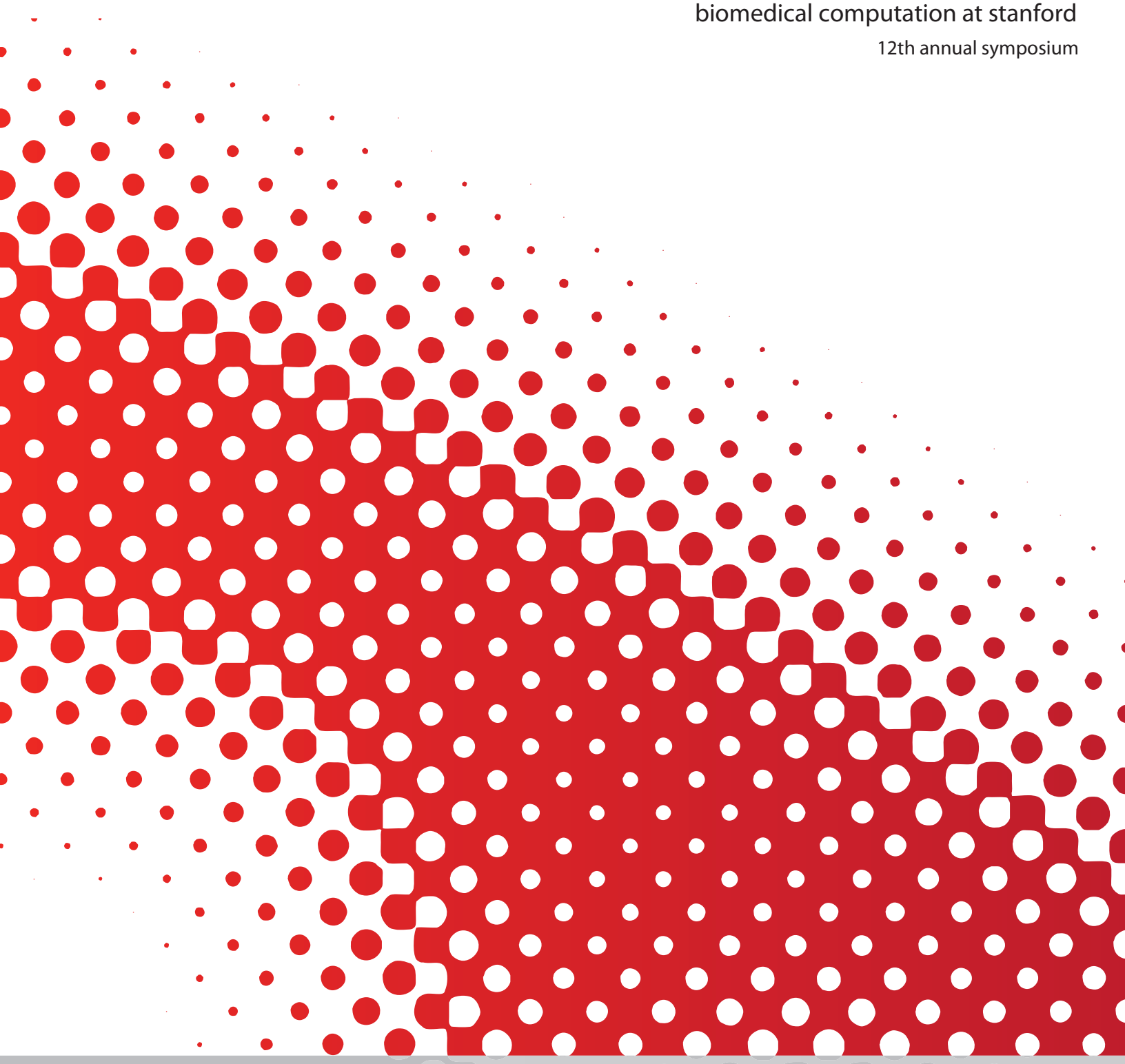
bcats 2011

biomedical computation at stanford

12th annual symposium



abstracts

Welcome

to the twelfth annual symposium on
Biomedical Computation at Stanford (BCATS)

This student-run one-day symposium provides an interdisciplinary forum for students and post-docs to discuss their latest work in computational biology and medicine with their peers at Stanford and other local universities. Since its inception in 1999, BCATS has seen growth and change in the field of biomedical computation and has evolved in concert. This year's schedule features cutting-edge research from one of the most diverse pools of participants in its 12 year history.

We thank our keynote speakers, student presenters, judges, sponsors, and all 2011 attendees.

The BCATS 2011 organizing committee

Harendra Guturu, Electrical Engineering
Nandita Garud, Genetics
Jaclyn Chen, Electrical Engineering
Lauren Chircus, Chemical and Systems Biology
Francisco Gimenez, Biomedical Informatics
Alicia Martin, Genetics
Sanna Ali, Biomedical Computation

NIH Center for Biomedical Computation

enabling groundbreaking research in physics-based simulations of biological structures

Interested in how biocomputation is changing biology and medicine?

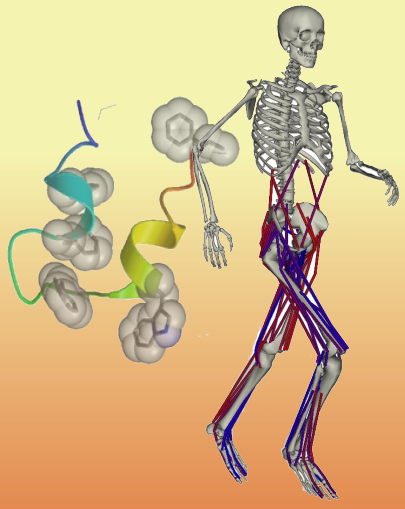
Sign up for a free subscription at:
www.BiomedicalComputationReview.org

Want to develop, share or find biosimulation software or data?

Explore the biosimulation repository and development environment at: www.simtk.org

Looking for high performance tools to simulate biological structure movement?

Download Simbody, an open-source library for rigid body dynamics: www.simtk.org/home/simbody



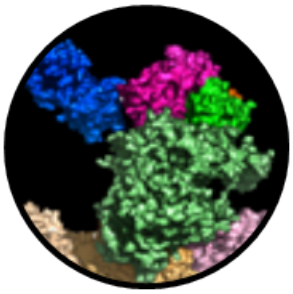
Interested in collaborating on important computational biological problems?

Visit us at: http://simbios.stanford.edu

Running image courtesy of Sam Hamner
NIH Roadmap Grant U54 GM072970

Bio-X at Stanford

To Discover … To Educate … To Invent



What is Bio-X?

Bio-X is Stanford's pioneering interdisciplinary biosciences program. Bio-X brings together biomedical and life science researchers, clinicians, engineers, physicists, and computational scientists to unlock the secrets of the human body. Rather than study cells and tissues in isolation, Bio-X investigators work to understand entire organ systems in all their complexity.

Who is affiliated with Bio-X?

Over 500 Stanford faculty from the Schools of Humanities & Sciences, Earth Sciences, Engineering, Law, and Medicine, representing 60+ departments.

What are Bio-X's programs?

- Seed Grants for Success (Interdisciplinary Initiatives Program)
- Bio-X NeuroVentures Program
- Bio-X Graduate Student Fellowships
- Bio-X Stanford Interdisciplinary Graduate Fellowships in Human Health
- Bio-X Undergraduate Research Awards
- Bio-X Travel Awards

What is the Bio-X Corporate Forum program?

A networking portal for industry companies to create stronger and lasting relationships with Stanford faculty. Opportunities for companies include symposia/seminars/mixers, customized technical summits, faculty liaison, and more!

Bio-X is facilitated by the James H. Clark Center, which comprises the equipment, resources and utilities required to conduct breakthrough research at the cutting edge of engineering, science and medicine.

To learn more about Bio-X, please visit our website at
http://biox.stanford.edu

ion torrent

by life technologies™

Powered by semiconductor technology.
Propelled by a global community.

Get on the map with the Ion Personal Genome Machine™ (PGM™) Sequencer.

In just one year, throughput has increased 100-fold, sequencing readlength has doubled to 200 bp—recently topping 400 bp internally—and we're on a path to clinical applications. We've also opened our protocols, datasets, and source code to the Ion Community—a network of scientists who are collaborating and developing applications. When the global community gets open access to transformative technology, you get a genomic revolution.

Watch the video at www.lifetechnologies.com/ionfirstnine

Register for your chance to win $1 million at www.lifetechnologies.com/grandchallenges

life technologies™

Stanford Biomedical Informatics Training Program

bmi.stanford.edu

Stanford: leading research and training in biomedical informatics since 1982

•Special expertise in key informatics research areas:

- •Translational Bioinformatics
- •Clinical Informatics
- •Imaging Informatics
- •Ontologies and the Semantic Web
- •Temporal Reasoning
- •Data Integration
- •Physics-based Simulation

•Established research collaborations with over 40 labs across campus
•A unified Bioinformatics and Clinical Informatics program
•Engineering, Life Sciences and the Medical School are all located on one campus

Full-Time On-Campus Programs

- •PhD
- •PhD Minor
- •Research-based Masters
- •Co-terminal Masters for Stanford undergraduates

Rigorous training programs including courses in informatics, computer science, probability and statistics, decision science, life sciences and ethics.

