

**BIOMEDICAL COMPUTATION @
STANFORD 2002
SYMPOSIUM PROCEEDINGS**

BCATS 2002 SYMPOSIUM PROCEEDINGS

Copyright © 2002 Biomedical Computation at Stanford (BCATS)

Printed in the United States of America

Editors: *Mike Hsin-Ping Liang, Yueyi Irene Liu, Michelle Green, Mehmet Serkan Apaydin, Madhup Gulati, Devshruti Pahuja*

Associate Editors: *Jill S. Higginson*

“Hands” artwork courtesy of *Biomedical Information Technology at Stanford (BITS)*

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limits of U.S. copyright law for private use of patrons.

Web Site: <http://bcats.stanford.edu/>

BIOMEDICAL COMPUTATION AT STANFORD 2002

Symposium Co-Chairs

Mike Hsin-Ping Liang
Yueyi Irene Liu
Michelle Green
Mehmet Serkan Apaydin
Madhup Gulati
Devshruti Pahuja

Administrative Help

Eva Elliott	Rosalind Ravasio
Tiffany Jung	Kelli Schreckengost
Candy Lowe	Fiona Sincock
Carol Maxwell	

Symposium Volunteers

Doris L. Cepeda	Jiong Ma
Jordan Cepeda	Brian Naughton
Jonathan Dugan	Rajiv Ramdeo
Jill S. Higginson	Sanjiv Ramdeo
Brett T. Kawakami	Serge Saxonov
Eric Ketchum	Jesse Tenenbaum

Symposium Sponsorship

Biomedical Information Technology at Stanford (BITS)
The National Institute of Health - BISTI
Bio-X

Full Sponsors

Affymetrix
Alloy Ventures
Apple Computer
Roche Bioscience
Sun Microsystems

Half Sponsors

BioScience Forum
Entelos
Genentech
Honda R&D Americas
NASA
Wilson Sonsini Goodrich & Rosati

TABLE OF CONTENTS

I.	Symposium Information.....	1
	a. Acknowledgements	
	b. About BITS	
	c. Symposium Schedule and Map	
II.	Keynote Speakers.....	5
	a. Mark Boguski, M.D. Ph.D.	
	b. Isaac Kohane, M.D. Ph.D.	
III.	Abstract List.....	9
IV.	Scientific Talks Session I.....	17
V.	Scientific Talks Session II.....	25
VI.	Poster Session.....	33
VII.	Symposium Participant List.....	83
VIII.	Symposium Sponsors.....	95

SYMPOSIUM INFORMATION

ACKNOWLEDGEMENTS

Numerous individuals and organizations have contributed to the 2002 symposium in Biomedical Computation at Stanford.

The organizing committee would like to thank the Biomedical Information Technology at Stanford (BITS) faculty who fostered the establishment of a forum through which Stanford researchers from across the university could share and discuss common interests and outline future directions in biomedical computation.

We would like to thank Dr. Mark Boguski and Dr. Isaac Kohane for setting the tone for a stimulating program and providing a roadmap through the interactions of biomedicine and computation in the new millennium.

We would also like to recognize Dr. Russ Altman, Dr. Charles Taylor, Dr. Clay Anderson, Dr. Oussama Khatib, and Dr. Parvati Dev for leading a series of Friday pizza dinner talks, in order to introduce the main topics and computational challenges at BCATS for those not working in the field.

We would like to acknowledge the generous financial support of the Biomedical Information Science and Technology Initiative from the National Institutes of Health and the additional support of the Bio-X initiative at Stanford.

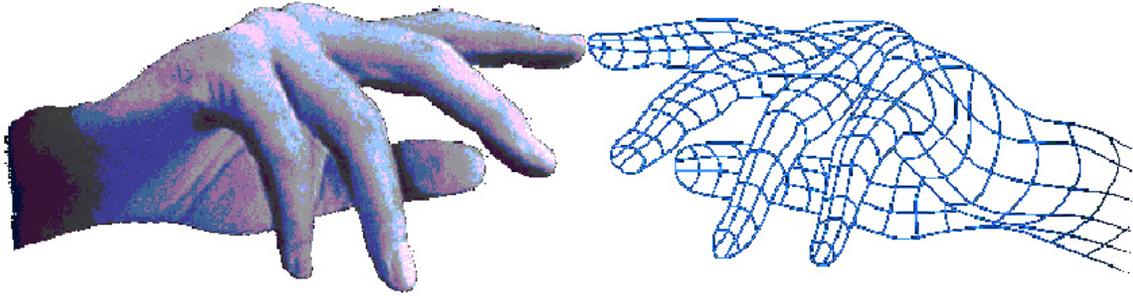
We are also grateful to the following corporate sponsors for their financial support and for promoting biomedical computation: Affymetrix, Alloy Ventures, Apple Computer, Roche Bioscience, Sun Microsystems, BioScience Forum, Entelos, Genentech, Honda R&D Americas, NASA, and Wilson Sonsini Goodrich & Rosati. In addition, we would like to thank the people at those organizations whose efforts made the sponsorships possible. We would especially like to thank Dr. Douglas Brutlag for his assistance in contacting sponsors.

We laud the BCATS 2000 committee for starting BCATS successfully. We are indebted to the BCATS 2001 committee, especially Brett Kawakami and May Liu for their guidance and assistance.

Finally, the organizing committee wishes to thank the many volunteers and department administrators, especially Tiffany Jung, for their tireless assistance with every aspect of this year's symposium.

Have a fantastic day. 😊

Your friendly BCATS 2002 committee



BIOMEDICAL INFORMATION TECHNOLOGY @ STANFORD

ABOUT BITS

The Biomedical Information Technology at Stanford (BITS) faculty group is the key supporter of BCATS.

BITS is an inter-connected, cross-disciplinary group of researchers who develop, share, and utilize computer graphics, scientific computing, medical imaging, and modeling applications in biology, bioengineering, and medicine. Our mission is to establish a world-class biomedical computing and visualization center at Stanford that will support joint initiatives between the Schools of Engineering, Medicine and Humanities and Sciences.

Participating labs promote the efficient development of new courses, programs, computational models, and tools that can be used in classrooms, clinical practice, and the biomedical research community. Our goal is to become an international resource for partners in the biotechnology, biomedical device, computing, medical imaging, and software industries.

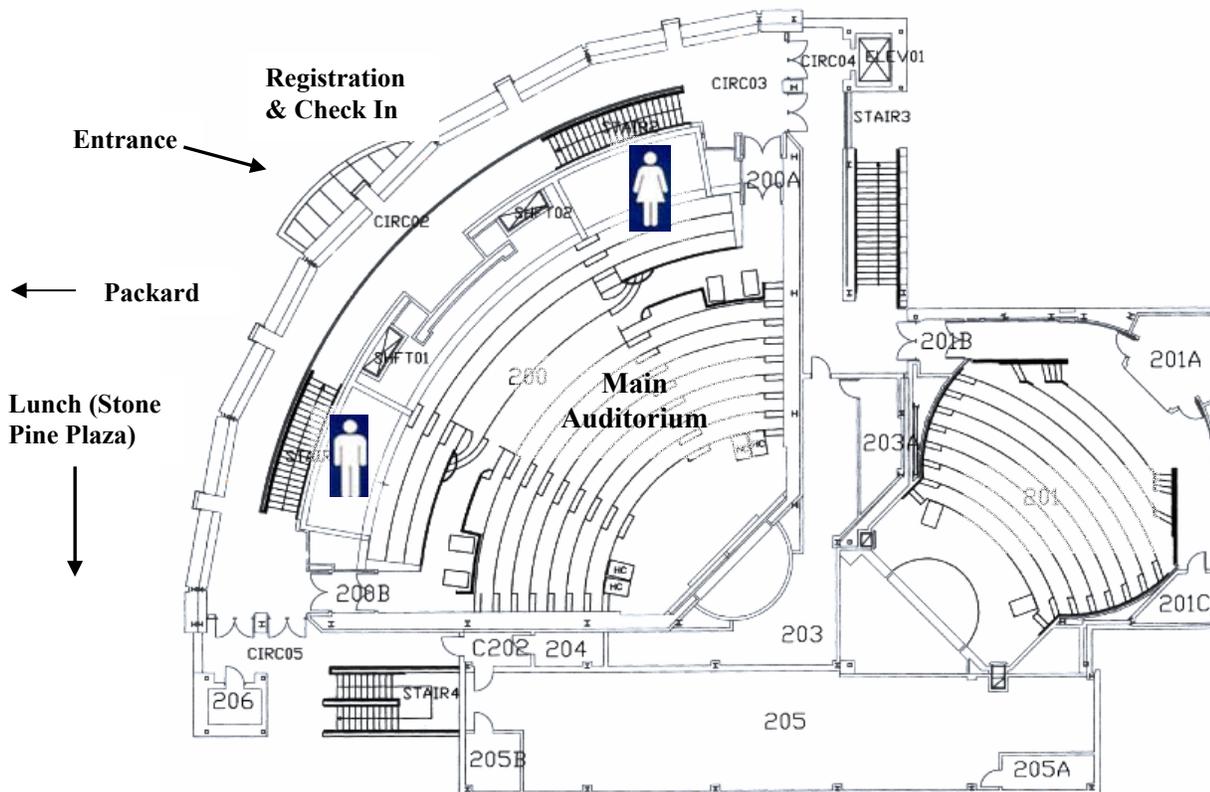
BITS faculty support teaching and training in the biomedical computing sciences and the creation of interdisciplinary biocomputational courses at the undergraduate, graduate, and post-graduate levels, both on-campus and at remote sites.

More information can be found at:
<http://neurosurgery.stanford.edu/bits/index.php>

SYMPOSIUM SCHEDULE AND MAP

Saturday, October 26, 2002

- 8:00am - 8:45am [On-Site Registration](#), Badge Pickup, and Breakfast (TCSEQ)
Poster Setup
- 8:45am - 9:00am Opening Comments (TCSEQ Lecture Hall 200)
- 9:00am - 9:45am [Keynote Address](#) - Dr. Mark Boguski (TCSEQ Lecture Hall 200)
- 9:45am - 10:00am Break
- 10:00am - 11:30am Scientific Talks Session I (TCSEQ Lecture Hall 200)
- 11:30am - 12:30pm Lunch (Stone Pine Plaza)
- 12:30pm - 1:30pm Poster Session I - Odd numbered posters (Packard Lobby)
- 1:30pm - 2:15pm [Keynote Address](#) - Dr. Isaac Kohane (TCSEQ Lecture Hall 200)
- 2:15pm - 2:30pm Break
- 2:30pm - 4:00pm Scientific Talks Session II (TCSEQ Lecture Hall 200)
- 4:00pm - 5:00pm Poster Session II - Even numbered posters (Packard Lobby)
- 5:00pm - 5:15pm Closing Presentation and Awards (Packard Lobby)
- 5:15pm - 6:00pm Informal Mixer (Hors d'oeuvres served)



KEYNOTE SPEAKERS

BCATS 2002 Symposium Proceedings

Keynote Speaker



Mark Boguski, M.D. Ph.D.

*Visiting Investigator
Fred Hutchinson Cancer Research Center
Seattle, WA*

Mark Boguski, M.D., Ph.D. is a well-known leader in informatics and genomics research. He has written and lectured extensively on bioinformatics and genomics, and developed the first publicly available database system for expression array data. Dr. Boguski leads bioinformatics specialists and experimental biologists in developing the interface between computational biology and functional genomics to gain new insights into systems biology. An original member of the U.S. National Center for Biotechnology Information (NCBI), Dr. Boguski has been involved with the development of a number of high-impact, enabling information resources. These include: dbEST, a critical resource for gene discovery and in silico SNP mining; applications of UniGene for creation of the first large-scale maps of the human genome and the design of gene chips from expression profiling; and ArrayDB for management and analysis of expression data. He has also made significant contributions to comparative genomics and pharmacogenomics.

Dr. Boguski is the author or co-author of over 100 articles and is the recipient of the Regents' Award from the National Library of Medicine and the NIH Director's Award. He is an organizer of the Cold Spring Harbor Symposium on Genome Sequencing and Biology and has served on grant review and advisory panels for a number of government and private funding agencies and as a consultant to industry. He is an adjunct Professor in the Department of Molecular Biology and Genetics at the Johns Hopkins University School of Medicine, a former Editor of Genome Research and serves on the Board of Reviewing Editors for Science magazine. Dr. Boguski is a member of the Scientific Advisory Board of the Merck Genome Research Institute, a member of the Genetics Advisory Group for the Wellcome Trust and an advisor to the Howard Hughes Medical Institute. He received his M.D. and Ph.D. degrees from the Medical Scientist Training Program at the Washington University School of Medicine in St. Louis and pursued specialty training in pathology.



Isaac S. Kohane, M.D. Ph.D.

*Associate Professor of Pediatrics
Harvard Medical School*

*Director
Children's Hospital Informatics Program
Boston, MA*

Isaac (Zak) Kohane is the director of the Children's Hospital Informatics Program and Associate Professor of Pediatrics at Harvard Medical School. Dr. Kohane is leading multiple collaborations at Harvard Medical School and its hospital affiliates in the elucidation of regulatory networks of genes and the interaction between genotype and phenotype using a variety of bioinformatics techniques. Application domains he is currently involved in include tumorigenesis, neurodevelopment, neuro-endocrinology and transplantation biology. Dr. Kohane's research builds on his doctoral work in computer science on decision support and subsequent research in machine learning applied to biomedicine. Dr. Kohane has also led the development of cryptographic health identification systems and automated personal health records. He has published over 50 papers in biomedical informatics.

Dr. Kohane has chaired several national meetings including the Spring Symposia on Artificial Intelligence in Medicine at Stanford University and the session on Linking Phenotype to Genotype at the Pacific Symposium on Biocomputing. He is also a founder of the Center for Outcomes and Policy Research at the Dana Farber Cancer Institute as well as founder and Associate Director for the Center for Genetic Epidemiology at Harvard Medical School. He is a Fellow of the American College of Medical Informatics and a Fellow of the Society for Pediatric Research. He is Associate Editor for Bioinformatics, Journal of Biomedical Informatics and on the editorial board of the Journal of the American Medical Informatics Associations. Dr. Kohane is also a practicing pediatric endocrinologist at Children's Hospital in Boston.

ABSTRACT LIST

SCIENTIFIC TALKS SESSION I

Using a Pathway/Genome Database for Plasmodium falciparum to Find Novel Drug Targets

Iwei Yeh, Theodor Hanekamp, Peter D. Karp and Russ B. Altman

Module Networks: Reconstruction of Molecular Modules and Their Regulation from Gene Expression

Eran Segal, Michael Shapira, Aviv Regev, Dana Pe'er, Daphne Koller and Nir Friedman

Interactive Simulation of the Human Hand

Leonard Sibille and Jean-Claude Latombe

Native-like Mean Structure in the Unfolded Ensemble of Small Proteins -- A Distributed Computing Study

Bojan Zagrovic, Siraj Khaliq, Michael R. Shirts, Christopher D. Snow and Vijay S. Pande

3D Differential Descriptors for Computer Aided Diagnosis of Colonic Polyps in CTC

Burak Acar, David S. Paik, Christopher F. Beaulieu, R.B. Jeffrey, Jr., Marta Davila and Sandy Napel

Multiple Genomic Sequence Alignment

Michael Brudno, Chuong Do, Michael Kim and Serafim Batzoglou

SCIENTIFIC TALKS SESSION II

Knowledge-Based Public Health Surveillance

David L. Buckeridge, Martin O'Connor, Justin Graham, Michael K. Choy, Zachary Pincus and Mark Musen

Enzyme Reaction Modeling of Hazardous Pollutant Transformation: Structural Basis of Biodegradability

Brett T. Kawakami, Sean Mooney, Charles B. Musgrave, Martin Reinhard and Paul V. Roberts

Using Scientific Literature Automatically to Understand Gene Expression Data

Soumya Raychaudhuri and Russ B. Altman

Comparison of Simulated and Experimental Protein Folding

Christopher D. Snow, Houbi Nguyen, Vijay S. Pande and Martin Gruebele

Targeted Clustering: A Method for Building Context-specific Clusters around Genes of Interest in DNA Microarray Data

Joshua M. Stuart, Kathy E. Mach, Stuart K. Kim and Art B. Owen

Automated Vessel Identification and Mobile Interactive Reporting for CT Angiography

Bhargav Raman, Raghav Raman, Sandy Napel and Geoffrey D. Rubin

POSTER SESSIONS

- P01 A Coupled Momentum Method for Modeling Blood Flow in Deformable Arteries**
C. Alberto Figueroa Alvarez and Charles A. Taylor
- P02 Decomposing Gene Expression into Cellular Processes**
Alexis Battle, Eran Segal and Daphne Koller
- P03 Collection and Transformation of Chromosomal Imbalances in Human Neoplasias for Data Mining Procedures**
Michael Baudis
- P04 Image Based Finite Element Modeling**
Erik Bekkers and Charley Taylor
- P05 Identification of Functional Regions in Proteins on the Basis of Evolutionary Sequence Analysis**
Jonathan Binkley and Arend Sidow
- P06 Creating an Online Dictionary of Abbreviations from MEDLINE**
Jeffrey T. Chang, Hinrich Schutze and Russ B. Altman
- P07 Direct Hydroxide Attack is a Plausible Mechanism for Amidase Antibody 43C9**
Lillian T. Chong, Pradipta Bandyopadhyay, Thomas S. Scanlan, Irwin D. Kuntz and Peter A. Kollman
- P08 The Draft Problem: Genomic Sequence Alignment Reconsidered**
Chuong B. Do, Michael Brudno and Serafim Batzoglou
- P09 Inferring Conformational Flexibility from Variations in Experimental Measurements**
Irene S. Gabashvili, Michelle Whirl-Carrillo, Michael Bada, D. Rey Banatao and Russ B. Altman
- P10 Establishing Standards in Medical Education: Use of Statistical Data Mining to Define Performance Characteristics**
Kunal S. Girotra and Carla M. Pugh
- P11 Large Multiple Sequence Alignments and Phylogenetic Analysis of the CFTR Locus**
Gregory Cooper and Arend Sidow (collaborations: Eric Green, Serafim Batzoglou)

- P12 Single Trial Classification of Independent Sources from MEG Task Related Recordings**
Marcos Perreau Guimaraes, Dik Kin Wong, E. Timothy Uy and Patrick Suppes
- P13 Predicting Growth-Generated Strains During Cranial Development**
James H. Henderson and Dennis R. Carter
- P14 Coherent Array Processing for Cost-effective Real-time Acoustic Imaging**
Jeremy Johnson, Ömer Oralkan, Mustafa Karaman, A. Sanli Ergun and Butrus T. Khuri-Yakub
- P15 Alternative Sensory Representations of the Visual World**
Neel Joshi, Dan Morris and Kenneth Salisbury
- P16 GeneXPress: Visualization and Analysis of Gene Expression and Sequence Data**
Amit Kaushal, Roman Yelensky, Tuan Pham, Eran Segal, Nir Friedman and Daphne Koller
- P17 Python Interface between PHYLIP and R**
Martina Koeva and Susan Holmes
- P18 Large-Scale Computational Protein Design: Sequence Space, Structure Prediction, and Protein Evolution**
Stefan M. Larson, Jeremy L. England, Amit Garg, John R. Desjarlais and Vijay S. Pande
- P19 CT Colonography Prone and Supine Data Registration**
P. Li, S. Napel, B. Acar, D. Paik, R. B. Jeffrey and C. F. Beaulieu
- P20 Applying Disclosure Control Methods to Biomedical Data**
Zhen Lin, Micheal Hewett and Russ Altman
- P21 Alternative Splicing: Data Storage and Analysis**
Shuo Liu and Russ B. Altman
- P22 Searching for Exons within Introns and Using Mutual Information to Help Find and Define DNA Regulatory Motifs**
Brian T. Naughton and Douglas L. Brutlag
- P23 Analysis of DNA Microarrays Using Knowledge-Based Algorithms**
Kuang-Hung Pan, Chih-Jian Lih and Stanley N. Cohen
- P24 Support for Guideline Development through Error Classification and Constraint Checking**
Mor Peleg
- P25 Modeling Mutations, Abnormal Processes, and Disease Phenotypes Using a Workflow/Petri Net Model**
Mor Peleg, Irene S. Gabashvili and Russ B. Altman

BCATS 2002 Symposium Proceedings

Abstract List

- P26 Teleoperable Dermatology: A Feasibility Study for Haptic Displays**
Kirk Phelps and Ken Salisbury
- P27 Like Standards but Better: Sharing Non-Traditional Health Data and Metadata**
Zachary Pincus, David L. Buckeridge, Michael K. Choy, Justin V. Graham, Martin J. O'Connor, Mark Musen
- P28 Automated Quantification of Arterial Calcification Using CT Angiography (CTA): Method and Evaluation**
Raghav Raman, Bhargav Raman, Sandy Napel and Geoffrey D. Rubin
- P29 Computational Modeling of T Cell Receptor Signal Integration**
Tim Reddy, Peter Lee, Arancha Casal, Cenk Sumen and Mark Alber
- P30 Protein Folding Simulation with Multiplexed Replica Exchange Molecular Dynamics Method**
Young Min Rhee and Vijay S. Pande
- P31 Fast DRRs Using Light Field Rendering**
Daniel Russakoff, Torsten Rohlfing, Daniel Rueckert and Calvin Maurer, Jr.
- P32 Structural Motif Discovery Using an Improved Structural Superposition Algorithm**
Jessica Shapiro and Douglas Brutlag
- P33 Interactive Segmentation and Visualization of Volumes**
Anthony Sherbondy, Mike Houston, Pat Hanrahan and Sandy Napel
- P34 SOURCE: The Stanford Online Universal Resource for Clones and ESTs**
G. Sherlock, T. Hernandez-Boussard, M. Diehn, A. Alizadeh, J.C. Matese, G. Binkley, H. Jin, J. Gollub, J. Demeter, J. Hebert, C.A. Ball, P.O. Brown and D. Botstein
- P35 An Elastic Model of the Malleus-incus Complex of the Human Middle Ear**
Jae Hoon Sim and Afraaz R. Irani
- P36 Identifying Structural Motifs in Proteins**
Rohit Singh and Mitul Saha
- P37 Characterizing Genomic Divergence between Two Urochordates, *Ciona savignyi* and *Ciona intestinalis***
Kerrin Small and Arend Sidow
- P38 Distances Between Phylogenetic Trees -- An Algorithm Using Bipartite Graphs**
Aaron Staple and Susan Holmes
- P39 Weighting Data Related by Phylogeny**
Eric A. Stone

- P40 CT Colonography: Does Improved Z Resolution Aid Computer-Aided Polyp Detection?**
Padmavathi Sundaram, Christopher F. Beaulieu, David S. Paik, Pamela K. Schraedley, R. Brooke Jeffrey and Sandy Napel
- P41 Confidence Regions and Topology of Phylogenetic Trees**
Henry Towsner and Susan Holmes
- P42 Dry Electroencephalography**
E. Timothy Uy, Dik Kin Wong, Marcos P. Guimaraes and Patrick Suppes
- P43 A Coupled Multi-domain Method for Computational Modeling of Blood Flow**
Irene Vignon, Brooke N. Steele and Charles A. Taylor
- P44 Supervised Neural Network for Brainwave Recognition Using Individual Trials**
Dik Kin Wong, Marcos Perreau Guimaraes, E. Timothy Uy and Patrick Suppes
- P45 From Promoter Sequence to Expression: A Probabilistic Framework**
Roman Yelensky, Eran Segal and Daphne Koller
- P46 Differential Gene Expression in a Partial Trisomy 16 Mouse Model of Down Syndrome**
Shuli Zhang, Farshid Oshidari, Ahmad Salehi and William C. Mobley
- P47 Classification of Gene Microarrays by Penalized Logistic Regression**
Ji Zhu and Trevor Hastie
- P48 Endoscope Calibration and Accuracy Testing for 3D/2D Image Registration**
Michael R Bax, Rasool Khadem, Jeremy A Johnson, Eric P Wilkinson, Ramin Shahidi

**BCATS 2002 Symposium Proceedings
Abstract List**

SCIENTIFIC TALKS SESSION I

USING A PATHWAY/GENOME DATABASE FOR *PLASMODIUM FALCIPARUM* TO FIND NOVEL DRUG TARGETS

Iwei Yeh, Theodor Hanekamp, Peter D. Karp and Russ B. Altman

Purpose

A pathway/genome database (PGDB) integrates pathway information with information about the complete genome of an organism. It is one of the first steps in assembling the “parts list” contained in the genome sequence into a working model of the cell. With the malaria genome nearly complete[1], we can identify possible gene products and their cellular functions and assemble metabolic pathways. A pathway database describes the interconnection of metabolites and enzymes within an organism.

