what extent the geometry of the domains and the initial noise in the '***her***' gene expression help determine a cell's fate. To study somitogenesis, we explored how transcription and translation delays coupled with notch pathway and independent moving wave-front activity of ***fgf*** may contribute to the oscillation and synchronization of gene expression during formation of somites. Lastly, we used our **models to examine how time and concentration of a morphogen gradient may signal gene activation** and eventually form patterns of cells with stable and differing gene expression. Although, there is much room to study in more detail the systems of equations and the numerical analysis we used, overall this method is a good addition to the traditional methods of studying how gene-expression patterns may develop.



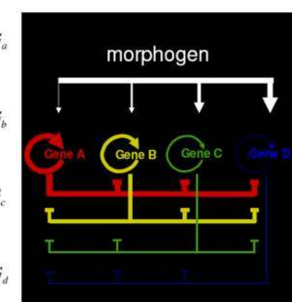


$$\frac{dG_a}{dt}=\frac{S_a(G_a^2+M_tM_a)}{G_a^2+G_b^2+G_c^2+G_d^2+1}-R_aG_a$$

$$\frac{dG_b}{dt}=\frac{S_b(G_b^2+M_tM_bG_a)}{G_a^2+G_b^2+G_c^2+G_d^2+1}-R_bG_b$$

$$\frac{dG_c}{dt}=\frac{S_c(G_c^2+M_tM_cG_b)}{G_a^2+G_b^2+G_c^2+G_d^2+1}-R_cG_c$$

$$\frac{dG_d}{dt}=\frac{S_d(G_d^2+M_tM_dG_c)}{G_a^2+G_b^2+G_c^2+G_d^2+1}-R_dG_d$$

