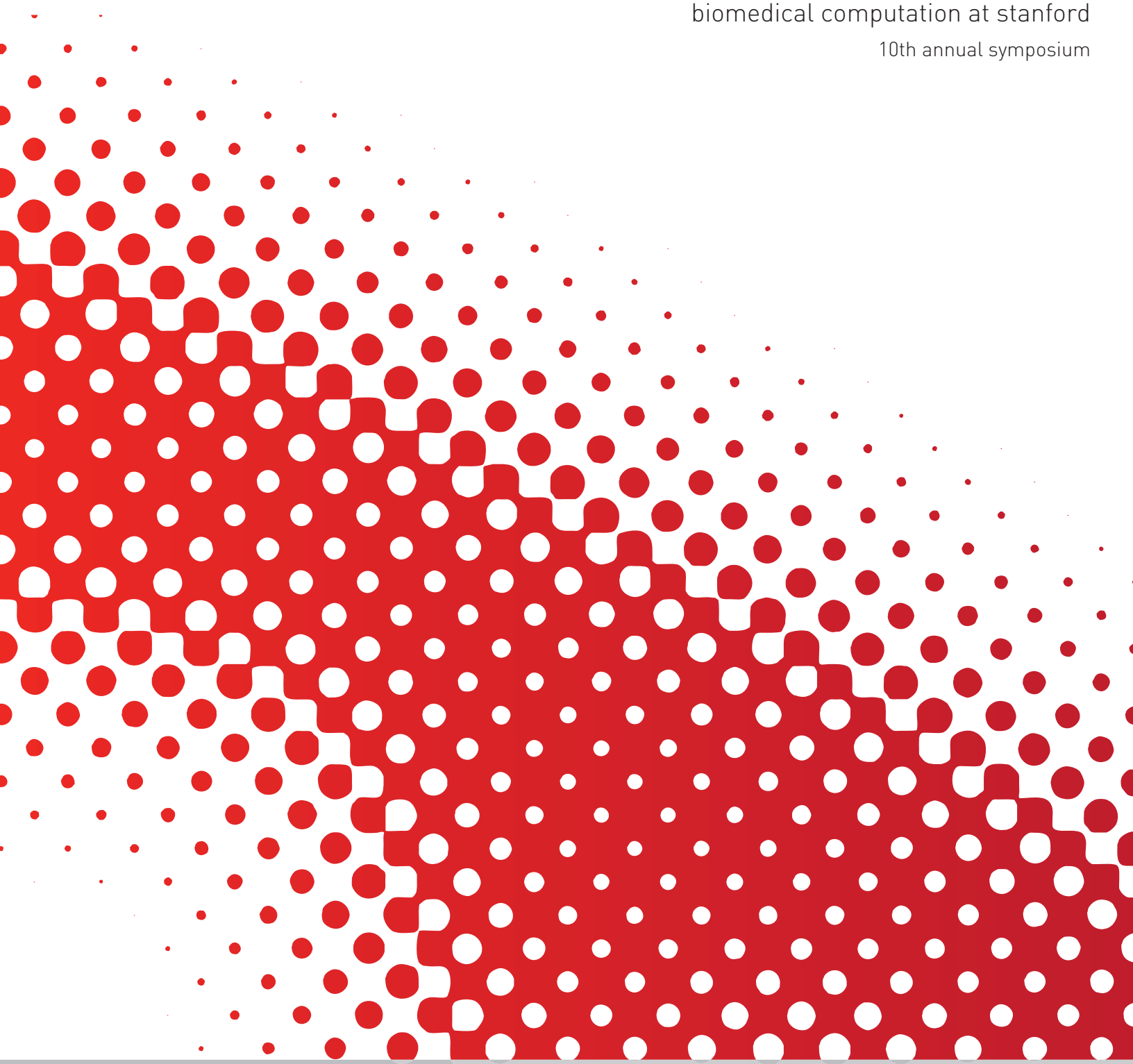
bcats2009

biomedical computation at stanford

10th annual symposium



abstracts

Welcome

to the tenth 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 diverse, cutting-edge research from the largest and most geographically diverse pool of participants in its 10 year history.

We thank our keynote speakers, student presenters, judges, sponsors, and all 2009 attendees.

The BCATS 2009 organizing committee

David Chen, Biomedical Informatics (Chair)

Sarah Aerni, Biomedical Informatics (Co-Chair)

Robert Bruggner, Biomedical Informatics

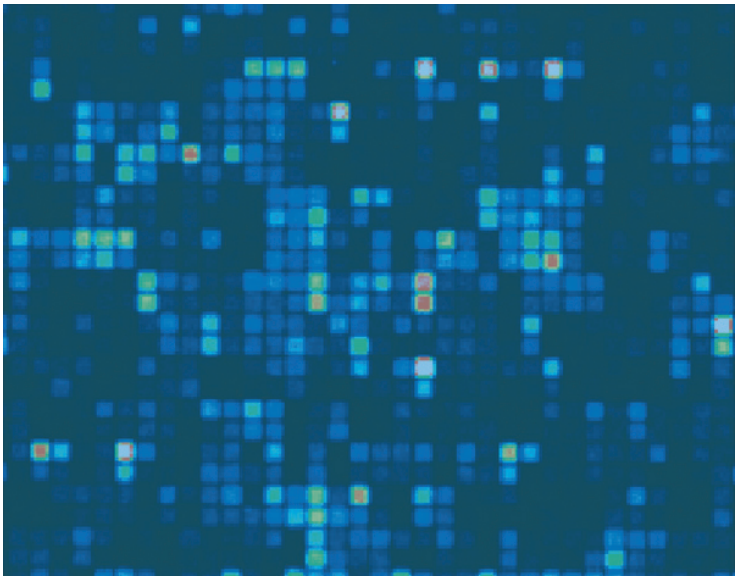
Samuel Hamner, Mechanical Engineering

Jonathan Karr, Biophysics

Linda Liu, Biomedical Informatics

Dan Newburger, Biomedical Informatics

Chirag Patel, Biomedical Informatics



BECAUSE

Our CAUSE is Milosh and his cancer.

For more than 30 years, Genentech has been at the forefront of the biotechnology industry, using human genetic information to develop novel medicines for serious and life-threatening diseases. Today, Genentech is among the world's leading biotech companies, with multiple therapies on the market for cancer and other unmet medical needs.

Our founders believed that hiring talented, enthusiastic people would make Genentech a success. Today, we still believe our employees are our most important asset.

Genentech's research organization features world-renowned scientists who are some of the most prolific in their fields and in the industry. Our more than 1,100 scientists and postdocs have consistently published important papers in prestigious journals and have secured approximately 7,400 patents worldwide (with another 6,250 pending). Genentech's research organization combines the best of the academic and corporate worlds, allowing researchers not only to pursue important scientific questions, but also to watch an idea move from the laboratory into development and out into the clinic. We are proud of our long history of groundbreaking science leading to first-in-class therapies, and we hope you will consider us as we continue the tradition.

Genentech is a proud sponsor of the 2009 Biomedical Computation at Stanford (BCATS) Conference & Symposium.

To learn more about our current opportunities, please visit **careers.gene.com**. Genentech is an equal opportunity employer.

In January 2009, Genentech was named to FORTUNE's list of the "100 Best Companies to Work For" for the eleventh consecutive year.

Genentech

IN BUSINESS FOR LIFE

MAKING A MEASURABLE DIFFERENCE.

As the world's premier measurement company, Agilent Technologies works with engineers, scientists and researchers around the globe to meet the challenges of today and tomorrow.

From home entertainment to forensics, from food safety to network reliability and from wireless communications to discovering the genetic basis of disease, Agilent provides the measurement capabilities that make our world more productive, safer, healthier and more enjoyable.

Agilent is committed to being an economic, intellectual and social asset to each and every nation and community where we operate right down to the neighborhoods where we work and live.

It's all part of making a measurable difference. For everyone.

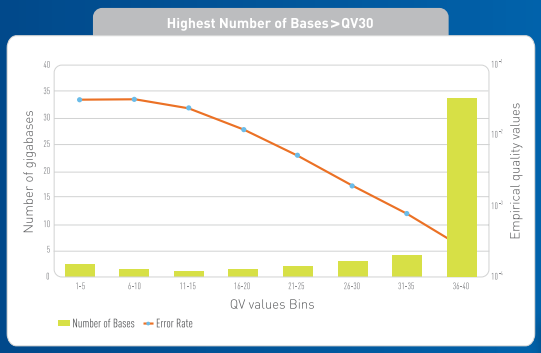
Can Your Next Gen Sequencer Tell The Difference?

Have confidence in your sequencing results with the new SOLiD™ 3 Plus System.

- Highest accuracy and sensitivity for variant discovery with less coverage
- Increased throughput and lower running costs
- Standard base sequence format with the quality of color space

The SOLiD™ 3 Plus System. The right answer, the first time.

For data demonstrating the superior accuracy of the SOLiD™ 3 Plus System go to **www.appliedbiosystems.com/solidaccuracy**



applied biosystems™

Bioscience Job Openings at Sandia

Sandia National Laboratories is a premier U.S. DOE national-security laboratory, where the best and the brightest partner on exciting projects that can literally change the world. We provide innovative, science-based systems engineering solutions to the most challenging problems that threaten peace and freedom for our nation and the globe.