We have built Plasmocyc (<http://plasmocyc.stanford.edu>), a PGDB for *Plasmodium falciparum* in the Pathway Tools framework. Using Plasmocyc, we perform systematic analyses over the entire system, such as determining the metabolic roles of individual proteins in order to identify promising drug targets.

Material and Methods

Pathways are identified using the PathoLogic program which takes as input the annotated genome of an organism [2]. The genomic sequences for *P. falciparum* were obtained from PlasmoDB [3] and the identified proteins were functionally annotated by GeneQuiz [4].

We identify potential drug targets by looking for metabolic enzymes with the following properties. (1) No known significant similarity between the *P. falciparum* enzyme and any human protein (this increases the chances of creating an inhibitor that does not interfere with human cellular processes). (2) The enzyme should be a “choke point” reaction, either consume a substrate that is not consumed by any other enzyme (to promote toxic accumulation), or produce a product that is not produced by any other enzyme (to cut off parasite access to essential compounds). (3) The potential target should be the only polypeptide that catalyzes the reaction in the parasite.

Results

We identify 25 drug targets in *P. falciparum*. Two of the proposed targets are inhibited by existing anti-malarial drugs. Our approach to finding drug targets can be improved with additional information, such as which genes are expressed during which cell cycle stages and where they are localized.

References

1. Gardner, M.J., et al., *Genome sequence of the human malaria parasite Plasmodium falciparum*. Nature, 2002. **419**: p. 498-511.
2. Paley, S. and P.D. Karp, *Evaluation of computational metabolic-pathway predictions for Helicobacter pylori*. Bioinformatics, 2002. **18**(5): p. 715-724.
3. Bahl, A., et al., *PlasmoDB: The Plasmodium Genome Resource*. Nucleic Acids Research, 2002. **30**: p. 87-90.
4. Andrade, M.A., et al., *Automated genome sequence analysis and annotation*. Bioinformatics, 1999. **15**(5): p. 391-412.

Web Page

<http://plasmocyc.stanford.edu>

MODULE NETWORKS: RECONSTRUCTION OF MOLECULAR MODULES AND THEIR REGULATION FROM GENE EXPRESSION

*Eran Segal, Michael Shapira, Aviv Regev,
Dana Pe'er, Daphne Koller and Nir Friedman*

Purpose

A living cell is a complex system that performs multiple functions and has to respond to a variety of signals. To achieve this, the cell is organized as a dynamic network of interacting functional modules, be those cellular compartments, protein complexes, or metabolic or signal transduction pathways. In this work we present *Module Networks*, a novel method to automatically infer functional modules directly from gene expression data. It simultaneously identifies the regulating genes that control gene expression and partitions the genes into groups whose behavior can be explained by common regulatory rules.

Material and Methods

We devised a fully automated procedure that discovers functional modules including their "regulation programs", directly from high throughput genome-wide expression profiles. A regulation program specifies the behavior of the genes in the module as a function of experimental conditions and the expression of transcriptional regulatory proteins (e.g., the module is up-regulated when YAP1 is up-regulated under heat shock). Our method initially clusters the genes based on their expression profile. For each "cluster" of genes, the procedure searches for a set regulation program that provides an explanation for the expression profiles. After the regulation programs are defined, the algorithm re-assigns each gene to the regulation program that best predicts that gene's behavior. The algorithm repeats until convergence, refining both the regulation program and the gene partition at each iteration.

Results

We applied Module Networks to a dataset of 173 *Saccharomyces cerevisiae* expression profiles under various stress conditions [1], automatically constructing 50 modules covering 2355 genes. We analyzed all 50 modules and for each examined all the components: biological coherence of the genes (38/50 scored well), the consistency of their regulators (supported by both known biology and binding site motifs) and the regulatory logic of the program (35/50 scored well). As an example, we discovered a respiration module consisting of (39/55) respiratory genes. HAP4 is the module's key activator, consistent with HAP4's known role in regulation of respiration. The HAP4 binding site was detected in (39/55) genes further supporting HAP4 regulation.

This analysis suggests both novel hypotheses and the experiments that can validate them. Each hypothesis has the form: gene A regulates the set of genes B under experimental condition C. We selected 3 hypotheses in which gene A was of unknown function and performed biological experiments (microarray experiments of gene A mutant strains under condition C) to test each hypothesis. In all 3 cases, the biological experiments confirmed the computational predictions, thereby suggesting a regulatory role for 3 uncharacterized genes.

References

1. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. Mol Biol Cell 2000 Dec;11(12):4241-57

INTERACTIVE SIMULATION OF THE HUMAN HAND

Leonard Sibille and Jean-Claude Latombe

Purpose

Virtual-reality techniques have an increasing number of applications in medicine. For instance, they can be used to teach anatomy to medical students. Surgeons can also train and develop their skills on virtual models before operating real patients. Another possible application consists in simulating digital actors for movies and video games. Here, we describe methods that we have integrated into a software system to simulate a generic human hand.

Material and Methods

We describe the hand skeleton as an articulated linkage. Visco-elastic soft tissue around the bones is represented using damped mass-spring meshes. A predictor-corrector method computes the deformation of the meshes. Incompressibility and collision constraints are taken into account. Tendons responsible for the movement of the fingers do not have a separate geometric model, but their mechanical effect is explicitly represented by means of forces and rotational springs at the joints between finger bones. In addition, our current software creates a “skin surface” wrapped around the bone structure and the soft tissue.

Results

The implemented software simulates a human hand, including bone movement, soft-tissue deformation, and collision handling, at interactive rates on a PC. Our software is written in C++, uses the SGI's Open Inventor library, and runs on a Linux PC.

Conclusion

Our current research aims at incorporating a better model of the skin, modeling additional degrees of freedom, enabling the grasping of rigid and flexible objects, taking into account reaction torques exerted by soft tissue on the joints of the bone structure, and generating patient-specific hand models.

NATIVE-LIKE MEAN STRUCTURE IN THE UNFOLDED ENSEMBLE OF SMALL PROTEINS – A DISTRIBUTED COMPUTING STUDY

Bojan Zagrovic, Siraj Khaliq, Michael R. Shirts, Christopher D. Snow and Vijay S. Pande

Purpose

A remarkable property of protein molecules is that they self-assemble (fold) into intricate shapes which are directly responsible for their function. Significant attention has been directed towards understanding this process, and much progress has been made. In particular, a lot is known about the structure and dynamics of the final folded state and the folding intermediates. However, the nature of the amorphous, unfolded state remains mysterious. Due to its transient nature and structural complexity, the unfolded state is extremely difficult to study experimentally. Furthermore, accurately studying the unfolded state with computer simulation is hampered by the great deal of sampling required. Using a super-cluster of over 10,000 processors we have performed close to 800 μ s of molecular dynamics simulation in atomistic detail of the folded and unfolded states of four polypeptides from a range of structural classes: the all alpha villin headpiece molecule, the all alpha tryptophan cage, the beta hairpin tryptophan zipper, and a designed alpha-beta zinc finger mimic. The central question that we ask is: what does the unfolded state look like on average?

Material and Methods

Using distributed computing techniques we have generated thousands of atomistic Langevin dynamics simulations of the native and the unfolded states of the four proteins, each tens of nanoseconds long. The simulations were run at room temperature and experimental solvent viscosity using the OPLS force field and implicit GB/SA solvent representation.

Results and Implications

A comparison between the folded and the unfolded ensembles generated in our simulations revealed a surprising fact. Even though virtually none of the individual members of the unfolded ensemble exhibit native-like features, the *mean* unfolded structure (averaged over the entire unfolded ensemble) has a native-like geometry. This observation suggests that protein folding may be viewed as narrowing of the structural variance around the mean which does not change throughout the process. Second, it suggests that the protein structures found through ensemble-averaged measurement may miss some of the underlying diversity. Finally, it suggest a way of performing structure prediction by looking at average features of unfolded ensembles.

References

1. Zagrovic B., Snow C., Khaliq S., Shirts, M. & Pande, V. Native-like mean structure in the unfolded ensemble of small proteins. *In Press. Journal of Molecular Biology*.

3D DIFFERENTIAL DESCRIPTORS FOR COMPUTER AIDED DIAGNOSIS OF COLONIC POLYPS IN CTC

*Burak Acar, David S. Paik, Christopher F. Beaulieu,
R.B. Jeffrey, Jr., Marta Davila and Sandy Napel*

Purpose

Computed Tomography Colonography (CTC) is a minimally invasive alternative to the conventional Fiber-Optic Colonoscopy (FOC). Its principal goal is to detect the colonic polyps (precursors of colon cancer) within the 3D CT images of human abdomen. The main problem is to detect and identify the colonic polyps in the segmented CTC data that are protruding structures that are generally spherical patches. A previously proposed method (HTD) to detect them relies on the increased concentration of intersecting surface normals in the vicinity of such patches. The basic source of false positives (FPs) is the existence of numerous other curved surface patches on the flexible colon wall, like haustral folds. We propose to characterize and use the distribution of the concentration maps (HT_maps) mentioned above to discriminate between spherical (polyps) and ellipsoidal (folds) structures. Currently, we have tested our method on an 8 patient data set.

Material and Methods

We used the 3D gradient vector field (V) of the HT_maps. Let $\lambda_3 \geq \lambda_2 \geq \lambda_1$ be the eigenvalues of the Jacobian of V at a suspicious point (detected by HTD). We computed: $\lambda_3, \lambda_1, \lambda_{3D} = \text{mean}(\lambda_3, \lambda_2, \lambda_1), \lambda_{2D} = \text{mean}(\lambda_3, \lambda_1), D_{\max/\min} = \max/\min(D_{32}, D_{31}, D_{21}), D_{3D} = \text{mean}(D_{32}, D_{31}, D_{21})$ where D_{mn} is the distance to circularly symmetric (star-like) topology in eigen-plane(mn). A Mahalanobis distance based classifier is trained and tested using different combinations of the new parameters and the HT_map values, in a 10-fold cross-validation experiment. There were 4946 suspicious locations with 19 true positives (7 polyps $\geq 10\text{mm}$) confirmed by FOC. We acquired the CTC data in the supine position from 8 patients (7 male, age 41-85). Typical acquisition parameters for single- (4-) detector row CT were 3mm(2.5mm) collimation, pitch 1.5-2.0(3.0), 1.5mm(1.0-1.5mm) reconstruction interval, 120KVp, 200MAs(56MAs).

Results

We performed FROC analysis on the data. Considering polyps $\geq 10\text{mm}$ only: At 6/7 (and 7/7) sensitivity, HT_map alone had 5 (and 5.5) FPs/patient, respectively, while [HT_map, D_{\max}, D_{3D}] had 1 (and 2.3) FPs/patient, respectively. Considering all 19 polyps: At 18/19 (and 19/19) sensitivity, HT_map alone had 155.8 (and 163.3) FPs/patient, respectively, while [HT_map, λ_{2D}] had 97.5 (and 151.5) FPs/patient, respectively. The results suggest that the small polyps need further characterization to be identified. Nevertheless the performance for clinically relevant polyps is within the applicability range.

Conclusion

We have shown that 3D differential parameters can be used to identify clinically relevant polyps within a large set of suspicious structures and thus improve specificity at high sensitivity levels. Currently we are conducting research on using such parameters on different 3D vector fields generated based on 3D CTC data.

References

1. Acar B, Beaulieu CF, et al. 3D Differential Descriptors For Improved Computer-aided Detection (CAD) of Colonic Polyps in Computed Tomography Colonography (CTC). To be presented at *RSNA 2002*, Chicago, USA.

MULTIPLE GENOMIC SEQUENCE ALIGNMENT

Michael Brudno, Chuong Do, Michael Kim and Serafim Batzoglou

Purpose

Several entire eukaryotic genomes, such as human, mouse, rat, fugu, and fly, have been completed or are currently being sequenced, providing us for the first time with the opportunity to compare the genetic code of humans with that of related organisms. Multiple alignments have proven effective in comparing biological sequences; however, there is currently no method both efficient enough to align long genomic sequences, and reliable enough to correctly align biological features between distant homologues.

We present LAGAN, a system for rapid alignment of two homologous genomic sequences, and Multi-LAGAN (M-LAGAN), the first practical method for multiple alignment of long genomic sequences. We tested our systems on syntenic genomic sequences totaling 13 Mbp from 12 vertebrate species and compared them against leading aligners. LAGAN and M-LAGAN produced the most accurate alignments, while requiring just 75 minutes on a PC to obtain the multiple alignment of all 12 sequences. M-LAGAN is the first practical multiple aligner for genomic sequences.

Algorithm

LAGAN employs an anchoring approach to quickly and accurately align two sequences. A rapid algorithm (CHAOS) is used to generate all of the local alignments between the two sequences. These are resolved into a strictly increasing set of anchors. A full Needleman-Wunsch[1] search is performed in the rectangles between the anchors and in bands around the anchors.

M-LAGAN is based on progressive alignment: A multiple alignment of K sequences is constructed progressively in K-1 pairwise alignment steps, where in each step two sequences or intermediate alignments, are aligned. LAGAN is used as the pairwise-alignment subroutine. M-LAGAN owes its strength to the strength of LAGAN, as well as to new methods that we introduce for scoring and refining a multiple alignment.

Results

In genomic sequences longer than 20 Kbp, we found that M-LAGAN is several orders of magnitude more efficient than current multiple aligners. We tested the accuracy of M-LAGAN in the CFTR gene region in 12 vertebrate species of total length 14 Mbp. That region is too long for existing multiple aligners, and accordingly we tested M-LAGAN and LAGAN against the leading pairwise aligner. In this data set M-LAGAN was able to align perfectly 96% of the exons, where the best pairwise aligners, LAGAN and BlastZ[2] aligned 95% and 94% respectively. A much more pronounced difference was present when comparing distant relatives, such as human & fish. Here, M-LAGAN aligned 76% of the exons, while LAGAN and BlastZ aligned 74 and 68%. Overall M-LAGAN was clearly the best aligner, and LAGAN the best pairwise aligner.

Conclusion

Multiple sequence alignment should become an increasingly important tool for biological discovery as several genomes suitable for cross-species comparison become available. Our systems demonstrate the ability of multiple alignments to properly align features where pairwise approaches fail. M-LAGAN will enable researchers for the first time to align long syntenic regions from several species, and design multiple alignment pipelines on a whole-genome scale.

References

1. Needleman S.B. and C.D. Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* **48**:443-453, 1970.
2. Schwartz S., et al. PipMaker – A web server for aligning two genomic DNA sequences. *Genome Research* **10**:577-586, 2000.

Web Page

<http://lagan.stanford.edu>

SCIENTIFIC TALKS SESSION II

KNOWLEDGE-BASED PUBLIC HEALTH SURVEILLANCE

David L. Buckeridge MD MSc, Martin O'Connor MS, Justin Graham MD MS, Michael K. Choy, Zachary Pincus, Mark Musen MD PhD, Stanford Medical Informatics, Stanford University, Stanford, CA, and VA Palo Alto Healthcare System, Palo Alto, CA.

Purpose

Current public health surveillance systems rely on manual reports of diagnoses. This limits how quickly an epidemic can be detected, and for epidemics such as those resulting from a bioterrorist attack, a detection delay of a single day could lead to thousands of deaths. In an effort to enhance the timeliness of epidemic detection, many surveillance systems are beginning to follow pre-diagnostic 'non-traditional' data sources (e.g., school absenteeism, pharmaceutical sales, emergency medical services calls) in addition to the usual diagnostic data sources. However, pre-diagnostic data are not as specific as diagnostic data, so multiple sources must be followed together in an attempt to reduce false positive detections.

Methods

The goal of our research is to identify effective approaches to piecing together multiple surveillance data sources to support public health decision-making. Combining data sources for this purpose is difficult because knowledge of how data sources relate to one another is often qualitative and uncertain. Moreover, the results of many analyses throughout the longitudinal surveillance process must be coherently applied to decision-making. Statistical methods have obvious application to this problem, but statistical methods do not readily and transparently incorporate uncertain and qualitative knowledge, and statistical models can become unwieldy as the number of parameters grows. We hypothesize that knowledge-based methods can be used to facilitate decision making in public health surveillance. In order to apply knowledge-based methods to this problem, surveillance tasks, knowledge and decisions must be explicitly modeled. 'Problem solving methods' (PSMs) can then be developed that operate on these models to accomplish surveillance tasks. This modular development approach enables controlled evaluation of different problem solving methods and knowledge representations in terms of epidemic detection, and impact on decision-making around interventions.

Results and Conclusions

Prototypes methods have been implemented within the BioSTORM (**B**iological **S**pace-**T**ime **O**utbreak **R**easoning **M**odule) system to: detect syndromes in individuals heuristically; provide forecasts of data streams using Kalman Filtering; and, detect epidemics using a dynamic Gaussian Bayesian belief network and other probabilistic methods for evidence combination. Based on our experience with the application of these prototype methods to simulated data, we are formally modeling surveillance knowledge requirements, implementing additional PSMs, and developing an evaluation framework. In addition, we are working closely with Public Health departments in San Francisco and Santa Clara counties to better model surveillance decision-making, and to evaluate our methods using real data.

ENZYME REACTION MODELING OF HAZARDOUS POLLUTANT TRANSFORMATION: STRUCTURAL BASIS OF BIODEGRADABILITY

*Brett T. Kawakami, Sean Mooney, Charles B. Musgrave, Martin Reinhard and
Paul V. Roberts*

Purpose

Environmental engineers are concerned with identifying microorganisms that can degrade hazardous pollutants. Enzymes involved in pollutant degradation pathways can often each transform a wide range of different compounds (substrates). In this talk, we present computational modeling and analysis of enzyme structure and reaction mechanism as a means to accelerate identification of potential targets for haloalkane dehalogenase - an enzyme important in the degradation of halogenated compounds. Our goal is to determine the potential substrate range of this enzyme, and to improve fundamental understanding of the enzyme mechanism and the molecular basis of biodegradability.

Material and Methods

We introduce a methodology that makes combined use of docking, quantum chemistry and molecular dynamics methods to explore the chemistry and 3-dimensional space of the haloalkane dehalogenase active site. We simulate the interaction of the enzyme with over 50 haloalkane structures so that all important substructures are represented. First, using docking analysis, we perform rapid screening for ability to achieve proper fit within the active site (Fig. 1). Next, intrinsic substrate reactivity is determined by quantum chemistry calculations. Finally, molecular dynamics simulation allows determination of the temporal distribution of conformations and orientations available to a mobile substrate molecule within the active site (Fig. 2). Our combined methodology allows us to account for the variety of enzyme-substrate interactions that affect enzyme activity. Applying this methodology across a series of structurally variant substrates allows efficient exploration of this interaction space and identification of important structural features and their corresponding effect on activity.

Results

We have identified a number of substrate structural features that significantly impact substrate transformation by haloalkane dehalogenase (Fig. 3). Halogen type and position are most influential, while size of the substrate molecule was critical after a certain cutoff. Based on these results, we also offer possible explanations for experimental observations and make predictions for untested haloalkane pollutants.

Conclusion

We have presented a novel application of biocomputational methods to an environmental engineering problem. Our results and methodology can assist environmental scientists in determining target pollutants for biological removal, and in developing possible protein engineering strategies for removal enhancement. Such applications will improve prospects for alleviation of the ongoing risk to human health presented by contaminated soils and groundwater.

Web Page

<http://www.stanford.edu/~brettk/bcats.html>

USING SCIENTIFIC LITERATURE AUTOMATICALLY TO UNDERSTAND GENE EXPRESSION DATA

Soumya Raychaudhuri and Russ B. Altman

Purpose

High throughput gene expression assays permit rapid assessment of the expression of thousands of genes. Recent application of these assays includes profiling of human cancer specimens, tracking gene expression during fruit fly development, and comprehensive measurement of yeast gene expression in response to specific gene deletions. In analyzing these complex gene expression data sets, investigators apply clustering methods to organize the data into tractable subgroups, or clusters, of genes sharing similar expression patterns. Careful examination of the genes that cluster together leads investigators to develop hypotheses about gene function and co-regulation. The most commonly used clustering method, hierarchical clustering, requires expert discretion to determine the exact cluster boundaries. Typically, cluster boundaries are drawn so that gene clusters contain functionally related genes; this is achieved by tedious manual examination of the potential clusters and the genes within them. Another challenge is separating the clusters that represent biological phenomena from the remaining spurious ones.

Methods

Here we propose that the peer-reviewed published literature in biology contains the necessary functional information about genes to automatically (1) define hierarchical cluster boundaries that respect biological function and (2) recognize biologically relevant clusters that contain functionally related genes. First, we present and evaluate the *neighbor divergence per gene (NDPG)* method that assigns a score to a given subgroup of genes indicating the likelihood that the genes share a biological property or function. To do this, it uses only a reference index that connects genes to documents, and a corpus including those documents. Then we present a method to partition hierarchical clusterings of gene expression data. The method begins by scoring all of the possible clusters with *NDPG* and then selecting a partitioning of the hierarchical tree that maximizes the weighted average of cluster *NDPG* scores. A successful partitioning, therefore, should identify clusters of genes data that share a common function and correlate with meaningful biology.