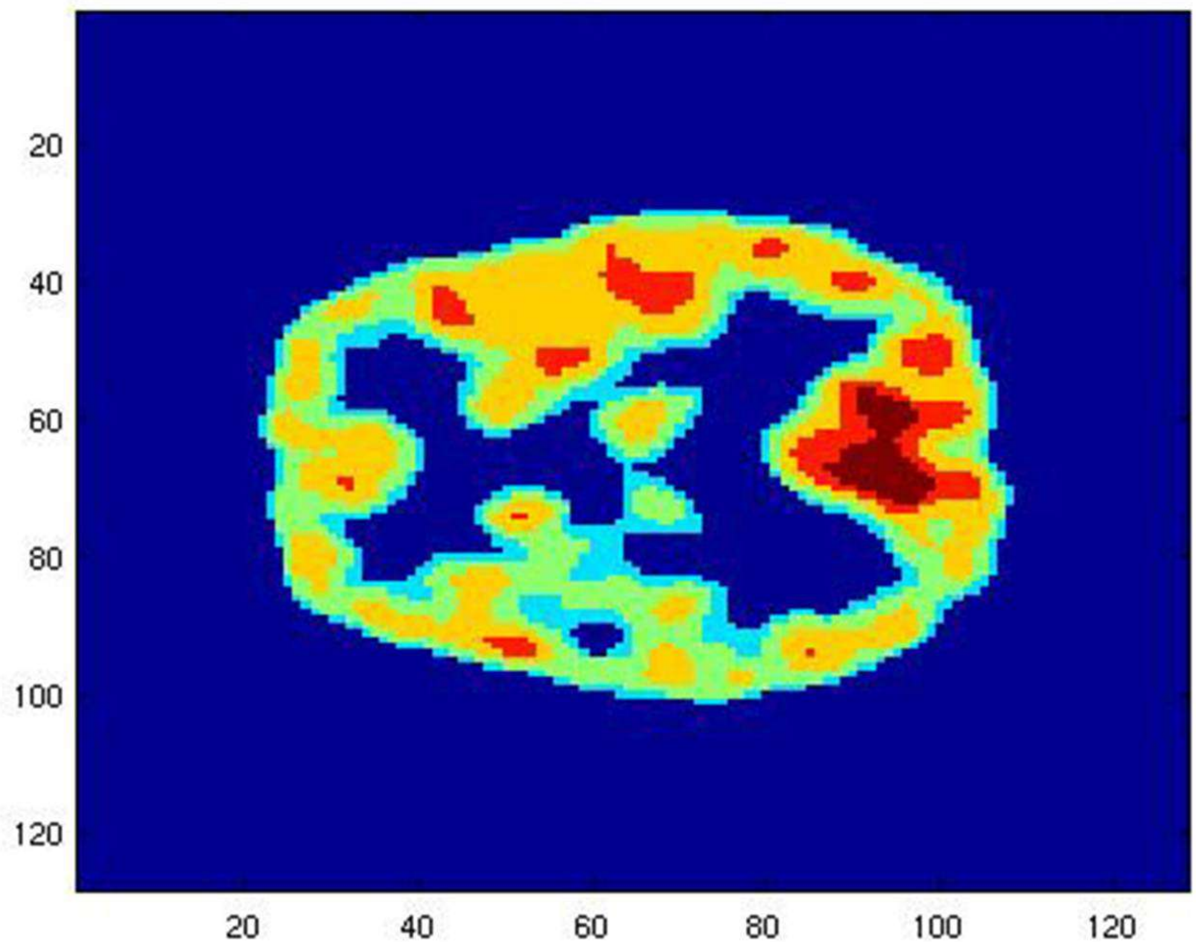


Robust Clustering of PET Data

Prasanna Velamuru

Advisors: Rosemary Renaut and Hongbin Guo, Arizona State University

Clustering has recently been demonstrated to be an important preprocessing step prior to parametric estimation from dynamic PET images. Clustering, as a form of segmentation, is useful in improving the accuracy of voxel level quantification in PET images. Classical clustering algorithms such as hierarchical clustering and K-means clustering can be applied to dynamic PET data using an appropriate weighting technique. **New variants of hierarchical clustering with different preprocessing criteria** were developed by Dr. Guo recently. Our research focus is to **validate these different algorithms with respect to their efficiency and accuracy**. Different inter and intra cluster measures and statistical tests are considered to assess the quality of the different cluster results.