Part-Time Distance Learning Programs

- •Professional Masters Program (Honors Cooperative Program)
- •Certificates in Bioinformatics and Clinical Informatics
- •Individual Graduate Classes – the Non-Degree Option

For more information, contact:

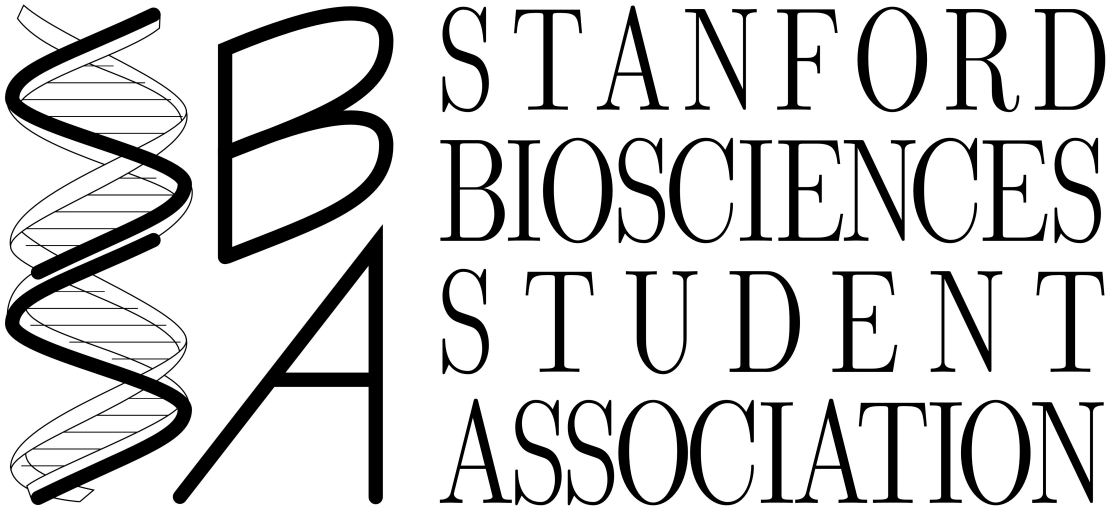
Student Services Officer
Stanford Biomedical Informatics Training Program
Medical School Office Building, Room X-215
1265 Welch Road, Mail Code: 5479
Stanford, CA 94305-5479

Phone: (650) 723-1398
Fax: (650) 725-7944
bmi-contact@lists.stanford.edu
bmi.stanford.edu

Nov. 2011

Additional Sponsors

http://forum.stanford.edu/



http://sbsa.stanford.edu/

http://buttelab.stanford.edu/

http://www.mendeley.com/

BCATS 2011 Schedule

(All events will take place in 230 LKSC)

8:30 On-Site Registration, Badge Pickup, and Breakfast

9:00 Opening Remarks

9:15 Keynote: **Elizabeth Burnside (University of Wisconsin, Madison)**

Data-driven Decision Support to Improve Breast Cancer Diagnosis and Outcomes

10:15 Break

10:30 Student Talks Session I

Ryan P. Jackson	*Stapes Three-Dimensional Vibration for a Biological Gear in the Middle Ear: Modeling and Measurements*
Katherine M. Steele	*How much muscle strength is required to walk in a crouch gait?*
Jonathan R Karr	*A Whole Cell Model of Mycoplasma genitalium Elucidates Mechanisms of Bacterial Growth and Replication*
Biao Li	*Automated Inference of Molecular Mechanisms of Disease from Amino Acid Substitutions*

11:30 Keynote: **Jason Myers (Ion Torrent)**

Democratizing Sequencing

12:30 Lunch

1:30 Keynote: **Tim Elston (University of North Carolina, Chapel Hill)**

Mathematical modeling of cell polarity

2:30 Poster Session

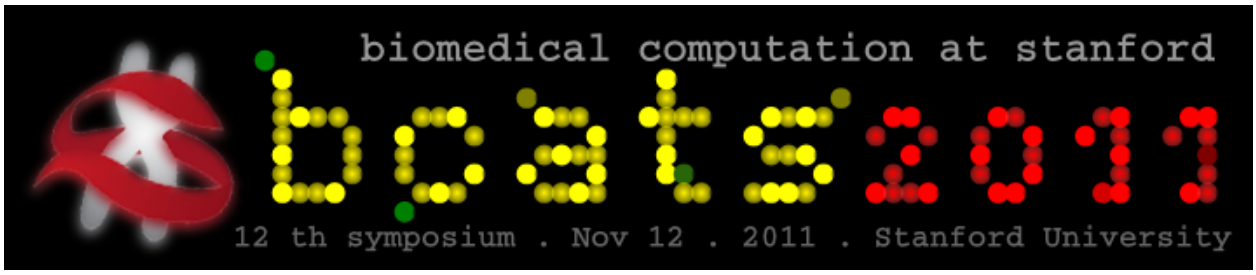
3:30 Student Talks Session II

Konrad J. Karczewski	*Assessing Functional and Clinical Significance of Regulatory Variants*
Zhengqing Ouyang	*SeqFold: Accurate genome-scale RNA structure reconstruction integrating experimental measurements provides insights into gene regulation*
Bethany Percha	*Discovery and Explanation of Drug-Drug Interactions via Text Mining*
Arnav Moudgil	*The Evolutionary Potential of Lattice Proteins*
Adam Wang	*Synthetic CT: Simulating arbitrary low dose single and dual energy protocols*
Tobias Wittkop	*STOP and DEFOG: Web applications for a comprehensive functional gene set analysis*

5:00 Awards, Closing Remarks and Reception

BCATS 2011 Posters

Poster		
1	Amsallem, David	Efficient Hodgkin-Huxley Simulations in Arbitrary Dendritic Networks Using Reduced-Order Modeling
2	Bhaduri, Aparna	Rapid Identification of Nonhuman Sequencing in High Throughput Sequencing Data Sets
3	Connolly, Jay	Science Exchange: An Innovative New Marketplace for Scientific Services
4	Cordero, Pablo	Two-dimensional Chemical Mapping: Inference and Falsification of Structure in Non-Coding RNA
5	Devabhaktuni, Arun	Novel Mass Tag De Novo Sequencing Approach for Peptide Identification
6	Gao, Hong	Stochastic Modeling of Blood Cancer Prognosis Revealed by Deep Sequencing of Immune Repertoire
7	Golden, Daniel	Heterogeneity in DCE-MRI Predicts Response of Breast Cancer Patients to Neoadjuvant Chemotherapy
8	Kaganovich, Mark	Phosphorylation of Yeast Transcription Factors Correlates with the Evolution of Novel Sequence and Function
9	Kehr, Birte	Computation of Pairwise Local Matches for Whole-Genome Alignment
10	Kim, Hojin	Dose Optimization with TFOCS-Based Total-Variation Minimization for Dense Angularly Sampled and Sparse Intensity Modulated Radiation Therapy (DASSIM-RT)
11	Mendez, Derek	High Resolution Structure Refinement of F-actin Using Correlated X-ray Scattering
12	Newton, Yulia	Evidence of Biased Gene Conversion in Analysis of 1000 Genomes Trio Data
13	Ng, Sam	Predicting the Impact of Mutations in Cancer Using an Integrated Pathway Approach
14	Pimentel, Samuel	Phylo-RLQ: Three Table Ordination for Microbial Community Data
15	Tsai, Fen-Chiao	Ca2+ Pulses Control Local Cycles of Lamellipodia Retraction and Adhesion Along the Front of Migrating Cells
16	Tsai, Tony	Dynamic Cytoskeleton Organization Couple Cell Shape Variations with Migration Phenotypes in HL-60 Cells
17	Wagoner, Jason	Smoothly Decoupled Boundaries in Hybrid Solvent Simulations
18	Wang, Rui	Using Phase to Recognize Phonemes in the Brain
19	Yao, Yuan	Efficacy of Fixed Filtration for Rapid kVp-Switching Dual Energy X-ray Systems
20	Zhang, Xiaomeng	CT Metal Artifact Reduction by Constrained Optimization with a Model-based Scanning Scheme
21	Zhao, Yiqiang	Functional Organization and Its Implication in Evolution of the Human Protein-Protein Interaction Network



Keynote Speakers

BCATS Keynote Speaker

Elizabeth S. Burnside, M.D., MPH

University of Wisconsin

Associate Professor of Radiology, Breast Imaging Section
Affiliate appointment in the Department of Biostatistics & Medical Informatics

Elizabeth Burnside is currently an Associate Professor of Radiology in the University of Wisconsin School of Medicine and Public Health. She got her MD degree combined with master's in Public Health as well as a master's degree in Medical Informatics. As a result her research investigates the use of artificial intelligence methods to improve decision-making in the domain of breast imaging in the pursuit of improving the population based screening and diagnosis of breast cancer. This multidisciplinary research is facilitated by affiliate appointments in the Departments of Industrial Engineering, Biostatistics and Medical Informatics and Population Health Science at UW. Dr. Burnside has published over 40 peer review articles and serves as a charter member on the Biomedical Imaging Technology (BMIT) Study Section at the NIH. Her research has been funded by the NIH (predominantly the National Cancer Institute) and the DOD. Dr. Burnside is a subspecialty trained breast imager with an active clinical practice providing all imaging and interventional procedure utilized for the early diagnosis of breast cancer. Dr. Burnside was elected a Fellow in the Society of Breast Imaging in 2004. Most recent career milestones include assumption of the role of Vice Chair of Research in the Department of Radiology in 2010 and selection as the Executive Leadership in Academic Medicine (ELAM) fellow for the UWSMPH for 2010-2011.

Talk Title: Data-driven Decision Support to Improve Breast Cancer Diagnosis and Outcomes

Talk Abstract: In the new era of "-omic"-based research, many scientists have shifted from the study of the individual parts of a system to the system itself. This new paradigm focuses on a comprehensive collection of a fundamental data type that can provide a platform for a myriad of research directions on a given level ranging from the subcellular to the population. However, developing methodologies that integrate these rich data sources to inform and improve healthcare decisions on the patient level remains challenging. Our team of physicians, computer scientists, and industrial engineers at the University of Wisconsin has collaborated for the last decade to develop methods to improve breast cancer diagnostic decision-making using inductive logic programming, statistical relational learning, and Markov Decision Processes. Our algorithms are designed to utilize the ever-expanding, multi-relational data that predicts breast cancer including: genetic, imaging, and epidemiologic risk factors. This talk will present an overview of our research programs and provide a vision of the future of computational methods in the domain of breast cancer risk prediction.

BCATS Keynote Speaker

Timothy Elston, PhD

University of North Carolina, Chapel Hill

Professor, Director, Graduate Program in Bioinformatics and Computational Biology
Professor, Department of Pharmacology

Tim Elston is currently a Professor of Pharmacology and the Director of the Graduate Program in Bioinformatics and Computational Biology at the University of North Carolina at Chapel Hill. He received his Ph.D. in Physics from Georgia Institute of Technology. His research interests focus on understanding the dynamics of complex biological systems, and developing reliable mathematical models that capture the essential components of these systems. The projects in his lab encompass a wide variety of biological phenomena including signal transduction, cell fate decisions and gradient sensing in yeast, noise in gene networks and signaling pathways, airway homeostasis, and energy transduction in motor proteins. His lab also is interested in developing computational tools for performing stochastic and spatiotemporal simulations of signaling networks and image analysis.