Results

We apply our technique to published gene expression data sets in yeast and fly containing close to 100 conditions; it discovers many biologically relevant clusters.

Conclusions

Literature can be used effectively to help guide experimental data analysis.

COMPARISON OF SIMULATED AND EXPERIMENTAL PROTEIN FOLDING

Christopher D. Snow, Houbi Nguyen, Vijay S. Pande and Martin Gruebele

Purpose

Essential events involved in protein folding can happen with astonishing speed. Small proteins have been shown to fold completely in tens of μs . Unfortunately, even the fastest folding proteins are difficult to simulate with classical molecular dynamics; performing a molecular dynamics simulation for 10 μs on even a small system would require decades for a typical CPU. Here we have extended the methods of molecular dynamics to observe a ~ 10 microsecond process using distributed computing, and directly compared the results to ultrafast experimental folding kinetics.

Material and Methods

To allow experiment and theory to meet on a microsecond timescale, we have chosen to study mutants of the fast-folding designed mini-protein BBA5. Laser temperature jumps of 11 ± 1 °C produced final temperatures between 293 and 303 K. Relaxation to the new equilibrium concentration was monitored at 100 ns intervals by following the temporal shift of the fluorescence spectrum with our newly developed submicrosecond real-time multichannel fluorescence wavelength detector¹. Our simulations² used the OPLS united atom parameter set and software adapted from the TINKER molecular modeling package. To model aqueous solution we used constant temperature stochastic dynamics and the generalized Born / surface area implicit solvent model.

Results

Experimentally the same 1.5 ± 0.7 μs relaxation rate was detected whether analyzing intensity changes or wavelength shifts of the time-resolved fluorescence spectrum. Standard definitions for the folding rate and the equilibrium constant of a two-state folder resulted in a folding rate of 7.5 ± 3.5 μs at 298 K. Considering all 32500 folding simulations, we observed the expected β -hairpin in over 1100 independent trajectories and α -helix in over 21000 independent trajectories. We estimate simulated folding times of 3 to 13 μs at 298 K.

Conclusion

10,000 trajectories on the tens of nanosecond time-scale seem sufficient to provide a reasonable statistical estimate of the folding rate for this rapidly folding two-state mini-protein. For the first time simulation and experiment can be compared directly.

References

1. Ervin, J., Sabelko, J. & Gruebele, M. Submicrosecond real-time fluorescence detection: application to protein folding. *J. Photochem. Photobiol.* B54:1-15, 2000.
2. Zagrovic, B.; Sorin, E.J.; Pande, V.S. Beta Hairpin Folding Simulations in Atomistic Detail. *J. Mol. Biol.* 313:151-169, 2001.

Web Page

<http://folding.stanford.edu/>

TARGETED CLUSTERING: A METHOD FOR BUILDING CONTEXT-SPECIFIC CLUSTERS AROUND GENES OF INTEREST IN DNA MICROARRAY DATA

Joshua M. Stuart, Kathy E. Mach, Stuart K. Kim and Art B. Owen

Purpose

We would like to use the large amount of gene expression data compiled in microarray databases to find genes that share common functions with a set of genes of interest. To do this, we search for genes that are expressed together in all informative experiments contained in the expression database. We present a method for collecting a highly co-expressed set of genes given a starting set of query genes. The method is based on the idea that some experiments will show coordinate regulation of the query set of genes and be useful for gene clustering, whereas other experiments will not show coordinate regulation and only add noise to the attempt.

Material and Methods

The method calculates which microarray experiments in the database show significant levels of coordinate regulation of the query set of genes, and then uses only these experiments to identify new genes whose expression is similar to the query set. Genes that score better than half of the query genes are considered as new candidates and included in the cluster.

Results

The method was applied to 5 *Caenorhabditis elegans* homologs of the human retinoblastoma complex. Biological confirmation was obtained by knocking out each candidate using RNAi and scoring the phenotype of the mutant worms. This resulted in the identification of 2 new genes that interact with genes in the retinoblastoma complex. We compared the method to simple clustering, in which all of the experiments were used to form the cluster, and found the method to be significantly more precise in identifying true candidates.

Conclusion

Preliminary results on several other diverse query sets indicate the method has the ability to propose new candidates for a variety of different processes. We are developing a web-based tool for biologists to submit their own genes of interest as queries (see <http://cmgm.stanford.edu/~kimlab/cassettes> for a version of the website still under development).

References

1. Friedman JH and Meulman JJ. Clustering Objects on Subsets of Attributes. *Stanford Tech Report*. June 2002.
2. Ihmels J, Friedlander G, Bergmann S, Sarig O, Ziv Y, Barkai N. Revealing modular organization in the yeast transcriptional network. *Nat Genet* 4:370-7, 2002.
3. Kim SK, Lund J, Kiraly M, Duke K, Jiang M, Stuart JM, Eizinger A, Wylie BN, Davidson GS. A gene expression map for *Caenorhabditis elegans*. *Science*. 293:2087-92, 2001.
4. Pavlidis P, Lewis DP, Noble WS. Exploring gene expression data with class scores. *Pac. Symp. Biocomput.* 474-85, 2002.

Web Page

<http://www.smi.stanford.edu/people/stuart/bcats2002>

AUTOMATED VESSEL IDENTIFICATION AND MOBILE INTERACTIVE REPORTING FOR CT ANGIOGRAPHY

Bhargav Raman, Raghav Raman, Sandy Napel and Geoffrey D. Rubin

Purpose

To develop and validate a fully automated post-processing and image review system by automatically determining the location and identity (ID) of major arteries in CT Angiography (CTA) studies, producing labeled measurements and images and presenting the data in an interactive web interface.

Material and Methods

The algorithm requires the user to provide one point identifying the root of the aorta. Vessel endpoints are then found using a distance map, and median paths through the aorta, celiac axis and hepatic, splenic, superior and inferior mesenteric (SMA & IMA) and renal arteries as well as the common (CIA), internal and external iliac arteries (IIA & EIA) and their branches are computed. Next, arteries and their origins and termini are identified with reference to an anatomical model comprised of branching patterns, relative cross-sectional areas, orientations, relative positions, and other anatomic landmarks. Artery cross-sectional area (CSA) profiles, lengths, angulation and curvature are calculated. Curved planar reformats (CPRs), curved slab maximum intensity projections (CS-MIPs), and Volume MIPs are generated at 18° rotational intervals and labeled. All data is output to a web – enabled database that additionally stores positional correlation information, allowing interactive review with confirmation of abnormalities seen in one image with any other image or with tabulated and graphed measurements. Server – side processing allows inexpensive wireless laptops to be used for report review. Validation was carried out using CTAs from 7 consecutive patients with abdominal aortic aneurysms. We compared automated processing time to standard manual processing by expert 3D technologists. Vessel ID and the number of vessels missed were scored by an experienced 3D technologist, and 4 radiologists independently localized arterial origins in space. The error in automated localization of origins was compared to inter-observer variability.

Results

Manual processing required 88.2 ± 13.6 min. Automated processing required 0.1 ± 0.1 min of user interaction and 15.1 ± 1.6 min of computer processing time. The identity of all detected vessel segments was correct and identified length was adequate. All patent major vessels were detected except for one 1.3 mm IMA, with a stenotic origin. The error for localizing arterial origins was 4.0 ± 0.9 mm. This was not significantly different from the mean interobserver variability of 3.3 ± 1.2 mm ($p = 0.33$).

Conclusion

Our system allows automated identification of arteries in CTA, decreases postprocessing time and allows interactive review of CTA reports on inexpensive mobile workstations and thus has the potential to improve the cost-effectiveness of CTA.

Web Page

<http://heartct.stanford.edu/AutoRad/VesselID.asp>
<http://heartct.stanford.edu/AutoRad/Autoreport.asp>

POSTER SESSIONS

A COUPLED MOMENTUM METHOD FOR MODELING BLOOD FLOW IN DEFORMABLE ARTERIES

C. Alberto Figueroa Alvarez and Charles A. Taylor

In order to understand the normal and pathologic behavior of the human vascular system, detailed knowledge of the blood flow and the response of blood vessels is required. Rigid wall models for blood flow computations have been used by a number of investigators to represent blood flow. However, these models do not provide a realistic framework to study pressure wave propagation phenomena in arteries, nor do they provide a good description of the blood motion close to the vessel walls. Since this formulation does not account for vessel wall motion, the subsequent pressure wave propagation speed is infinite, which is far from being physiologically meaningful.

In this work, a new formulation called the *Coupled Momentum Method for Fluid-Solid Interaction problems* (CMM-FSI) is presented. The main features of this formulation are summarized below.

Linearized description of the wall mechanics.

The assumption throughout the three dimensional models that the geometry of the arteries and surrounding structure are fixed but the deformations of the structure are accounted for through the use of linearized kinematics is proposed. The displacements of the arterial wall are assumed small so that the geometry of the vessel does not need to be updated throughout the cardiac cycle. Likewise, the contiguous blood domain is held fixed but the velocity at the blood vessel wall is not zero – there are small components in the plane of the lumen as well as the axial direction, compatible with the deformations of the artery. The in-plane component of velocity creates the possibility of flow “storage” that is impossible in a fixed geometry with no-slip boundary condition under the assumption of flow incompressibility.

Membrane formulation for the wall.

Arteries tend to respond primarily in membrane mode rather than in bending mode. Therefore, a linear membrane element may be used rather than a more complex shell element accounting for bending behavior.

Elastodynamics equations as a special boundary condition for the fluid.

The physics of the vessel wall and the blood are coupled at the variational stage, by considering the motion of the vessel wall (i.e., the classical elastodynamics equations) as a special boundary condition for the fluid domain. The same idea was used by Womersley on his derivation on an analytical solution for pulsatile blood flow in an elastic vessel.

The approach delineated here will result in a tractable, efficient and robust procedure, while at the same time respecting essential physics and enabling realistic simulation of wave-propagation phenomena in the arterial system.

DECOMPOSING GENE EXPRESSION INTO CELLULAR PROCESSES

Alexis Battle, Eran Segal and Daphne Koller

Purpose

A living cell is a complicated system that performs multiple functions and has to respond to a variety of signals. To organize this complex web of activity, the cell tends to compartmentalize its activity into distinct processes, or modules. This global organization cannot be discerned by studying the properties of isolated components. Genome-wide measurements of mRNA expression level across multiple experimental conditions provide us with a global picture of the cell's activities, and provide the potential for a high-level understanding of its behavior. In this work, we propose a probabilistic model for cellular processes, and an algorithm for discovering them from gene expression data.

Material and Methods

In our model, a *process* is associated with a set of genes that participate in it; unlike clustering techniques, our model allows genes to participate in multiple processes. Each process may be active to a different degree in each experiment. The expression measurement for gene g in array a is a sum, over all processes in which g participates, of the activity levels of these processes in array a . Our goal is to learn both the processes in which each gene participates and the activity levels for each process in each array. To do so, we devised an iterative procedure, based on the EM algorithm, which decomposes the expression matrix into a given number of processes.

Our model resembles the Plaid model proposed by Lazzeroni and Owen [1]. The Plaid model also decomposes the expression data as a sum of overlapping "layers"; each layer is associated with a set of genes and experiments that define the expression of that layer. There are several differences between our model and Plaid. Of these, the most important is that Plaid uses a greedy sequential approach to perform the decomposition, attempting, in each step, to explain as much of the unexplained expression data as possible. In contrast, our model is trained as a unified whole, allowing the association of genes with processes to change as the process models become more refined, and the process models to change as the set of assigned genes changes.

Results

We applied our method to a dataset of 173 *Saccharomyces cerevisiae* expression profiles under various stress conditions [2], and learned a model with 30 processes for 1010 genes. Once a model has been learned, we can read the processes directly from it. For each process, we read both the genes that participate in it as well as the levels of activity of each process in all experiments. Overall, we discovered processes in which the set of genes associated with the process often contained a very high fraction of genes that are known to share a functional role (135 different Gene Ontology annotations). We measured the significance of the enrichment of each annotation in each process. When comparing to Plaid, we found that most annotations (122 of 135 cases) had higher significant enrichment in our model. The same result was achieved when we compared our model to standard clustering. Finally, we show cases where our learned activity levels for processes had extremely high correlations with the expression levels of known regulators of those processes; importantly, these regulators were *not* part of the input data given to our program, indicating that our program reconstructed the levels of activity of these processes.

References

1. L Lazzeroni and A. Owen. Technical report, Stanford, 1999.
2. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. Mol Biol Cell 2000 Dec;11(12):4241-57

COLLECTION AND TRANSFORMATION OF CHROMOSOMAL IMBALANCES IN HUMAN NEOPLASIAS FOR DATA MINING PROCEDURES

Michael Baudis

Purpose

Although the deciphering of the human genome has been pushed forward over the last years, little effort has been made to collect and integrate the treasure trove of clinical tumor cases analyzed by molecular-cytogenetic methods into current data schemes. progenetix.net [1] has been developed as the largest public source of chromosomal imbalance data with band-specific resolution.

Material and Methods

Chromosomal aberration data of more than 4800 cases from 168 publications of CGH experiments [2] were collected. Minimal requirements were the analysis of clinical tumor samples and report of the results on a case by case basis, resolved to the level of single chromosomal bands.

Data was transformed to a two-dimensional matrix with code for the aberration status of each chromosomal band per case. Graphical representations and cluster images were generated for all different subsets (Publications, ICD-O-3 entities, meta-groups).

Results

Of the current 4896 tumor samples, 3862 (79%) showed imbalances by CGH. Of all cases, the average per band probability was 4.5% for a loss (maximum 12.9% at 13q21) and 6.5% for a gain (maximum 15.6% at 8q23).

Distinctive genomic aberration patterns became visually apparent in the different neoplastic entities. Differences showed in the average frequency and distribution pattern of imbalanced chromosomal regions. Tumor entities with the strongest hot spots for losses were small cell lung carcinomas (max. 96.2% at 3p26) and pheochromocytomas (max. 92.7% at 3p26); prominent gain maxima were found in infiltrating duct carcinomas (max. 91.7% at 11q13), T-PLL (max. 81.8% for whole 8q) and liposarcomas (81.8% at 12q13) among others. Also, different clusters of aberrations could be shown to occur in some entities.

Conclusion

progenetix.net project was able to collect a large dataset of genomic aberration data; to develop the software tools to transform those data to a meta format based on genomic interval descriptions, and to produce graphical and numerical output from those data for hot spot detection and statistical analysis.

For future approaches, the data collection will be valuable for filtering data from expression array experiments for relevant dysregulated genes, and possibly for the description of common and divergent genetic pathways in the oncogenetic process of different tumor entities.

References

1. Baudis M. and Cleary M. (2001) Progenetix.net - An online repository for molecular cytogenetic aberration data. *Bioinformatics* 17(12):1228-1229
2. Kallioniemi, A., O.-P. Kallioniemi, et al. (1992). Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* **258**: 818-821.

Web Page

<http://www.progenetix.net/>

IMAGE BASED FINITE ELEMENT MODELING

Erik Bekkers and Charley Taylor

Purpose

The conversion of three dimensional (3D) images to finite element (FE) models is known as reverse engineering. In medicine and biology, the resulting FE models are useful for studying processes such as blood flow, muscle and bone movement, and cellular transport. The objective of our research is to convert dense, discrete 3D data into a compact, analytic, continuous (C0 or C1) surface description. This surface description is topology independent, easily refined, and largely user independent, making it useful for a variety of applications.

Materials and Methods

The input data can come from any volumetric imaging modality, including MR, CT, or confocal laser microscopy. The primary goal is to automatically convert the input data into a surface consisting of rectangular patches; the rectangular surface representation is a standard one for CAD and FE packages. The input data is first converted to a connected, triangulated surface via the Marching Cubes algorithm. A Voronoi diagram is then constructed using the Fast Marching Method solution to the Eikonal equation, from randomly selected interior vertex points. The dual structure to this Voronoi diagram is the Delaunay triangulation of the surface. Triangle edges are generated by streamline integration along the gradient of the Eikonal solution and then least squares fit to cubic lines. The control points of these lines are input to a cubic tensor product surface patch. Pair-wise combination of triangles produces the desired rectangular patch network.

Results

We present results from the application of the above method to several vascular imaging problems. The output cubic Bezier network surface representation is suitable for many CAD and FE packages; these packages allow surface shape modification and can generate arbitrarily dense volumetric triangulations required for finite element analysis.

Conclusions

Our method is a useful tool for medical and biological reverse engineering. One of many possible applications is surgical planning. With an accurate geometric model of a vascular or bone network and an advanced CAD package, one could recommend optimal size and placement of stents, grafts, muscle attachments, and surgical connections. Second, the geometric model could be used to study functional problems such as stent/ vessel wall interaction forces, the required flexibility of a device, and flow through an altered vascular network. A major existing impediment to widespread clinical use of 3D models has been the time required to construct them; methods requiring little user interaction, such as the one described here, overcome this problem, making them clinically useful.

References

1. Matthias Eck, Tony DeRose, Tom Duchamp, Hugues Hoppe, Michael Lounsbery, Werner Stuetzle. Multiresolution Analysis of Arbitrary Meshes. Conference Proceedings, Annual Conference Series, pages 173--182. ACM SIGGRAPH, Addison Wesley, August 1995

IDENTIFICATION OF FUNCTIONAL REGIONS IN PROTEINS ON THE BASIS OF EVOLUTIONARY SEQUENCE ANALYSIS

Jonathan Binkley^{} and Arend Sidow^{*†}*

Departments of ^{*}Pathology and [†]Genetics

The principle that the rate of molecular evolution is inversely correlated with the strength of selective constraints forms the basis for identifying functionally important regions in a protein by their conservation over long time spans. The Evolution-Structure-Function (ESF) protocol (A. L. Simon, E. A. Stone, and A. Sidow 2002; *PNAS* **99**:5, pp. 2912–2917) is a more sensitive measure of the strength of selective constraint, one based on the rate of protein evolution, estimated locally in small sections of the protein. The method is quantitative in that it allows a ranking of the predicted regions on the basis of their evolutionary rate, guiding the investigator directly to the key regions of any protein. We present the ESF analyses of several known protein functional interactions, and predict several novel interactions, to demonstrate the power and utility of ESF analysis in structure: function studies.

CREATING AN ONLINE DICTIONARY OF ABBREVIATIONS FROM MEDLINE

Jeffrey T. Chang, Hinrich Schutze and Russ B. Altman

The immense volume and rapid growth of biomedical literature present special challenges for humans as well as computer programs analyzing it. One such challenge comes from the common use of abbreviations that effectively augments the size of the vocabulary for the field. To cope with this, we have developed an algorithm to identify abbreviations in text. It uses a statistical learning algorithm logistic regression to score abbreviations based on their resemblance to previously identified ones, achieving up to 84% recall at 81% precision. We then scanned all of MEDLINE and found 781,632 high-scoring abbreviation definitions. We are making these available as a public abbreviation server at <http://abbreviation.stanford.edu/>.

Web Page

<http://abbreviation.stanford.edu/>

**DIRECT HYDROXIDE ATTACK IS A PLAUSIBLE MECHANISM FOR
AMIDASE ANTIBODY 43C9**

*Lillian T. Chong, Pradipta Bandyopadhyay, Thomas S. Scanlan, Irwin D. Kuntz
and Peter A. Kollman*

Direct hydroxide attack on the scissile carbonyl of the substrate has been suggested as a likely mechanism for esterase antibodies elicited by phosphonate haptens, which mimic the transition states for the alkaline hydrolysis of esters. The unique amidase activity of esterase antibody 43C9 has been attributed to nucleophilic attack by an active-site histidine residue. Yet, the active site of 43C9 is strikingly similar to those of other esterase antibodies, particularly 17E8. We have carried out quantum mechanical calculations, molecular dynamics simulations, and free energy calculations to assess the mechanism involving direct hydroxide attack for 43C9. Results support this mechanism and suggest that the mechanism is operative for other antiphosphonate antibodies that catalyze the hydrolysis of (p-nitro)phenyl esters.

THE DRAFT PROBLEM: GENOMIC SEQUENCE ALIGNMENT RECONSIDERED

Chuong B. Do, Michael Brudno* and Serafim Batzoglou*

Purpose

To date, most successful methods for inferring biological function from genomic data involve comparison of nucleotide sequences from different organisms, relying on common evolutionary descent to suggest common function. Despite the recent interest in DNA sequencing, obtaining the complete genetic code is extremely difficult and prohibitively expensive for many organisms, so many genomes are currently available only in “draft” form, short unordered chunks of DNA called “contigs.” Classical sequence comparison alignment methods, however, require finished syntenic sequence pairs as input; clearly, efficient algorithms are needed for the alignment of draft sequences.

Material and Methods

We present a novel method for globally aligning a set of unfinished draft sequences from one organism (e.g. mouse) to the finished sequence of another organism (e.g. human) which extends the Multi-LAGAN sequence alignment toolkit (Brudno et al, 2002). The main features of the algorithm include incorporation of:

- 1) a scalable $O(n \log n)$ chaining algorithm with gap penalties for contig placement
- 2) a Pair Hidden Markov model for identifying regions of significant conservation
- 3) an $O(2^M MK)$ dynamic programming algorithm for optimal contig ordering on the finished sequence.

Results

Preliminary results are promising: the multiple draft alignment of 99 lemur, mouse, rabbit, and chicken contigs against human finished sequence from the APOA1 region was computed in under 9 minutes on a 2.3 GHz Pentium IV machine. Comparison of annotated human exons with the alignment showed significant conservation over the entire length of the exon in 94% of 236 cases.

Conclusion

Our proposed algorithm for extending the functionality of the Multi-LAGAN global sequence aligner is an effective though rudimentary method for aligning draft genomic sequences against a finished genomic sequence. We will discuss algorithmic improvements currently being developed, implications of multiple alignment for exon discovery, and direction for future work.

References

1. Brudno M, Do CB, Kim M, Davydov E and S Batzoglou. Multiple Alignment of Genomic Sequences. In preparation.

Web Page

<http://www.stanford.edu/~chuongdo/draft.jpg>

*These authors contributed equally.

INFERRING CONFORMATIONAL FLEXIBILITY FROM VARIATIONS IN EXPERIMENTAL MEASUREMENTS

*Irene S. Gabashvili, Michelle Whirl-Carrillo, Michael Bada, D. Rey Banatao and
Russ B. Altman*

Purpose

The techniques used to study topological relationships in complex folds and supramolecular systems differ greatly in their protocols and implementations, and appropriate combination of individual measurements is challenging. Fortunately, for the ribosome, the cellular organelle synthesizing proteins, there are high-resolution structures and hundreds of low-resolution measurements reported with independent sources of error. Accordingly, there may be an opportunity to detect statistically significant signals by combining these multiple noisy sources of structural information.