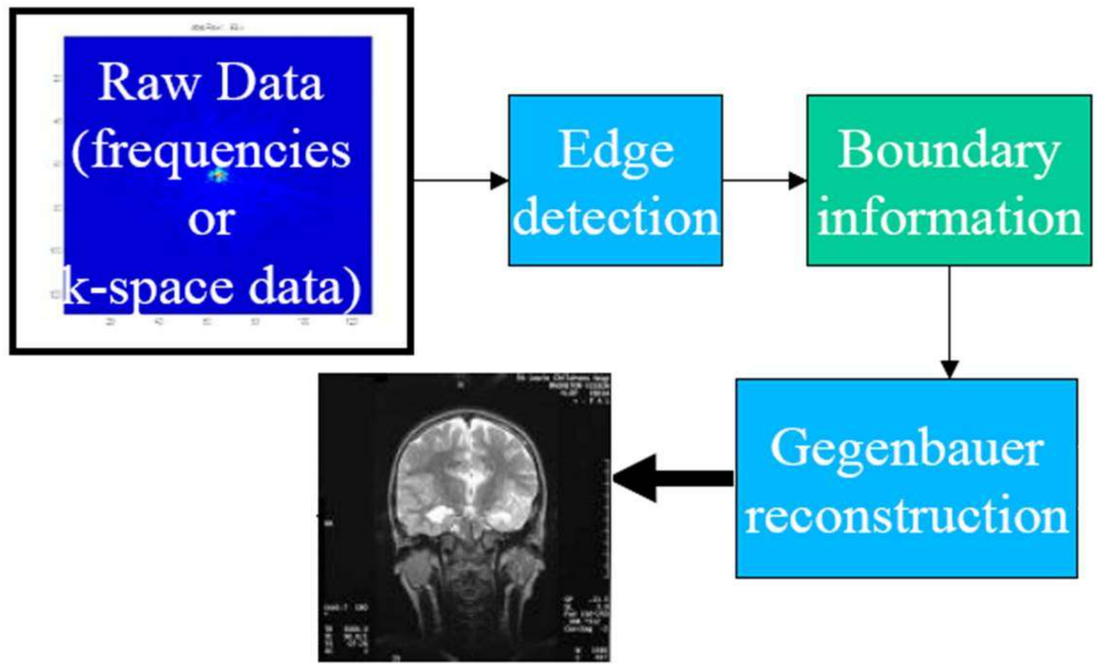


Gegenbauer High Resolution Reconstruction of Magnetic Resonance Imaging

Jim Estipona and Prasanna Velamuru

Advisors: Rick Archibald and Rosemary Renaut, Arizona State University

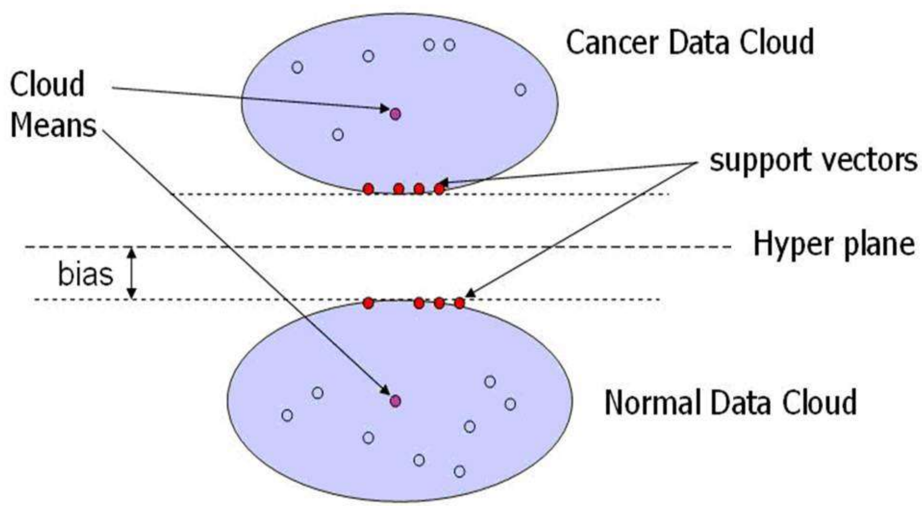
A variety of **image artifacts are routinely observed on MRI images**. We concentrate on the Gibbs Ringing that manifests itself as bright or dark rings seen at the borders of abrupt intensity change on the images. **Gegenbauer High Resolution Reconstruction method has been previously shown to eliminate the undesirable ringing** at the jump discontinuities in MRI. Prior work concentrated on applying this reconstruction method on frequency data obtained from reconstructed images and not on raw K-space data obtained from the MRI scanning machine. Our **project work concentrates on using the raw k-space data for reconstruction** and aims at comparing the reconstructed images obtained to those reconstructed from other commonly used reconstruction methods.

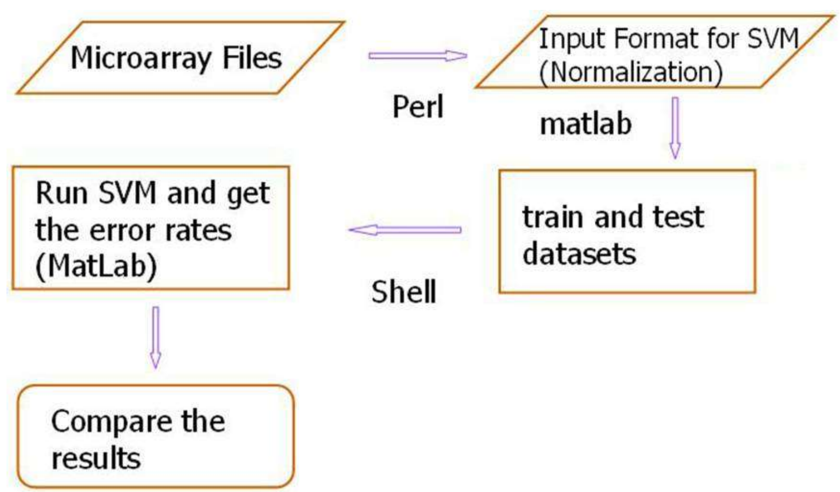


Regularized Total Least Squares in a Support Vector Machine

Sting Chen, Beryl Liu and Carol Barner
Advisor: Rosemary Renaut, Arizona State University