Talk Title: Mathematical modeling of cell polarity

BCATS Keynote Speaker

Jason Myers, PhD

Ion Torrent

Director of Genomic Applications and Collaborations

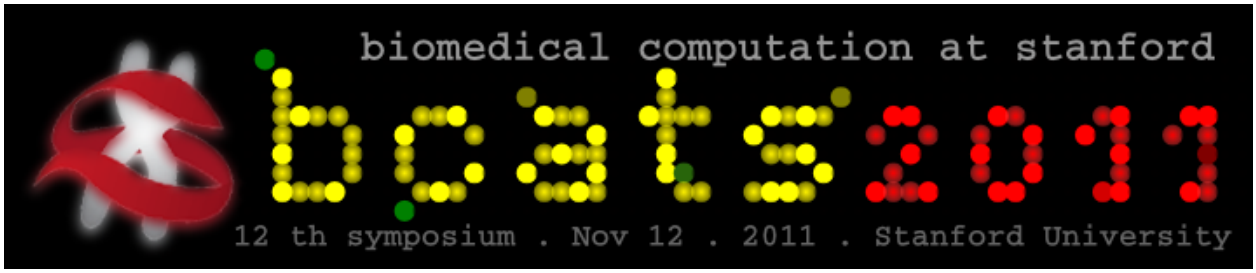
Jason was mentored by Jim Ferrell at Stanford University. During this time, he developed, patented, licensed and facilitated commercialization of the In Vitro Dicing Technology. In a collaborative effort with Tobias Meyer, Jason developed high-throughput and high-content screening methods to understand cellular signaling processes. Recombinant Dicing Kits have been used extensively by researchers to elucidate gene function.

As a postdoctoral fellow in Pat Brown's lab Jason developed microarray-based methods to investigate genome-wide post-transcriptional gene expression. These methods have been used to understand the regulation of gene expression in yeast and humans, including how miRNAs regulate the translation of mRNAs.

Jason was one of the first employees at Ion Torrent and helped to develop the biochemical aspects of the Ion Torrent Technology. Currently he leads an R&D group responsible for developing sequencing-based applications for the Ion Torrent PGM.

Talk Title: Democratizing Sequencing

Talk Abstract: Ion Torrent has pioneered an entirely new approach to sequencing. Instead of using light as an intermediary to sequence DNA, Ion Torrent uses semiconductor technology and simple chemistry. The technology translates chemical signals into digital information, much as a CMOS chip in a digital camera translates photons into pixels. Since it's availability in late 2010, the Ion PGM™ sequencer has expanded its capabilities to allow researchers to choose between different Ion Semiconductor Sequencing chips to produce from 10Mb to greater than 1Gb of highly accurate sequence (>99.999% consensus accuracy and >99.5% raw accuracy) in addition to expanding read lengths growing from 200 base pairs to a goal of >400 base pairs reads. Combining the read depth and read length flexibility with the speed and cost effectiveness of the platform, the Ion PGM Sequencer has enabled a rapid expansion in its application portfolio such as variant detection, RNA-Seq, ChIP-Seq, and de-novo genome assembly.



Talk Abstracts

talk 1

Stapes Three-Dimensional Vibration for a Biological Gear in the Middle Ear: Modeling and Measurements

Ryan P. Jackson
Stanford University

Hongxue Cai
Suni Puria

Purpose:
We hypothesize that the malleus-incus joint found in larger mammals may reduce overall middle ear input impedance at high input frequencies, which would increase output to the cochlea. The malleus and incus articulate through a mobile, saddle-shaped joint. Morphometry suggests that the malleus and incus can rotate or "twist" about independent axes in this configuration. This twisting would function like a torsional "gear" converting rotational motion about one axis to another axis. In doing so, the motion of the stapes, which connects the incus to the cochlea, and indicates system output, is predicted to change from a primarily one-dimensional, piston-like motion, to a complex three-dimensional motion. Anatomical considerations make comprehensive vibration measurements of the malleus and incus prohibitive, so stapes dynamics are examined to confirm gear-like motion.

Materials and Methods:
We have constructed a finite element model of the human middle ear using geometric data acquired from micro-CT scans and material data from literature. This model was compared to several experimentally measured parameters from the literature: middle ear input impedance, middle ear reflectance, and stapes velocity versus input pressure. We characterized the dynamics of the malleus, incus, and stapes with the malleus-incus joint normally articulated and when rigidly fused. In addition to our modeling work, we have measured sound-driven, stapes vibration for fresh human temporal bones (n=5) using both three-dimensional and one-dimensional laser vibrometry

Results:
Our model shows good agreement with literature values for middle ear input impedance, reflectance, and stapes velocity versus input pressure. In our model, we observed "twisting" of the malleus and incus in a gear-like fashion at high frequency. Twisting was eliminated when the joint was rigidly fused, though overall output appears to be higher. Preliminary experimental results show that stapes motion has complex three-dimensional motions at high-frequency, which correspond well with model predictions.

Conclusions:
We constructed a realistic human middle ear model in good agreement with existing and new middle ear experimental results. The model predicts twisting of the malleus and incus in a gear-like mechanism at high frequency as hypothesized. Stapes vibration measurements indicate complex three-dimensional motions that support a gear mechanism.
[Work supported by R01 DC005960 and ARRA supplements to SP/CRS and F30DC010305 to RPJ from the NIDCD of NIH.]

References:
O'Connor, KN, Tam, M, Blevins, N, and Puria, S. Tympanic Membrane Collagen Fibers: A Key to High-Frequency Sound Conduction. Laryngoscope March 2008; 118: 483-490

talk 2

How Much Muscle Strength Is Required to Walk in a Crouch Gait?

Katherine M. Steele
Stanford University

Marjolein van der Krogt
Michael Schwartz
Scott Delp

Purpose:
Individuals with cerebral palsy who walk in a crouch gait often participate in strength training programs; however, outcomes from these programs are inconsistent [1,2]. The purpose of this study was to determine how robust crouch gait is to muscle weakness compared to unimpaired gait. Identifying which muscle groups require more or less force during crouch gait can provide guidance for designing effective strength training programs.

Materials and Methods:
Musculoskeletal simulations were created for 3 unimpaired subjects and 9 subjects with cerebral palsy and crouch gait. A generic model [3] with 19 degrees of freedom and 92 musculotendon actuators was scaled to each subject and the maximum isometric force of all muscles was scaled by each subject's height squared [4]. The muscle force required to drive the model according to each subject's kinematics was estimated using the computed muscle control algorithm (CMC) [5]. The strength of each muscle group of interest was iteratively decreased by reducing the maximum isometric force until the model could no longer track the subject's kinematics. The muscle groups of interest included the ankle plantarflexors, anterior tibialis, vasti, bi-articular hamstrings, gluteus maximus, and gluteus medius.

Results:
Crouch gait was less sensitive to weakness of the gluteus medius and ankle plantarflexors than unimpaired gait. During crouch gait, the hip abduction moment and the ankle plantarflexor moment were reduced in terminal stance and, thus, required less gluteus medius and ankle plantarflexor strength. In contrast, the vasti strength required increased quadratically with crouch severity. During crouch gait, greater vasti muscle strength was required due to larger knee flexion moments during stance. Crouch gait also required greater gluteus maximus strength than unimpaired gait to extend the hip during stance. The maximum isometric force of anterior tibialis could be reduced to zero in all cases except severe crouch gait.

Conclusions:
Current strength training programs designed for individuals with crouch gait commonly target the hip and knee extensors; however, the results of this study demonstrate that crouch gait requires greater force from the hip and knee extensors than unimpaired gait and strengthening these muscles may not produce a more upright posture. In contrast, crouch gait could be a compensation for weak ankle plantarflexors or gluteus medius and strength training programs that target these muscles may produce better outcomes.

References:
[1] Mockford et al, Pediatr Phys Ther, 2008. [2] Damiano et al, Phys Ther, 2010. [3] Steele et al, J Biomech, 2010. [4] Scott et al, Muscle Nerve, 1982. [5] Thelen et al, J Biomech, 2003.

talk

3

A Whole Cell Model of Mycoplasma Genitalium Elucidates Mechanisms of Bacterial Growth and Replication

Jonathan R. Karr
Stanford University

Jayodita C Sanghvi
Jacobs M Jacobs
Derek N Macklin
Markus W Covert

A central challenge of biology is to understand how complex phenotypes are controlled by individual molecules and their interactions. We report the first computational model which explains the life cycle of an entire organism, Mycoplasma genitalium, including metabolism, macromolecule synthesis, and cytokinesis, from the level of individual molecules and their chemical interactions. The hybrid computational model consists of submodels of 28 cellular processes integrated through 16 cellular states, accounts for the specific function of every annotated gene product, and predicts the dynamics of every molecule. Using the model we identified the molecular determinates of cellular growth and replication. We found the M. genitalium cell cycle is 9.0 ± 0.5 h, and that most of its variance is due that of metabolism and of thymidylate and acetate kinase expression. Additionally, we found that replication initiation and DnaA binding dynamics are a significant source of cell cycle length variation among fast growing cells. We examined the genetic requirements of single cell growth and found four distinct classes of single-gene deletion strains: strains indistinguishable from wild-type, strains with early growth cessation, strains with slowly decaying growth, and non-dividing strains. The model correctly predicts the experimentally observed essentially of > 80% of genes. We believe that gene-complete models will accelerate biomedical discovery and bioengineering by enabling rapid, low cost in silico experimentation, facilitating experimental design and interpretation, and guiding rational engineering of biological systems and medical therapies.

talk

4

Automated Inference of Molecular Mechanisms of Disease from Amino Acid Substitutions

Biao Li
The Buck Institute for Research on Aging

Vidhya G. Krishnan
Matthew E. Mort
Fuxiao Xin
Kishore K. Kamati
David N. Cooper
Sean D. Mooney
Predrag Radivojac

Purpose:
Single nucleotide substitutions within protein coding regions are of particular importance owing to their potential to give rise to amino acid substitutions that affect protein structure and function which may ultimately lead to a disease state. Over the last decade, a number of computational methods have been developed to predict whether such amino acid substitutions result in an altered phenotype. Although these methods are useful in practice, and accurate for their intended purpose, they are not well suited to providing probabilistic estimates of the underlying disease mechanism.

Materials and Methods:
We have developed a new computational model, MutPred, that is based upon protein sequence, and which models changes of structural features and functional sites between wild-type and mutant sequences. These changes, expressed as probabilities of gain or loss of structure and function, can provide insight into the specific molecular mechanism responsible for the disease state.

Results:
MutPred also builds on the established SIFT method but offers improved classification accuracy with respect to human disease mutations. Given conservative thresholds on the predicted disruption of molecular function, we propose that MutPred can generate accurate and reliable hypotheses on the molecular basis of disease for ~11%of known inherited disease-causing mutations.