Sandia provides employees with comprehensive benefits packages that include competitive salaries; medical, dental, and vision benefits; 401(k) savings plans; and paid time off. We value our employees as our greatest asset and support them in their efforts to achieve a personal balance among home, work, and the community. To learn more about career opportunities at Sandia, visit **www.sandia.gov/careers**.

Sandia National Laboratories

Sandia National Laboratories has multiple job openings for outstanding graduates in the biosciences and related disciplines. These positions are part of Sandia's Bioscience Research Foundation, one of six science-and-technology competencies underpinning Sandia's national-security mission work—helping to secure a peaceful and free world through innovative technology research and development.

Sandia's bioscience activities are focused on two strategic thrusts in national security: (1) biofuels and (2) biodefense and emerging infectious diseases (BEID):

- Sandia is responding to the call to develop clean, green, and renewable sources of energy. We are working to minimize climate change and reduce U.S. dependency on foreign oil by studying the sustainability and environmental impact of biofuel production. We are also investigating biomass structure and decomposition, the large-scale cultivation of algal biomass, the breakdown of biomass into constituents for fuel conversion, and the conversion process itself.
- Sandia's BEID Program helps the nation anticipate and defend against biothreats, such as biological weapons and emerging infectious diseases. We are researching the molecular mechanisms of pathogenesis and host–pathogen interactions and using our knowledge to develop assays, novel materials, and platforms for detecting pathogens and discovering therapeutic targets. We are also exploring the interrelationships between infectious diseases and the human microbiome.

Sandia's biological-sciences personnel work at three sites: our R&D laboratory in Livermore, California; our corporate headquarters in Albuquerque, New Mexico; and the Joint BioEnergy Institute, a U.S. Department of Energy (DOE) Bioenergy Research Center in Emeryville, California. Sandians at each facility are equipped with state-of-the art research instruments and collaborate with experts in a wide variety of technical fields.

Opportunities for biological scientists and bioengineers at Sandia abound in many areas: biochemistry, biofuels, biomaterials design, carbohydrate chemistry, computational biology/chemistry, medical diagnostics, metabolic/protein engineering, microfluidics, photosynthetic microorganisms, protein structure/function determination, proteomics, single-cell host–pathogen studies, and therapeutics screening.

Candidates for bioscience positions must have an advanced degree (PhD/MS) in biology, chemistry, or a related field (e.g., bioinformatics, engineering, or materials science). In addition, U.S. citizenship is needed for positions that require DOE security clearances.

Visit **www.bio.sandia.gov** to learn more about Sandia's bioscience programs. To browse and apply for current job openings, go to **www.sandia.gov/careers**.

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 for solving equations?

Download the open-source SimTK Core libraries at: www.simtk.org/home/simtkcore

Interested in collaborating on important computational biological problems?

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

Stop by our table at the industry reception

See how much we can speed up molecular dynamics via GPUs
Try OpenSim, our freely available musculoskeletal simulation software
and more...

NIH Roadmap Grant U54 GM072970

BIO-X

Bio-X is a Stanford University program supporting interdisciplinary work related to biology and medicine. The program is a joint effort by the Schools of Humanities and Sciences, Engineering, Earth Sciences, and Medicine.

The Bio-X Program encourages interdisciplinary work through various programs, including a seed grant program (the Interdisciplinary Initiatives Program), Bio-X Graduate Student Fellowships and Bio-X Stanford Interdisciplinary Graduate Fellowships in Human Health, Bio-X Undergraduate Research Awards, and Bio-X Travel Awards.

The Bio-X Program, which reaches across the university to nearly 450 faculty members in over 60 departments, is facilitated by the James H. Clark Center, which was completed in 2003 thanks to the enormous generosity of Jim Clark and Atlantic Philanthropies.

The Clark Center comprises the equipment, resources and utilities required to conduct breakthrough research at the cutting edge of engineering, science and medicine.

For more information on Bio-X,
visit our website at:

http://biox.stanford.edu

The Stanford Biomedical Informatics Training Program

•Well established program created in 1982

•Special expertise in key informatics research areas:

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

•Established research collaborations with over 30 labs across campus

•One of the few programs where Bioinformatics and Clinical Informatics are combined

•Engineering, Life Sciences and the Medical School are all located on one campus

Full-Time On-Campus Programs

- •PhD
- •Research-based Masters
- •Co-terminal Masters

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

Part-Time Distance Learning Programs

- •Professional Masters Program (Honors Coop Program)
- •Certificates in Bioinformatics and Clinical Informatics
- •Individual Graduate Classes – the non-degree option

Offered in conjunction with the Stanford Center for Professional Development. All distance education programs offer classes drawn from our rigorous on-campus classes for full academic credit.

For more information, contact:

Stanford Biomedical Informatics Training Program
Medical School Office Building, Room X-215
251 Campus Drive, Mail Code: 5479
Stanford, CA 94305-5479

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

BCATS 2009 SCHEDULE

8.00 On-Site Registration & Breakfast

8.45 **Opening Remarks: Russ Altman, MD, PhD**

9.00 **Keynote Address: Matthew Cooper, PhD, DAPT** (pg 14)
Informatics Trends in the Biopharmaceutical Industry

9.45 **Karen Yan** (pg 18)
Quantitative characterization of multi-signal activation of cellular responses in multiple biological systems

10.00 **Peng Qiu** (pg 19)
Discovering Biological Progression underlying Microarray Samples

10.15 **Richard Roettger** (pg 20)
Estimating the size and completeness of gene regulatory networks

10.30 **Spotlight Presentations** (odd-numbered presentations)

10.45 **Poster Session I (odd-numbered posters)**

11.45 **Ying Wang** (pg 21)
A belief propagation algorithm for network alignment

12.00 **Richard W. Lusk** (pg 22)
Evolutionary mirages: selection on binding site composition creates the illusion of conserved grammars in Drosophila enhancers

12.15 **Cory Y. McLean** (pg 23)
Searching for the Genetic Basis of Human-specific Traits

12.30 Lunch

1:30 **Keynote Address: Zoran Popović, PhD** (pg 15)
Visual computing methods for discovering natural animal motion and solving protein and drug design problems

2.15 **Gaurav Chopra** (pg 24)
Remarkable Patterns of Surface Water Ordering Around Nonpolar Solutes

2.30 **Katherine M. Steele** (pg 25)
Contributions of the ankle plantarflexors to joint and mass center accelerations during crouch gait

2.45 **Edith M. Arnold** (pg 26)
Fiber Operating Ranges of the Vasti and Implications for Isometric Force Generation

3.00 **Spotlight Presentations** (even-numbered presentations)

3.15 **Poster Session II (even-numbered posters)**

4.15 **Fraser Cameron** (pg 27)
Closed-Loop Control of Diabetes through Meal Detection and Hypoglycemia Risk Management

4.30 **Alexander Schoenhuth** (pg 28)
Constrained Mixture Estimation for Analysis and Robust Classification of Clinical Time Series

4.45 **Martin Tall** (pg 29)
Mapping Eye Movements in Three-Dimensions: Analyzing Gaze Paths when Interpreting Volumetric Chest CT Data

5.00 **Hoameng Ung** (pg 30)
Detecting Chronic Pain in the Brain

5.15 Awards & Closing Remarks

5.30 Reception

BCATS 2009 Posters

BCATS 2009 Posters

BCATS 2009 Posters

INVITED PRESENTATIONS

bcats2009

Keynote Speakers

BCATS Keynote Speaker

Matthew Cooper, PhD, DAPT

Roche

Head of Non-clinical Safety

Matthew Cooper, Ph.D., DABT, serves as the Head of Non-Clinical Safety Information at Roche. He is currently responsible for spearheading innovation in the information space, as well as coordinating all strategic and operational aspects associated with the informatics portfolio supporting the global NCS organization. Matthew also serves on the global leadership team for NCS. Previously, Matthew lead discovery safety efforts for the biologics drug portfolio at Roche Palo Alto and has served as an investigative toxicologist for the site. Prior to Roche, Matthew worked for Biogen-Idec in the Biomarker Development and Validation department where he led the Transcript Profiling and Computational Biology group.

Matthew is board certified in toxicology, obtained his Ph.D. in Toxicology at the University of Kentucky College of Medicine and his BS in Chemistry from the University of Tulsa College of Engineering. An avid wine collector, Matthew was recently knighted into the Confrérie des Chevaliers du Tastevin, an ancient order dedicated to the appreciation of Burgundy.

Keynote Address
Informatics Trends in the Biopharmaceutical Industry

The rate of data generation in the biopharmaceutical industry continues to increase at a rapid pace. As a result, there is a tremendous need for advanced tools and techniques that effectively mine and capture the underlying knowledge in the data. I will share my perspective on the current gaps and trends in knowledge management, including data integration, workflow automation, and semantic networks. In addition, I will share my experiences and opinions on opportunities for students in the biopharmaceutical industry.