Material and Methods

We analyzed 2647 biochemical proximity measures, including cross-link, footprint and cleavage data stored in the RiboWeb knowledge base (<http://sophia.stanford.edu/kb-pub/browser/login.asp>). Experimentally determined distances were compared with 38 crystal structures. 1199 measures that did not agree with any of these structures were represented as 2398 vectors in multidimensional space. The space coordinates characterized (i) movements needed to “satisfy” biochemical data, (ii) regions where these movements would occur, (iii) uncertainties of experimental results and (iv) probabilities of variations for not having the conformational dynamical origin. Clustering approaches were used for finding reproducible and functionally valuable groups.

Results

We find that biochemical proximity results that do not match crystallographic data correlate with functionally conserved and relatively flexible or mobile regions within the structure. We found coherency in both distribution and direction of mismatches that suggest conformational states not observed in crystals. Clustering and analysis of local geometries revealed a possible global movement within both the 30S and 50S ribosomal subunits. All contributing local movements occur in the same plane, parallel to the interface of the two subunits. The main direction of movements is approximately 45 degrees to the reversed tRNA translocation path. If the conformational changes in the two subunits are shifted in phase or localized in smaller areas, the global movement might appear as a ratchet-like reorganization.

Conclusion

Our findings show that judicious combination of multiple noisy solution measurements from different sources can provide novel biological insights into aspects of structural dynamics within molecular complexes

ESTABLISHING STANDARDS IN MEDICAL EDUCATION: USE OF STATISTICAL DATA MINING TO DEFINE PERFORMANCE CHARACTERISTICS

Kunal S. Girotra and Carla M. Pugh

Purpose

Recent mandates from the Association of American Medical Colleges (AAMC) call for greater accountability in assessing clinical performance. The female pelvic exam is just one example of several clinical skills doctors in training must learn but there is a lack of standard, objective methodology to assess performance. We have developed a method of simulation using sensors and data acquisition technology that enables us to collect performance data for several clinical skills. We will discuss the use of statistical data mining to extract and evaluate performance measures from simulated clinical examinations.

Materials and Methods

From our database of 1196 simulated female pelvic examinations and 500 simulated prostate examinations, Matlab was used to extract five pre-defined performance variables including: 1) Time to perform a complete exam, 2) Number of Sensors touched during an exam, 3) Frequency each sensor was touched, 4) Maximum Pressure used and 5) the Sequence in which the sensors were touched. As the female pelvic exam database consists of both students and clinicians, comparative analysis of these variables was possible using a statistical classification tool, CART (Classification and Regression Trees) as well as ANOVA (Analysis of Variance).

Results

The Matlab code, was verified as a reliable variable extraction method by comparing the outputs with manual extractions previously performed on a subset of 250 examinations. Descriptive statistics performed on these variables reveal obvious patterns in examination technique for both the female pelvic and prostate examinations, including, for example, the anatomical areas most commonly touched by 90% of the clinicians, the areas that weren't touched as frequently and the average pressures used. Results from the ANOVA's show significant differences in exam techniques when comparing student and clinician data. The results of the CART analysis identified specific characteristics and patterns in exam technique that account for the differences observed when comparing students and clinicians. CART also highlighted and ranked the importance of our pre-defined variables, enabling us to understand certain characteristics about exam technique. In addition, results from the CART analysis showed cross validation accuracies greater than 90% for the variables we defined.

Conclusion

Our preliminary analyses show that clinical performance can be quantitatively defined and that there are specific, objective differences in performance when comparing students and experienced clinicians.

Web Page

<http://www.stanford.edu/~cpugh/bcats.htm>

LARGE MULTIPLE SEQUENCE ALIGNMENTS AND PHYLOGENETIC ANALYSIS OF THE CFTR LOCUS

Gregory Cooper and Arend Sidow

(collaborations: Eric Green, Serafim Batzoglou)

Purpose

Comparative sequence analyses present a paradigm that not only increases our understanding of evolution, but also provides a powerful tool for discovering and characterizing noncoding regulatory elements that might be elusive to other techniques. We present an exploration of a sequence data set covering in excess of 1 Mb of genomic DNA containing and surrounding the CFTR gene. The sequences (courtesy Eric Green, NHGRI) are derived from 9 mammals, chicken, and two fish. mLAGAN, a large-scale multiple sequence aligner developed by Batzoglou's group (Stanford CS Dept), was used to align the sequences.

Materials and Methods

'Sliding Window' Rate Calculations

Briefly, this analysis consists of using successive, overlapping windows of the alignment and a known species topology in a maximum likelihood procedure to estimate branch lengths; it is similar in nature to previous work used to quantify local rates of evolution in proteins (Simon *et al*, 2002). These branch lengths are summed across the entire tree to generate window-specific rates. Such rates are analyzed to highlight regions of significant conservation to identify potential regulatory sites.

Mammalian Rate Estimates

To obtain accurate estimates of branch length within and between mammalian lineages, specific high-quality portions of the alignment, including non-coding, UTR, and protein-coding sequence, were extracted, concatenated, and used for maximum likelihood tree estimation. Estimates of 'background' rates of evolution can also be generated.

Results

Varying numbers and lengths of conserved non-coding sites are found. Estimates range in size up to in excess of 100 bases at levels of conservation as stringent as 0.05 subs/site. A minority of such sites can be annotated with both literature-based binding sites and simple consensus motif matching.

Use of large (3kb to 1.1Mb) subsets of the global alignment allows accurate mammalian relative rate estimates. Our data suggest that relative rates are highly stable: estimates of branch length are tightly correlated across all of our data subsets, and also correlate well with other data sets in the literature.

Conclusions

Early results of the analysis of the CFTR locus show great promise for comparative genomics analyses. We are able to rapidly generate hypotheses about potential regulatory sites, as well as confirm or make suspect previously generated hypotheses. Additionally, our phylogenetic analysis suggests stable relative rates of evolution in mammals across various genomic sites and types of sequence. Such rates will be useful when exploited to determine the relative information value of a genome.

References

1. Alexander L. Simon, Eric A Stone, and Arend Sidow. Inference of functional regions in proteins by quantification of evolutionary constraints. *PNAS* 2002 99: 2912-2917

SINGLE TRIAL CLASSIFICATION OF INDEPENDENT SOURCES FROM MEG TASK RELATED RECORDINGS

Marcos Perreau Guimaraes, Dik Kin Wong, E. Timothy Uy and Patrick Suppes

Purpose

We used Independent Component Analysis (ICA) [1] to improve single trials classification results on MEG recordings of event-related tasks [2].

Material and Methods

Seven different words are presented to the subject while 146 MEG channels are continuously recorded at a sampling frequency of 1kHz. Two sessions are performed, the first with a visual presentation of the words, one at a time, the second with an auditory presentation. Each word is presented 50 times in a random order. Some EEG sensors, EKG and ocular movements are also recorded but are not used in this study. The MEG data is first down-sampled from 1kHz to 31.25Hz. Then an unmixing matrix is estimated and this matrix is used to recover 146 independent sources. The classification is done on each source using a simple nearest neighbor method with 4 parameters as described in [2] but with a 3 bins cross-validation setup. A third of the trials are averaged to build 7 prototypes. A second third is used to find the best set of parameters in a 4 dimensional grid. The last third, the cross validation bin, gives the actual classification rate. Two parameters define a pass-band filter and the two other the time clipping of the trials before the distance is computed. In order to have significant statistical results the rate is computed 20 times with different random permutations to build the 3 bins and the results given below are the averages.

Results

The classification performed directly on the down-sampled version without ICA gives very poor results in both cases visual and auditory with rates close to the chance level ($1/7=14\%$). After ICA the classification rate for the visual presentation is 39% which is significantly above chance level. The rate for the auditory presentation is lower, 30%, but still significant.

Conclusion

These results not only confirm results previously obtained on EEG recordings of the same experiment but the finer spatial resolution of the MEG recordings gives us a better way to localize the sources relevant to the tasks performed by the subject. As far as we know, these classification results for MEG recordings of verbal stimuli are among the best yet reported in the literature

References

1. Analysis and Visualisation of Single-Trial Event-Related Potentials. T.P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, T. J. Sejnowski. *Human Brain Mapping* 14:166-188, 2001.
2. Suppes P., Lu Z. L., Han B.: Brain wave recognition of words. *PNAS*. Vol 94. pp 14969-14969, Dec. 1997.

PREDICTING GROWTH-GENERATED STRAINS DURING CRANIAL DEVELOPMENT

James H. Henderson and Dennis R. Carter

Purpose

It is widely believed that rapid growth of the mammalian brain creates tensile strain in developing cranial sutures.¹ Despite interest in these growth-generated strains as possible regulators of sutural osteogenesis, there currently exists neither experimental substantiation of their existence nor model-based calculation of their magnitude. In this study, we implement a linear-elastic thin-walled pressure vessel model of cranial development to predict sutural strain magnitude and provide an analytic understanding of the developmental factors that influence strain generation.

Methods

Using an iterative representation of cranial growth and development, we calculate sutural strain magnitude as a function of age using equations developed from a thin-walled pressure vessel simplification of the cranium. To further gain insight into the generation of these strains, we introduce a method for calculating sutural strain based on areal brain changes. Brain volume data and geometrical analysis reveal sutural apposition rate and intracranial calvarial surface resorption rate.

Results

The model predicts residual (instantaneous) growth-generated sutural strains to be greatest during early development, on the order of $20\mu\epsilon$ in 6-month-old humans and $300\mu\epsilon$ in week-old rats. Calculated residual strains in humans decrease rapidly with age: during the period from 6 months-of-age to 4 years-of-age residual strains are estimated to decrease between 69% and 98%.

The magnitude of growth-generated sutural strains is shown to depend primarily on: (1) the ratio of actual intracranial volume (the volume of blood, brain, and cerebrospinal fluid) to intracranial volume of the unstrained cranium; and (2) the ratio of calvarial tensile modulus to sutural tensile modulus.

Human cranial growth generated by the special case of sutural apposition alone requires growth rates ranging from $89\mu\text{m}/\text{day}$ at birth to $39\mu\text{m}/\text{day}$ at 6 months and $3\mu\text{m}/\text{day}$ at 4 years. Metopic fusion in infants is predicted to increase sutural growth rates without sutural strains changing significantly. Growth rates increase with increasing strain/strain rate.

Conclusion

We have presented a computational approach for calculating sutural strain magnitude during cranial development. This approach reveals the relative importance of mechanical, geometrical, and structural factors in the generation of sutural strains. Our model provides a basis for determining the contribution of biomechanical factors to normal and pathological cranial development.

References

1. Cohen, M. M., Jr. Sutural biology and the correlates of craniosynostosis. *Am J Med Genet* 47:581-616; 1993.

COHERENT ARRAY PROCESSING FOR COST-EFFECTIVE REAL-TIME ACOUSTIC IMAGING

*Jeremy Johnson, Ömer Oralkan, Mustafa Karaman, A. Sanli Ergun and Butrus T.
Khuri-Yakub*

Abstract

In coherent array imaging, the complexity of the front-end electronics scales directly with the number of active parallel channels. For full phased array (FPA) imaging, all transducer elements (N) are active during both transmit and receive. Certain applications limit the number of processing channels such that there must be fewer channels than transducer elements. For these applications, we propose a new phased subarray (PSA) imaging method that employs a subset of M adjacent transducers multiplexed over the full transducer array in consecutive firings. A small number of beams are acquired from each active subarray. The low-resolution images are upsampled, filtered, weighted, and summed coherently to form a final high-resolution image. The PSA reduces the number of front-end parallel channels by a factor of N/M and achieves a resolution equivalent to FPA imaging. Both simulations and experimental data agree with the theoretical predictions. PSA imaging can easily be extended for use with 2D arrays for 3D imaging.

ALTERNATIVE SENSORY REPRESENTATIONS OF THE VISUAL WORLD

Neel Joshi, Dan Morris and Kenneth Salisbury

Purpose

Given the current trend toward miniaturization of electronics, it is feasible that in the near future a person will easily carry any number of powerful electronic devices on his or her body. The availability of powerful computers in increasingly small packages will allow algorithms in computer vision - a field that has traditionally been constrained to high-end desktops - to run on laptops and handhelds.

Consequently, advances in computer vision will soon be able to contribute to a traditionally "low-tech" field: assistive devices for the blind. The goal of this project is to develop a system that extracts information from a moving camera and presents important or especially salient visual features to the user's non-visual senses.

When we visually explore a scene, we quickly obtain a tremendous amount of information. We are able to identify people, navigational obstacles, textures, etc. We ask the question: "how much of this information can be rendered using alternative sensory representations?"

Material and Methods

We have used two cameras and a Sensable Technologies "Phantom" force-feedback haptic display to visually and haptically render a three-dimensional surface that accurately represents key aspects of a visual scene. In particular, our system can be used to "feel" a depth map of the visual scene or to "feel" the contours defined by edges in the visual scene. The system runs in real-time; the user presses a button to trigger the capture and rendering of either a depth map or an edge-detected image.

Once an image has been rendered, the user can use the Phantom to explore the entire surface. See Sensable's website for a more detailed description of the Phantom's capabilities; in short, the sensation of using a Phantom is similar to running one's finger over an object. The force-feedback is extremely convincing, although it is currently limited to a single point of contact.

In addition to rendering depth and edges with the Phantom, we capture optic flow information in real-time and present this to the user using sound cues: a tone pans from left to right according to the horizontal location of maximum optic flow and scales in volume according to the magnitude of optic flow (the system is silent for sub-threshold flow).

Conclusion

We propose that with further development, this system could be used as an assistive device that would allow blind or visually impaired individuals to explore the "visual" world.

Web Page

<http://www.stanford.edu/~neel/haptic.scene.representation/>

GENEXPRESS: VISUALIZATION AND ANALYSIS OF GENE EXPRESSION AND SEQUENCE DATA

*Amit Kaushal, Roman Yelensky, Tuan Pham,
Eran Segal, Nir Friedman and Daphne Koller*

DNA microarray technology is currently producing a wealth of gene expression data on a genome-wide scale. Much work has focused on clustering genes and experiments with similar expression level, since identifying genes with similar expression profiles may indicate participation in common cellular processes. In addition, sequencing projects have produced vast amounts of sequence data, and many efforts have been devoted to the task of identifying common regulatory sequence motifs in promoters of genes, as these motifs may be sites to which molecule bind and induce or inhibit expression. Computational tools and algorithms that address these challenges are rapidly being developed. However, the output of such clustering and motif finding programs rarely provides us with complete information and additional analysis steps are usually required to interpret the quality of the output. Moreover, the outputs are usually complex and therefore requires good visualization tools. For these two purposes, we developed GeneXPress, an analysis and visualization tool for gene expression and sequence data.

GeneXPress has two main components. The first handles visualization of gene expression profiles in several different customizable and interactive views. The second component addresses promoter sequence data and displays the motifs as well as the locations in the promoters of genes where the motif appears. In addition, GeneXPress provides automatic analysis tools. One analysis allows loading external predefined gene annotation files and checking the clusters for enrichment of these annotations. Another analysis tests the uniqueness of appearances of binding sites in promoters of genes in the cluster compared to appearances elsewhere in the genome.

As an example, assume we obtained expression data under some conditions. The standard approach is to cluster the data using one of the many available clustering algorithms. Then, we might take the clustering result and apply some motif finding technique to search for motifs in each of the clusters found by the clustering program. GeneXPress can then be used to analyze and visualize both the clustering and motif finding results in a few easy automatic steps:

- Load the clustering result and visualize the resulting clusters.
- Load gene annotation files (e.g. Gene Ontology) and perform a global analysis resulting in a sortable excel-like sheet displaying for each cluster, the fraction of genes that have each annotation. Statistical significance measures are provided. Thus, in a few minutes we can determine which clusters are enriched for which gene annotations.
- Load the motifs found and visualize the consensus of each motif, its information content, the location within the genes' promoter where the motif is present.
- Perform a global analysis of motifs resulting in a sortable excel-like sheet displaying for each cluster, the fraction of genes with the motifs in their promoter. In addition, GeneXPress tests the uniqueness of the motif appearance in the cluster and provides a statistical significance measure for each such test.

To make the tool broadly applicable, the input is in XML format and can readily display clustering results and motif binding sites from many types of algorithms.

GeneXPress is freely available for academic use at <http://GeneXPress.stanford.edu>.

PYTHON INTERFACE BETWEEN PHYLIP AND R

*Martina Koeva and Susan Holmes**

*Department of Statistics

Purpose

Biologists are often interested in assessing bootstrap confidence level values for phylogenetic trees. A method for estimating such values, which is based on two-level bootstrap analysis, as described in (Efron, Halloran and Holmes, 1996) has been implemented through a Python interface between R and PHYLIP (phylogenetic software). The input required by the program is a DNA sequence data set and a clade of interest, for which the user would like a confidence level value calculated, and the output generated by the program is a single value, corresponding to the bootstrap confidence level. Some additional information needed from the user also includes the tree-building algorithm to be used by the program for the construction of the trees, the number of replicates at 1st level of bootstrap, the number of replicates at 2nd level of bootstrap. The program uses the PHYLIP package and its implementations of the tree-building algorithms for the automatic construction and analysis of the phylogenetic trees. It allows the use of maximum-likelihood (DNAML, DNAMLK), distance-based (FITCH, KITSCH, NEIGHBOR), and parsimony (DNAPARS) tree-building algorithms.

Results

The program then provides a corrected bootstrap level of confidence for the clade entered by the user.

References

1. Efron, B. Halloran E. and Holmes S. (1996) Bootstrap confidence levels for Phylogenetic Trees, PNAS , 13429--13434.
2. Felsenstein, J. Phylip computer package: URL:
<http://evolution.genetics.washington.edu/phylip.html>

LARGE-SCALE COMPUTATIONAL PROTEIN DESIGN: SEQUENCE SPACE, STRUCTURE PREDICTION, AND PROTEIN EVOLUTION

Stefan M. Larson, Jeremy L. England, Amit Garg, John R. Desjarlais and Vijay S. Pande

Purpose

Defining the sequence-structure relationship in proteins requires a clear understanding of the nature of sequence space, the set of all amino acid sequences for a specific protein structure. We apply an all-atom protein design algorithm on a very large scale to thoroughly sample the sequence space of hundreds of existing protein structures.

Material and Methods

We have established a public distributed computing cluster of several thousand processors (Genome@home) for large-scale sequence design to protein structural ensembles¹.

Results

The diversities of designed sequences for similar protein structures cluster tightly together. The similarity of sequence entropies within a given fold and the separation of entropy distributions for different folds suggest that the diversity of a structure's sequence space is determined by its overall fold and that the designability principle from simple models holds true in real proteins¹.

Designing to a structural ensemble produces a much greater diversity of sequences than the traditional method of designing to a fixed backbone. Homology searches against natural sequence databases show that the quality of these sequences is not diminished. We have used the increased coverage of sequence space to develop the "Reverse BLAST" method, which more than doubles the number of detected structural templates for comparative modeling of genomic sequences².

A comprehensive comparison of designed and natural SH3 domain sequences showed that optimization for native state structure accounts for the sequence conservation patterns at 85% of the positions in natural SH3 domains. A strong link between optimization for the native state structure and a consistent role in the transition state (as modeled by phi-value analysis) is seen. We propose that fast folding through a particular transition state requires optimization of residues for the native state structure, particularly in the highly structured portions of the transition state.

Conclusion

All-atom protein design on a very large scale has allowed for fundamental insights into the nature of sequence space, with immediate applications in structure prediction and protein evolution.

References

1. Larson SM, England JL, Desjarlais JR, Pande VS. "Thoroughly sampling sequence space: large-scale protein design of structural ensembles." *Prot Sci*. 11(12), 2002.
2. Larson SM, Garg A, Desjarlais JR, Pande VS. "Increased detection of structural templates using alignments of designed sequences." submitted to *Proteins*, 2002

Web Page

<http://genomeathome.stanford.edu/>

CT COLONOGRAPHY PRONE AND SUPINE DATA REGISTRATION

P. Li, S. Napel, B. Acar, D. Paik, R. B. Jeffrey and C. F. Beaulieu

Purpose

The primary goal of Computed Tomographic Colonography (CTC) is to detect colonic polyps. A patient is typically scanned both prone and supine to overcome the deficiencies due to (1) insufficient cleansing and air-insufflation that can hide a polyp in one view; (2) retained stools that mimic polyp structures. However, observable deformations could be expected

from prone to supine images. Also, a radiologist's work doubles for reading images and interpreting results from Computer Aided Detection (CAD). Therefore, an algorithm that registers prone and supine CTC data could be very beneficial.

Material and Methods

We developed a two-step registration algorithm: (1) path registration, to registers prone and supine colon central paths; (2) polyp registration, to match prone and supine polyps.

The path registration algorithm is based on the morphological similarity between prone and supine colon central paths. The high curvature sections close to the colon boundary are the most stable sections and can be used as the landmarks, which are simplified as the local extreme points. The algorithm recursively and iteratively matches the landmarks, and linearly shrinks/extends the points in between landmarks.

The polyp registration algorithm is statistics-based. Three parameters are identified: (1) a polyp's path distance (after applying path registration algorithm); (2) a polyp's orientation on the cross-sectional plane; (3) a polyp's CAD intensity score. The joint density distribution of three parameters, which are each modeled to follow a "Folded-Normal Distribution", is used as the score function for polyp matching. The optimum matching scheme is the one that maximizes the sum of all matching scores. A Dynamic Programming algorithm is developed to efficiently solve the optimization problem in linear time.

Results

The registration algorithm was evaluated using a dataset of 24 patients. For each case, a radiologist manually identified 5 pairs of identical points from prone and supine images. After path registration, the average misalignment distance over 120 pairs was reduced from 47.08 mm to 12.66 mm.

The polyp registration algorithm was applied to 5 cases, whose polyp matching schemes were identified by a radiologist. The results showed all true polyp pairs were successfully matched.

Conclusion

Initial results shows that the CTC data registration algorithm is promising in automatically matching the colon central path and polyps. The algorithm can enhance the efficiency and accuracy of radiologists' reading and CAD algorithm.

Web Page

<http://www.stanford.edu/~pingli98/BCATS2002/registration.ppt>

APPLYING DISCLOSURE CONTROL METHODS TO BIOMEDICAL DATA

Zhen Lin, Micheal Hewett and Russ Altman

Tremendous amounts of patient data are collected through clinical research. These data include information about diseases, treatments, and patient responses. With the recent development of inexpensive DNA sequencing and analysis technologies, these data are enriched with genomic information containing an unprecedented amount of very specific information about individuals. To maximize the use of these data for research analysis, central data repositories, such as PharmGKB, are established to store and distribute biomedical data collected from other research centers. Because advances in science are more likely to benefit from wide availability of research data, we would like to make these pharmacogenomic data publicly available as detailed as possible to facilitate studies of correlations between genomic variations and drug responses.