Conclusions:
We also note that the proportion of changes of functionally relevant residues in the sets of cancer-associated somatic mutations is higher than for the inherited lesions in the Human Gene Mutation Database which are instead predicted to be characterized by disruptions of protein structure.

talk

5

Assessing Functional and Clinical Significance of Regulatory Variants

Konrad J. Karczewski
Stanford University

Joel T. Dudley
Nicholas P. Tatonetti
Russ B. Altman
Atul Butte
Michael Snyder

Many genomic variants are discovered outside of genes, where their functional consequences are more difficult to characterize. However, as many of these variants are associated with disease, it is likely that they affect molecular physiology at the level of gene regulation. We investigate the role of variants in regulatory regions, on both transcription factor cooperativity as well as disease pathophysiology. First, we developed the Allele Binding Cooperativity (ABC) test and the ALPHABIT pipeline, which utilizes variation in transcription factor binding among individuals to discover combinations of factors and their targets. We find some factors that have been known to work with NFκB (E2A, STAT1, IRF2), but whose global co-association and sites of cooperative action were not known, and discover one co-association (EBF1) that had not been reported previously. Second, we demonstrate a systematic approach to combine disease association, transcription factor binding, and gene expression data to assess the functional consequences of variants associated with hundreds of human diseases. We find that disease-associated SNPs are enriched in NFκB binding regions overall, and specifically for inflammatory mediated diseases, such as asthma and atherosclerosis. Using genome-wide binding variation information, we find regions of NFκB binding correlated with disease-associated variants in an allele-specific manner. Furthermore, we show that this binding variation is often correlated with expression of nearby genes, which are also found to have altered expression in independent profiling of the variant-associated disease condition. In this systematic approach, we close a major loop in biological context-free association studies and assign putative function to many disease-associated SNPs.

talk

6

SeqFold: Accurate Genome-Scale RNA Structure Reconstruction Integrating Experimental Measurements Provides Insights Into Gene Regulation

Zhengqing Ouyang
Stanford University

Michael P. Snyder
Howard Y. Chang

Purpose:
RNA structure is essential for nearly every step in the biogenesis, function, and regulation of coding and noncoding RNAs [1]. The recent advent of experimental methods to measure RNA accessibilities at a genomic-scale has raised the promise of genome-scale reconstruction of RNA structures, but the optimal strategy toward this goal is unclear. Important metrics for success include the accuracy of the reconstruction, as well as computational efficiency such that it can be scaled up to an entire transcriptome. Here we introduce a new integrative strategy, called SeqFold, that allows accurate and automated reconstruction of RNA structure genome-wide given input experimental RNA accessibility data.

Materials and Methods:
SeqFold integrates computational modeling of RNA secondary structures with the massive amount of deep sequencing signals from the parallel analysis of RNA structure (PARS) [2], which measures the genome-wide cleavage sites of structure-specific enzymes. It statistically infers the structure preference profiles from sequencing reads, and uses it to prioritize computationally sampled RNA secondary structures. Using machine learning algorithms, SeqFold efficiently identifies the most likely centroid of RNA structure clusters in the high-dimensional space of sampled structures.

Results:
We use SeqFold to reconstruct the secondary structures of over 3,000 yeast RNAs. It accurately predicts the known structures of both mRNAs and non-coding RNAs. Analysis of the structural profile output of SeqFold reveals the diverse roles of RNA secondary structure in gene regulation. Comparing to previous studies, SeqFold-derived RNA accessibility is much more widely correlated with ribosome density, a proxy of translation efficiency. SeqFold also reveals a hitherto unknown correlation of 5' RNA accessibility with Poll II density and the localization of histone modifiers. Further more, it demonstrates the effectiveness of RNA structure information in identifying RNA binding protein targets, validated by genome-wide RIP-chip data.

Conclusions:
We show that SeqFold enables highly accurate reconstruction of known RNA structures not possible with previous tools, and that the accurate structural landscape reveals significant roles of RNA structures in controlling transcription initiation, translation, and RNA-protein interactions. SeqFold should be widely applicable to analyze and compare RNA structures in any transcriptome.

References
1. Wan, Y., Kertesz, M., Spitale, R.C., Segal, E. & Chang, H.Y. Nat Rev Genet 12, 641-655 (2011).
2. Kertesz, M., Wan, Y., Mazor, E., Rinn, J.L., Nutter, R.C., Chang, H.Y. & Segal, E. Nature 467, 103-107 (2010).

talk

7

Discovery and Explanation of Drug-Drug Interactions via Text Mining

Bethany Percha
Stanford University

Yael Garten
Russ B. Altman

Purpose:
Drug-drug interactions (DDIs) can occur when two drugs interact with the same gene product. Most available information about gene-drug relationships is contained within the scientific literature, but
is dispersed over a large number of publications, with thousands of new publications added each month. In this setting, automated text mining is an attractive solution for identifying gene-drug relationships and aggregating them to predict novel DDIs. In this work, we hypothesize that we can combine data about gene-drug relationships, drawn from a very broad and diverse literature, to infer both known and novel DDIs.

Materials and Methods:
In previous work, we have shown that gene-drug interactions can be extracted from Medline abstracts with high fidelity - we extract not only the genes and drugs, but also the type of relationship expressed in individual sentences (e.g. metabolize, inhibit, activate and many others). We normalize these relationships and map them to a standardized ontology. Equivalent relationships are mapped to common standards in a context-sensitive manner. Using a training set of established DDIs, we have trained a random forest classifier to score potential DDIs based on the features of the normalized assertions extracted from the literature that relate two drugs to a gene product.

Results:
The classifier recognizes the combinations of relationships, drugs and genes
that are most associated with the gold standard DDIs, correctly identifying 79.8% of assertions relating interacting drug pairs and 78.9% of assertions relating non-interacting drug pairs. Most significantly, because our text processing method captures the semantics of individual gene-drug
relationships, we can construct mechanistic pharmacological explanations for the newly proposed DDIs.

Conclusions:
Our classifier can be used both to explain known DDIs and to uncover new DDIs that have not yet been reported.

talk

8

The Evolutionary Potential of Lattice Proteins

Arnav Moudgil
Stanford University

Michael Palmer

Purpose:
Lattice proteins are a simplified model of protein folding where each of the 20 amino acids are constrained to occupy points on a two-dimensional lattice. This model permits inexpensive estimation of the folded conformation of a sequence, allowing simulations of populations of thousands of proteins over thousands of generations of evolution. Previous studies by Bloom et al. suggest that protein stability is important to a protein's capacity to evolve, because it allows a protein to tolerate deleterious mutations on its way to gaining beneficial mutations. We set out to verify these studies and discover other possible factors promoting protein evolution.

Materials and Methods:
We used Python simulations of lattice protein evolution and ran multiple replicates on a parallel computer cluster. We screened a large number of randomly generated proteins to generate 1600 stably-folding proteins (those with some folded conformation with delta G < 0). We divided the metapopulation into 32 demes of 50 individual proteins each. Each deme contained a unique target ligand; the fitness of a protein is a function of its strength of binding to the target ligand. The mutation rate per residue and the migration rate among demes were held constant. The descendants of the initial proteins were tagged and tracked throughout the experiments. At the end of the evolutionary runs we evaluated the expected count of the descendants of each initial protein after 2000 generations. This is a metric of ""long term fitness"", as opposed to the standard one-generation fitness used in population genetics. Each set of runs were replicated 264 times.

Results:
We analyzed several sets of proteins whose initial stabilities (i.e., whose free energy of folding) ranged uniformly between 0 and -3.5Joules. A handful of the initial 1600 proteins were the reliable long-term winners, in terms of their expected count of descendants after 2000 generations. The ranking of the top 10 proteins was very repeatable. We first found a weak but significant negative correlation between initial stability and the 2000-generation fitness of that protein lineage (Spearman correlation coefficient rho =- 0.053, $p < 2.2e-16$). We next looked at the 1-generation fitness as a predictor of the 2000-generation fitness, finding a stronger correlation than initial stability (rho= 0.234, $p < 2.2e-16$).

Conclusions:
The initial stability of a protein is only a weak predictor of protein's ability to evolve binding towards several ligands. In contrast, a protein's initial (one-generation) fitness is a good predictor of a protein's success. We plan to investigate what additional factors contribute to the long-term fitness; for example, robustness of a fold to mutations.

talk 9

Synthetic CT: Simulating Arbitrary Low Dose Single and Dual Energy Protocols

Adam Wang
Stanford University

Yuan Yao
Charles Feng
Norbert Pelc

Purpose:
A primary disadvantage of medical x-ray imaging is the radiation dose delivered to the patient. However, the minimum dose needed for diagnostic quality images is a complicated function of the imaging protocol, diagnostic task, and patient anatomy. In an attempt to study this trade-off, existing methods can synthesize images of lower exposure (mAs) but are limited to the same single incident spectrum (kVp and filtration).

Our work allows users to retrospectively study how low dose single or dual energy protocols affect image quality by accurately synthesizing radiographs or CT scans of the same subject that would be acquired with some other protocol – including arbitrary kVp and filtration. Therefore, a tool incorporating this method can be used both to teach practitioners about imaging protocols and to select lower dose protocols for future clinical scans.

Materials and Methods:
Low and high energy scans are used to form a pair of material decomposition images. Assuming knowledge of what the simulated protocol's spectrum is, we synthesize the CT measurements using forward projection through the full field of view material decomposition. The target noise level is calculated, and noise is added to simulate low dose protocols. The synthesized CT measurements are then reconstructed.

These techniques were tested using a cylindrical acrylic phantom with contrast inserts imaged with a GE HD750 CT scanner. The original dual energy scan was done at 80 kVp and 140 kVp, and we synthesized images at 100 kVp and 120 kVp over a wide range of exposures (mAs). The synthesized images were compared with actual images acquired at the same protocols (e.g., 100 kVp, 400 mAs).

Results:
The synthetic CT algorithm accurately matches both signal and noise properties in the projections and decompositions, so the synthetic reconstructed images have the same signal and noise levels as an actual image from the simulated protocol, which can have a different kVp and mAs from the original dual energy scans. Note that the synthesized images reproduce the polychromatic effect of x-ray tubes and even beam hardening, just as realistic CT images have.

Conclusions:
Synthetic CT generates realistic single energy or dual energy images of a patient as if they were scanned at a different protocol and offers radiologists the remarkable capability of retrospectively studying the effects of protocol changes on dose and image quality. As a teaching aid, this provides realistic and potentially real-time feedback to teach users how scanning parameters affect image quality and patient dose. As a clinical tool, it can be used to select minimum dose protocols for future scans.

talk 10

STOP and DEFOG: Web Applications for a Comprehensive Functional Gene Set Analysis

Tobias Wittkop
The Buck Institute for Research on Aging

Emily Teravest
Ari E. Berman
Uday Evani
Matthew Fleisch
Corey Powell
Nigam Shah
Sean D. Mooney

Purpose:
One of the most common outcomes of high-throughput biological experiments is a list of genes or proteins of interest. In order to explain the observed changes of these specific genes and to create new hypotheses one needs to understand the functions and roles of the genes in the lists under the condition studied in the experiment.