BCATS Keynote Speaker

Zoran Popović, PhD

University of Washington

Associate Professor, Department of Computer Science and Engineering

Zoran Popović is an Associate Professor in computer science at University of Washington. He received a Sc.B. with Honors from Brown University, and M.S. and Ph.D in Computer Science from Carnegie Mellon University. He has held research positions at Sun Microsystems and Justsystem Research Center and University of California at Berkeley and Electronic Arts. Zoran's research focuses on computer animation (physically based modeling, high-fidelity control of animal natural phenomena), and games that solve real problems (protein design, STEM education). His contributions to the field of computer graphics have been recently recognized by a number of awards including the NSF CAREER Award, Alfred P. Sloan Fellowship and ACM SIGGRAPH Significant New Researcher Award.

Keynote Address

Visual computing methods for discovering natural animal motion and solving protein and drug design problems

This talk will cover two computer graphics methods that have implications not just to the entertainment industry, but also in science. The first part of the talk will focus on new models for generating realistic, high-dimensional locomotion and behavior controllers that can model a wide range of animal movement. I will describe methods for exploring the coupled space of animal motions and morphology, the robust realtime controllers that cover the full range of human locomotion, and an optimal control framework to combine basic motion skills into complex behaviors.

The second part of the talk will cover recent findings of the foldit project, a game that enables people all over the world to take active part in protein design problems towards finding new vaccines, drug design and other key problems in biochemistry. With over 100,000 people taking part in the social game over the past year, foldit has produced some surprisingly good outcomes. I will describe aspects of human collective game play that tend to produce results that outperform all known methods, and point to some exciting current and future directions for this new computation paradigm.

bcats2009

Talk Abstracts

talk no.

1

Quantitative characterization of multi-signal activation of cellular responses in multiple biological systems

Karen Yan
Stanford University

Ibrahim Al-Shyoukh

Purpose:
Current drug therapy views patients that have the same diagnosis as a homogeneous group. However, multiple factors, including genetic and environmental variations, are thought to contribute to the large discrepancy in patient response to the same drug therapy. Thus combinational drug treatment may offer a huge improvement in overall patient response levels. In order to determine which combination may be most effective, we take a mathematical approach using artificial neural networks (ANN) to assess the effect of multiple drug combinations under different pathophysiological scenarios, such as non-small cell lung cancer (NSCLC) and Kaposi's Sarcoma-associated Herpesvirus (KSHV).

Materials and Methods:
Using in vitro cell culture models of NSCLC and KSHV, single and combination drug profiles were determined with respect to cell survival and transcription factor activity, respectively. Three drugs (AG490, a tyrosine kinase inhibitor; indirubin-3'-monoxime, a cyclin dependent kinase inhibitor; and U0126, a MEK inhibitor) were selected based on high tumor selectivity and good dose curves from over thirty clinically approved inhibitors in NSCLC. Cellular ATP level was used as a direct readout for cell survival and proliferation in NSCLC. Reactivation of KSHV was induced by a combination of bortezomib, db-cAMP, prostratin, and valproate. The activity of PAN, AP1, CREB, NF-κB, RBP-Jk, and E2F transcription factors was measured using a dual luciferase assay at three different timepoints. To build the model, we used varying percentages of experimental datapoints to train the ANN to generate models for both cell lines. The remaining datapoints were used to test the accuracy of the model.

Results:
ANN created drug response models that had correlation coefficients greater than 0.9 and a mean absolute error of less than 10% to predict cell survival.

Conclusion:
By taking a quantitative approach to characterize the effect of multiple inputs in three biological systems, we are able to reduce the size of drug combination studies in vitro. The advantage of our AI-based approach to predict a full data set with limited data set acquisition is attractive in situations in which there are restrictions on the number of drug combinations that can be tested in a clinical setting. Ultimately, this research could potentially lead to the development of mathematical tools that can aid in the design of personalized drug therapy for cancer and viral infections.

References:
(1) Mao J; Jain, AK. Artificial neural networks for feature extraction and multivariate data projection. IEE Transactions of Neural Networks 1995; 6: 296-317

talk no.

2

Discovering Biological Progression underlying Microarray Samples

Peng Qiu
Stanford University

Andrew Gentles
Sylvia Plevritis

Purpose:
In many biological systems, a clear concept of progression exists, often described by changes over time. For example, cellular differentiation is known to be an ordered and highly orchestrated process in which distinct transcriptional programs are activated or repressed at appropriate stages. When a microarray study takes samples at known points during a known progression, there exist a variety of methods to identify genes that change in a manner consistent with the progression. However, consider the situation where microarray samples are generated but the progression is not known, that is, the correct order of the experimental samples is not known. We present a novel method, Sample Progression Discovery (SPD), to discover the progression order and identify the genes that drive the progression.

Methods:
SPD assumes the microarray data are sampled from an unknown progression which can be captured by the gradual shift in expression of subsets of genes, and each sample represents one unknown point during the progression. SPD discovers the progression order in three steps: clustering genes, constructing minimum spanning trees (MST), and comparing modules vs. MSTs. Gene clustering is needed to reduce the number of expression patterns to be tested. For each gene module, we build an MST, whose nodes are the samples and whose edges are weighted by the distance between nodes in terms of their gene expression. The MST describes the sample progression order defined by this gene module. The progression order is not necessarily linear; branching is allowed. SPD then evaluates the statistical fit between all the modules and all the MSTs. If multiple modules fit well with the same tree, these modules are similar in the sense that they describe a common progression order. The common progression order supported by multiple modules is likely to be biologically meaningful.

Results:
We applied SPD to time series microarray samples taken during cell cycle. Without any prior knowledge of the samples, SPD identified the correct time order and many genes associated with cell cycle. When applied to B-cell differentiation data, SPD recovered the correct order of stages of normal B-cell differentiation and the linkage between preB-ALL with its cell origin preB. We also applied SPD to follicular lymphoma, and uncovered a molecular signature that was recently shown to be associated with the transformation from indolent to aggressive disease stages.

Conclusion:
SPD can identify the progression order underlying microarray samples, while simultaneously selecting genes that are associated to the progression.

talk no.

3

Estimating the size and completeness of gene regulatory networks

Richard Roettger
Technical University of Munich and
University of California, Berkeley

Jan Baumbach

Purpose:
The increasing amount and diversity of sequenced species spread over all domains of life suggest that the number of genes does not reflect our perception of relative organism complexity. Recently, Stumpf et al. estimated the size of the human protein-protein interaction network (interactome). Here, we investigate a similar topic: the size and completeness of bacterial and eukaryotic gene regulatory networks, a related but more challenging task due to specific characteristics of these networks.

Materials and Methods:
Stumpf et al. estimated the size of the human interactome by basically calculating the expected edge-probability between each pair of nodes in the known subnetwork (sampled by biologists in wet labs). Afterwards, they used this likelihood for predicting the size of the entire (nature-given) network. We use a similar approach to estimate the size of gene regulatory networks for several species from different domains of life. In contrast to the interactome, gene regulatory networks are more complex. Thus, we enhanced the method and differentiate in our approach between three different kinds of transcriptional regulatory interactions: (1) transcription factor (TF) regulations among themselves: TF-->TF, (2) TF self regulations, and (3) TF-->non-TF genes. Having calculated these three edge-probabilities, we are able to predict the size of the underlying whole-organism regulatory network. To validate the reliability of our approach, we sample 5000 sub-subnetworks of different sizes (defined by a node-pick probability p) and use these samples to predict the size of our known subnetworks.

Results:
First results show that Escherichia coli uses about 21K gene regulatory interactions to control its gene expression. Whereas, for Bacillus subtilis and we predict only 9K regulations, 10K for Corynebacterium glutamicum. The results for complex mammals like human or mouse indicate that the expected size of their gene regulatory network lies in the order of magnitude between 500K and 1500K regulatory connections. In addition we were able to show, that the expected network size does not directly correlate with the ratio between TFs and non-TF genes.

Conclusion:
We designed a robust, unbiased statistical method to predict the size of regulatory networks. However, our simulations show that the amount of publicly available data is still too limited to allow reliable estimations for higher mammals. For instance, less than 10% of the regulatory network of human is known, which may result in comparably high levels of uncertainty.

References:
Stumpf MP, Thorne T, de Silva E, Stewart R, An HJ, Lappe M, Wiuf C (2008) Estimating the size of the human interactome. Proc Natl Acad Sci. 2008 May 13;105(19):6959-64.

talk no.

4

A belief propagation algorithm for network alignment

Ying Wang
Stanford University

Mohsen Bayati
Margot Gerritsen
David Gleich
Amin Saberi

Purpose:
Network alignment is a computational technique for finding similar structures in multiple networks. For bioinformatics applications, it is used to identify orthologous protein interactions between protein interaction networks from different species. Mathematically, the problem is posed as a matching problem between the vertices of two graphs. The objective is to maximize the weight of the matching and an additional term for the overlap (common edges in both graphs) induced by the matching. Our first contribution is to investigate sparse alignment problems where a protein in one species (a vertex in one graph) can only match with a small set of proteins in the other species (a few vertices in the other graph). We also propose a new algorithm based on a belief propagation (BP) heuristic [Bayati et al. to appear] that handles problems with 10,000 vertices in problems from biology and over 100,000 vertices in problems from other domains.