Even though we only accept de-identified patient data at PharmGKB, with no name, address, SSN and other explicit personal identifiers, these data are not anonymous. It is possible to re-identify research subjects by matching their combination of unique sensitive information with knowledge outside of the central data repository. Therefore, to allow these data be accessible from anonymous researchers without comprising research subjects' privacy and data confidentiality, we must develop techniques to assure the anonymity of the data and minimize the risk of individual re-identification.

We intend to build a confidentiality module and dynamic query system so that query results from anonymous researchers are filtered and modified, and if necessary, queries are reformed to ensure that data confidentiality is appropriately maintained. We build upon the previous idea that if records are binned into groups, they are more difficult to trace back to any single individual. Our system ensures that only values shared by at least a predefined minimum number, bin size, of records are available to the query users. We represent symbolic and numeric data hierarchically, and bin them by generalizing the record. The bin size parameter controls the tradeoff between confidentiality and data integrity. We measure the information loss due to disclosure control techniques using an information theoretic measure called mutual information.

Previously, we decomposed the binning problem into two steps: (1) single-attribute binning, then (2) multiple attribute binning. An attribute is a type of sensitive data, e.g. diagnosis. The first step bins each attribute individually so that each value is shared by at least the bin size of records. The second step bins each combination of attributes so that each combination also fulfills its bin size requirement. The bin size for single attributes may be different than that for multiple attributes. Also, the user may specify the order of binning to preserve some attributes over others. The preliminary results show that we can bin the data to different levels of precision, yet the information loss is enormous. To minimize the information loss and maximize the research potential, we are currently exploring an alternative approach to the two-step binning method. We propose to combine machine learning methods to bin attributes concurrently.

ALTERNATIVE SPLICING: DATA STORAGE AND ANALYSIS

Shuo Liu and Russ B. Altman

Abstract

Alternative splicing is gaining recognition as an important way to modulate the expression and function of genes. However, many questions remain unanswered or only partially answered: How does the splicing machinery carry out alternative splicing? How can all splice variants be identified? What functionality is regulated by alternative splicing? In order to answer these questions on a genome scale, a data repository that is designed to host alternative splicing information is necessary. Due to the relatively recent realization of the importance of alternative splicing, most of the current data repositories are not adequately equipped to store the various facets of data for alternative splicing. In this presentation, I will describe a relational database that I have designed from the ground up to house alternative splicing data, discuss the analyses that could be carried out with the database, and share some of my findings.

SEARCHING FOR EXONS WITHIN INTRONS AND USING MUTUAL INFORMATION TO HELP FIND AND DEFINE DNA REGULATORY MOTIFS

Brian T. Naughton and Douglas L. Brutlag

Purpose

Ematrix is a software package developed in the Brutlag lab which uses minimal risk scoring matrices to find motifs within protein sequences. We are using this software to search for protein motifs within introns, translated in all three frames. We are using the entire NCBI Reference Sequence database of human genes for the intron data. Refseq is a stable, non-redundant database. This task is computationally very intensive and has taken several weeks running on a supercomputer. This experiment has given us a broad idea of how many exons are currently unidentified in the human genome within introns and has shown motif-finding to be an effective way to identify these exons.

There are several programs available for finding conserved DNA motifs in upstream regulatory regions of co-expressed genes, including BioProspector, which was developed in the Brutlag lab. The mutual information between nucleotides within the motif is not taken into account in these programs, however. Mutual information is a measure of how much information is granted about a certain nucleotide given the identity of another nucleotide. Work by Rozkot et al. suggests that this information may be important for defining many motifs in *E.coli*. We posit that this information may also be useful for a significant fraction of motifs throughout nature. I am writing a software program which uses Gibb's sampling similarly to BioProspector to find motifs, using as the scoring function the sum of the information in the motif plus the mutual information within the motif and possibly between neighbouring motifs.

Material and Methods

Ematrix¹ was used in the first experiment.

Results

We have compiled a table of putative motifs that we have found within introns. Approximately 1000 hits have been hand-annotated, with a high percentage shown to be functionally related to the gene in which they were found. Many exons previously known to be alternatively spliced and rediscovered with our technique act as a positive control in this experiment.

Conclusion

There are many unidentified exons in the RefSeq database. Using motif-finding algorithms can significantly aid in the discovery of these exons.

References

1. Thomas D. Wu, Craig G. Nevill-Manning, and Douglas L. Brutlag, "Minimal-risk scoring matrices for sequence analysis," *Journal of Computational Biology* 6: 219-235, 1999

ANALYSIS OF DNA MICROARRAYS USING KNOWLEDGE-BASED ALGORITHMS

Kuang-Hung Pan, Chih-Jian Lih and Stanley N. Cohen

Purpose

DNA microarrays have been widely used to measure the relative abundances of mRNA for thousands of genes in various experimental conditions. Previously, methods of unsupervised classification of gene expression profiles have been applied, such as hierarchical clustering and self-organizing maps. However, researchers employ DNA microarrays not merely to monitor the mRNA levels but to address their biomedical questions, such as mechanisms for a phenomenon, pathways involved in a phenotype, fingerprints of a biological state, and the relationships between genes. Post-hoc manual application of knowledge through visual inspection was used, but such a mode was not systematic or efficient. To address this issue, we develop GABRIEL (Genetic Analysis By Rules Incorporating Expert Logic), a platform that incorporates knowledge into computer software to assist users apply knowledge in microarray analysis.

Material and Methods

To develop such a platform for incorporating knowledge, two main challenges are: 1. What knowledge to incorporate. To solve this, GABRIEL can determine the relevancy of knowledge by extracting statistically robust relationships between pieces of information. 2. How to incorporate knowledge. Domains of application for microarrays are numerous and diverse; traditional knowledge-based systems cannot handle such a situation. Thus, we separate knowledge into modules that are reusable to make GABRIEL a scalable system.

Results

GABRIEL has been applied to three different biological questions: 1. Antibiotics synthesis pathways in *Streptomyces*. With GABRIEL, novel genetic pathways were identified. 2. Human replicative senescence. Using GABRIEL, molecular fingerprints of senescence were characterized. 3. *Streptomyces* deletion mutants with circularized chromosomes. By GABRIEL, the deletion and amplification regions of these mutants were characterized.

Conclusion

The ability of GABRIEL to incorporate multiple pieces of knowledge into microarray analysis expands the utility of microarrays from monitoring the ups or downs in gene expression level to addressing a variety of biological questions, such as pathways, cellular state characterization, and deletion/amplification detection.

GABRIEL is a proposed scalable platform to incorporate transferable knowledge through knowledge-based algorithms. With GABRIEL, the utilization of knowledge is made more systematic, organized, and efficient.

References

1. Pan K.-H, Lih C.-J., Cohen S. N., *PNAS*. 99:2118-23, 2002.
2. Huang J., Lih C.-J., Pan K.-H., and Cohen S. N., *Genes & Development*. 15:3183-3192, 2001.

Web Page

<http://gabriel.stanford.edu/>

SUPPORT FOR GUIDELINE DEVELOPMENT THROUGH ERROR CLASSIFICATION AND CONSTRAINT CHECKING

Mor Peleg

Purpose

Clinical guidelines aim to eliminate clinician errors. Computer-interpretable guidelines (CIGs) can deliver patient-specific advice during clinical encounters, which makes them more likely to affect clinician behavior than narrative guidelines. To reduce the number of errors that are introduced while developing narrative guidelines and CIGs, I studied the process used by the ACP-ASIM to develop clinical algorithms from narrative guidelines.

Material and Methods

I observed and recorded ACP-ASIM experts as they created flowchart versions of clinical algorithms based on two narrative guidelines that they had created previously, for treating migraines. I used a classification scheme proposed by Knuth to classify changes between narrative guideline text and the clinical algorithm produced from it. This scheme was developed to classify changes between requirement documents and resulting software.

I used Protégé-2000 to develop an authoring and validation tool for CIGs. Protégé enables defining allowed data types and checking them, establishing cardinality constraints, and setting limits on numerical values. I used Protégé's axiom language to define integrity constraints in a subset of first-order predicate logic

Results

The medical expert who created the algorithm made the following types of changes between algorithm versions: (1) logic changes, (2) addition of details, (3) complexity management, and (4) omissions.

The changes between the original narrative guidelines and the final version of the clinical algorithms were classified as: (1) better user interaction, (2) clarity, (3) quality improvement, (4) omission, (5) generalization/specialization, and (6) changes that were a result of confusion.

I used the authoring/validation tool to author guidelines and to check them for errors. The tool identified errors such as decision steps that were linked to single decision options and instances of nodes that were not part of any algorithm.

Conclusion

Using the CIG authoring and validation tool might help the ACP-ASIM team create algorithms that are valid and clear. By looking specifically at the process of algorithm creation and following it closely, we can recommend that the ACP-SIM team should: (1) guarantee that all relevant information is carried from the narrative guideline to all versions of the clinical algorithm, (2) provide all the information necessary to rank treatment options, and (3) consider different patient scenarios.

References

1. M Peleg, VL Patel, V Snow, et al., "Support for Guideline Development through Error Classification and Constraint Checking", to appear in Proc AMIA Symp. 2002.

MODELING MUTATIONS, ABNORMAL PROCESSES, AND DISEASE PHENOTYPES USING A WORKFLOW/PETRI NET MODEL

Mor Peleg, Irene S. Gabashvili and Russ B. Altman

Purpose

Predicting the molecular- and cellular-level effects of genetic mutations is a challenging task. It calls for models that integrate different data sets, and represent the interactions of mutated gene products with other cellular components, in order to understand their effects on molecular, cellular, and organism-level processes. We have developed a graphical knowledge model for representing molecular functional information as a first step towards modeling the relationship between molecular structure and disease phenotypes.

Material and Methods

Our model is based on a Workflow model that can be mapped to Petri Nets, and is implemented as a frame-based knowledge base using the Protégé-2000 tool. We use TAMBIS and the UMLS as to describe biological and medical concepts that can be mapped to the participants, roles, and processes in the workflow model.

Results

The formal nature of our model allows us to write queries about structural and functional aspects of biological systems –such as relationships between defective processes and the clinical phenotype of the mutation that is causing it. To illustrate the power of this model, we have used it to represent mutations in tRNA and their affects on the process of translation. Mapping the workflow model to Petri Nets enabled us to verify the soundness of some dynamic aspects of tRNA biology and to simulate system behavior in the presence of different mutations.

Conclusion

The results presented here are a first step in which we demonstrate that the knowledge model developed in another context (malaria biology) is capable of capturing a qualitative model of tRNA function. We have shown that the resulting qualitative model can be queried to show the relationships between mutation and their sequelae.

References

1. Mor Peleg, Irene S. Gabashvili, and Russ B. Altman, “Qualitative models of molecular function: linking genetic polymorphisms of tRNA to their functional sequelae“, submitted to P. IEEE

Web Page

<http://smi-web.stanford.edu/projects/helix/pubs/process-model/>

TELEOPERABLE DERMATOLOGY: A FEASIBILITY STUDY FOR HAPTIC DISPLAYS

Kirk Phelps and Ken Salisbury

Purpose

There are many tasks for which humans require both kinesthetic and tactile sensing. For instance, a dermatologist attempting to remotely palpate a patient's skin depends on both these modalities in order to correctly diagnose certain skin diseases. Current haptic display of remote and simulated environments falls far short of this requirement, largely due to hardware limitations. Although there have been perceptual experiments which have investigated the interaction between kinesthetic and tactile displays, they invariably employ limited dof, custom-built hardware. Therefore, we set out to study the feasibility of creating a robust and generalized haptic display capable of displaying both kinesthetic and tactile information. Once built, the system would allow simulation of virtual and remote environments in a way previously unexplored in the field of haptics.

Material and Methods

A haptic device was assembled from a 3-dof force display and a 96-pin piezoelectric tactile display, both of which are commercially available products. Basic haptic environments were built using OpenGL and the Robotics Laboratory's SAI library.

Results

We have successfully constructed a haptic display capable of displaying a translation vector at 1000 Hz and an 8 bit tactile profile on 3 fingers at 15 Hz. Software and hardware interfaces are such that the device can interact with a wide range of simulated environments, including medical deformables. We look forward to using this display to investigate the interplay between kinesthetic and tactile information, especially as it relates to task performance in simulated and remote environments.

**LIKE STANDARDS BUT BETTER:
SHARING NON-TRADITIONAL HEALTH DATA AND METADATA**

*Zachary Pincus, David L. Buckeridge MD MSc, Michael K. Choy,
Justin V. Graham MD MS, Martin J. O'Connor MSc, Mark Musen MD PhD*

Purpose

“Non-traditional” health data, such as school absenteeism and pharmacy sales, are used in many syndromic surveillance systems. Such data is thought to increase the sensitivity and timeliness of their output on the assumption that aberrations in population health states can be detected from such data before confirmed clinical diagnoses become available. A standard description scheme for non-traditional data and their sources will greatly increase the utility of such data by facilitating their use, communication, and processing. For example, reliable data aggregation, important to minimizing the noise and non-specificity inherent in non-traditional, pre-clinical data, requires a standard way to describe the format of the data and the context of the data sources. (E.g. knowing the location and enrollment of a school is necessary to combining absentee counts with those from other schools.) Data sharing and intelligent automated data processing also require standard formats and a mechanism to describe the data sources.

Standards such as those proposed in the National Electronic Disease Surveillance System and the Health Level 7 Reference Information Model enable interchange of traditional epidemiological and clinical data. Unfortunately, such standards were not designed to accommodate the needs of non-traditional data and their sources. To meet these needs, we have developed a flexible, extensible ontology (metadata model) for characterizing non-traditional information sources and the specific data that they provide.

Material and Methods

Most data standards attempt to pre-enumerate fields and formats for every needed piece of information. However, it is not yet established what types of nontraditional data and metadata are necessary or useful. Additionally, heterogeneity in the format and structure of nontraditional data makes it difficult to construct a single, comprehensive “standard.” Instead, our ontology allows users to construct detailed, customized, machine-readable descriptions of data sources, data formats, and relationships between data sources, by building them up from simple terms and attributes provided by the ontology. Instead of a specified standard format, we propose a common structure and set of elements for use in constructing custom data descriptions.

Results

New classes of data or data sources, and individual instances of these classes, can be entered through Protégé-2000, an open-source knowledge acquisition application. In an initial evaluation, we successfully used the ontology to produce a model of output data from a simulated outbreak. Currently, we are testing the ontology as a system to share nontraditional health data between several allied research groups across the country.

Conclusion

We are continuing to validate the ontology as a robust, flexible, and extensible structure for rapidly characterizing, describing, and communicating nontraditional data.

Web Page

<http://smi-web.stanford.edu/projects/biostorm/>

AUTOMATED QUANTIFICATION OF ARTERIAL CALCIFICATION USING CT ANGIOGRAPHY (CTA): METHOD AND EVALUATION

Raghav Raman, Bhargav Raman, Sandy Napel and Geoffrey D. Rubin

Purpose

To develop and validate an automated method of detecting and quantifying calcium in the aorta and its branches in order to quantify the distribution and severity of systemic atherosclerosis.

Material and Methods

A branched centerline from the aortic root to the femoral arterial bifurcations was defined. The arterial lumen was segmented and calcium detected using a longitudinally varying threshold based on luminal and surrounding soft tissue intensity. Fragment mass was calculated using a conversion factor derived from standards of known calcium density included in the scan field. A surface map of the aortoiliac wall showing mural calcium was created, and calcium load, distribution, and fragment statistics were quantified for each artery segment. To validate this method, virtual aortic phantoms were scanned using a CT simulator and physical aortic phantoms and 21 patients were scanned on a 16-row CT scanner using low and standard dose protocols. Scans were repeated twice to assess interscan variability. Phantoms contained mural calcium fragments of known mass (0.3-40 mg elemental calcium). Actual vs. measured mass in virtual and physical phantoms was compared using linear regression. Aortoiliac calcium was quantified in CTAs of 21 patients without clinical manifestations of atherosclerosis. Interscan variability was compared using paired t-tests.

Results

For actual vs. measured mass, combined R^2 and std. error were 0.99 and $4.5 \pm 5.1\%$ for virtual phantom and 0.99 and $7.1 \pm 6.3\%$ for physical phantom studies, respectively. Fragments less than 4 mg in mass had a std. error of $11.4 \pm 8.8\%$. The mean calcium mass difference between high and low dose scans was 0.07 mg (0.76 %, $p = 0.26$) and interscan variability was 0.22 mg (1.2 %, $p = 0.23$). Four patients had no detected calcium. In the remaining 17 patients, the mean calcified mass was 319.7 mg (19.1-1344.4 mg). Mean calcified surface area of the arterial wall was 1.45% (0.1-3.3). The mean number of fragments was 10.5 (1-22) and the mean fragment size was 42.1 ml (30.3 mg). The infrarenal aorta accounted for 67.8% of all detected calcium. Average interscan variability per fragment was 2.49 mg (3.5%) and per scan was 11.3 mg (7.4%). Fragments <4mg had an interscan variability of 19.4%. Calcium identification and quantification required a mean (\pm sd) of 1.6 ± 0.3 minutes for user interaction and 6.2 ± 1.5 minutes for processing time per patient.

Conclusion

Our method accurately and reproducibly quantifies the mass and distribution of calcification in simulated arteries and can be applied in vivo. It requires minimal user interaction and has the potential to be applied to both research and clinical evaluations of systemic atherosclerosis.

Web Page

<http://192.168.1.48/AutoRad/Calcium.asp>

COMPUTATIONAL MODELING OF T CELL RECEPTOR SIGNAL INTEGRATION

*Tim Reddy, Peter Lee, Arancha Casal, Cenk Sumen and Mark Alber**

*Notre Dame

T lymphocytes play a central role in orchestrating the adaptive immune response. In particular, they exhibit the critical ability to recognize the presence of foreign antigens, such as viruses or bacteria, in the context of infected cells or specialized antigen-presenting cells (APCs).

To initiate the T cell recognition process, a 'naive' T cell must first come in direct contact with an APC, "reading" and scanning for any abnormal patterns in the array of surface molecules presented by the APC. Specifically, T cell receptors (TCRs) on the surface of the T cell interact with peptide-MHC (pMHC) complexes on the APC surface. While a T cell expresses a homogeneous population of roughly 50,000 TCRs, an APC presents an extraordinarily heterogeneous and vast population of pMHCs - up to 40,000 unique pMHC combinations, each of which may be present 1 to over 5,000 times. The T cell is thus confronted with an extremely intricate, composite picture on the APC surface which it must rapidly decode.

Within the first 30-60 seconds of contact with the APC, the T cell makes a crucial decision to either sustain the interaction, leading to potential activation, or to disengage and move on. Learning how a T cell quickly surveys the diverse repertoire of pMHCs and how it integrates this vast array of signals into its critical activation decision is key to a thorough understanding of T cell recognition. Equally as important, it may also lead to a better understanding of certain disease evasion mechanisms. For instance, HIV 'quasi-species', and perhaps even cancer, successfully elude T cell recognition by inducing subtle changes in the pMHC populations presented at the surface of the infected cell.

Over the past decade the nature of the interaction between several pairs of TCR-pMHC has been elucidated. However, only small numbers and combinations of peptide-MHCs can be studied experimentally. It is not feasible to address the tens of thousands of different peptide-MHC combinations typically present at the T cell-APC interface. Mathematical models and computer simulations can be used to overcome such experimental limitations. If biologically-sound and experimentally verified whenever possible, a computational model such as the one we propose here can result in valuable insights into two of the central questions in T cell recognition:

- How does a T cell integrate the vast array of signals it receives from diverse TCR-pMHC interactions into an activation decision?
- How would subtle changes within the pMHC repertoire of an APC or a target cell (such as an HIV-infected cell) be perceived by the T cell?

PROTEIN FOLDING SIMULATION WITH MULTIPLEXED REPLICA EXCHANGE MOLECULAR DYNAMICS METHOD

Young Min Rhee and Vijay S. Pande

Purpose

In a molecular dynamics simulation study of thermodynamics, a representative sampling over the entire phase space is needed to obtain an accurate canonical distribution at a given temperature. For large molecules such as proteins, this sampling is usually difficult, especially at physiological temperature, because molecules tend to be trapped in a large number of local energy minima. Recently, a replica exchange method has been proposed to overcome this kinetic trapping problem, where trajectories from different temperature levels are exchanged periodically. However, this method can only be efficient with well synchronized parallel computers.

Here, we present an improved replica exchange algorithm. To enhance the efficiency of the original method, multiplexed replicas are adopted in each temperature level and exchanges between these multiplexed replicas are also tried. These exchanges eliminate the necessity for synchronization of processors, rendering the algorithm to be well-suited for a large scale distributed computing such as Folding@Home.

Methods

Using Langevin dynamics and an implicit solvent model, we have obtained 200 microseconds of aggregate simulated time of model protein BBA5 with a screensaver program distributed over the world. The simulation was started from a fully extended structure. The folding of this extended structure was monitored to measure the convergence of the simulation.

Results

With our method, extended BBA5 reaches the folded state through the configuration space within a reasonable amount of simulation time. Moreover, the rate at which the extended configuration reaches this folded state is considerably faster compared to a constant temperature molecular dynamics simulation. Indeed, this enhanced sampling results in (to our knowledge) the first example of a replica exchange algorithm that samples a folded structure starting from a completely unfolded state.

Conclusion

Compared to regular constant temperature simulation, multiplexed replica exchange method shows an enhanced convergence. However, only with exchanges using high temperature simulation, this enhancement is found to be limited. Therefore, further development is required to attain a more reliable convergence within a reasonable simulation time.

References

1. Young Min Rhee and Vijay S. Pande, Multiplexed Replica Exchange Molecular Dynamics Method for Protein Folding Simulation. *Biophys. J.* (accepted for publication).

Web Page

<http://folding.stanford.edu/>

FAST DRRs USING LIGHT FIELD RENDERING

Daniel Russakoff, Torsten Rohlfing, Daniel Rueckert and Calvin Maurer, Jr.

Purpose

Calculating synthetic X-ray projection images, called digitally reconstructed radiographs (DRRs), from CT images is an important step in 2D-to-3D fluoroscopy-to-CT image registration. Unfortunately, the standard techniques to generate DRRs involve ray casting and are generally $O(n^3)$, where we assume that n is both the size of one dimension of the DRR and the path length through the CT volume. We address this issue by extending light field rendering techniques to generate DRRs instead of standard images. Using light fields allows us to perform much of the work in a preprocessing step so that we may generate very accurate DRRs in time $O(n^2)$.