Materials and Methods:
Here we present two novel applications that facilitate the extraction of functional content of lists of genes. With STOP we overcome the limitations of manually annotated ontologies (like GO) and use automatic annotation techniques using descriptive text as an annotation source. This resource utilizes the National Center for Biomedical Ontology's database(1), which includes over 200 biomedically related ontologies. These automatic annotations are then processed and used for enrichment analyses against submitted gene lists. Our second application, DEFOG, eases the functional analysis of gene sets by hierarchically organizing them into functional related groups using data fusion of high-throughput experimental data. The underlying computational pipeline utilizes the state-of-the-art applications GeneMANIA(2), Transitivity Clustering(3), and Ontologizer(4) for gene set specific network fusion, non-agglomerative hierarchical clustering, and GO term enrichment respectively.

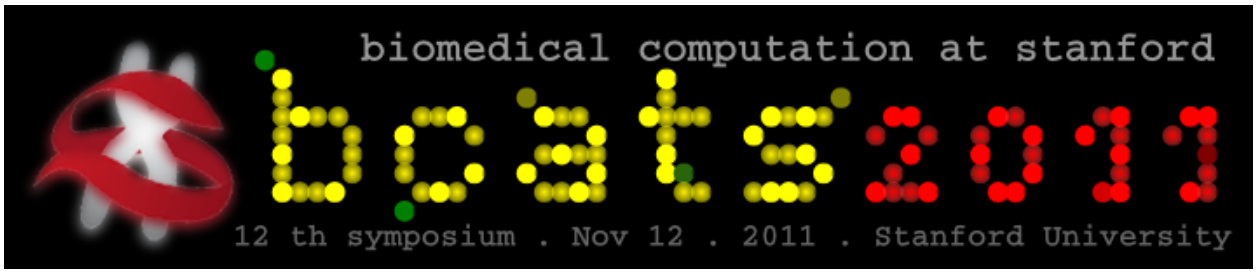
Results:
In recent studies, we demonstrated how accurate the automatic annotations in STOP can be. STOP ranked 7th in a recent comparison of functional annotation tools (CAFA, http://biofunctionprediction.org/) when trying to predict novel annotations in the biological process category of GO. We further demonstrate the usability of both our methods by applying them on publicly available gene sets, such as the primary interactors of the Huntingtin protein in the human PPI network and a list of human genes known to be involved in the aging process.

Conclusion:
With STOP and DEFOG we present two easy-to-use web applications whose novel approaches on functional gene set analysis aid in the discovery of potentially important biological mechanisms and assist in the generation of new hypotheses from gene lists.

Webpage: http://mooneygroup.org/stop and http://mooneygroup.org/defog

References:

1. D. L. Rubin et al., Omics 10, 185–198 (2006).
2. D. Warde-Farley et al., Nucleic Acids Res. 38, W214–20 (2010).
3. T. Wittkop et al., Nat. Methods 7, 419–420 (2010).
4. S. Bauer, S. Grossmann, M. Vingron, P. N. Robinson, Bioinformatics 24, 1650–1651 (2008).



Poster Abstracts

Poster 1

Efficient Hodgkin-Huxley Simulations in Arbitrary Dendritic Networks Using Reduced-Order Modeling

David Amsallem
Stanford University

Jaijeet Roychowdhury

Purpose:
The goal of this project is to develop a fast an accurate computational method for predicting the potential propagation in complex dendritic trees of a neuronal cell. This is a first step towards simulating potential propagation in a large network of multiple, interconnected neurons. Starting from a very large multi-compartmental model, it is proposed to reduce the number of equations, thereby reducing the computational cost, while preserving accuracy.

Materials and Methods:
The Hodgkin-Huxley [1] equations are considered for a computational domain encompassing the dendritic branches and the soma. Current is injected at one or several locations of the dendritic tree and the potential propagation in the whole tree computed. After a discretization of the equations by the finite differences methods, one obtains a very large set of nonlinear differential equations. The proposed method proceeds by reducing the number of degrees of freedom involved in those equations. Proper Orthogonal Decomposition (POD) [2] first reduces the number of unknowns for a single branch. Then, multiple such reduced unknown vectors are assembled for the whole dendritic tree.
Reducing the number of unknowns however does not reduce the complexity of the equations and the computational cost is not reduced. Therefore, a second level of approximation is considered so that the computational cost decreases dramatically: a Gappy POD-based technique [3,4] is applied so that only a subset of the equations of interest is evaluated in the simulation process. Similarly to the first level of approximation, the Gappy POD technique is applied to a single dendritic branch, resulting in a simpler model for that branch. Then, such simple models are assembled for the whole tree. Furthermore, computations in multiple tree configurations can be conducted after constructing the reduced-order model for a single branch.

Results:
The proposed method is shown to dramatically reduce the simulation cost for several different scenarios of dendritic trees. For a single branch, the number of unknowns can be reduced from more than 1400 to 30, which is 46 times less. As the reduced model for a dendritic tree is obtained by assembling multiple reduced-order models built for one branch, the number of unknown for the whole tree is also reduced by a factor 46.

Conclusions:
Reducing the computational cost associated with the simulation of potential propagation in a neuronal cell is a first step towards the simulation of more complex networks of neuronal cells. In this work, it is shown that reduced-order modeling techniques are effective in reducing the computational cost associated with the Hodgkin-Huxley equations while preserving a good accuracy of the predictions. Furthermore, since the proposed method builds a model for a single branch, arbitrary networks of such branches can be considered.

References:
[1] E.M. Izhikevich. Dynamical Systems in Neuroscience - The Geometry of Excitability and Bursting. MIT Press, Cambridge, Massachusetts, 2007.

[2] L Sirovich. Turbulence and the dynamics of coherent structures. part i: Coherent structures. Quarterly of applied mathematics, 45(3):561–571, 1987.

[3] AR Kellems, S Chaturantabut, DC Sorensen, and SJ Cox. Morphologically accurate reduced order modeling of spiking neurons. Journal of computational neuroscience, pages 1–18, 2010.

[4] K Carlberg, C Bou-Mosleh, and C Farhat. Efficient nonlinear model reduction via a least squares petrov–galerkin projection and compressive tensor approximations. International Journal for Numerical Methods in Engineering, 86(2):155–181, 2011.

Poster

2

Rapid Identification of Nonhuman Sequencing in High-Throughput Sequencing Data Sets

Aparna Bhaduri
Stanford University

Kun Qu
Carolyn S. Lee
Alexander Ungewickell
Paul Khavari

Purpose and Methods:
Rapid Identification of Nonhuman Sequences (RINS) is an intersection-based pathogen detection workflow that utilizes a user provided custom reference genome set for identification of nonhuman sequences in high throughput sequencing datasets.

Results:
In less than 2 hours, RINS correctly identified the known virus in the dataset SRR73726, and is compatible with any computer capable of running the pre-requisite alignment and assembly programs. RINS is capable of accurately identifying sequencing reads from intact or mutated nonhuman genomes in the dataset, and robustly generating contigs with these nonhuman sequences.

Poster 3

Science Exchange: An Innovative New Marketplace for Scientific Services

Jay Connolly
Science Exchange

Ryan Abbot
Dan Knox
Elizabeth Iorns Ph.D.

Purpose:
The goal of Science Exchange is to promote efficiency and collaboration in scientific research. Scientific research is becoming increasingly specialized, necessitating greater use of collaborations and outsourcing to draw on experimental expertise. To facilitate this, many research institutions have established core facilities to help reduce the inefficiencies associated with the historical system of 'bartering' to gain access to specialized equipment and expertise. However, the current core facility system is fragmented: researchers have varying degrees of access to services, quality is hard to evaluate and pricing is not transparent. Researchers who need to collaborate outside of their institution lack an easy way to find providers, evaluate them, coordinate logistics and pay for work. Similarly, core facilities lack a robust mechanism for marketing their expertise and services in order to attract new business from outside their institution.

Materials and Methods:
Science Exchange, an online marketplace accessible at www.ScienceExchange.com, was launched in August 2011 to address these issues. Science Exchange works by signing up core facilities as "providers" of different experiment types. Researchers can search for an experiment type they wish to outsource and choose a facility to perform the work, or post an open project and receive bids from qualified facilities. Science Exchange acts as a centralized hub for provider information and reviews, and assists with project management, including billing and payment. The goal of Science Exchange is to create a vibrant and accessible marketplace for experimental services.

Results:
As of October 2011 more than 3,000 scientists from over 500 research institutions have registered with Science Exchange. On average, experiments outsourced through the platform have saved users 54% of the value of their experiment ($4,664 per experiment).

Conclusions:
Science Exchange addresses the barriers to effective experiment outsourcing. As with any marketplace, awareness and participation by key players is essential. Our team is committed to responsiveness and making the platform as useful for researchers and providers as possible.

Poster

4

Two-Dimensional Chemical Mapping: Inference and Falsification of Structure in Non-Coding RNA

Pablo Cordero
Stanford University

Wipapat Kladwang
Chris VanLang
Rhiju Das

Although it is generally acknowledged that structure is a key determinant to RNA function, current state-of-the-art phylogenetic and biophysical tools cannot always confidently reveal RNA structure at nucleotide resolution. To address this problem, we have developed a high-throughput two-dimensional extension of classic chemical mapping approaches, called the 'mutate-and-map' strategy. Systematic mutation of each nucleotide leads to perturbations of the chemical accessibilities of the nucleotide's base pairing partners, giving an experimental readout of an RNA's 'contact map'. Testing the approach on a benchmark of riboswitch, ribosomal, and ribozyme domains, the method recovers the RNAs' secondary structures with unprecedented accuracy. In addition, the mutate-and-map measurements permit the generation and falsification of hypotheses for structural rearrangements in three ligand-binding RNAs, including a cooperative glycine riboswitch with currently unknown mechanism. Finally, in silico simulations suggest that the method can discriminate RNA sequences that have robust structures from a random background, a prospect we are now testing in vitro.

Poster 5

Novel Mass Tag De Novo Sequencing Approach for Peptide Identification

Arun Devabhaktuni
Stanford University

Josh Elias

Purpose:
Tandem mass spectrometry is a powerful tool for high-throughput proteomics, but peptide identification from mass spectra typically requires an annotated sequence database [1]. This approach generally cannot detect unknown modifications, uncharacterized splice isoforms, mutations, and unknown proteins. Through de novo sequencing, peptides are inferred directly from mass spectra with none of these drawbacks. However, such algorithms have hereto been limited by the quality of spectral data [2]. Here, we present a combined experimental/computational approach, in which complex peptide mixture is isotopically labeled to elucidate the identify of ions in resultant fragmentation spectra and analyzed on high mass accuracy instruments. The resulting spectra are deisotoped and denoised, and peptides-of-origin are inferred by our Label Assisted De novo Sequencing (LADS) software.