Methods:
We compare our new belief propagation heuristic with two other heuristics (IsoRank [Singh et al. 2007] and a subgradient algorithm [Klau 2009]) on a series of synthetic problems where an exact alignment is known. We also evaluate the results on two sets of protein-protein interaction networks: one is aligning Drosophila melanogaster with Saccharomyces cerevisiae, and the other is aligning Homo sapiens with Mus musculus. Finally, we explore the algorithms on a large ontology alignment problem between the Library of Congress subject headings and its French National Library equivalent: Rameau.

Results:
Our computational results indicate that our BP heuristic performs as well as the best results from the existing algorithms at optimizing the network alignment objective functions when the set of potential matches is extremely sparse (around 5 choices per items). When the set of potential matches grows (›10), then the objective values of BP are higher. Also, when an ideal alignment is known (in our synthetic experiments and the ontology alignment problem), however, we find that the BP heuristic identifies more of the correct alignment than the other algorithms.

Conclusion:
The results demonstrate that our BP algorithm for network alignment is promising for new network alignment problems in bioinformatics. Towards this end, we make all our codes and examples publicly available for others to use in a “netalign” software package [Bayati et al. 2009a].

References:
Klau. BMC Bioinformatics, 10:S59, 2009.
Singh, Xu & Berger. RECOMB2007 / LNCS 4453:16-31, 2007.
Singh, Xu & Berger. PNAS 105:12763-12768,2008.
Bayati, Gerritsen, Gleich, Saberi & Wang. http://www.stanford.edu/~dgleich/publications/2009/netalign, 2009a.
Bayati, Gerritsen, Gleich, Saberi & Wang. ICDM 2009, to appear.

talk no.

5

Evolutionary mirages: selection on binding site composition creates the illusion of conserved grammars in Drosophila enhancers

Richard W Lusk
University of California, Berkeley

Michael B Eisen

Purpose:
The clustering of transcription factor binding sites in developmental enhancers and the apparent preferential conservation of clustered sites have been widely interpreted as evidence that specific physical interactions between transcription factors are required for regulatory function. However, alternative means of generating these signals have not, until now, been explored.

Materials & methods:
We use simulations of idealized enhancers from Drosophila melanogaster. Our simulations approximate the neutral mutation patterns in Drosophila and incorporate selection steps that require binding site composition, but not spatial arrangement, to be maintained.

Results:
We show here that selection on only the composition of enhancers, and not their internal structure, leads to the accumulation of clustered and overlapping sites with evolutionary dynamics that suggest they are preferentially conserved. In our simulations, mutations to an enhancer that destroy an existing binding site are tolerated only if a compensating site has emerged somewhere else in the enhancer. We find that point mutations and deletions that affect more than one site are accepted far less frequently than those affecting single sites, leading to a significant increase in the density and evolutionary half-lives of overlapping sites. As our simulations incorporate the known Drosophila bias for deletions over insertions, sites tend to become closer over time. This leads to a clustering of sites in the absence of any selection for it, and creates the false impression that proximal sites are more conserved. In simulations of enhancer conservation following speciation, sites tend to be closer together in descendent species than in their common ancestors, violating the common assumption that apparent conservation of a feature in existing species reflects its ancestral state. Finally, we show that there exist plausible models for the evolution of existing Drosophila enhancers that generate the observed number of overlapping and closely neighboring sites in the absence of any functional interaction between those sites.

Conclusion:
This study calls into question the common practice of inferring “cis-regulatory grammars” from the organization and evolutionary dynamics of binding sites in developmental enhancers.

talk no.

6

Searching for the Genetic Basis of Human-specific Traits

Cory Y. McLean
Stanford University

Philip L. Reno
Alex A. Pollen
Vahan B. Indjeian
Gill Bejerano
David M. Kingsley

Purpose:
The availability of many whole genome sequences has spurred great excitement for the prospect of understanding the molecular basis of unique human traits. Recent investigations have discovered dozens of proteins that show evidence of positive selection within the human lineage [1], as well as conserved non protein coding genomic loci which have experienced accelerated base pair changes in the human lineage [2,3]. These genome surveys have focused on small base pair changes in otherwise well-aligned sequences. Here we expand these studies to look for a type of event particularly likely to produce functional effects: complete deletion in humans of sequences that are otherwise highly conserved in other organisms.

Materials and Methods:
We used the whole genome sequences of human, chimpanzee, and other species to identify regions of the chimpanzee genome highly conserved over mammalian evolution that are uniquely missing in humans. We used PCR and computational methods to validate the human deletions. For a subset of the sequences, we performed transgenic mouse experiments to determine the functional consequences of the human loss of the sequence.

Results:
By searching for regions of the chimpanzee genome highly conserved over mammalian evolution that are clearly missing in humans, we discover 583 human-specific losses of putatively functional ancestral DNA. Roughly 80% of the deletions are fixed in human populations, while others are polymorphic in different individuals. Most of the deletions removed conserved non-coding sequences rather than protein-coding regions, and many lie in proximity to genes involved in neural functions, axon guidance, and steroid hormone signaling.

We have functionally tested a subset of human-specific deletions in transgenic mice, and have found intriguing examples of regulatory alterations in humans that appear to be associated with evolution of specific anatomical differences between humans and other animals.

Conclusion:
Loss of conserved regulatory sequences may have contributed to the evolution of several different human-specific traits, by changing the tissue-specific expression of key developmental control genes that flank the deleted regions. I will present one example of a human loss that correlates with human-specific anatomical differences. The full set of 583 human-specific deletions may well contain additional loci associated with evolution of human-specific traits, a possibility that can now be tested by detailed functional studies of other conserved non-coding sequences that are surprisingly missing from the human genome.

References
1.Sabeti P.C. et al., Nature, 2007.
2.Pollard K.S. et al., Nature, 2006.
3.Prabhakar S. et al., Science, 2008.

talk no.

7

Remarkable Patterns of Surface Water Ordering Around Nonpolar Solutes

Gaurav Chopra
Stanford University

Michael Levitt

Purpose:
Water is a medium for many important biological interactions, which occur in nature. An accurate description of the structure of water around the solute of interest could be used to improve our understanding of various biological processes like protein folding. We study the hydration structure of hydrophobic solutes using Molecular Dynamics (MD) simulations. We ask how well is hydrophobic affect represented in classical force fields compared to a general-purpose quantum mechanical polarizable (QMPFF) force field?

Materials and Methods:
A recently introduced state-of-the-art quantum mechanical polarizable forcefield (QMPFF3) fitted solely to high-level quantum mechanical data at MP2/cc-pVTZ level [1] with a simple model correction using CCSD(T) data for higher accuracy of aromatic carbon atom type was used to obtain detailed spatially resolved surface maps of the density of the water O and H atoms surrounding the nonpolar solutes. We ran 100ns of all-atom MD simulation with explicit solvent in a periodic box to determine the behavior of these nonpolar solutes (methane, benzene, cyclohexane and C60) in water and present interesting insights from correlation analysis of these surface maps.

Results:
QMPFF water [2] has less orientation entropy around large hydrophobic solutes like C60, in that water dipole has preferred orientations around C60 compared to SPC or TIP4P water. The VDW is stronger for QMPFF (Dispersion and Exclusion effect) compared to the classical case. All other solutes have more orientational entropy. As expected, most of the waters point away from the hydrophobic solutes but in many cases one OH bond faces the hydrophobic solute surface.

Conclusion:
Polarization causes highly ordered water structure within a water layer, i.e. the hydrophobic surface imprints itself on water and this is a long range effect (seen at 10Å) for QMPFF water. These affects are less for classical force fields. One major conclusion from this study is that QMPFF increases hydrophobic effect, which could have a profound affect on protein folding.

References:
[1] Donchev AG, et al, Quantum mechanical polarizable force field (QMPFF3): Refinement and validation of the dispersion interaction for aromatic carbon. J Chem Phys, 2006. 125(24)
[2] Donchev AG, et al, Water properties from first principles: simulations by a general-purpose quantum mechanical polarizable force field. Proc Natl Acad Sci U S A, 2006. 103(23)

talk no.

8

Contributions of the ankle plantarflexors to joint and mass center accelerations during crouch gait

Katherine M. Steele
Stanford University

Ajay Seth
Jennifer Hicks
Michael Schwartz
Scott Delp

Purpose:
Crouch gait is a common movement disorder in patients with cerebral palsy characterized by excessive hip and knee flexion that compromises walking efficiency and can lead to joint pain and degeneration. The ankle plantarflexors - gastrocnemius and soleus muscles – are common targets for surgical lengthening to treat crouch gait although the outcomes of these surgeries are highly variable. During unimpaired gait these muscles play an important role supporting and propelling the body. Understanding how they contribute to motion during crouch gait can aid in treatment planning. The purpose of this study was to quantify the accelerations generated by the gastrocnemius and soleus at the hip, knee, ankle, and mass center (COM) during crouch gait.