The goal was learn about Support Vector Machines, and explore **use of the Regularized Total Least Squares statistical method in Support Vector Machine classification of microarray data**. SVM is a special case of a Neural Net algorithm. It eliminates the rows (patient cases) of data items least valuable in determining the hyperplane until only key data items are left. Then these points are weighted, with heavier weights being given to those points that are close to the hyperplane boundary. These points are called Support Vectors. The new reduced, weighted space is called Feature Space. Finally, the trained program is given test data to see how well it can classify new patients as sick or normal.

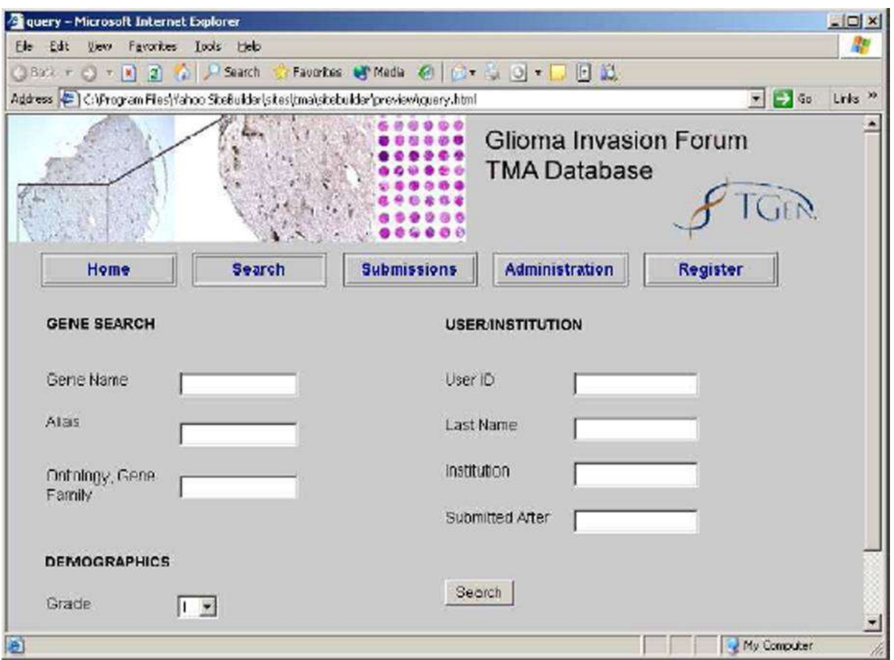


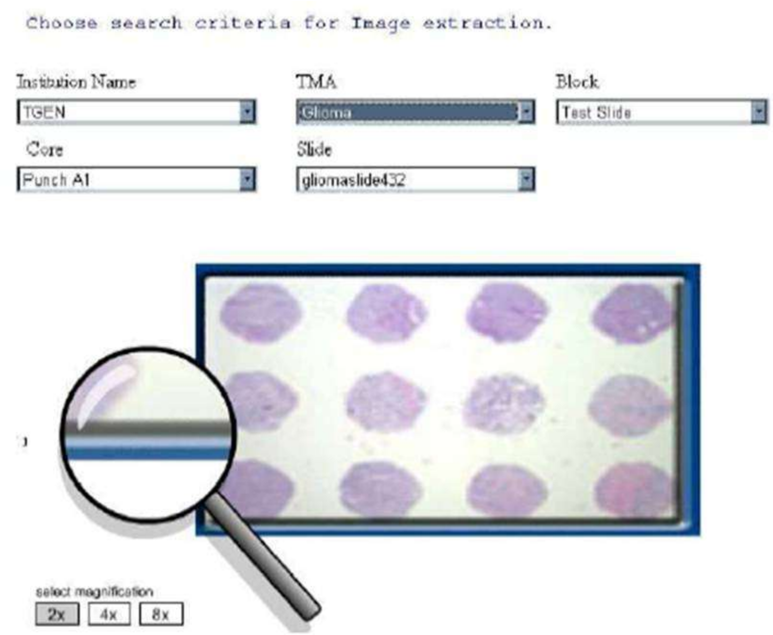


Database Construction and Mining of Pathology Specimens

Charu Gaur, Jennifer Szeto and Sotiris Mitropanopoulos
Advisor: Dominique Hoelzinger, Translational Genomics Research Institute, Phoenix