Materials and Methods:
Mouse embryonic stem cells (LF2) were partially metabolically labeled with isotopically heavy (+8 Da) lysine [3]. Cells were lysed, digested with the protease LysC, acidified, and analyzed via a data dependent acquisition method, alternating between CID and HCD fragmentation for the top five most intense ions observed in an LTQ Velos Orbitrap. The resulting spectra were analyzed by LADS.

Results were compared to output by SEQUEST and Mascot [1], two database search programs, and evaluated for accuracy and precision. Accuracy and precision scores are determined from the number of fragmentation sites correctly predicted by LADS and dividing by the number predicted by SEQUEST or Mascot and predicted by LADS, respectively.

Results:
For paired peptides predicted by SEQUEST, LADS has an accuracy of 87% and a precision of 85%. Compared to MASCOT, LADS has an accuracy of 90% and a precision of 90%. In addition, a number of peptides unsuccessfully identified by SEQUEST and MASCOT returned high LADS scores, including several previously uncharacterized peptides.

Conclusions:
LADS presents a viable alternative to traditional database searches when analyzing uncharacterized samples, and can complement database methods when analyzing complex protein samples.

References:
1. Geveart, K and Vandekerckhove, J. Electrophoresis 21: 1145-1154 (2000).
2. Pevtsov, S et al. J. Proteome Res. 5(11): 3018-3028 (2006).
3. Mann, M et al. Mol Cell Proteomics 1(5): 376-86 (2002).

Poster

6

Stochastic Modeling of Blood Cancer Prognosis Revealed by Deep Sequencing of Immune Repertoire

Hong Gai
Stanford University

Chunlin Wang
Carlos D. Bustamante
Michael Mindrinos
David Miklos
Ronald W. Davis
Wenzhong Xiao

In patients with chronic lymphocytic leukemia (CLL), one or a few transformed B cells experiences a rapid expansion and dominates in B cell population of patients' peripheral blood. This drastic expansion of transformed B cells suggests that they might evolve through processes distinct from those of the normal B cells. To understand the underlying mechanism of CLL, we sequenced the V-J junction segments of immunoglobulin heavy chain amplified from isolated B cells collected from both CLL patients and their corresponding allogeneic hematopoietic cell transplantation (allo-HCT) donors using the Roche/454 sequencing platform. The samples of CLL patients were collected at diagnostics, and after allo-HCT at 56 days, 180 days, 365 days, and 550 days. From those sequencing data, we modeled the reconstruction of immune repertoire of CLL patients post transplantation in the absence of antigen stimulations using a linear birth process with immigration, where abnormalities occurring in an immune repertoire such as cancer can be detected as substantial deviations from this null model. We applied the Ewens sampling test (EST) derived from this stochastic process to distinguish the immune repertoires of the CLL patients from healthy controls and to monitor their recovery or relapse accurately. Extensive simulations based on repertoire sequencing data show that the EST is sensitive in detecting cancer clones in CLL patients. In addition, we estimated the immigration rate of donor's B cell clones entering the circulatory system and the average time to reconstruct the normal immune repertoire after transplantation, which are useful clinical parameters. We anticipate that our stochastic model will be useful in diagnostics of CLL and prognosis after transplantation and it is straightforward to extend it to other immune-related diseases.

Poster

7

Heterogeneity in DCE-MRI Predicts Response of Breast Cancer Patients to Neoadjuvant Chemotherapy

Daniel Golden
Stanford University

Daniel Rubin

Purpose
Breast cancer is the most common type of fatal cancer in women in the United States today, with a 5-year survival rate of 89%. An unsolved critical challenge to treating patients with breast cancer is predicting who will respond to therapy, as breast cancer is a diverse disease and there is currently insufficient information to guide tailored treatment in individual patients. A model which uses minimally-invasive techniques and is able to predict whether patients will respond to therapy could be used clinically to significantly improve the ability to choose the optimal treatments for individual patients, which will greatly increase their ability to combat the disease.

Materials and Methods
This work uses dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI), which has the ability to reveal information about tissue kinetic properties, to predict response to therapy for breast cancer patients. We make use of data from patients with triple-negative breast cancer who are involved in a current clinical trial at Stanford to test treatment by neoadjuvant chemotherapy with PARP inhibitors. Using pre-therapy DCE-MRI images, we developed quantified measures of the kinetic heterogeneity of breast lesions. These measures of heterogeneity were used as features of a multiple regression model to predict the response, based on post-therapeutic residual cancer burden (RCB), of individual patients to chemotherapy. We designed a linear regression model to predict post-therapeutic RCB based on patient age, number of distinct kinetic regions within the lesion, and kinetic lesion texture.

Results
In preliminary results using nine patients involved in the clinical trial, the model performs quite well, and is able to explain 85\% of the variance of post-therapeutic RCB, with a correlation of approximately 0.95 between the modeled and observed RCB. In the future, we will extend this model to include the majority of the 93 patients involved in the clinical trial.

Conclusions
Heterogeneity of breast tumors assessed using dynamic contrast-enhanced magnetic resonance imaging shows great potential for predicting whether patients with triple-negative breast cancer will or will not respond to neoadjuvant chemotherapy with PARP inhibitor treatment. If a model may be developed which shows similar success on the entire 93-patient clinical trial cohort, it may be used clinically to allow oncologists to personalize breast cancer treatment on an individual basis and improve patient health and clinical outcomes.

Poster 8

Phosphorylation of Yeast Transcription Factors Correlates with the Evolution of Novel Sequence and Function

Mark Kaganovich
Stanford University

Michael Snyder

Gene duplication is a significant source of novel genes and the dynamics of gene duplicate retention vs. loss are poorly understood, particularly in terms of functional and regulatory specialization of their gene products. We compiled a comprehensive dataset of S. cerevisiae phosphosites to study the role of phosphorylation in yeast paralog divergence. We found that proteins coded by duplicated genes created in the Whole Genome Duplication (WGD) event and in a period prior to the WGD are significantly more phosphorylated than other duplicates or singletons. Though the amino acid sequence of each paralog of a given pair tends to diverge fairly similarly from their common ortholog in a related species, the phosphorylated amino acids tend to diverge in sequence from the ortholog at different rates. We observed that transcription factors (TFs) are disproportionately present among the set of duplicate genes and among the set of proteins that are phosphorylated. Interestingly, TFs that occur on higher levels of the transcription network hierarchy (i.e. tend to regulate other TFs) tend to be more phosphorylated than lower-level TFs. We found that TF paralog divergence in expression, binding, and sequence correlates with the abundance of phosphosites. Overall, these studies have important implications for understanding divergence of gene function and regulation in eukaryotes.

Poster 9

Computation of Pairwise Local Matches for Whole-Genome Alignment

Birte Kehr
International Max Planck Research School for Computational Biology and Scientific Computing, Berlin, Germany

Aaron E. Darling
Knut Reinert

Purpose: Multiple whole-genome alignments aim to capture all homologies between a set of sequences. Due to the size of genomes and the resulting complexity of the problem, genome alignment programs usually start by generating a set of local matches. Therefore, it is important that the method for computing local matches does not miss significant similarities. The decision of what subset of local matches is most likely to display most of the homologies, is left to the process of multiple whole-genome alignment that generally applies some optimization function.

We have developed the local pairwise alignment program STELLAR [1] that has full sensitivity for ε-alignments, i.e. guarantees to report all local alignments of a given minimal length and maximal error rate ε. We are planning to use STELLAR matches for whole-genome alignment with an optimization function similar to the breakpoint score of the widely used whole-genome alignment program progressiveMauve [3].

Methods: The program STELLAR is composed of two steps, filtering and verification. The first step of STELLAR implements the lossless filtering algorithm SWIFT [2]. In the second step, STELLAR verifies SWIFT hits by a specific seed-and-extend strategy that we have proved to be exact [1].

Results and Conclusions: We have compared STELLAR to widely-used local alignment programs in terms of running time and sensitivity. Our results on simulated and real genomic data confirm and quantify the conjecture that heuristic tools like BLAST or BLAT miss a large percentage of significant local alignments. STELLAR is very practical and fast on very long sequences which makes it a suitable new tool for finding local matches for whole-genome alignments.

Outlook: Sets of local matches can be summarized to locally col-linear blocks with respect to one or more other genomes. These blocks define rearrangement breakpoints for which progressiveMauve's optimization function applies a breakpoint penalty. We are currently developing an optimization function that generalizes the pairwise breakpoint score of progressiveMauve to a breakpoint score for triplets or even larger subsets of genomes. With this extended score and the application of STELLAR for computing the initial set of local matches we are hoping to obtain improved whole-genome alignments.

References:
[1] Kehr B, Weese D, Reinert K (2011). STELLAR: fast and exact local alignments. BMC Bioinformatics, 12(Suppl 9):S15.
[2] Rasmussen KR, Stoye J, Myers EW (2006): Efficient q-gram filters for finding all ε-matches over a given length. J Comput Biol, 13(2):296-308.
[3] Darling AE, Mau B, Perna NT (2010): progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One, 5(6):e11147.

Project Webpage: http://www.seqan.de/projects/stellar.html

Poster

10

Dose Optimization with TFOCS-Based Total-Variation Minimization for Dense Angularly Sampled and Sparse Intensity Modulated Radiation Therapy (DASSIM-RT)

Hojin Kim
Stanford University

Ruijiang Li
Lei Xing

Purpose:
In radiation therapy (RT), it has been focusing on increasing the beam angular sampling for better dose conformity, which would be problematic in IMRT due to substantially prolonged delivery time. Recently, with the aid to digital accelerators, the prolonged delivery time in high angular beam frequency, called Dense Angularly Sampled and Sparse Intensity Modulated Radiation Therapy (DASSIM-RT), has been successfully compensated1. However, critically increased variables in large number of beams add more difficulties to fluence-map optimization with the existing techniques2. This study proposes an incorporation of a new practical L1-solver, called templates for first-order conic solver (TFOCS)3, into the varied beam configuration to simplify the fluence-map with total-variation minimization for better delivery efficiency, while improving the conformal dose distribution.

Methods and Material:
Given the dose matrices computed by Monte-Carlo simulations for uniformly distributed beams, the fluence-map optimization is performed by TFOCS based total-variation minimization to simplify the fluence-map. Compared with the traditional quadratic programming for L1-norm, TFOCS assures faster convergence with the first-order method, and less memory usage for iterates. Thus, it enables the computational time elapsed in fluence-map reconstruction to be linearly increased over the angular frequency, instead of quadratically. To evaluate the dose conformity to the target volume, conformation number (CN)4 was used, while modulation index (MI)5 was adopted to identify the degree of delivery efficiency for the head-neck patient data with 7, 15, and 30 beams assigned.