Methods:
Subject-specific dynamic simulations were created for ten subjects who walked flat-footed in a mild crouch gait and had no previous surgeries. Simulations of the single-limb stance phase of gait were generated for each subject using a model with 19 degrees of freedom and 92 musculotendon actuators [1]. The model was scaled to each subject and kinematic and ground reaction force data were used to estimate muscle activations [2]. Estimated muscle activations were compared to experimentally-measured electromyography signals to ensure that the timing and magnitude of the estimated activations reflected the subject's muscle activity. A perturbation analysis [3] was used to determine the contributions of the gastrocnemius and soleus to hip, knee, ankle, and COM accelerations.

Results:
In all subject simulations, the gastrocnemius and soleus were highly active and made large contributions to joint and COM accelerations. The soleus generated extension accelerations at all three joints. The gastrocnemius generated flexion accelerations of both the hip and the knee, but also accelerated the ankle into extension. Both muscles contributed to support by accelerating the mass center upward throughout single-limb stance. The soleus provided the largest contribution to support of all muscles. (Please see figures on website.)

Conclusion:
These results demonstrate that the gastrocnemius and soleus have different functions and should be treated separately in crouch gait. The soleus is a critical crouch-countering muscle and surgically lengthening it could compromise a patient's ability to support their body. The gastrocnemius, meanwhile, could be contributing to crouch gait by flexing the hip and knee; however, due to its effect at the ankle, it still contributes to support of the mass center and therefore presents a trade-off for surgical lengthening.

References:
[1]Delp et al. (2007) IEEE BME. [2]Thelen et al. (2003) J Biomech. [3]Liu et al. (2005) J Biomech.

talk no.

9

Fiber Operating Ranges of the Vasti and Implications for Isometric Force Generation

Edith M. Arnold
Stanford University

Samuel R. Ward
Richard L. Lieber
Scott L. Delp

Purpose:
Computer models that estimate the force generation capacity of lower limb muscles have become widely used to simulate the effects of musculoskeletal surgeries and create dynamic simulations of movement [1]. However, the use of these models to study relationships between muscle structure and function has been limited by the experimental data used to create them. Most models have been based on two classic studies of muscle fiber lengths, physiological cross sectional areas, and pennation angles measured in five cadaver subjects [2, 3]. These studies did not link fiber length and joint position, so this relationship was approximated to produce adequate overall moment generation. Such models are not appropriate for examining functional fiber operating ranges of muscles.

Methods:
Recently, Ward et al. conducted a study of lower limb muscle architecture in twenty-one cadaver subjects [4]. In addition to the parameters measured in past studies they measured the length of sarcomeres (the smallest functional unit of muscle) in each muscle at a known joint angle. From this parameter we can extrapolate a relationship between muscle fiber length and joint angle. With this data we created a model of the lower limb that provides reliable representations of muscle moment arms, force generation capacities, and fiber operating ranges. We used this model to examine the relationships among architecture, moment arms, and joint moment in the knee extensors.

Results:
We found that joint moments computed with the model for the knee extensors deviated from experimental results at high flexion angles. The model overestimated passive moment and underestimated active moment, compared to experimental results. This occurred because vastus intermedius, lateralis, and medialis reached optimal fiber length at 32°, 36°, and 34° of knee flexion respectively and stretched to 1.40, 1.35, and 1.35 times optimal fiber length at 100°. Vastus lateralis had the greatest impact due to its large PCSA.

Conclusions:
This behavior may be a symptom of altered passive properties of the vasti compared to other muscle groups or complex fiber arrangements that are not captured by a lumped parameter model of muscle that assumes all fibers are the same length. The fibers of these larger muscles may be distributed over a range of lengths, leading to a more gradual change in maximum force output with knee flexion.

References:
1.Delp, S. L., et al. (1990) IEEE Transactions on Biomedical Engineering, 37(8), pp. 757-767.
2.Friederich, J. A., and Brand, R. A. (1990) J Biomech, 23(1), pp. 91-95.
3.Wickiewicz, T. L., et al. (1983) Clin Orthop, (179), pp. 275-283.
4.Ward, S. R., et al. (2009) J Bone Joint Surg Am, 91(1), pp. 176-185.

talk no.

10

Closed-Loop Control of Diabetes through Meal Detection and Hypoglycemia Risk Management

Fraser Cameron
Stanford University

Bruce Buckingham, MD
Günter Niemeyer, PhD

Purpose:
Type I Diabetes Mellitus is a disease characterized by a complete loss of natural insulin production and thus control of the blood glucose (BG) level. This work uses recent advances in insulin pumps and continuous glucose monitors (CGMs) to propose an autonomous BG controller.

BG controllers experience significant uncertainty due to BG dynamics, exercise, sensor noise, and particularly meals. Our only corrective action, insulin, is both slower than meals and can only reduce BG levels. While such action is necessary to prevent chronically dangerous high BG levels, they may cause acutely dangerous low BG levels.

Consequently, our control strategy explicitly considers uncertainty, allowing it to be aggressive by regulating the risk of low BG levels. Our meal detection supports this by being adaptive and estimating uncertainty.

Materials and Methods:
The controller is based around maintaining a dynamic, minimal set of most probable meal models to capture meals as they occur. Kalman Filters recursively update state and accuracy estimates for each meal model. Using the innovation signals and Bayes' Rule we calculate the probability of each meal model, allowing quick detection of meals.

The set of meal models provides future BG predictions and certainties. Combined with the assumed insulin impulse response, and a desired future hypoglycemic risk level, these allow the controller to regulate hypoglycemic risk by selecting the maximum safe insulin bolus to administer.

Results:
We validate this approach by comparing three controller, manual (continuous basal equilibrating at 130 mg/dL, with optimal meal bolus at meal time), with this controller in feedback only with no meal detection, and with this controller with meal detection using the FDA approved University of Virginia Type I Diabetes Simulator. The scenario includes 11 adults and 10 adolescents and lasts 33 hours, with 6 meals and only one evening. The manual/feedback only/full control had a mean glucose of 189/157/142 on adolescents and 206/157/137 on adults. Similarly, the percent time within the target zone of 70-180 mg/dL was 48/61/72 on adolescents and 37/65/81% on adults. Also, the number of minutes spent below 60 mg/dL per day was 0/0/0 minutes on adolescents, and 0/0/12 minutes on adults.

Conclusion:
We have developed a controller that provides acceptable performance that is significantly improved with the addition of our meal detection algorithm. We have shown an improvement with respect to a strict, optimal basal control with optimal boluses at meal times. We are one step closer to freeing diabetics from their disease.

talk no.

11

Constrained Mixture Estimation for Analysis and Robust Classification of Clinical Time Series

Alexander Schoenhuth
University of California, Berkeley

Ivan Costa
Christoph Hafemeister
Alexander Schliep

The work was presented at this year's ISMB/ECCB 2009 in Stockholm. See [1] for the respective paper with more detailed explanations and guiding figures.

Purpose:
Personalized medicine based on molecular views of diseases, predominantly supported by gene expression profiling methods, has become increasingly popular. However, when analyzing clinical gene expression data one faces multiple challenges; most of the well-known theoretical issues such as high-dimensional feature spaces, few examples, noise and missing data apply. Special care is needed when designing classification procedures that support personalized diagnosis and choice of treatment. Here, we particularly focused on classification of interferon-beta (IFN-beta) treatment response in Multiple Sclerosis (MS) patients which has attracted special attention in the recent past. While still being standard, about half of the patients remain unaffected by IFN-beta treatment. In such cases, treatment should be timely ceased to mitigate the side effects.

Materials and Methods:
We propose constrained estimation of mixtures of hidden Markov models as a methodology to classify patient response to IFN-beta treatment. The advantages of our approach are to take the temporal nature of the data into account and its robustness with respect to noise, missing and mislabeled data. Moreover, mixture estimation enables to explore the presence of response sub-groups of patients on the transcriptional level.

Results:
We clearly outperformed all prior approaches in terms of prediction accuracy, raising it, for the first time, above 90%. Moreover, we were able to identify potentially mislabeled samples and to subdivide the good responders into two subgroups that exhibited different transcriptional response programs. This is supported by recent findings on MS pathology and therefore may raise interesting clinical follow-up questions.

Conclusion:
We have developed a sound and reliable methodology to analyze and classify clinical time series. We outperformed all prior approaches on MS treatment response data. We found that subdividing the positive responders into two subclasses yielded best results, which is supported by recent findings on MS pathology as it indicates that MS patients display distinct expression profiles signatures. We were also able to identify a mislabeled patient. A general explanation for the superiority of our approach is its capacity of handling noisy, missing and mislabeled data which are notorious issues in clinical classification settings.

References:
[1] I.Costa, A.Schoenhuth, C.Hafemeister, A.Schliep, Bioinformatics 25 (ISMB/ECCB 2009), i6-i14.

talk no.

12

Mapping Eye Movements in Three-Dimensions: Analyzing Gaze Paths when Interpreting Volumetric Chest CT Data

Martin Tall
Stanford University

Martin Tall
Tae Jung Kim
Justus E. Roos
Sandy Napel
Geoffrey D. Rubin

Purpose:
Previous studies investigating eye movements of radiologist during their image analysis have recorded gaze data on single images containing one or more objects, e.g., nodules in a CT scan. Typically the images are displayed for a fixed duration and the results are illustrated through aggregated visualizations e.g., fixation plotting, scan paths or heat maps. However, when analyzing search patterns while performing standard interpretation of volumetric chest CT data, paging through a stack of transverse sections warrants three-dimensional analysis of eye movements. To our knowledge this has not been studied.