New technology allows hundreds of pathology specimens from human diseases to be sampled as .6mm punches of tissues that are arrayed into new “TMA” paraffin blocks; these blocks are then sectioned with microtomes to produce **hundreds of slides containing hundreds of human tissue specimens (tissue microarrays, TMAs).** Databases to support analysis of these high throughput TMAs will include information on diagnosis, treatment, disease response, and multiple images from follow-on studies linked to the coordinates of each of the hundreds of punches on the TMA. **Data mining from the results of TMA experiments will allow text mining and image feature extraction.** In this project, we present the requirements, design, and a prototype of a web based TMA database application.



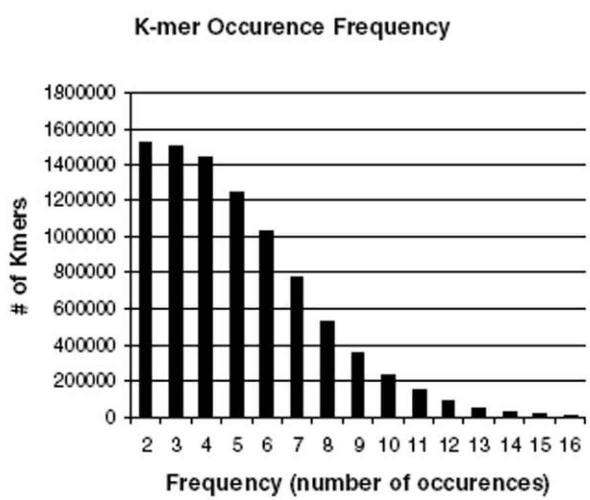


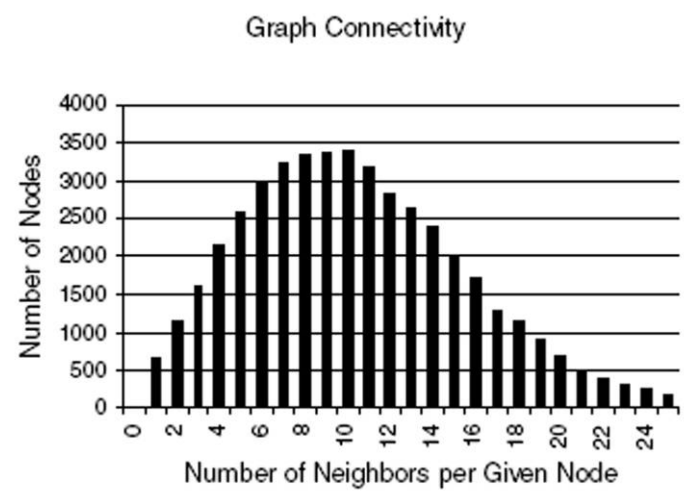
A Novel Genome Assembler: Using K-mers to Indirectly Perform N^2 Comparisons in O(N)

Ho-Joon Lee, Stephanie Rogers and Maulik Shah

Advisor: Jeffrey Touchman, Arizona State University and Translational Genomics Research Institute, Phoenix