Results and Discussion:
For the given data with quite complicated target shape, the increase in the beam angular frequency was proven to yield the higher conformation number: 0.6582,0.6967, and 0.7095 with 85,85, and 90 apertures corresponding to 7, 15, 30 beams, respectively. Also, modulation index representing the complexity of the fluence-map of 30 beams was significantly reduced by 60%, compared with that in 7 beams. This implies that the increased beam angular sampling producing better conformal dose distribution would not harm the delivery time in actual treatment.

Conclusions:
Increased beam angular sampling in IMRT inverse planning with TFOCS based total-variation can outperform the conventional IMRT in terms of dose conformity while maintaining high delivery efficiency.

Poster

11

High Resolution Structure Refinement of F-Actin Using Correlated X-ray Scattering

Derek Mendez
Stanford University

Jongmin Sung
Daniel Ratner
James Spudich
Sebastian Doniach

By using the ultra-high flux and ultra-short exposure times characteristic of the LCLS (Linac Coherent Light Source) we plan to refine current models of F-actin. These beam properties preserve angular fluctuations in the scattering, which manifest in the spatial correlation functions of such scattering patterns on area detectors. [1] Using correlated x-ray scattering analysis, we will refine the low resolution fiber diffraction model of F-actin [2] using high resolution LCLS scattering data of F-actin in solution, circumventing the complications involved with F-actin crystallization. As a preliminary study, we have imaged F-actin embedded in trehalose glass at the Stanford Synchrotron Radiation Lightsource (SSRL). In an attempt to emulate LCLS beam line parameters, we used zero beam attenuation and an optimal exposure time that insured adequate sampling and minimal damage. In an attempt to extract the F-actin correlations from artifacts introduced by the detector geometry we computed the correlation function between pixels at different scattering vectors (Q), as the sample correlations vary less rapidly as a function of Q than do the detector gap correlations.

References:

[1] P. Wochner et al., “X-ray cross correlation analysis uncovers hidden local symmetries in disordered matter,” PNAS 106 (28), 11511 (2009).

[2] K. C. Holmes et al., “Atomic model of the actin filament,” Nature 347, 44 (1990)

Poster 12

Evidence of Biased Gene Conversion in Analysis of 1000 Genomes Trio Data

Yulia Newton
University of California, Santa Cruz

Purpose:
Biased gene conversion (BGC) is a phenomenon that can replace A/T (weak) alleles with C/G (strong) alleles during meiosis. Classical Mendelian genetics tells us that a cross of two heterozygotes, WS x WS, for example results in a 1:2:1 ratio of WW:WS:SS in the offspring. However, biased gene conversion could disrupt this ratio, if the alleles are weak and strong in the parents. BGC is thought to be a recombination-mediated phenomenon and is observed to be more prevalent in recombination hotspots. In this project we investigate BGC in the 1000 Genomes [1] trio data to determine if it can be detected in a single generation.

Methods:
We used 1000 genomes trio data for two families (of European and African descent) and analyzed it for evidence of biased gene conversion. For each family, these data include two alleles at ~3.5 million selected SNP sites for the European family, and ~4.3 million sites for the African family. We took the parental genotypes for various subsets of SNP's for which the child's genotype ratios should be trivially predicted by Mendelian genetics. We then analyzed the children's genotypes to see if expected outcome was met.

Results:
We tested ~490,000 and ~560,000 sites in which both parents were heterozygous for a weak and strong allele in the European and in African couples, respectively. Without the BGC effect we would expect to see a 1:1 ratio of WW:SS sites in the child. However, we observed an increase of SS sites in the children, and found the effect to be even more pronounced in recombination hotspots.

Conclusion:
We suspect that biased gene conversion is driving an increase in C/G allele frequency at polymorphic sites in at which both parents are heterozygous for W/S alleles. Our findings are consistent with the earlier research into BGC, especially an increase of strong alleles in recombination hotspots. BGC appears to be a phenomenon that is detectable within even a single generation.

References:
[1] The 1000 Genomes Project Consortium, A map of human genome variation from population-scale sequencing, Nature (2010) 467, 1061–1073 (http://www.nature.com/nature/journal/v467/n7319/full/nature09534.html)

Poster

13

Predicting the Impact of Mutations in Cancer Using an Integrated Pathway Approach

Sam Ng
University of California, Santa Cruz

Janita Thusberg
Stephen Benz
Charlie Vaske
Kyle Ellrott
Jing Zhu
Christina Yau
Sean D. Mooney
Christopher Benz
David Haussler
Joshua M. Stuart

The major mechanism by which cancer arises is through somatic mutations. These mutations can lead to alterations in gene regulation as well as changes in protein structure and function. Individual tumors can contain hundreds to thousands of mutations. It is critical to distinguish mutations that have an important role defining the cancer – driver mutations – from mutations that are unimportant to the tumor – passenger mutations. Differentiating driver and passenger events is essential for understanding cancer disease mechanisms, which can help guide treatment decisions as well as identify novel targets for treatment. Genomic probing with technologies such as expression arrays and high-throughput RNA sequencing provide insight into changes in gene regulation in cancer, but determining the tumorigenic role of a coding mutation is less clear. Genomic data coupled with pathway information may provide insight into the functional impact of a mutation to particular genes.

We are developing a mutation prediction method based on integrated pathway analysis to discriminate loss-of-function, neutral, and gain-of-function mutations. Utilizing the set of regulatory interactions annotated for a given gene, we can detect a discrepancy in the downstream effects of an altered gene compared to what is expected from its upstream influences. We show that a score based on this discrepancy is highly predictive of the presence of a mutation and that the directionality of this discrepancy also reflects the gain- or loss-of-function in a gene.

Application of our method to a set of known driver mutations reveals that there is a significantly strong signal for loss- and gain- of functional mutations in the surrounding network, demonstrating the sensitivity of this approach. In addition, when applied to the negative control of passenger mutations, the method predicts little pathway impact, indicating this approach also has high specificity. Application of this approach to all recurrent mutations in ovarian and glioblastoma multiforme cancers from the TCGA project identifies several important driver mutations across these cohorts. We also highlight the novel utility of this specific approach by comparison to earlier published approaches including MutSig, CHASM, and MutPred.

Poster 14

Phylo-RLQ: Three Table Ordination for Microbial Community Data

Samuel Pimentel
Stanford University

Yana E. Hoy
David A. Relman
Susan Holmes

Purpose:
Recent developments in sequencing technology have made it possible generate species abundance data for microbial communities within human and animal bodies. These abundance counts may exhibit dependencies on environmental factors measured at individual sites and on phylogenetic relationships between bacterial species. These dependencies must be described and accounted for in order to appropriately analyze diversity in these datasets.

Methods:
Specifically, a three-table ordination method incorporating species abundances, environmental variables, and phylogenetic association is needed. A new method called phylo-RLQ is proposed, based on the RLQ method of Dolédec and Chessel (1996) which chooses axes maximizing co-inertia between two tables linked by a third. Phylogenetic dissimilarity data is adapted for use in phylo-RLQ by constructing a lag matrix, a concept from spatial statistics.

Results:
Phylo-RLQ is demonstrated on simulated data and on a real data set from fecal samples of salmonella-infected mice. In these examples it selects axes that distinguish species according to the main branches of the phylogenetic tree and helps identify environmental variables of interest for further study.

Conclusions:
While the current form of phylo-RLQ makes limiting assumptions and uses potentially sub-optimal weights for phylogenetic data, it provides a starting point for exploratory diversity analysis of a microbiome.

References:
S. Dolédec, D. Chessel, C.J.F Ter Braak, and S. Champely. Matching species traits to environmental variables: a new three-table ordination method. Environmental and Ecological Statistics, 3(2):143–166, June 1996.

S. Dray, D. Chessel, and J. Thioulouse. Co-Inertia Analysis and the Linking of Ecological Data Tables. Ecology, 84(11):3078–3089, November 2003.

S. Dray, S. Saïd, and F. Débias. Spatial ordination of vegetation data using a generalization of Wartenberg's multivariate spatial autocorrelation. Journal of Vegetation Science, 19:45-56, 2008.

Poster

15

Ca2+ Pulses Control Local Cycles of Lamellipodia Retraction and Adhesion Along the Front of Migrating Cells

Feng-Chiao Tsai
Stanford University

Tobias Meyer

Purpose:
Ca2+ signals regulate polarization, speed as well as turning of migrating cells. However, the mechanism by which Ca2+ controls moving cells is not understood. So we aimed at studying the temporal-spatial change of Ca2+ in the front of migrating cells.
Materials and Methods: Human umbilical vein endothelial cells (HUVEC) were used in the two-dimensional wound healing model [1]. The cell sheet loaded with Fura-2/AM received live-cell imaging for migration activity monitoring and Ca2+ measurement. Image processing, cell edge tracking, and temporal-spatial cross-correlation analyses were subsequently performed for data acquisition.

Results:
Our findings reveal periodic local Ca2+ pulses along the front of migrating HUVEC. These pulses trigger cycles of protrusion and retraction locally in lamellipodia and, concomitantly, local adhesion to the extracellular matrix. These Ca2+ release pulses have small amplitudes (< 80 nM) and diameters (median diameter 3.8 μm). They are triggered repetitively along the leading plasma membrane with only little coordination between different regions. We further demonstrate that each Ca2+ pulse triggers contraction of actin filaments by activating myosin light chain kinase and myosin II in a region behind the leading edge. The cyclic force generated by myosin II operates locally, causing partial retraction of the nearby protruding lamellipodia membrane and strengthening of paxillin-based adhesion within the same lamellipodia. Photo-release of NP-EGTA-caged Ca2+ in the cell front confirms the direct role of Ca2+ in triggering retraction and adhesion.

Conclusion:
Spatial sensing, forward movement and turning of cells can be regulated by a spatially confined regulatory circuit comprised of local Ca2+ signals that drive local lamellipodia protrusion, retraction and adhesion cycles along the leading edge.

Reference:
1. Vitorino, P., and Meyer, T. (2008). Modular control of endothelial sheet migration. Genes Dev 22, 3268-3281.