Materials and Methods:
Our solution for recording volumetric gaze data, using a remote corneal reflection system, relies on two-way communication between the eye tracker and our custom DICOM viewer. This enables real-time mapping between on-screen fixations and the 3D DICOM coordinates of the viewed images at a rate of 50 Hz, thereby linking our eye tracker data and the physical space in which the 3D image data were obtained.

Results:
Our approach to tracking and visualizing gaze data recordings within the acquired volumetric space utilizes the DICOM coordinate system for rendering graphics in 3D space. It produces several interesting visualizations, such as time-stamped 3D gaze paths, and 3D heat maps revealing direct and foveal dwell time per voxel, all fused with the acquired volume data for simultaneous viewing.

Conclusion:
Our approach offers insight into the relationship between recorded gaze data and the volumetric space of the viewed medical images. We are employing it in an investigation of how paging through volumetric lung CT data affects nodule detection though pop-out feature detection in the peripheral visual field. More generally it will facilitate studies across a wide variety of 3D modalities and illuminate specific aspects of volumetric inspection that can be used for training and diagnostic support.

talk no.

13

Detecting Chronic Pain in the Brain

Hoameng Ung
Stanford University

Justin Brown
Sean Mackey

Purpose:
An estimated 50 million Americans suffer from chronic pain. Since pain is subjective, self-reported pain cannot be verified. Because of this, an objective measure of pain could assist in the management of patients who are suspected of malingering or drug seeking and of patients unable to communicate such as the critically ill. Previous research has shown that chronic pain is associated with gray matter atrophy, detectable by magnetic resonance imaging [1]. Therefore, we hypothesized that machine learning could be used to identify persons with chronic back pain from magnetic resonance images of the brain, and in doing so further elucidate the relationship between brain atrophy and pain.

Materials and Methods:
High resolution magnetic resonance images of the brain from 66 healthy controls and 60 chronic low back pain (CBP) patients were included in this study. Segmentation, normalization and volumetric modulation of these images were performed using the VBM toolbox for SPM8 to generate gray matter (GM) density images, which were then segmented into 5x5x5 contiguous volumes. This reduced the number of image features from over 300,000 voxels to 1700 volumes. Average GM density was computed for each volume and was used to train a binary linear support vector machine (SVM), which then classified test images as having CBP or not having CBP. The corresponding weights highlight the areas of the brain most important in driving the classifier. Performance of the model was assessed with a leave-one-out cross-validation scheme and a more general bootstrap paradigm iterated 1000 times, where successive evaluations were performed on a new, separate test set. Statistical significance was determined with a Monte Carlo permutation test.

Results:
The SVM classifier identified the presence or absence of chronic pain correctly 74.4% of the time ($p<0.01$). Sensitivity and specificity were 77.8% and 72.5% respectively. One of the most important regions for classification was the left thalamus and areas of the ventral medial prefrontal cortex.

Conclusion:
Our results show that an SVM can identify persons with chronic low back pain from high resolution images of the brain with greater-than-chance accuracy, and that there are thalamic as well as prefrontal cortical changes in gray matter density associated with chronic low back pain. These findings provide a promising outlook for using more powerful machine learning techniques to classify pain based on structural brain data.

References:
[1] Apkarian A., et al. (2004) J Neurosci 24(46):10410-15

bcats2009

Poster Spotlight Abstracts

poster

S1

Prediction of protein protein interaction sites and their impact on genetic disease

Angshuman Bagchi
Buck Institute for Age Research

Eunseog Youn
Matthew E. Mort
David N. Cooper
Sean D. Mooney

Purpose:
Protein-protein interactions play pivotal roles in many biological processes, like hormone–receptor binding, antigen-antibody interactions etc. Furthermore, there is increasing evidence that disease-causing mutations lead to disruption of protein interactions (1). There are several experimental methods available for the analysis of protein interactions, however, they are time consuming and expensive. A number of computational algorithms has been developed to predict residues in protein-protein interfaces with accuracies in the range of 70% (2). In the present work, we used machine learning tools to discriminate between interface and non-interface residues of proteins using both structure and sequence information.

Materials and Methods:
We generated sequence and structure based features from a non-redundant set of protein hetero-complexes from the protein data bank. The training dataset was divided into two categories; (A) interface residues and non-interface surface residues for structure based prediction and (B) interface residues and non-interface surface and core residues for sequence based prediction. The datasets were used to build classification algorithms using random forest (RF) and support vector machine (SVM) coupled with 10 fold cross-validation for evaluation (2).

Results:
Overall, RF outperformed SVM in most cases. The best performing sequence-based classification tool achieved an accuracy of 73% and when protein structure was included the accuracy was 75%. The predictors were then used to analyze mutation data including somatic mutations in cancer from tumor resequencing projects, the Human Gene Mutation Database (HGMD), and common human polymorphisms. The results show an enrichment of protein interaction sites in the disease datasets compared to the neutral set (Seattle SNPs).

Conclusion:
Overall our results indicate that disease mutations are enriched in disruption of PPI interfaces and these interfaces can be predicted using bioinformatic methods.

References:
1. David C. Fry, Peptide Science, 84, 535-582, 2006
2. Janin et al., Quaterly Reviews of Biophysics, 41, 133-180, 2008

poster

S2

New MR Pulse Sequences for Imaging Metallic Hardware in the Knee

Christina A. Chen
Stanford University

C.A. Chen
W. Chen
S.B. Goodman
B.A. Hargreaves
G.E. Gold

Introduction:
The metal in surgical implants distorts the main magnetic field in magnetic resonance imaging (MRI), producing artifacts(1) that limit the diagnostic value of musculoskeletal MRI for post-operative complications(2). We have developed 2 three-dimensional MRI prototypes that correct for these artifacts, Slice Encoding for Metal Artifact Correction (SEMAC)(3) and Multi-Acquisition Variable-Resonance Image Combination (MAVRIC)(4). We compared artifact size measured on SEMAC and MAVRIC images to that measured on conventional two-dimensional fast spin echo (FSE)(5) images of the knee.

Methods:
Nine knees of 8 volunteers with total knee replacements (TKR) were imaged in the sagittal plane using a GE Signa Exite HDx 1.5T MRI scanner and an 8-channel knee coil (Fig. 1; please see URL for figures and tables). Table 1 lists all imaging parameters.

For each knee, the medial tibial plate, anterior medial femoral component, and posterior medial femoral component were evaluated for artifact. For each method, 3 slices equally spanning the medial-lateral dimension were chosen from which to measure the lengths of the artifacts and compared among methods with paired t-tests.

To evaluate the accuracy of the methods in measuring geometry in the presence of metal, a post-operative knee model consisting of plastic femoral and tibial bones fitted to a TKR cobalt-chromium femoral component, plastic spacer, and stainless steel tibial component (Fig. 2) was scanned. The known anterior/posterior (A/P) and medial/lateral (M/L) dimensions of the metal components and plastic spacer were compared to the lengths measured by the 3 methods through percent deviations.

Results:
In all metal joint compartments, SEMAC and MAVRIC were significantly better at artifact reduction than FSE (all $P<.03$), while being statistically equivalent to each other (all $P>0.1$, Fig. 3). Table 2 lists the percent deviation between the measured and true A/P and M/L dimensions of the TKR.

Conclusion:
Results from the volunteers and TKR knee model show that SEMAC and MAVRIC correct for metal-induced artifact, allowing them to accurately measure metal implant geometry. FSE images suffered from statistically larger artifacts. MAVRIC and SEMAC are promising MR imaging methods that may allow for improved musculoskeletal follow-up imaging of metallic implants and soft tissue structures surrounding metal in the knee.

References:
1. Wendt RE et al., Radiology 1988;168:837-41.
2. Sofka et al., Semin Musculoskelet Radiol 2002;6:79-85.
3. Lu W et al., Magn Res Med 2009;62(1):66-76.
4. Koch KM et al., Magn Reson Med 2009;61(2):381-90.
5. Butts et al., Magn Reson Med 2005;53:418-424.

Acknowledgements:
Wallace H. Coulter Foundation; NIH grant 1R21EB008190

poster

S3

A Comparison of Landing Strategies and ACL Injury Risk Using an Inertial-Based System

Ariel V. Dowling
Stanford University

Julien Favre
Thomas P. Andriacchi

Purpose:
One of the most common anterior cruciate ligament (ACL) injury mechanisms is a single limb landing with the knee at or near full extension1. Additionally, increased knee abduction angle during a jump-landing task has been suggested to increase the risk of injury among female athletes3. It has previously been demonstrated that inertial sensors, such as gyroscopes and accelerometers, can be used to accurately measure knee kinematics during walking2. The objective of this study was to utilize an inertial- based system to assess ACL injury risk among healthy subjects by measuring landing kinematics during single support drop jumps.