The novel, exhaustive approach to genome assembly aims to **eliminate traditional heuristics and indirectly compare each sequence fragment to all the others in less than O(N^2) running time**. The first step in the algorithm involves building a k-mer library; accomplished by scanning through each sequence fragment using a sliding window of size k and cataloging each of these k-mers and the sequence fragment in which it occurs. It is assumed that neighboring fragments from the genome will share k-mers. From this k-mer table, an adjacency table is built; cataloging each sequence fragment and its neighbors. This adjacency table is generated in O(N) and represents all N^2 comparisons. Finally, with the information in the adjacency table, multiple breadth-first searches to collect and separate the connected fragments are performed. It is assumed that there exist disjoint graphs in the adjacency table and that each such graph represents an entire contig. This process was performed on two datasets, a simulated set and a real set: both having >40,000 sequence fragments of ~1,000 kb and 9-fold coverage. In both instances, the majority of the fragments were assigned to one contig.





ARIZONA STATE UNIVERSITY

Computational Biosciences • Bioinformatics • Genomics

PSM PROFESSIONAL SCIENCE MASTER'S

Technical Reports

The evolution of Runx genes I. A comparative study of sequences from phylogenetically diverse model organisms *Rennert, J.*, Coffman, J., Mushegian, A., Robertson, A.

Simulating Gene Expression Patterns of Neurogenesis, Somitogenesis, and Morphogenesis During Zebrafish Embryo Development *Gaw, E.*, Chitnis, A.

Evaluation of Gene Selection Using Support Vector Machine Recursive Feature Elimination *Huynh, J*

Development of an online database tool for locating bioinformatics-related software, *Kennedy, E.*

CpG Island Search Tool Application Development *Szeto, J.*

Application of GIS technology in bio sciences *Vuppaladadium, N*

Estimating the Divergence Time of Molecular Sequences using Bayesian Techniques *Gupta, S.*

Exploring and Exploiting the Biological Maze *Edupuganti, V.*

Exploiting Multiple Paths to Express Scientific Queries Lacroix, Z. *Morris, T., Parekh, K.*, Raschid, L., Vidal, M-E.

Automating the Biological Data Collection Process with Agents Lacroix, Z., *Parekh, K.*, Davulcu, H., Ramakrishnan, I.V., Julasana, N.,

Navigating through the Biological Maze Lacroix, Z., *Parekh, K.*, Rashid, L., Vidal, M-E.

How Biological Source Capabilities may Affect the Data Collection Process Lacroix, Z., *Edupuganti, V.*

ARIZONA STATE UNIVERSITY

Computational Biosciences • Bioinformatics • Genomics

PSM PROFESSIONAL SCIENCE MASTER'S

Technical Reports

Exploiting Agent and Database Technologies for Biological Data Collection

Lacroix, Z., *Parekh, K.*, Davulcu, H., Ramakrishnan, I.V., Julasana, N.

Evaluation Paths to Express Scientific Queries

Lacroix, Z., *Parekh, K.*, Vidal, M-E.

Techniques for Optimization of Queries on Integrated Biological Resources

Lacroix, Z., Raschid, L., Eckman, B.

Data Mining for Microsatellites in expressed Sequence Tags (ESTS) from Arabidopsis Thaliana and Brassica Species, For Use in Lesquerella (Brassicaceae)

Salywon, A., *Barber, M., Herling, N., Stewart, W.*, Dierig, D.A.

Feature Extraction From Synechocystis sp. PCC 6803 Cell Images

Kokoori, S.

Comparison of denovo predictions with known structure GcMAF

Ganta, S.

Protein Interaction Mapping: Use of Osprey to Map Survival of Motor Neuron Protein Interactions

Barnhart, M

Computational Investigation of Gene Regulatory Elements

Weddle, Ryan

Supertree Analysis of the Plant Family Fabaceae

Morris, Tiffany

Modeling Neural Patterning; Exploring the Gene Regulatory Networks in Developing Zebrafis Embryos

Gaur, Charu