Material and Methods

Light fields were originally proposed by Levoy & Hanrahan [1] as a means of performing fast 3D rendering. Essentially, light fields are a way of parameterizing the set of all rays that emanate from a static scene. In particular, any ray can be represented by a point in 4D space, namely its intersection with two parallel planes. So, during preprocessing, the idea is to calculate a large number of images using traditional ray tracing and to map each ray's pixel value to its corresponding 4D point. Since this process creates a large amount of spatially redundant information, the light field can be compressed using vector quantization so that it will fit into main memory. Extension to DRRs is a relatively simple matter of associating each ray not with a pixel value generated by a ray tracer, but rather with a projection sum generated by a conventional DRR calculation. Then, given a new focal point and imaging geometry, each ray's projection sum can be approximated by a 4D interpolation around that ray's corresponding point in the light field. Since this interpolation can be done in constant time, the asymptotic time for a DRR calculation from a pre-calculated light field is $O(n^2)$.

Results

Using a light field generated from 4,096 initial images of resolution 256x256, we are able to generate DRRs which look essentially identical to ones generated by ray casting. However, the DRRs generated by the light field are computed over 100 times faster than those using ray casting. Specifically, on a Sun Blade 1000, we can generate a DRR with resolution 256x256 in approximately 100 milliseconds. Moreover, we've performed actual 2D-3D rigid registrations using this technique to generate the DRRs and compared the results with those from registration using DRRs from conventional ray casting. While the final results are comparable to within about 1 millimeter in translation and 1 degree in rotation, the registration using light field rendering was on the order of 100 times faster (<1 min. vs. almost 2 hours).

Conclusion

Fast DRR calculation is a very important step in 2D-to-3D image registration. Since similarity measures are generally calculated in $O(n^2)$ time, the DRR construction is often the rate determining step in the speed of the registration algorithm. With modest preprocessing time, we present a method for the calculation of very accurate DRRs in $O(n^2)$ using an extension of light field rendering techniques.

References

1. Levoy M, Hanrahan, P. Light field rendering. *Computer Graphics*. 30:31-42, 1996.

STRUCTURAL MOTIF DISCOVERY USING AN IMPROVED STRUCTURAL SUPERPOSITION ALGORITHM

Jessica Shapiro and Douglas Brutlag

Purpose

Accurate structural alignments of distantly related structures are those that have the greatest potential to yield insight into questions such as the conservation of protein structure, the types of structural interactions “preferred” in nature, and the relationships among sequence, structure, and function. While many structural alignment algorithms are capable of aligning structurally similar proteins, they tend to produce widely varying results in the more interesting cases of distantly related structures.^{1,2} We report an improved structural alignment algorithm, LOCK 2.0, which is able to detect distant structural relationships by placing increased emphasis on the alignment of secondary structure elements

Results

The LOCK 2.0 algorithm obeys three properties essential in automated analysis of protein structure: it is symmetric, it is nearly transitive at the level of secondary structure elements (SSEs), and its scoring system is empirically found to behave as a metric. Statistical significance scores are calculated using background distributions of alignment scores that are specific to the query structure’s fold. We validate LOCK 2.0 by comparison to two widely used alignment algorithms. Furthermore, LOCK 2.0’s structural similarity search not only identifies the query protein’s structural neighbors, but also finds the structural motif that is the basis for the similarity among the proteins. This information is used to increase discrimination between homologs and analogs by the region of the query to which they align, thereby increasing the sensitivity and specificity of the search. The motif is described probabilistically in terms of the conservation of each secondary structure element’s position and orientation in the query protein. We find that when two proteins of the same fold are aligned, highly conserved secondary structure elements in one structure tend to align to the highly conserved elements in the other structure, suggesting that LOCK 2.0 consistently identifies the same motif in multiple members of a structural fold.

References

1. Feng, Z. K. & Sippl, M. J. (1996). Optimum superimposition of protein structures: ambiguities and implications. *Fold Des* 1, 123-32.
2. Godzik, A. (1996). The structural alignment between two proteins: is there a unique answer? *Protein Sci* 5, 1325-38.

Web Page

<http://dlb3.stanford.edu/LOCK>

INTERACTIVE SEGMENTATION AND VISUALIZATION OF VOLUMES

Anthony Sherbondy, Mike Houston, Pat Hanrahan and Sandy Napel

Purpose

Segmentation of anatomical structures in 3D medical images is a challenging task. The current promising techniques range from fully automated snakes, level-set methods, classification, and knowledge based probability networks to the completely manual region boundary drawing by highly trained human operators. While many of the automated methods have shown great promise with specific applications, some people believe that a fully automated approach will always need a human observer. In these automated approaches, the human is often called in to verify the results or “step-in” on pathological or even healthy unique cases.

We propose an interactive segmentation method, which allows the operator to observe the segmentation in real-time and augment the algorithm when necessary. The algorithm specifically leverages the speed and programmability of today’s easily accessible PC graphics cards.

Material and Methods

By utilizing the Stanford Real Time Shading language [1], we were able to write a parallel nonlinear diffusion based segmentation algorithm that is implemented as a shader program on the ATI 9700 or NVidia Geforce 4 graphics hardware. We chose to implement a stochastic representation of the Perona and Malik heat diffusion equation [2], which favors diffusion in contiguous regions and inhibits stepping across areas of high gradients. The user is able to interact with the algorithm by choosing the initial starting point, observing the segmentation “grow” and stopping the segmentation as well as many algorithm parameters that control the diffusion process. The user is given multiple volume rendering views that utilize non-photorealistic rendering techniques as well as traditional tomogram views in order to portray the segmentation within the image volume.

Results

Currently, we are able to run at interactive rates up to a 256x256x256 volume. Images available on the website display results of segmentations and different volume rendering visualization techniques in a CT dataset of a healthy lung.

Conclusion

We have presented an algorithm for real-time segmentation of volumes that includes the user in making difficult segmentation decisions based on complicated anatomy. Our system provides good feedback to the user by allowing the user to watch the segmentation happen live and adjust and interact with the segmentation to get the desired result. We believe our method is timely in exploiting the incredible parallel processing power available in easily accessible graphics cards in order to provide a viable interactive segmentation technique.

References

1. Kekoa Proudfoot, et. al. A real-time procedural shading system for programmable graphics hardware. In *Computer Graphics Annual Conference Series (Proceedings of SIGGRAPH 2001)*, Los Angeles, CA, August 2001.
2. Hamid Krim and Yufang Bao. Nonlinear diffusion: A probabilistic view. *ICIP*. 1999

Web Page

http://graphics.stanford.edu/~mhouston/Segmentation/BCATS_figures.html

SOURCE: THE STANFORD ONLINE UNIVERSAL RESOURCE FOR CLONES AND ESTS

G. Sherlock, T. Hernandez-Boussard, M. Diehn, A. Alizadeh, J.C. Matese, G. Binkley, H. Jin, J. Gollub, J. Demeter, J. Hebert, C.A. Ball, P.O. Brown and D. Botstein

Purpose

The Stanford Online Universal Resource for Clones and ESTs (SOURCE) dynamically compiles scientific information from many biological public databases that can be used to annotate human, mouse or rat clones and genes.

This project originated as a resource to summarize the diverse sources of information into simple reports in order to annotate clones that were spotted onto cDNA microarrays. These 'gene reports' facilitate the analysis of the large datasets that biologists are now producing using genome-scale experimental approaches.

Results

SOURCE displays a summary of gene expression data gleaned from select public microarray datasets. In this way, SOURCE's gene report provides a portal for published microarray data. In addition, SOURCE is linked to the human genome assembly, to allow one to see where a clone lies on the genome sequence. The process of mapping products to an organism's genome can provide valuable information not only for comparative genomic hybridization (CGH) experiments, but also for gene expression analysis and promoter identification. Effective genome annotation display is therefore becoming critical for biological analysis.

We plan to further implement a physical map of where the gene and clones mapped to the gene via Unigene lie on the genomic sequence. In addition to the single gene or clone reports, SOURCE supports batch queries and will return a file of annotations for a list of gene names or clones.

Conclusion

SOURCE represents the leading edge of development as a complete annotation source for large-scale experimental research in mammalian systems.

Web Page

<http://genome-www.stanford.edu/source>

AN ELASTIC MODEL OF THE MALLEUS-INCUS COMPLEX OF THE HUMAN MIDDLE EAR

Jae Hoon Sim and Afraaz R. Irani

Purpose

Mathematical and experimental analyses of three-dimensional motion of the temporal bones of the human middle ear have been done based on the anatomy and mechanics of the human middle ear. The purpose of these studies is to understand transfer functions of the human middle ear and apply these results to surgical prostheses used to reconstruct damaged temporal bones. In most of the models of the human middle ear, temporal bones were described as rigid bodies and the joints were fixed to simplify the motion of temporal bones. However, this representation, which was widely accepted in the past, is not enough to describe the complicated sound transmission for frequencies above about 2 kHz.

Our elastic model, where each temporal bone is considered a compliant body and the slippages at the joints are allowed, has been developed to more accurately describe sound transmission through the full audible range. Our model will lead to a more comprehensive understanding of the temporal bone function and help develop prostheses that more efficiently transmit sound to the inner ear.

Material and Methods

We introduce a mathematical model of the malleus-incus complex, which includes the effect of tension, bending, and twisting of temporal bones and the relative motion at the malleus-incus joint. To verify this mathematical model, the three dimensional motions of the malleus-incus complex are measured using a laser interferometer. Measurements are made at three specific angles to calculate xyz components of the velocity vector. A coil around the tympanic annulus and a tiny magnet are used to drive the malleus-incus complex.

Results

Several different vibration modes of the malleus-incus complex will be presented at the frequencies where the corresponding modes are dominant. The three dimensional motion of the end point of incus handle will be measured with the umbo driven by the magnet attached to it. The real motion of this point will be compared with the motion of the mathematical model calculated. The mathematical model and measurement data will be used together to estimate unknown parameters.

Conclusion

Our elastic model makes it possible to describe the sound transmission of the malleus-incus complex through the whole audible frequency range. The measurement method is developed to verify the mathematical model and estimate parameters.

References

1. Decramer WF, Khanna SM, Feunnel WRJ. A method for determining three-dimensional vibration in the ear. *Hearing Research*. 77:19-37, 1994
2. Schön F, Müller J. Measurements of ossicular vibrations in the middle ear. *Audiol Neurootol*. 4:142-149, 1999
3. Eiber A et al. Dynamics of the middle ear prostheses simulations and measurements. *Audiol Neurootol*. 4:178-184, 1999

IDENTIFYING STRUCTURAL MOTIFS IN PROTEINS

*Rohit Singh^{*ab} and Mitul Saha[†]*

^{*}Accelrys Inc, San Diego

[†]Department of Mechanical Engineering, Stanford University

In biological macromolecules, structural patterns (motifs) are often repeated across different molecules. Detection of these common motifs in a new molecule can provide useful clues to the functional properties of such a molecule. We formulate the problem of identifying a given structural motif (pattern) in a target protein (example) and discuss the notion of complete matches vis-a-vis partial matches. We describe the precise error criterion that has to be minimized and also discuss different metrics for evaluating the quality of partial matches. Secondly, we present a new polynomial time algorithm for the problem of matching a given motif in a target protein. We also use the sequence and (if available) secondary structure information to annotate the different points in motif and the target protein, thus reducing the search space size. Our algorithm guarantees the detection of a perfect match, if present. Even otherwise, the algorithm computes very good matches. Unlike other methods, the error minimized by our algorithm directly translates to root mean square deviation (RMSD), the most commonly accepted metric for structure matching in biological macromolecules. The algorithm does not involve any preprocessing and is suitable for the detection of both small and large motifs in the target protein. We also present experiments exploring the quality of matches found by the algorithm. We examine its performance in matching (both full and partial) active sites in proteins.

Web Page

<http://robotics.stanford.edu/mitul/psb03.ps>

^a This work was done while the author was at the Department of Computer Science, Stanford University

^b Corresponding Author: rohitsi@cs.stanford.edu

CHARACTERIZING GENOMIC DIVERGENCE BETWEEN TWO UROCHORDATES, *CIONA SAVIGNYI* AND *CIONA INTESTINALIS*

Kerrin Small and Arend Sidow

We undertook an exploratory comparative genomic analysis to characterize the divergence between two sequenced urochordate sea squirts, *Ciona savignyi* and *Ciona intestinalis*. Coding sequences from 4 *intestinalis* annotated cosmids were used to identify orthologous regions in the *savignyi* genome assembly. These regions were then aligned to allow sequence comparison between the two species. The *C. savignyi* and *C. intestinalis* genomes are shown to exhibit considerable conservation of gene order and exons. Intergenic regions demonstrated extremely limited sequence conservation.

DISTANCES BETWEEN PHYLOGENETIC TREES – AN ALGORITHM USING BIPARTITE GRAPHS

Aaron Staple and Susan Holmes

Purpose

Phylogenetic trees derived from biological sequence data can be useful in analyzing this data, but it is important to consider the accuracy of the derivation process employed. If we have multiple trees, it is important to be able to calculate distances between trees using a reasonable distance metric. This allows a more meaningful analysis of bootstrapped trees, and can also be used to compare trees derived using different methods. Using this distance metric, we can define an average tree that represents the consensus of a set of trees.

Material and Methods

We represent distances between trees using a metric tree space, where a point in this space represents both a tree's topology and its edge lengths. In this paper we present an algorithm to calculate such distances, taking into account issues like normalization. We also present two notions of an average tree and algorithms to search for these.

Conclusion

In the worst case the distance algorithm is intractable, but for most biologically useful cases (including bootstrapping) it runs quickly and is suitable for general use. In trial runs, we analyzed differences between phylogenetic trees describing Australia's bower birds and also calculated distances between bootstrapped samples of a malaria tree.

References

1. Billera, Holmes, Vogtmann. Geometry of the Space of Phylogenetic Trees. *Advances in Applied Mathematics*. 73, 4202-4211 (2001).
2. Jost, Jurgen. *Nonpositive Curvature: Geometric and Analytic Aspects*. Birkhauser Verlag: Basel, Switzerland. 61-66 (1997).

WEIGHTING DATA RELATED BY PHYLOGENY

Eric A. Stone

Purpose

Evolutionary relationships are a common source of correlation among biological data sets. A typical problem involves the computation of a summary statistic, such as an average, from observations of a property common to various organisms. When a phylogeny of the organisms is available, evolutionary relationships can be used to model the dependence structure of the observations toward a meaningful estimate. We propose a new phylogeny-based model and derive an optimal weight for each observation under this framework.

Numerous approaches have been suggested to account for biological correlations, and several make use of phylogeny in some way to construct a set of weights. Our method distinguishes itself among the latter in circumventing the dependence of the weights on a fixed root position. Consequently, the relative importance of an observation is based on gross dependency among the observations rather than a priori differences in comparative significance.

As an example, we apply our weighting scheme to aligned biological sequences. In this case, scores (e.g. hydrophathy) “averaged” over alignment columns are calculated as the weighted mean of the individual scores. The method corrects for skewed sequence representation within the alignment and mitigates score similarity due to phylogenetic relatedness.

Material and Methods

Given a phylogeny (rooted or unrooted), we treat the root position as a random variable and define the joint distribution of the observations conditionally. Conditional on the root position, the observations are assumed to come from the model discussed in (Felsenstein, 1973) and (Altschul, Carroll, & Lipman, 1989). We derive the best unbiased linear estimator of the mean under this framework and suggest the coefficient of each observation as its weight. Some mathematical manipulation leads to an $O(n^2)$ algorithm for computing the weights.

Results

Comparisons with other weighting schemes are discussed, and we argue for the effectiveness of our method.

Conclusion

While there does not exist an objectively correct weighting scheme, our weights do satisfy a model-based optimality property. Our method also meets the criteria proposed in (Vingron & Sibbald, 1993) for evaluation of a weighting scheme.

References

1. Vingron M, Sibbald PR. Weighting in sequence space: a comparison of methods in terms of generalized sequences. *PNAS*. 90:8777-81, 1993.
2. Altschul SF, Carroll RJ, Lipman DJ. Weights for data related by a tree. *J Mol Biol*. 207:647-53, 1989.
3. Felsenstein J. Maximum-likelihood estimation of evolutionary trees from continuous characters. *Amer J Hum Genet*. 25:471-92, 1973.

CT COLONOGRAPHY: DOES IMPROVED Z RESOLUTION AID COMPUTER-AIDED POLYP DETECTION?

Padmavathi Sundaram, Christopher F. Beaulieu, David S. Paik, Pamela K. Schraedley, R. Brooke Jeffrey and Sandy Napel

Purpose

In helical computed tomography (CT), patient translation and data acquisition occur simultaneously through the use of a continuously rotating gantry. In order to increase the volume coverage speed, helical CT systems are now equipped with multiple rows of detectors. They are called multidetector-row CT (MDCT) scanners.

At a given detector setting, MDCT offers many choices of reconstruction slice thickness (z resolution). The goal of MDCT of the colon (CT colonography) is to detect polyps in the colon. We sought to determine the effect of z resolution on the computer-aided detection (CAD) of polyps.

Material and Methods

A 56 year old female underwent prone CT colonography with 4 detector-row CT. The data set was reconstructed with (slice width (SW), slice spacing (SS)) of (1.25mm, 1.25mm) and (2.5mm, 1.25mm). An expert radiologist determined the locations and sizes of 48 polyps (4 > 10mm, 28 between 5 and 9mm and 16 < 5mm).

We also simulated CT scans of a colon phantom with folds and polyps (61 in total, 20 between 5 and 9 mm, 21 > 10mm) with the following (SW, SS): (1.25mm, 0.625mm), (2.5mm, 1.25mm), (3.75mm, 1.875mm), (5mm, 2.5mm) and (10mm, 5mm). The phantom was scanned at various angles with the z axis (0°, 30°, 45°, 90°) in order for the detection results to be independent of the lesion orientation.

All the volume data sets were trilinearly interpolated to isotropic voxel size of 0.6 x 0.6 x 0.6 mm³. We ran a Hough transform-based CAD algorithm on all the data sets and studied its performance for various ranges of sensitivity and polyp sizes.

Results

For the patient data, improvement in z resolution from 2.5mm to 1.25mm does not significantly improve the detection of polyps larger than 5mm. It does improve the detection of polyps smaller than 5mm. Relative to the 1.25mm data, the number of false positives in the 2.5mm data at sensitivity levels of 50%, 75% and 94% increased by 48%, 115% and 206% respectively.

For the phantom data, improvement in resolution from 10mm to 1.25mm does not significantly affect the detection of polyps larger than 5mm. For polyps smaller than 5mm, the 1.25mm data achieved a sensitivity of 90% with 0 false positives compared to 18, 39, 67 and 120 false positives on the 2.5mm, 3.75mm, 5mm and 10mm data respectively.

Conclusion

We found that improvement in z resolution improves the performance of the CAD algorithm. The improvement is significant in the detection of polyps smaller than 5mm. The detection of polyps larger than 5mm are not significantly affected by the z resolution.

CONFIDENCE REGIONS AND TOPOLOGY OF PHYLOGENETIC TREES

Henry Towsner and Susan Holmes

Purpose

Maximum likelihood and maximum parsimony are two of the primary methods used for estimating a phylogenetic tree from genetic data. While current applications are effective at the actual estimation, finding confidence regions is a much more difficult task.

We examined confidence regions generated by the maximum likelihood method to better understand their behavior. We also looked at the behavior of the maximum parsimony method on data generated by different topologies.

Material and Methods

We generated contour maps of the likelihood of different trees (given many different data sets). Two-dimensional slices of this map could be visualized using Matlab, and the effect on the isocontours of choosing different trees to generate gene sequences was observed.

The apparent length (as calculated by the maximum parsimony method) of random data (generated according to a particular tree) was calculated for each eight leaf topology. Smaller topologies were then used to examine the effect of using different trees to generate data and estimate the length.

Results

For the maximum likelihood method, three different types of confidence regions were observed, determined by the relative sizes of the inner and pendant branch lengths. Depending on the relative size of the inner lengths, confidence regions either bend to include many other shapes, or to avoid them entirely. When the inner lengths vary dramatically, the third type was observed. In this case different pendant lengths make the confidence regions appear to be contiguous or islands.

Different topologies strongly affected the maximum parsimony method. In particular, even when the underlying data is random, some tree topologies appear to be much more parsimonious than others.

Conclusion

This data suggests that the number of neighbors an estimate has--the number of tree shapes with a significant likelihood of being the correct shape--can increase drastically over relatively small changes in the likelihood. This provides some support for the argument that adding more species to break up large branches can improve confidence regions.

For maximum parsimony, this data suggests that, while the parsimony method is certainly biased, it may not be so simple as merely balance or imbalance. This bias gets worse as lengths get longer, and is severe enough to cause parsimony to converge incorrectly for large enough branch lengths.

References

1. Louis J. Billera, Susan P. Holmes, and Karen Vogtman. Geometry of the space of phylogenetic trees. *Advances in Applied Mathematics*, 27:733-767, 2001.
2. D. M. Hillis, Inferring complex phylogenies. *Nature*, 383:247-264, 1996.
3. Andrey Zharkikh and Wen-Hsiung Li. Inconsistency of the maximum-parsimony method: the case of five taxa with a molecular clock. *Systematic Biology*, 42(2):113-125, 1993.

DRY ELECTROENCEPHALOGRAPHY

E. Timothy Uy, Dik Kin Wong, Marcos P. Guimaraes and Patrick Suppes

Purpose

In conventional electroencephalography (EEG), scalp abrasion and use of electrolytic paste are needed to insure low-resistance between sensor and skin. By replacing this wet process with dry sensors, setup and cleanup time were drastically reduced. Also, the number and frequency of sessions were no longer limited by the amount of abrasion the scalp could tolerate. However, the unavailability of amplifiers with sufficiently low noise and high input impedance held back the development of dry EEG sensors until the 90s¹. While single-sensor comparisons of wet versus dry exist², there has yet to be a multi-sensor study. Our goals are to quantitatively evaluate dry EEG sensors under multi-sensor conditions and to develop specifications for high fidelity recording.

Material and Methods

Eight dry 1x Softsens™ sensors from IBT (Redwood City, CA) were used to record EEG from five subjects. For each subject, 3000 sentences, 100 sentences displayed 30 times each, were presented over 12 sessions. Prototype and test waveforms for each of the 100 sentences were generated by averaging 15 trials for the prototype and the remaining 15 for the test. Each test waveform was compared to the set of prototypes, amongst which the best fitting prototype was selected. Classification rate is the rate at which the correct prototype is selected for a given test. Frequency and temporal windows were adjusted to maximize classification rate³.

Also, noise from dry 5x Softsens™ sensors was quantified using the SRS780 network analyzer and digitized for analysis in Matlab.