Methods:
10 healthy subjects were evaluated in an IRB approved study with informed consent. The jumping task for the study was a drop jump, where subjects were asked to drop directly off of a box and then immediately perform a maximum vertical jump4. Two inertial measurement units (IMUs) were used to record the subject's movement and were placed on the subjects' legs with elastic straps. A fusion algorithm was used to track the orientation of the IMUs4, and two calibration sequences were completed to align the IMUs on the leg to the segment bone anatomical reference frames2. 3D knee angles were calculated according to the joint coordinate system defined by Grood and Suntay. Student's T-Tests (two tail, equal variance, κ< 0.05) were used to determine statistical significance between the variables of interest.

Results:
Two distinct landing strategies were observed: 1) a peak knee adduction and external rotation angles during landing (aDd, 3 subjects), or 2) a peak knee abduction and internal rotation angle (aBd, 7 subjects). As a result, the subjects were classified into two groups based on the chosen landing strategy. An 18° difference in peak abduction/adduction angle and a 17° difference in peak external/internal rotation angle were measured between the groups.

Conclusion:
The results of this study indicate that distinctive and important differences in the landing strategies for a single support drop jump can be identified with an inertial-based system. A strategy that couples a high abduction angle with internal rotation could put subjects at greater risk for ACL injury4. Therefore, subjects in the aDd group could be stratified into "low risk" and subjects in the aBd group could be classified as "high risk" based on landing strategy.

References:
1. Boden BP et al. (2000) Orthopedics, 23(6): 573-578.
2. Favre J et al. (2009) J Biomech, Epub Aug 7.
3. Hewett TE et al. (2005) Am J Sports Med, 33(4): 492-501.
4. Myer GD et al. (2007) BMC Musculoskelet Disord, 8:39.

poster

S4

Computational and Molecular study of Intron-Mediated Enhancement of Gene Expression

Shahram Emami
University of California, Davis

Shahram Emami
Genis Parra
Keith Bradnam
Ian Korf
Alan Rose

Purpose:
Introns have been shown to significantly enhance gene expression in diverse organisms in a process called intron-mediated enhancement (IME)(1). We sought the intron sequences that are involved as one route towards understanding the mechanism of IME. Unlike an enhancer, IME boosts expression only when intron is located in transcribed sequences near the promoter. Introns clearly differ in their ability to enhance, suggesting that stimulating sequences are more abundant in some. Advancing our understanding of IME, will not only increase our knowledge of gene expression, but it will also benefit biotechnology applications that require high gene expression.

Materials and Methods:
Intron is added between exons 1 & 2 of a native A. thaliana gene, which is fused to a reporter gene. Protein expression and mRNA accumulation data is obtained from homozygous single copy stable integrated transgene in A. thaliana.

An algorithm called IMEter(1), uses word frequencies to describe intron's composition and provides a log odds score. IMEter score describes similarity of an intron to the set of promoter proximal introns in the genome, and it correlates with its effect on expression. To identify the most significant motifs within the top 100 IMEter scoring introns in A. thaliana, a probabilistic motif finding software (NestedMICA) was used to discover several motifs that were overrepresented in enhancing introns. Modified introns formed by changing the abundance of the top IME motif have also been tested.

Results:
A) IMEter predictions of introns' enhancing capabilities were validated experimentally
B) Intron has to be close to the 5' end of the transcription unit to maximally boost expression
C) IME motifs can boost expression when their orientation is reversed (the reversed sequences have been fused to 5' and 3' sequences that ensure efficient splicing of the intron)
D) Deleting parts of enhancing intron decreases, but does not destroy the IME effect
E) Hybrid intron made from parts of enhancing and non-enhancing introns, elevates expression as compared to no intron control
F) TTNGATYTG was identified as one IME motif, and its role was confirmed by changing its abundance in different introns

Conclusion:
1) Efficiently spliced introns do not boost expression equally
2) IME motifs are dispersed throughout enhancing introns and can operate in either orientation
3) IMEter detects compositional differences in promoter-proximal introns that are related to their effects on expression
4) One IME motif has been verified

References:
1) Rose A.B., Elfersi T., Parra G., Korf I.; Promoter-Proximal Introns in Arabidopsis thaliana Are Enriched in Dispersed Signals that Elevate Gene Expression; The Plant Cell, 20: p. 543–551, March 2008

poster

S5

Independent component analysis of large microarray data compendium identifies fundamental modules of human biology

Jesse Engreitz
Stanford University

Bernie Daigle
Jon Marshall
Russ Altman

Purpose:
Cellular physiology, including disease states and drug responses, results from the combined influences of many genes. Experimentalists have now sampled many conditions and cell types, contributing vast amounts of microarray data that represent many key biological modules and pathways. While most methods of microarray analysis utilize only a small portion of this data, we use a 9,395-microarray compendium to derive human functional modules that together span human biology. Modeling gene expression as a combination of these fundamental modules, we identify novel effects of the pre-clinical anticancer drug parthenolide (PTL).

Materials and Methods:
We apply independent component analysis (ICA) to a large compendium of publicly available microarray data. To control for the effects of rare or over-represented conditions in the compendium, we apply a hierarchical clustering step and filter the input arrays. To functionally annotate the resulting fundamental components, we 1) look for enriched Gene Ontology (GO) codes among the top genes, 2) associate each module with the experiments that express it, and 3) identify enriched pathways using Ingenuity Pathway Analysis [1]. To elucidate the mechanism of PTL, we project a 24-array data set [2] into fundamental component space and identify the top differentially expressed components in PTL-treated compared to untreated AML cells.

Results:
We identify 423 reproducible, fundamental components that describe a wide range of human cellular biology. GO annotation suggests that while some represent known biological modules, some may describe biology not well characterized by existing manually-curated ontologies. We find that our method correctly identifies the NF-κB signaling and oxidative stress pathways as treatment responses for PTL, yielding more statistically significant results than analysis of the PTL data set in isolation. We also identify a number of highly-connected hub genes, including HES1 and GTF2B, that may play an important role in cellular response to PTL.

Conclusion:
ICA yields biologically relevant, reproducible gene modules that aid in the analysis of new microarray experiments. Leveraging a large gene expression compendium, our method outperforms conventional differential expression analysis using single data sets. The ability to identify differentially expressed gene modules and link new arrays to existing experiments provides a powerful tool for data mining and hypothesis generation, and in the arena of drug response allows for the identification of novel drug targets and mechanisms.

References:
1. Ingenuity Systems, www.ingenuity.com
2. Hassane et al (2008) Blood 111 5654-5662

poster

S6

Quantifying Margin Characteristics of Lesions from CT Images for Content Based Image Retrieval

Ankit Gupta
Stanford University

Sandy Napel
Hayit Greenspan
Christopher F. Beaulieu
Daniel L. Rubin

Purpose:
To develop a method to quantify the sharpness of the margin of liver lesions and to evaluate its performance for retrieval of images with similarly appearing lesions.

Materials and Methods:
We compiled 30 portal venous phase CT images of liver lesions: cysts, metastases, and hemangiomas. We created a reference standard of similarity, in which two radiologists rated all pairs of images by consensus in terms of similarity on a 3-point scale. One radiologist manually circumscribed each lesion, creating a region of interest (ROI). We also created 100 simulated circular liver lesions with varying margin sharpness by applying Gaussian blur, adding Gaussian noise, and suppressing noise with either a gradient anisotropic diffusion filter or salt and pepper median filter. The intensities of the lesion and liver, and the margin sharpness were varied keeping the noise level close to that in abdominal CT scans. We computed a margin sharpness feature for each lesion by dilating the ROI to include a small rim of normal liver beyond the lesion, suppressing noise with a median filter, and measuring intensities along radial lines starting at the center of the ROI to points on its boundary. We fit a sigmoid function to these values, computing a 2-valued feature vector to describe the sigmoid, and averaged the feature vectors over all radii. We evaluated this feature in the reference standard, selecting one image as a query image and rank ordering the remaining images according to Euclidean distance between feature vectors. Mean Normalized Discounted Cumulative Gain, (NDCG), a standard Information Retrieval evaluation score (best=100%, worst=0%), was then used to compare this ordering with the expected ordering based on the reference standard as a function of K, the number of images retrieved.

Results:
Image retrieval performance was good in both simulated and clinical images. For K= 30 the NDCG score for retrieval was 84.59% for the simulated images and 95.34% for the clinical images. In simulated images, retrieval performance was excellent for both blur and intensity components of the margin sharpness feature. Both gradient anisotropic diffusion and salt and pepper median filters used for noise reduction while preprocessing were not significantly different in their impact on the results.

Conclusion:
We have developed a method to quantify the sharpness of the margin of liver lesions. In simulated and clinical datasets, this image feature has good performance in retrieving similar images of lesions having different types of margin characteristics, suggesting potential utility in content-based image retrieval.

poster

S7

Electron Tunneling in Complex I of the Electron Transport Chain

Tomoyuki Hayashi
University of California, Davis

Alexei Stuchebrukhov

Purpose:
Complex I (NADH-quinone oxidoreductase) is one of three energy-transducing enzyme complexes of the respiratory chain in mitochondria, which catalyzes the oxidation of NADH and the reduction of ubiquinone in mitochondria and respirating bacteria. This reaction involves the transfer of electrons along nine redox centers over approximately 90 Å from NADH bound to the hydrophilic domain to ubiquinone in or near the hydrophobic membrane bound domain of complex I 1. Here we apply our Tunneling Current Theory 2 to explore the atomistic detail of the whole electronic wiring in Complex I.