Poster

16

Dynamic Cytoskeleton Organization Couple Cell Shape Variations with Migration Phenotypes in HL-60 Cells

Tony Tsai
Stanford University

Michael Davidson
James E. Ferrell Jr
Julie Theriot

Under constant conditions, the neutrophil-like HL60 cell exhibits highly heterogeneous migration phenotypes, as characterized by their speed, shape, and frequency of polarity switching. These natural variations allow us to investigate the mechanism underlying shape determination and its correlation with migration phenotypes. We apply principal component analysis on thousands of cell contours and identified biologically meaningful principal shape modes. Variations in cell length, leading-edge width, and left-right asymmetry can explain close to 70% of total shape variation. Cells with relatively wider leading-edge migrate faster and switch polarity less frequently. Left-right asymmetry of cell shape correlates well with asymmetric distribution of myosin, as well as turns in cell trajectory. We also use the movement and position of cell nucleus to probe for the cell's internal mechanical properties. Nocodazole treatment increases persistence of protrusion, and reduces the distance between leading-edge and the nucleus, suggesting a possible contribution of microtubule to the phenotypic heterogeneity.

Poster

17

Smoothly Decoupled Boundaries in Hybrid Solvent Simulations

Jason Wagoner
Stanford University

Vijay Pande

Biomolecular modeling often requires a compromise between computational efficiency and model resolution. Hybrid solvent models attempt to alleviate this problem by combining the computational efficiency of simple continuum solvent methods with the finer resolution of explicit solvent molecules in regions where such details are expected to be important. Unfortunately, such hybrid methods are subject to significant artifacts at the boundary between explicit and implicit solvent. Here, we introduce a hybrid model that incorporates a boundary that gradually transitions from fully interacting particles to a continuum solvent, effectively removing any boundary artifacts. In addition, this model allows for the incorporation of a dynamic explicit region that changes shape and size throughout the course of a simulation.

Poster

18

Using Phase to Recognize Phonemes in the Brain

Rui Wang
Stanford University

Marcos Perreau Guimaraes
Patrick Suppes

Purpose:
The purpose of this study is to explore the neural mechanism used by the human brain to process phonemes using electroencephalograph (EEG). We addressed the problem of whether frequency analysis can extract attributes of EEG associated to auditory phoneme perceptual activities.

Materials and Methods:
This study focused on 8 English consonants and 4 English. We analyzed the EEG data from two experiments. The auditory stimuli of the first experiment were 32 consonant-vowel syllables. The EEG data were recorded using the 128-channel Geodesic Sensor Net. For each of the 32 syllables, we made 672 brain recordings. In the second experiment, we used the same experimental setup to record 1792 trials for each of the 4 isolated, non-syllabic vowels.
We extended the EEG classification methods in our previous work to build an EEG phoneme recognition model based on Support Vector Machine (SVM). The statistical methods such as bootstrap aggregating (Bagging), Independent Component Analysis (ICA), Principal Components Analysis (PCA) and cross-validation were used in the model. We used the model to recognize the averaged EEG trials represented by the Discrete Fourier Transform (DFT) coefficients, only their amplitudes or only their phases.

Results:
The 8 initial-consonant recognition results showed that the model using only the phase of the DFT coefficients achieved a significant 39.2% recognition rate, comparable to that using the time-domain signal (41.5%). The DFT amplitude-only model only had a near chance level rate. We also looked for the optimal frequency range and improved the recognition rate to 51.4% by using the phase of the frequency range from 2Hz to 9Hz. The finding generalizes to the isolated vowels. Furthermore, the qualitative analysis of the similarities between the EEG representations, derived from the confusion matrices, illustrates the invariance of brain and perceptual representation of phonemes. Inspired by these results, we proposed a new distinctive-feature-based recognition model which can be easily extended for recognizing more phonemes.

Conclusion:
Our results showed that the phase pattern of brainwave oscillations in the frequency range from 2Hz to 9 Hz is highly related to phoneme processing. We also have shown that the EEG representations reflect the temporal distinctions of the auditory stimuli more accurately than the spectral distinctions. The results provide a significant support for the importance of phonological distinctive features in the neural mechanism of phoneme perception, which in turn suggests a computationally efficient brain mechanism for recognizing phonemes.

Poster

19

Efficacy of Fixed Filtration for Rapid kVp-Switching Dual Energy X-ray Systems

Yuan Yao
Stanford University

Adam S. Wang
Norbert J. Pelc

Purpose:
Dose efficiency of dual kVp imaging can be improved if the beams are filtered to remove photons in the common part of their spectra, thereby increasing spectral separation. While there are a number of advantages to rapid kVp-switching for dual energy, it may not be feasible to have two different filters for the two spectra. Therefore, we are interested in whether a fixed added filter can improve the dose efficiency of kVp-switching dual energy x-ray systems.

Materials and Methods:
We hypothesized that a K-edge filter would provide the energy selectivity needed to remove overlap of the spectra and hence increase the precision at constant dose. Preliminary simulations were done to decompose known phantoms into basis materials of aluminum and water, using 80 and 140 kVp x-ray spectra. Precision of the decomposition was evaluated based on the propagation of the Poisson noise in the detected intensities through the decomposition function. The optimal filter material depends somewhat on the phantom composition and ranges across the lanthanide series. Considering availability and cost, we finally chose a commercial Gd2O2S screen as our filter for experiment validation.
To gain more comprehensive understanding of our selected filter material on different thickness of object, we made acrylic-copper step wedge phantom, with various linear combinations of each basis material. During the experiment, 70kVp and 125 kVp x-ray spectra were generated by a table-top system. We kept the phantom exposure to be roughly the same with and without filtration by adjusting the tube current. The filtered and unfiltered raw data of both low and high energy was decomposed into basis material and the variance of the decomposition was calculated using statistical methods for each thickness pair. To evaluate the filtration performance, we attained the variance reduction ratio of filtered and unfiltered result by simple division. Simulation was done with the same experimental settings and we could compare the expected variance reduction from simulation with the real value from experiment for validation.

Results:
Simulation result shows that the variance reduction monotonically increases as the object becomes more attenuating and the spectra are more separated. The experimental result validates this presumption, yet it is overall worse than expectation. For thinner object, the filtration will induce higher variance rather than reduction. However, at more clinical relevant thickness region, the experiment shows a promising precision improvement with the tested filter.

Conclusions
This study demonstrates the potential of fixed Gd2O2S filtration to improve the dose efficiency and material decomposition precision for rapid kVp-switching dual energy systems.

Poster

20

CT Metal Artifact Reduction by Constrained Optimization with a Model-based Scanning Scheme

Xiaomeng Zhang
Stanford University

Lei Xing

Purpose:
The streak artifacts caused by metal implants have been recognized as a “missing data” problem that limits various applications of CT imaging, such as target delineation and accurate dose calculation. How to deal with the missing data is essential in metal artifact problems. In this work we want to investigate a method that can minimize the missing information and reconstruct images with significantly reduced metal artifacts.

Methods:
A penalized-weighted-least-squares method is first used to accurately identify the metal objects in image space. Based on this prior knowledge, a new model-based scanning scheme is designed by shifting the object center during a CBCT scan to avoid the metal regions and reduce the missing projections. An iterative algorithm based on constrained optimization is then used for the image reconstruction. It minimizes a quadratic edge-preserving smoothness measure function of the image, subject to the constraint that the estimated projection data is within a specified tolerance of the available metal-shadow-excluded projection data, with image non-negativity enforced. The algorithm is evaluated using a numerical QA phantom (350x350x16, 1 mm3, only central slice considered) with simulated Poisson noise in the projections. The new scanning scheme is modified over a conventional half-fan scanning geometry with source-to-axis and source-to-detector distances of 100 cm and 150 cm, respectively. Total 339 views projection data are simulated over 360o rotation.

Results:
Studies showed that the constrained optimization with the model-based scanning data has superior performance compared with analytical FDK reconstruction and other iterative reconstructions. It significantly suppressed metal artifacts in the presence of noise. Profile comparisons and RMSE measurements also suggested that the model-based scanning scheme can effectively reduce the missing information and yield better images.

Conclusion:
The proposed algorithm can be used to significantly reduce metal artifacts to produce clinically acceptable image for current on-board CBCT image systems.

Poster

21

Functional Organization and Its Implication in Evolution of the Human Protein-Protein Interaction Network

Yiqiang Zhao
The Buck Institute on Research for Aging

Sean Mooney

Purpose:
Based on the distinguishing properties of the PPI network such as power-law degree distribution and modularity structure, several stochastic models for the evolution of the PPI network were purposed, given the idea that a validated model should reproduce similar topological properties of the empirical network. However, being able to capture topological properties does not necessarily mean it correctly reproduces how a network emerged and evolved. More importantly, there is already evidence suggesting functional organization and significance of the PPI network, however, the current stochastic models grow the network in the absence of biological function and natural selection. Because functionality is an important aspect of molecular evolution, it is important to clearly address this question in order to have a better understanding of how the PPI network evolves.

Materials and Methods:
Nucleotide sequences used in this study were collected from two sources: the NCBI Reference Sequence (RefSeq) database for human and mouse and the Unigene database for all other species. All human genes were classified into six temporal groups based on a nucleotide sequence similarity search using BLAST against several clades in the known evolutionary tree with an E-value threshold set to e-20. Human protein-protein interaction data was integrated from three source BioGrid, HPRD and REACTOME. Functional annotations for human genes were retrieved from the PANTHER database. The functional distance between genes in the PPI netowork is calculated as the Euclidean distance based on gene annotations.Function enrichment/overrepresentation of specific functional annotations was determined by the hypergeometric test. The Z-score test is used to determine if proteins in some functional categories are significantly high or low in network properties.

Results:
we examined the evolution of the PPI network both the topological and functional level, by dividing human proteins into temporal groups using known phylogenetic information. The human PPI network is shown to be both functionally organized and function evolves with the topological properties of the network. Our analysis suggests that function most likely affects the local modularity. Consistently, we further found that the topological unit is also the functional unit of the PPI network.

Conclusion:
We have demonstrated the functional organization of the PPI network. Given our observations, we suggested that significance should not be overlooked when studying PPI network evolution.

Thank You

Guidance and Help

Claudia McClure
Felipa Raul
Patrick More
Katherine Daval-Santos
Jennifer J Wachter

Previous Organizers

Konrad Karczewski
Rob Tirrell
Keyan Salari
Matt DeMers
Jessica Faruque
Amir Ghazvinian

2011 Organizing Committee

Harendra Guturu
Nandita Garud
Jaclyn Chen
Lauren Chircus
Francisco Gimenez
Alicia Martin
Sanna Ali

Platinum Sponsors

Simbios
Bio-X
Ion Torrent

Gold Sponsors

Stanford Biomedical Informatics Training Program (BMI)

Silver Sponsors

Stanford Computer Forum
Stanford Biosciences Student Association (SBSA)

Other Sponsors

Butte Lab

Startup Gold Sponsors

Mendeley