Results

Classification rates were highly significant. The rate for the best subject was 93/100. The next best subject rated 27/100. These results are qualitatively similar to experiments using wet sensors³. Inspection of single trial data revealed artifacts that were not present in wet sensor recordings. Analysis of individual sensors suggested that the artifacts could have been due to defective sensors, and that they were not typical of dry sensors.

Conclusion

Preliminary results are promising. For use in single trial analysis, further work is required to understand and eliminate dry sensor artifacts.

References

1. B. Alizadeh-Taheri, R. L. Smith, and R. T. Knight, "An active, microfabricated, scalp electrode array for EEG recording," *Sensors and Actuators A*, vol. 54, pp. 606-611, 1996.
2. A. Searle and L. Kirkup, "A direct comparison of wet, dry and insulating bioelectric recording electrodes," *Physiological Measurement*, vol. 21, no. 2, pp. 271-283, May 2000.
3. P. Suppes, B. Han, Z. Lu, "Brain-wave recognition of sentences," *PNAS*, vol. 95, pp. 15861-15866, 1998.

A COUPLED MULTI-DOMAIN METHOD FOR COMPUTATIONAL MODELING OF BLOOD FLOW

Irene Vignon, Brooke N. Steele and Charles A. Taylor

Purpose

We have previously proposed a new paradigm of predictive medicine in which the physician utilizes computational tools to construct and evaluate a combined anatomic/physiologic model to predict differential changes in blood flow for alternative treatment plans for an individual patient. The specification of appropriate "post-operative" boundary conditions for the computational analyses is a critical element of predictive modeling. Current methods for modeling blood flow require the specification of flow rate or pressures as boundary conditions. These quantities are part of the solution and not known *a priori*.

We have developed a space-time finite element method for solving the one-dimensional equations of blood flow including a resistance and an impedance boundary condition. An in-Vivo validation has been performed showing good agreement between the experimental and the numerical studies. However the impedance boundary condition assumes flow periodicity and thus cannot model dynamic changes downstream. Furthermore, there is no mathematical framework to understand the relationship between several types of outlet boundary conditions. Therefore, we have developed a new approach for specifying outflow boundary conditions in general using a "coupled multidomain method".

Material and Methods

We couple the region of interest (3D or 1D) with other downstream regions that influence it. The coupling is made through an analytical solution for the downstream domain, which provides a boundary condition for the upstream domain in the form of an integro-differential term. We have first started by looking at the coupling of a 1D domain to downstream domains with various representations (lumped or 1D). For each of them, we derived the integro-differential operators. In particular, we solved for a 1D downstream domain with linear equations using Green's functions.

Results

The different operators will be presented on the poster. Furthermore, we compared the flow waves at the root of a left coronary tree for constant and time-varying outlet resistances: only the time varying downstream boundary condition gave a flow curve that is comparable to the physiologic one, thus showing the importance of using proper boundary conditions.

Conclusion

We have developed a method to model blood flow with a relationship between pressure and flow at the outlet that represents more accurately the influence of the downstream domain. Current efforts are directed to incorporate this special boundary condition into a Space-Time finite element formulation and to program the Coupled Multidomain Method.

Web Page

<http://www-med.stanford.edu/school/vascular/> - Cardiovascular Biomechanics Research Laboratory

SUPERVISED NEURAL NETWORK FOR BRAINWAVE RECOGNITION USING INDIVIDUAL TRIALS

Dik Kin Wong, Marcos Perreau Guimaraes, E. Timothy Uy and Patrick Suppes

Purpose

EEG recordings of brain waves were made under several different experimental conditions in previous work. Significant classification rates were reported both for the averaged signals and individual trials.¹ This presentation serves as our initial exploration of using supervised neural networks as a tool for doing extensive individual trials recognition.

Material and Methods

Unlike the previous analysis, no explicit averaging is used to create the prototypes for classification. Instead, individual trials are presented to the neural network in order to capture their characteristics. Instead of using the neural network as a black box tool, techniques of optimization and signal processing are incorporated by introducing specific topology and initialization.

The neural network takes the temporal data as input, and then classifies it into one of a number of categories. Each sentence or word trigger is defined as a class. The result is compared to those results obtained in earlier investigation using filtering method and least squares criteria.

Two sections are presented, one involving a no-hidden layer network and the other involving a one-hidden-layer network. For no-hidden layer, a few optimization techniques are being explored.² These techniques convert the problem into a convex one in which only one local minimum exists. For the hidden layer, simple temporal networks are studied and further examined by combining with bandpass filters, general optimal FIR filters and time-varying filters.

Results

While using optimization techniques can be much faster and sometimes yield better results in our analyses of the mandarin data, its performance is generally worse than filtering method and backpropagation network in several of our 24 and 48 sentences analyses. On the other hand, a multi-channel neural network, with weights computed by backpropagation, turns out to be fairly successful in increasing robustness while improving results in several cases. For S34, we are able to recognize more than 90 out of 240 individual test samples in a cross-validation setting.

Conclusion

We see that the flexibility of neural network increases the size of the search space which makes the problems harder. On the other hand, given some of the better results, we are encouraged to explore further use of neural networks in analyzing individual trials.

References

1. Suppes P., Lu Z. L., Han B.: Brain wave recognition of words. *Proc. Natl. Acad. Sci.*, **94**, 14965-14969.
2. Castillo E. et al.: A Global Optimum Approach for One-Layer Neural Networks. *Neural Computation*, **14**, 1429-1449.

FROM PROMOTER SEQUENCE TO EXPRESSION: A PROBABILISTIC FRAMEWORK

Roman Yelensky, Eran Segal and Daphne Koller

Purpose

A central goal of molecular biology is the discovery of the regulatory mechanisms governing the expression of genes in the cell. The expression of a gene is controlled by many mechanisms. A key junction in these mechanisms is mRNA transcription regulation by various proteins, known as *transcription factors (TFs)*, that bind to specific sites in the promoter region of a gene and activate or inhibit transcription. Loosely speaking, we can view the promoter region as an encoding of a “program,” whose “execution” leads to the expression of different genes at different points in time. To a first-order approximation, this “program” is encoded by the presence or absence of TF binding sites within the promoter. In this work, we attempt to construct a unified model that relates the promoter sequence to the expression of genes, as measured by DNA microarrays.

Material and Methods

We present a probabilistic framework that models the process by which transcriptional binding explains the mRNA expression of different genes. Our joint probabilistic model unifies the two key components of this process: the prediction of gene regulation events from sequence motifs in the gene’s promoter region, and the prediction of mRNA expression from combinations of gene regulation. Our method initially clusters the genes based on their expression profile. For each “cluster” of genes, the procedure searches for binding site motifs that appear in the promoter region of the genes in the cluster and do not appear in other genes. After the motifs are found, the algorithm identifies which combinations of motifs best predict the current partition of genes to clusters, allowing us to detect combinatorial interactions among the different motifs. Given the motifs that activate each cluster, the algorithm re-assigns each gene to the cluster that best matches it using a probabilistic score composed of two components: the first examines whether the cluster binding sites appear in the gene’s promoter and the second scores the extent to which the gene’s expression matches the cluster expression profile. The algorithm repeats until convergence, refining all components: the binding site motifs, the motif combinations that activate each cluster, and the gene partition into clusters at each iteration.

Results

We applied our method to a dataset of 77 *Saccharomyces cerevisiae* cell cycle expression profiles [1], automatically constructing 12 clusters over 795 genes and 12 motif binding sites. We analyzed all clusters and for each examined both the biological coherence of the genes and the quality of the binding sites. All clusters were significantly enriched for genes sharing the same annotation in Gene Ontology (GO), and there were very few cases in which an annotation spanned more than one cluster, indicating that the method converged to coherent functional clusters. Of the 12 motifs discovered, 7 are known cell cycle motifs in the literature. Finally, there were several clusters for which activation by combinations of motifs was predicted, consistent with the literature.

References

1. P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein, and B. Futcher. *Mol. Biol. Cell*, 9(12):3273–97, 1998

DIFFERENTIAL GENE EXPRESSION IN A PARTIAL TRISOMY 16 MOUSE MODEL OF DOWN SYNDROME

Shuli Zhang, Farshid Oshidari, Ahmad Salehi and William C. Mobley

Purpose

Down syndrome (DS) is caused by the presence of three copies of human chromosome 21, and is the most frequent genetic cause of mental retardation. Segmental trisomy 16 mice are considered as models of DS, as a portion of mouse chromosome 16 is homologous to the long arm of human chromosome 21. The Ts65Dn mouse is trisomic for the region of mouse chromosome 16 that is homologous to the Down syndrome critical region. Previous work in our lab has shown neurological abnormalities in both the hippocampus and septum of the Ts65Dn mouse brain. To identify genes that are associated with these defects, we used oligonucleotide microarray to study differential gene expression in Ts65Dn vs. wild-type 2N mice.

Material and Methods

RNA from eighteen-month Ts65Dn and wild-type mouse hippocampus and septum were isolated, labeled and hybridized onto the Affymetrix MG-U74Av2 arrays. Gene expression analysis was done using three different methods: Affymetrix Micoarray Suite 5.0/Data Mining Tool 3.0, dChip, and Cluster. Real-time quantitative reverse transcription-polymerase chain reaction (Q-PCR) was performed to confirm the microarray results.

Results and conclusion

As expected, most of the differentially expressed genes in both hippocampus and septum are those present in three copies with a statistically significant trend towards a 1.5 fold overexpression. Interestingly, several genes on mouse chromosome 9 and chromosome 17 are also up-regulated in Ts65Dn mice by about 1.5 fold. Only two genes, calbindin2 and T-box brain gene 1, are consistently found to be down regulated in septum of Ts65Dn mouse brain by both array and Q-PCR analysis. This study represents an essential step for understanding the genetic basis of this complex disorder. The results of three different analysis methods were compared and will be discussed.

CLASSIFICATION OF GENE MICROARRAYS BY PENALIZED LOGISTIC REGRESSION

Ji Zhu and Trevor Hastie

Purpose, Methods and Results

Classification of patient samples is an important aspect of cancer diagnosis and treatment. The support vector machine (SVM) has been successfully applied to microarray cancer diagnosis problems. However, one weakness of the SVM is that given a tumor sample, it only predicts a cancer class label but does not provide any estimate of the underlying probability. We propose penalized logistic regression (PLR) as an alternative to the SVM for the microarray cancer diagnosis problem. We show that when using the same set of genes, PLR and the SVM perform similarly in cancer classification, but PLR has the advantage of additionally providing an estimate of the underlying probability. Often a primary goal in microarray cancer diagnosis is to identify the genes responsible for the classification, rather than class prediction. Identifying the relevant genes which contribute most to the classification is another challenge in microarray cancer diagnosis. We consider two gene selection methods in this work, univariate ranking (UR) and recursive feature elimination (RFE). Empirical results indicate that PLR combined with RFE performs consistently better than the SVM with either UR or RFE, in both cross-validation and test samples. A fast algorithm for solving PLR is also described.

References

1. Zhu J, Hastie T. Classification of gene microarrays by penalized logistic regression. *Submitted*. 2002.
2. Zhu J, Hastie T. Kernel logistic regression and the import vector machine. *Advances in Neural Information Processing Systems*. 14, 2001.

ENDOSCOPE CALIBRATION AND ACCURACY TESTING FOR 3D/2D IMAGE REGISTRATION

*Michael R Bax, Rasool Khadem, Jeremy A Johnson, Eric P Wilkinson,
Ramin Shahidi*

Background and Theory

Today lightweight endoscopes are used in small body cavities, but they display only visible surfaces and cannot view the interior of opaque tissue. Combining endoscopic video with overlaid volumetrically-reconstructed CT or MRI images permits surgeons to see beyond visible surfaces and provides "on-the-fly" 2D and 3D information for planning and navigation. Precise endoscope calibration and accuracy testing are necessary for surgical quality. Following calibration, tracking the position and rotation of the endoscope enables the rendering of virtual images to match the endoscope video. A ray from the center of projection (the optical origin) to a point in physical space passes through the image projection plane (the CCD). Errors such as residual lens distortion or tracking error cause this intersection to be some distance e from the image point. Adding full non-linear radial lens-distortion compensation to the rendering engine reduces performance; an alternative method, here named "constant-radius linear compensation", is to linearly scale the virtual image such that the radii of the predefined region of interest in each image are made equal.

Method

A STORZ Tricam endoscope with a 50200A telescope, Traxtal universal passive trackers, and an NDI Polaris hybrid optical tracking system were used. Tsai's calibration algorithm was used. The calibration target was a planar grid pattern of black dots 1 and 2 mm in diameter, with the grid rows and columns spaced 3 mm apart. This was mounted in a jig with an attached tracker such that the telescope of the endoscope in calibration is positioned 15 mm away from the center of the target and angled 30° from the normal of the target plane. The physical target for accuracy testing was a version of the calibration target large enough to fill the field of view of the endoscope at a distance of 65 mm, and mounted on a tracked moveable plate. The grid position was localized relative to the tracker. The grid was stepped through distances from 5 to 65 mm from the tip of the endoscope's telescope. At each position the image error e was found for each dot in the endoscope video image. This was repeated for 20 separate calibrations.

Results

The mean error between the rendered location of the dots in the virtual image and the corresponding dots in the endoscope video image, normalized to the diameter of the FOV, increases towards the image edge to nearly 8% of the image diameter without compensation. With non-linear compensation it remains below 2%, and linear compensation reaches 5%. Normalized error as a function of distance of the target plane from the endoscope tip is roughly constant at below 5% without compensation, about 1.5% with full compensation, and just over 3% with linear compensation. Constant-radius linear compensation is clearly an effective scheme for overcoming lens distortion, as it provides half the improved accuracy performance of non-linear compensation (compared to no compensation) without the associated rendering performance loss.

Web Page

<http://igl.stanford.edu>

SYMPOSIUM PARTICIPANT LIST

BCATS 2002 Symposium Proceedings
Symposium Participant List

Ali El-Sayed Abbas
Management Science and
Engineering
aliabbas@stanford.edu

Laurent Abi-Rached
Structural Biology
laurent.abi-rached@stanford.edu

Burak Acar
Radiology
bacar@stanford.edu

Angelie Agarwal
Mechanical Engineering
angelie@stanford.edu

Sameer Hasan Al-Sakran
Scientific Computing and
Computational Mathematics
malkave@stanford.edu

David R Almassian
Bioinformatics
dalmassian@stanford.edu

Russ Biagio Altman
Genetics
Russ.Altman@stanford.edu

Joshua Alwood
Aeronautics/Astronautics
alwood@stanford.edu

Amitpal
Santa Clara University
adhillon@scu.edu

Andrew Chongmin An
Biology
aan@stanford.edu

Kirk Anders
Genetics
anders@genome.stanford.edu

Clay Anderson, Ph.D.
Mechanical Engineering
fca@stanford.edu

Jack C. Armstrong
Jack C. Armstrong & Associates
jackcarmstrong@mindspring.com

Allison S. Arnold, Ph.D.
Mechanical Engineering (BME)
asarnold@stanford.edu

Navneet Aron
Electrical Engineering
navneet@stanford.edu

Blake Ashby
Mechanical Engineering
bmashby@stanford.edu

Orna Avsian-Kretchmer
GYN/OB
orna.avsian-
kretchmer@stanford.edu

Brian Babcock
Computer Science
babcock@cs.stanford.edu

Christopher Baer
Human Biology
nobilis3@stanford.edu

Alvin Barlian
Mechanical Engineering
barlian@stanford.edu

Jeremy L. Baryza
Chemistry
jbaryza@stanford.edu

Serafim Batzoglou
Computer Science
serafim@cs.stanford.edu

Michael Baudis
Pathol;olgy
mbaudis@stanford.edu

Michael Raymond Bax
Electrical Engineering

Stephen Bay
CSLI, Stanford
sbay@apres.stanford.edu

Asa Ben-Hur
biochemistry
asa.benhur@stanford.edu

Aviv Bergman
Biological Science
Aviv@Stanford.edu

Elliot Berkman
CSLI & Psychology
eberkman@stanford.edu

Gail A. Binkley
Genetics
gail@genome.stanford.edu

Jon Binkley
Pathology
binkley@genome.stanford.edu

Roy A. Bohenzky
Roche Molecular Systems
roy.bohenzky@roche.com

Bill Bosl
bosl@inreach.com

Onn Brandman
Molecular Pharmacology
onn@stanford.edu

Relly Brandman
Molecular Pharmacology
rbrandman@stanford.edu

Andrew Broderick
SRI Consulting - Business
Intelligence
abroderick@sric-bi.com

Igor Eric Brodsky
Microbiology and Immunology
ibrodsky@stanford.edu

Chrissy Brown
Chemistry
cdbrown9@Stanford.edu

Christina Dawn Brown
Chemistry
cdbrown9@Stanford.edu

James Alan Brown
Radiation Oncology
james.brown@stanford.edu

Michael Brudno
Computer Science
brudno@cs.stanford.edu

David Buckeridge
Stanford Medical Informatics
david.buckeridge@stanford.edu

Carol (Cheng) Cain
Biomedical Informatics
cheng@smi.stanford.edu

Sandra Caldeira
Medicine Oncology
sandrac@stanford.edu

Kaizad Cama
Human Biology
kaizad@stanford.edu

Sophie Candille
Genetics
sophie.candille@stanford.edu

BCATS 2002 Symposium Proceedings
Symposium Participant List

Lillian L. Cao
GSB
Cao_Lei@gsb.stanford.edu

Paolo Carnevali
Protein Mechanics
pcarnevali@ProteinMechanics.com

Dana Carpenter
BME
danac40@hotmail.com

Dennis R Carter
Mechanical Engineering
dcarter@stanford.edu

Erin Catto
Protein Mechanics
ecatto@proteinmechanics.com

Albert S Chan, MD
Biomedical Informatics
achan@smi.stanford.edu

Alvin Chan
San Jose University
a_chan03@hotmail.com

Desmond Chan
West Valley College
dkhchan@yahoo.com

Jeffrey Tien-Hao Chang
Biomedical Informatics
jefftc@stanford.edu

Elizabeth D. Chao
Department of Biochemistry
lizchao@stanford.edu

Rowan Chapman
Mohr, Davidow Ventures
rchapman@mdv.com

Anjun Chen
Callida Genomics
achen@sbh.com

Chung-Chu Chen
Electric Engineering
Chung-Chu.Chen@stanford.edu

Hao Chen
Developmental Biology
hc49@cornell.edu

Rong Chen
Chemistry
rongchen@stanford.edu

Betty Cheng
SMI
bcheng@smi.stanford.edu

Mike Cherry
Genetics
cherry@stanford.edu

Yen Lin Chia
Management Science &
Engineering
yenlin@stanford.edu

Sou Cheng Choi
Scientific Computing and
Computational Mathematics
scchoi@stanford.edu

Ashmi Ashokkumar Chokshi
Computer Science
ashmi@stanford.edu

Lillian T. Chong
Chemistry
ltchong@stanford.edu

Sanders Chong
Computer Science
sanders.chong@stanford.edu

Jim Chow
Computer Science
jchow@cs.stanford.edu

Marina Su Yin Chow
Chemistry
marinac@stanford.edu

Lonnie Chrisman
Institute for the Study of Learning
and Expertise
lonnie@apres.stanford.edu

John Chuang
UCSF
chuang@alumni.duke.edu

Hediye Nese Cinar
University of California Santa Cruz
cinar@biology.ucsc.edu

Hulusi Cinar
University of California Santa Cruz
hcinar@biology.ucsc.edu

Gregory Michael Cooper
Genetics
coopergm@stanford.edu

Glenn Cornett, MD, PhD
Metastrat, Inc.
cornett@metastrat.com

Monica Crubezy
Stanford Medical Informatics
crubezy@smi.stanford.edu

Dr. Behzad Dariush
Honda R&D Americas
bdariush@hira.com

Debbie Dasker
debbiedasker@yahoo.com

Wyn Davies
Apple Computer
wyn@apple.com

Eugene Davydov
Computer Science
edavydov@stanford.edu

Rob DeConde
Biomedical Informatics
rdeconde@smi.stanford.edu

Daphne Dembo
d_dembo@yahoo.com

Janos Demeter
Dept. of Genetics
jdemeter@stanford.edu

Kathy DeRiemer
Infectious Diseases
kathy@molepi.stanford.edu

Parvati Dev
Information Resources and
Technology, and SUMMIT
parvati.dev@stanford.edu

Chuong B. Do
Biology
chuong.do@stanford.edu

Holger Doege
Molecular Pharmacology
holgerjdoege@hotmail.com

Kyle Jonathan Doerksen
Neuroscience
doerksen@stanford.edu

Kara Dolinski
Genetics
kara@genome.stanford.edu

Stan Dong
Genetics
qdong@genome.stanford.edu

Dr Manali Doshi
Graduate School of Business
doshi_manali@gsb.stanford.edu

Peter Douglas
CSLI
peteysan@stanford.edu

BCATS 2002 Symposium Proceedings
Symposium Participant List

Carl Dowds
Scios
dowds@sciosinc.com

Matt Downs
downs_matt@gsb.stanford.edu

Katerina Athena Drouvalakis
Medicine
drouvaka@stanford.edu

Dr. Chris Duffield
Pediatrics
cd@iptq.com

Jonathan Dugan
Biomedical Informatics
dugan@smi.stanford.edu

Alex Dunn
Caltech
dunn@caltech.edu

Seth Andrew Eatinger
Electrical Engineering
seat@stanford.edu

Lawrence M Fagan
Medicine
FAGAN@smi.stanford.edu

Ali Faghfuri
Electrical Eng.
faghfuri@stanford.edu

Zhenbin Fan
Urology
Zhenbin.Fan@stanford.edu

Scott Farrar, PhD
farrar@stanfordalumni.org

Bernard Martin Fine
Hematology/Biochemistry
bfine1@stanford.edu

Anthony Fodor
HHMI - Aldrich Lab
afodor@stanford.edu

Pavlina Fonville
Pathology
pavlinaf@stanford.edu

Kirk E. Fry
XDx, Inc.
kirkfry@xdxinc.com

Vincent Anthony Funari
surgery
vfunari@STANFORD.EDU

Irene Gabashvili
Medicine and Genetics
serene@stanford.edu

Yashar Ganjali
Electrical Engineering
yganjali@stanford.edu

Tia Gao
Electrical Engineering
tiagao@stanford.edu

Kunal Girotra
SUMMIT
kgirotra@stanford.edu

David Goldberg
Palo Alto Research Center
goldberg@parc.com

Ambarish Goswami
Honda R&D Americas
agoswami@hra.com

Carlos Ernesto Guestrin
Computer Science
guestrin@stanford.edu

Marcos Perreau Guimaraes
CSLI
marcospg@csl.stanford.edu

Danwei Guo
SCCM
dguo@stanfordalumni.org

Rohit Gupta
rohit_bmw@hotmail.com

Zoltan Gyongyi
Computer Science
zgyongyi@stanford.edu

Dr. Edward Olof Haeggstrom
electrical engineering
edward@piezo.stanford.edu

Gail M. Hashimoto
Hopkins & Carley
ghashimoto@hopkinscarley.com

James Henry Henderson
Mechanical Engineering
(Biomechanics)
jayh@stanford.edu

Silvia Herrero
Apple Computer
herrero1@apple.com

Micheal Hewett
Department of Genetics
hewett@cs.stanford.edu

Jill Startzell Higginson
Biomechanical Engineering
jstartz@stanford.edu

Matthew Hill
Genetics
mmhill@stanford.edu

Bess Ho
Health Highway
bizbunny@aol.com

Susan P Holmes
Statistics
susan@stat.stanford.edu

Wan-Jen Hong
Medicine
wanjen@stanford.edu

Alison Hottes
Electrical Engineering
ahottes@stanford.edu

John Howard
Center for Integrated Systems
jkh@Stanford.edu

Ron Howard
Management Science and
Engineering
rhoward@stanford.edu

Che Hsu
Computerists United Inc.
cjhsu@computeristsunited.com

Bo Huang
Chemistry
bohuang@stanford.edu

Jianqiang Huang
Genetics
jqhuang@stanford.edu

Joseph Huang
Gyn/Ob
huang3@stanford.edu

Kuo-Wei Huang
Chemistry
hkw@stanford.edu

Yang Huang
Stanford Medical Informatics
yhuang@ieee.org

Angela Kwan-Yee Hung
Management Science and
Engineering
ahung@stanford.edu

BCATS 2002 Symposium Proceedings
Symposium Participant List

Juanita Hung
UCSF
jhung@itsa.ucsf.edu

Patrick Hung
Electrical Engineering
hung@arith.stanford.edu

Tony Hung
Bioengineering
tonyhung@stanford.edu

Deborah M. Hussey
Law
hussey@stanford.edu

Stuart Hwang
Millennium Pharmaceuticals, Inc.
stuart.hwang@mpi.com

Samuel Sze Ming Jeong
Computer Science
sieong@stanford.edu

John James Shedletsky III
Computer Science
shedletsky@stanford.edu

Farhad Imam
Biochemistry, Medicine
hod@stanford.edu

Jared Jacobs
Computer Science
jaredjacobs@stanford.edu

Hemanth Jagannathan
Electrical Engineering
jhemanth@stanford.edu

Guha Jayachandran
Computer Science
guha.jayachandran@stanford.edu

Nuwan Jayasena
Electrical Engineering
jayasena@stanford.edu

Thusitha Jayawardana
Medical Informatics
twj@smi.stanford.edu

Michael Christopher Jewett
Chemical Engineering
mjewett@stanford.edu

Youngran Ji
surgery
youngran.ji@stanford.edu

Jeremy Aaron Johnson
Electrical Engineering
jeremyj@stanford.edu

Sunita Jones
Urology
sbjones@stanford.edu

Sonal Josan
Electrical Engineering
sjosan@stanford.edu

Neel Joshi
Computer Science
nsj@cs.stanford.edu

Titania Juang
UC Berkeley
kameko@uclink4.berkeley.edu

Urania Juang
UC Berkeley
mahamari@uclink4.berkeley.edu

Andrew Kacsmar
Computer Science
andyk@db.stanford.edu

Joseph Kahn
MS&E
jkahn@stanford.edu

John Kang
Alpha Innotech Corporation
jj-kang@aicemail.com

Peter Kasson
Medicine
kasson@cmgm.stanford.edu

Manisha Kathuria
Computer Science
manisha_kathuria@hotmail.com

Amit Kaushal
CS
akaushal@stanford.edu

Brett Taketsugu Kawakami
Civil and Environmental
Engineering
brettk@stanford.edu

Eric Ketchum
ketch22@cs.stanford.edu

Viktoria Kheifets
Molecular Pharmacology
vky@stanford.edu

Sami Khuri
San Jose State University
khuri@cs.sjsu.edu

Michael F. Kim
mfk@cs.stanford.edu

Stuart K Kim
Dev. Bio./Genetics
kim@cmgm.stanford.edu

Teri Klein
Genetics and Stanford Medical
Informatics
teri.klein@stanford.edu

Tod Klingler
tod@pacbell.net

Jeff Klingner
Computer Science
klingner@stanford.edu

Pete Klosterman
UCSF
petek@accesscom.com

Dennis Chun-Yone Ko
Dev Bio
dcko@stanford.edu

Kwok Ko
Stanford Linear Accelerator Center
kwok@slac.stanford.edu

Willie Koh
SCPD
weelee@stanford.edu

Robert Kohlenberger
Celera Diagnostics
kohlenrw@appliedbiosystems.com

Tamar Rachel Kolodny
Computer Science
trachel@cs.stanford.edu

Yves Konigshofer
Immunology program
yvesk@stanford.edu

Seunghum Koo
ME
koosb@stanford.edu

Charles Kou
CS
charlesk@stanford.edu

Lampros Kourtis
ME:BME
lakourtis@stanford.edu

John Koza
Biomedical Informatics
koza102@tmolp.com

Lee G. Kozar
Molecular and Genetic Medicine
kozar@stanford.edu

BCATS 2002 Symposium Proceedings
Symposium Participant List

Daniel Kraft M.D.
Hematology/Oncology
daniel.kraft@stanford.edu

Tom Krummel, MD
Surgery
tkrummel@stanford.edu

Anand Kumble
Pointilliste, Inc
kumble@pointilliste.com

Steve Laderman
Agilent Technologies
laderman@labs.agilent.com

Jacob S Lai
Chemistry
jakelai@stanford.edu

Chuck Lam
EE
chucklam@stanford.edu

Giora Landesberg MD, DSc
Stanford Medical Informatics
gxl@smi.stanford.edu

Stefan M. Larson
Biophysics
smlarson@stanford.edu

Denny Lau
radiology
denny.lau@stanford.edu

Alice Y. Lee
Computer Science
alice.lee@cs.stanford.edu

Kunwoo Lee
Mechanical Engineering
kunwoo12@stanford.edu

Su-In Lee
EE
silee@stanford.edu

Denise Leong
Pathology
deleong@stanford.edu

Harrison Leong, Ph.D.
Celera Diagnostics
leonghm@appliedbiosystems.com

Steve Lew
Physics
lew@ccrma.stanford.edu

Natasha Kaleta Li
Mathematics
natasha.li@stanford.edu

Ping Li
Computer Science
pingli98@stanford.edu

DeYong Liang
Anesthesiology
dyliang@stanford.edu

Wenchuan Liang
Biochemistry
wliang@cmgm.stanford.edu

Jason Lih
Department of Genetics
jasonlih@stanford.edu

Connie Lin
Biological Sciences
cslin@stanford.edu

Richard Lin
Genetics
richlin1@stanford.edu

Shaowei Lin
Mathematics
swlin@stanford.edu

Steve Lin
Genetics, School of Medicine
slin@smi.stanford.edu

Wei-Te Lin
Radiology
weite.lin@stanford.edu

Zhen Lin
Stanford Medical Informatics
zlin@stanford.edu

Maria Jamsa Linawaty
sci91099@nus.edu.sg

May Liu
Mechanical Engineering
mql@stanford.edu

Shuo Liu
Biomedical Informatics
shuo.liu@stanford.edu

Yueyi Liu
Biomedical Informatics
yliu@stanford.edu

Yuhong Liu
G & ES
yuhong@stanford.edu

Itay Lotan
Computer Science
itayl@stanford.edu

Ching-Huang Lu
MSE
ratiug@stanford.edu

Haige Lu
Chemistry
hlu@stanford.edu

Hui-Ling Lu
Electrical Engineering
vickylu@ccrma.stanford.edu

Yin Lu
MS&E
yin.lu@stanford.edu

Jiong Ma
Biology
jiongma@stanford.edu

Sanket Biren Malde
Computer Science
sanket@stanford.edu

Anisha Malhotra
Computer Science
anisham@stanford.edu

Mars Mallari
Pathology
mars.mallari@stanford.edu

Patrick Kingsley Mang
Applied Physics
mmk@stanford.edu

Penka Vassileva Markova
CS
penka@stanford.edu

Peter Markstein
Hewlett-Packard Labs
peter.markstein@hp.com

Vicky Markstein
in silico Labs
vicky@insilicolabs.com

Andy Marshall
Chemistry
andrew.marshall@stanford.edu

John Charles Matese
Genetics
jcmatese@stanford.edu

Calvin Maurer
IGL/Neurosurgery
calvin.maurer@igl.stanford.edu

Jennifer Maynard
Micro and Immunology
jamaynard@stanford.edu

BCATS 2002 Symposium Proceedings
Symposium Participant List

Matt McMullen
Aero/Astro Department
mcmu@stanford.edu

Nipun Mehra
Computer Science
nmehra@stanford.edu

Minal Mehta
Computer Science
minal@stanford.edu

Kambiz Merati
Biomedical informatics
merati@smi.stanford.edu

Sia Meshkat
Protein Mechanics
smeshkat@proteinMechanics.com

Hongcheng Mi
Computer Science
hmi@stanford.edu

Melissa Miller
Biophysics
melissad@stanford.edu

Nesanet Senaite Mitiku
Genetics
nesanet@stanford.edu

Helen Moore
Mathematics
moore@math.stanford.edu

Daniel Scott Morris
Computer Science
dmorris@cs.stanford.edu

Mark Musen
Stanford Medical Informatics
musen@stanford.edu

Chris J. Myers
Computer Science
cjmyers@stanford.edu

Sandy Napel
Radiology
snapel@stanford.edu

Brian Thomas Naughton
Biomedical Informatics
briannaughton@stanford.edu

Wing-On Ng
Civil & Environmental Engineering
wng@stanford.edu

Han Duy Nguyen
han.nguyen@stanford.edu

David A. Nix, PhD
Medicine
dnix@stanford.edu

Steve O
bub@io.com

William Ogle
Biological Sciences
wogle@stanford.edu

Serene Ong
Management Science and
Engineering
ongserene80@yahoo.com

David S. Paik
Biomedical Informatics
paik@smi.stanford.edu

Adam Thomas Palermo
Molecular Pharmacology
palermo55555@hotmail.com

Kevin Pan
Biomedical Informatics
kpan@stanford.edu

Dr. Vijay S. Pande
Chemistry and Structural Biology
pande@stanford.edu

Monita Pau
Chemistry
mpau@stanford.edu

Boris Peker
Biophysics
boris1@stanford.edu

Mor Peleg
Stanford Medical Informatics,
Medicine
peleg@smi.stanford.edu

Nicolas Peyret
Applied Biosystems
peyretnn @appliedbiosystems.com

Kirk Phelps
Computer Science
kphelps@cs.stanford.edu

Jan Benjamin Pietzsch
Management Science &
Engineering
pietzsch@stanford.edu

Zachary Pincus
Stanford Medical Informatics,
Department of Internal Medicine
zpincus@stanford.edu

Carla D. Pinon
Computer Science
cpinon@cs.stanford.edu

Adam Eugeniusz Piotrowski
Materials Science & Engineering
adamatic@stanford.edu

Sabine Pokutta
Structural Biology
spokutta@stanford.edu

Shyam Prabhakar
Mathematics
shyam@stanford.edu

Carla M. Pugh, MD. PhD.
SUMMIT
cpugh@stanford.edu

Ting Pun
EKP LLC
tpp@menlogroup.com

Xiaoli Qin
Microbiology&Immunology
xlqin@stanford.edu

Liliana Quintanar
Chemistry
lilianaq@stanford.edu

Rauch Rami
Gyn OB/Op
rami.rauch@stanford.edu

Soumya Raychaudhuri
Medicine
tumpa@stanford.edu

Haleh Razavi
University of California, Berkeley
haleh@eecs.berkeley.edu

Hedi Razavi
University of California, Berkeley
hrazavi@uclink.berkeley.edu

Hoofar Razavi
Computer Science
hoofar@hoofar.com

Haleh Razavo
University of California, Berkeley
haleh@eecs.berkeley.edu

Young Min Rhee
Chemistry
ymrhee@stanford.edu

Kerri Elyse Rieger
Medicine/Cancer Biology
rieger@stanford.edu

BCATS 2002 Symposium Proceedings
Symposium Participant List

Viola Rieke
Electrical Engineering
vrieke@stanford.edu

Torsten Rohlfing
Neurosurgery
torsten.rohlfing@stanford.edu

Michal Ronen
Genetics
michal.ronen@stanford.edu

Brian Rossi
New Focus, Inc.
brossi@newfocus.com

Daniel L Rubin
SMI
dlrubin@stanford.edu

Daniel Benjamin Russakoff
Computer Science
dbrussak@stanford.edu

Mitul Saha
ME
mitul@stanford.edu

Garima Sahai
Computer Science
gsahai@stanford.edu

Alok Jerome Saldanha
Genetics
alok.saldanha@stanford.edu

Peter Salzman
MS&E
psalzman@stanford.edu

Yaseen Samara PhD
Radiology
yaseen.samara@med.ge.com

Sukhwinder Sangha
Bayer Diagnostics
sukhwinder.sangha.b@bayer.com

Peter N. Saurugger
Agilent Technologies
peter_saurugger@agilent.com

Serge Saxonov
BMI
saxonov@stanford.edu

Deborah Schechtman
molecular pharmacology
deborah.schechtman@stanford.edu

Diane Schroeder
Biomedical Informatics
dschroed@stanford.edu

Stephan Schulz
Microbiology and Immunology
stephan.schulz@stanford.edu

George Christopher Scott
Medicine/Biomedical Informatics
gcscott@stanford.edu

Matthew P Scott
Developmental Biology
scott@cmgm.stanford.edu

John Scoville
Scientific Computation and
Computational Math
scoville@stanford.edu

Hossein Sedarat
EE
sedarat@isl.stanford.edu

Jonathan Seely, MD
jon.seely@iatrix.com

Eran Segal
Computer Science
eran@cs.stanford.edu

Anil Sehgal
Oleiotropica Biotech Inc.
cancersci@hotmail.com

Anand Sethuraman
Genetics
anand@genome.stanford.edu

C Shachaf
Oncology
cshachaf@stanford.edu

Jessica Carrie Shapiro
biophysics
jessicas@stanford.edu

Roded Sharan
International Computer Science
Institute
roded@icsi.berkeley.edu

Anthony Jacob Sherbondy
Electrical Engineering
sherbond@stanford.edu

Gavin Sherlock
Genetics
sherlock@genome.stanford.edu

Jing Shi
BMI
js2400@stanford.edu

Michael Shirts
Chemistry
mrshirts@stanford.edu

Irina Shiryayeva
Chemistry
irinas@stanford.edu

L Shrihari
Celerity Solutions, Inc
shrihari_l@yahoo.com

Leonard Niels Sibille
Computer Science
leonard.sibille@stanford.edu

Brent Silver
Materials Science and Engineering
bssilver@hotmail.com

Natalie Renee Simmons
Biology
natalie.simmons@stanford.edu

Nita Singh
EE
nital@stanford.edu

Nina Singhal
Computer Science
nsinghal@cs.stanford.edu

Kerrin Small
Genetics
kerrin@stanford.edu

Natalia P. Snarskaya
Developmental Biology
ntl@stanford.edu

Olivia Soesanto
Computer Science
soesanto@stanford.edu

Eric J. Sorin
Chemistry
esorin@stanford.edu

Ryan Spilker
Mechanical Engineering
rspilker@stanford.edu

Kunju Joshi Sridhar
Medicine, hematology
kunju@stanford.edu

Brooke Noelani Steele
Mechanical Engineering
bnsteel@stanford.edu

Charles R Steele
Mechanical Engineering
chasst@stanford.edu

BCATS 2002 Symposium Proceedings
Symposium Participant List

Richard Samuel Leopold Stein
Chemistry
rslstein@stanford.edu

Hilary Stockton
Law
hilary.stockton@stanford.edu

Eric Alan Stone
Statistics
eastone@stanford.edu

Matt Streeter
Genetic Programming Inc.
mjs@tmlp.com

Joshua Michael Stuart
Biomedical Informatics
stuart@smi.stanford.edu

Wonjoo Suh
electrical engineering
suhwonjoo@hotmail.com

JingLucy Sun
structural biology
lucys@stanford.edu

Padma Sundaram
EE
padma@stanford.edu

Arun Swami
Adobe Systems
arun@stanfordalumni.org

Serene Tan
Business Administration
chewy_seru@yahoo.com

Andres Bayani Tellez
BioMedical Informatics
tellez@stanford.edu

Wyatt Tellis
wtellis@radiology.ucsf.edu

Jessie Dale Tenenbaum
Biomedical Informatics
jessiet@stanford.edu

Chandra Theesfeld
Genetics
chandra@genome.stanford.edu

Daryl Thomas
University of California at Santa Cruz
daryl@soe.ucsc.edu

Matt Thompson
MBA
matt.thompson@stanford.edu

Lai A. Tin
ungraduate advising center
atin@stanford.edu

Rabindra Marie-Jean Tirouvanziam
Genetics
tirouvan@stanford.edu

T Kalee Tock
CSLI
kaleeg@stanford.edu

Mark Tong
Computer Science
mktong@stanford.edu

Gergely Toth
Protein Mechanics
gtoth@proteinmechanics.com

Garrick Toubassi
Protein Mechanics, Inc.
garrick@proteinmechanics.com

Henry P Towsner
Biology
htowsner@stanford.edu

Olga G Troyanskaya
Biomedical Informatics & Genetics
olgat@stanford.edu

Sheryl Tsai
Chemical Engineering
sheryl@stanford.edu

Jim Turner
Amersham Biosciences
jim.turner@am.amershambiosciences.com

Elwyn Timothy Uy
Applied Physics
torque@stanford.edu

J. Sonja Uy
sandwich@stanford.edu

Michael Vagell
Psychology
vagell@psych.stanford.edu

Abhi Vase
Mechanical Engineering
abhinand@stanford.edu

Henry E Neuman de Vegvar
Neurology
devegvar@cmgm.stanford.edu

Irene Vignon
ME, bioengineering division
vie@stanford.edu

Neal Suresh Vora
nealvora@stanford.edu

Kay Walter
UCSC Extension & Foothill College
klwalter@alumni.princeton.edu

Chia-wei Wang
Chemical Engineering
cwwang@stanford.edu

Dongmei Wang
Urology
dongmw@yahoo.com

Fu-Min Wang
Electrical Engineering
fumin@stanford.edu

Haidong Wang
Computer Science
hdwang@stanford.edu

Wei Wang
Genetics
wangwei@stanford.edu

Saad Aziz Waqas
Electrical Engineering
sawaqas@stanford.edu

Steven Waslander
Aeronautics and Astronautics
stevenw@stanford.edu

Nick Webb
Genetics
nickwebb@stanford.edu

Marion Webster
Applied Biosystems
webstemn@appliedbiosystems.com

Jacqueline Nerney Welch
Medicine / Mechanical Engineering
welch@stanford.edu

Wendy
Mechanical Engineering
vwcheng@mindspring.com

Evan Tang Williamson
Computer Science
ettang@stanford.edu

Cyrus A Wilson
Biochemistry
cyrus@stanford.edu

Benjamin Wong
Self
benwong@attglobal.net

BCATS 2002 Symposium Proceedings
Symposium Participant List

Catherine Wong
cywong@stanford.edu

Dik Kin Wong
CSLI and EE
dkwong@stanford.edu

Kim Woodrow
Chemical Engineering
kwoodrow@stanford.edu

Adam Thackeray Wright
Math and Computer Science
adamatw@stanford.edu

Xiaolin Wu
University of Chicago
xiwu@delphi.bsd.uchicago.edu

Matthew Xia
Genencor International, Inc.
mxia@genencor.com

Grace Xiao
graduate school of business
xiao_grace@gsb.stanford.edu

Chengpei Xu
Surgery
cx11@stanford.edu

Chunxia Xu
Geology
yongjun@pangea.stanford.edu

Eric Xu
ericcxu@stanford.edu

Ying(Janet) Xu
Petroleum Engineering
bxy116@stanford.edu

Beverly Yang
Computer Science
byang@stanford.edu

Danny Yang
CS
dbyang@stanford.edu

Mehmet Fatih Yanik
Applied Physics
fatih@stanford.edu

Iwei Yeh
BMI
iwei.yeh@stanford.edu

Roman Yelensky
Computer Science
ryelensky@stanford.edu

Vijaykalyan Yeluri
Radiology
yeluri@stanford.edu

Sung-Won Yoon
Electrical Engineering
yoons1@stanford.edu

Kwong Hiu Yung
statistics
Kwong.Yung@stanford.edu

Bojan Zagrovic
Biophysics Program and Chemistry
Department
zagrovic@stanford.edu

Peng Zhang
cardiovascular medicine
pengz@stanford.edu

Shuli Zhang
Neurology
szhang@stanford.edu

Genghong Zhou
Aclara BioSciences Inc.
gzhou@aclara.com

An Zhu
Computer Science
anzhu@cs.stanford.edu

Hongyun(June) Zhu
Division of Nuclear Medicine
junezhu@stanford.edu

Ji Zhu
Statistics
jzhu@stanford.edu

Jing Zhu
jzhu@itsa.ucsf.edu

Shirley Zhu
Pathology
szz@stanford.edu

Tiancheng Zhu
Apple Computer
mrt@mac.com

Feng Zhuge
EE
zhugef@stanford.edu

SPONSOR PARTICIPANTS

Affymetrix

Ann Loraine
Gangwu Mei
Tina Chen

Apple Computer

Silvia Herrero
Wyn Davies
Tiancheng Zhu

Alloy Ventures

J. Leighton Read
Craig Taylor
Doug Kelly

BioScience Forum

Shauna Farr-Jones

Entelos

MariaLuisa Ruiz

Genentech

Thomas Wu
Jianjun Zhang

Honda R&D Americas

Behzad Dariush
Ambarish Goswami

NASA

Karl Schweighofer

Roche Bioscience

Randal Walker
Lada Markovtsova
Doug Fenger

Sun Microsystems

Jon Arikata
William Smith
Stefan Unger

Wilson Sonsini Goodrich and Rosati

Ian Edvalson
David West
Chris Byrd

SYMPOSIUM SPONSORS

Full Sponsors

- **Affymetrix**
- **Alloy Ventures**
- **Apple Computer**
- **BioX**
- **NIH BISTI**
- **Roche Bioscience**
- **Sun Microsystems**

Half Sponsors

- **BioScience Forum**
- **Entelos**
- **Genentech**
- **Honda R&D Americas**
- **NASA**
- **Wilson Sonsini Goodrich & Rosati**