Methods:
The tunneling current method implemented at spin-unrestricted ZINDO level is applied for mapping electron transfer tunneling paths from FMN to N2 of Complex I of the electron transport chain. We found that the spin-unrestricted ZINDO calculation provides the anti-ferrromagnetic electronic spin states of Fe/S redox centers proposed by BS-DFT calculations 3.

Result:
The calculations show that the Cys ligands of the iron-sulfur cofactors provide the exit and entrance pathways of the electron tunneling of most pairs (except for N4-N5 and N6b-N2). Electron tunneling pathways are well defined in all pairs with 1-3 significant protein residues providing the main pathway in addition to Cys ligands. There are 1 or 2 through-space jumps involved in the electron tunneling of all pairs and their distances are in the range of 1.9 – 3.7 Å. We found a reasonable linear fitting of with respect to the tunneling distance, which is consistent with the well-known exponential decay of the tunneling matrix element over the electron transfer distance.

Conclusion:
The atomistic detail of electronic wiring in Complex I is revealed. Together with the structure, this is a major step forward in characterization of Complex I. The electron tunneling takes nearly the shortest path along the protein bridge between each redox center pair of the wiring with one or two through-space jumps. The dependence of the electron tunneling matrix elements on the tunneling distance is close to exponential.

References:
1 L. A. Sazanov and P. Hinchliffe, Science 311 (5766), 1430 (2006).
2 A. A. Stuchebrukhov, Theoretical Chemistry Accounts 110 (5), 291 (2003).
3 R. A. Torres, T. Lovell, L. Noodleman, and D. A. Case, Journal of the American Chemical Society 125 (7), 1923 (2003).

poster

S8

Bone Conduction Analyses in a Finite Element Model of the Human Middle ear and Cochlea

Namkeun Kim
Stanford University

Kenji Homma
Charles R. Steele
Sunil Puria

Purpose:
A three-dimensional finite element model of the human middle ear and cochlea, with a tapered box geometry, was developed to allow calculations of the basilar membrane (BM) velocity, as well as the round window (RW) to oval window (OW) volume displacement ratio (Urw/Uow). The model was used to explore the effects of bone conduction (BC) on cochlear function.

Materials and Methods:
To investigate the effects of BC on cochlear response, we specifically simulate the following three cases: 1) Inserting a "third window" (TW) in the cochlear outer wall that might account for a reported imbalance in Urw/Uow under BC excitation (Stenfelt et al., 2004); 2) Applying BC excitation through translational acceleration (TA) of the ear, versus through compression of the bony capsule; and 3) Allowing the primary osseous spiral lamina (OSL) to be mobile versus immobile for both air conduction (AC) and BC.

Results and Conclusion:
1) When the TW was modeled as a simple compliant membrane, the model produced an imbalance of Urw/Uow that was consistent with experimental measurements, whereas when modeling the TW as a tube representing the cochlear or vestibular aqueducts, the results were inconsistent with experimental data. Terminating the tube with a compliant membrane produced better experimental agreement than with the tube alone, but poorer agreement than with the membrane alone.

2) For TA without TW, BM velocities, normalized by the OW velocity, were found to depend on the direction (x, y, or z - longitudinal, transverse, and normal to BM, respectively) of the acceleration at high frequencies, but not at low frequencies. By contrast, the normalized BM velocities generated by compression of the bony capsule were inconsistent with experimental measurements (Stenfelt et al., 2003). A tentative conclusion is that, relative to compression, TA is the more dominant mechanism for BC, at least in the tapered box model geometry for the cochlea presently used.

3) For AC stimulation, adding a mobile OSL decreased the BM velocity (normalized by the OW velocity) at high frequencies, while at low frequencies a mobile OSL had a negligible effect. In addition, for AC responses, a mobile OSL has a significant influence on the cochlear input impedance, but not on the cochlear pressure, volume displacement ratio (Urw/Uow), or transfer function between the middle ear and the cochlea. On the other hand, the normalized BM velocities decreased with a mobile OSL under BC stimulation, for all frequencies.

References:
Stenfelt et al., (2004) J. Acoust. Soc. Am. 115(2): 797-812
Stenfelt et al., (2003) Hear. Res. 181: 131-143

poster

S9

Estimating the Likelihood of Cure from Lung Cancer

Ray Lin
Stanford University

Bronislava M. Sigal
Sylvia K. Plevritis

Purpose:
To predict the mortality reduction in lung cancer due to screening by estimating the relationship between the size of the primary tumor and likelihood of cure from lung cancer.

Materials and Methods:
A mathematical model of the natural history of lung cancer was developed to estimate the relationship between the size of the primary tumor and the likelihood of cure. The model is applied separately to lung adenocarcinoma, squamous cell carcinoma, large cell carcinoma, and small cell carcinoma, and separately for males and females. Model parameters were estimated using data from the Surveillance, Epidemiology and End Results (SEER) cancer registry. The model reproduces SEER, validates against an external dataset and produces estimates of tumor volume doubling times that are consistent with empirical data.

Results:
An individual patient who would have been clinically detected in the absence of screening has a 50% probability of cure if their primary tumor were diagnosed and treated by 6 - 8 mm. Among a population annually screened with a test that can detect tumors 6-8 mm in size, the reduction in lung cancer mortality is 25-30%, provided there was no delay in diagnosis and treatment. The reduction in lung cancer mortality is less than 10% if the mean tumor size at diagnosis and treatment increases to 15mm or greater. Because CT screen detected tumors in the mm-range are often followed for growth to reduce the risk of unnecessary biopsies, the opportunity to cure the patients with deadly disease may be significantly diminished by the time the tumor is diagnosed and treated.

Conclusion:
Because mm-sized lung nodules are often followed for growth to reduce the risk of unnecessary biopsies, the opportunity to cure lung cancer patients may be significantly diminished.

bcats2009

Poster Abstracts

poster

1

Automatic Segmentation and Structural Study of the Bacterial Cell Wall in Cryo-Electron Tomography

Fernando Amat
Stanford University

Farshid Moussavi
Luis R. Comolli
Kenneth H. Downing
Daphne Koller
Mark Horowitz

The advent of high throughput tomography is key to structural studies of cellular and subcellular assemblies, and remains an elusive goal. In recent years, there have been tremendous advances in the automatic acquisition of electron microscopy data and generation of tomograms. The last remainingbottleneck in the pipeline is the automatic segmentation of cellular structures. Due to severe signal to noise limitations and missing data, established image processing computer vision programs have significant challenges in these datasets. For example, apparent boundaries can be inconsistent, broken, and sometimes just nonexistent due to the missing data wedge. The segmentation of such datasets often involves days of manual effort of an expert.

To address these problems, we have developed a probabilistic framework based on conditional random fields (CRF's) that makes significant use of context and shape as well as physical features. A weak shape prior is assumed, and gradually refined as the inference progresses through the 3D volume. So far we have obtained encouraging results on several cryo-EM datasets of Caulobacter crescentus, achieving automatic segmentation of a membrane in less than 2 hours on a desktop computer. We are extending the model to be more robust to shape changes and different datasets.

One of our first applications of this segmentation is the study of the S-layer structure in its native state. While bacterial S-layers have been studied for over 30 years, most of the studies have been performed on isolated S-layer sheets or proteins. We used datasets from cryo-EM on whole cells of the gramnegative bacterium Caulobacter crescentus to obtain quantitative information on the S-layer structure and its interactions with the outer membrane in its native state. We need robust and efficeint pattern recognition techniques to process efficiently the large volume of low SNR data that results from this method. Using the automatic membrane segmentation described above as a starting point, we search for the S-layer in a thin volume around the models estimated surface, by locally maximizing the characteristic S-layer hexagonal signature in the Fourier domain.

Once we have the S-layer location we can register individual S-layer patches to improve the structural resolution. We use a metric based on sparse representation of 3D images in spectral domain to register 3D volumes with SNR well below odB. Registration between thousands of patches is key to identify different conformations and obtain an averaged structure for each class. Following this procedure we have identified two different conformations of the S-layer and obtain its structure at resolution of few nanometers.

poster

2

A Multiscale Model of Breast Tumor Growth

Aravindakshan Babu
Stanford University

Alexander Anderson
Sylvia Plevritis
Shih-jui Lin

We present a multiscale model of tumour growth whose parameters are estimated from epidemiology data. The tumour is modeled at two levels: 1) A partial differential equation model of the densities of various cell types(normoxia, hypoxia, apoptotic cells & vasculature) and the chemical factors(nutrient & drugs) 2) A stochastic model of the tumour natural history. These two models are integrated, allowing us to make inferences at the cellular level from epidemiology data.

The PDE model is of the reaction-diffusion type and models several complex dynamics of tumour development such as cellular diffusion, angiogenesis, metastasis etc at the cellular level. The model is controlled via parameters such as the diffusion and proliferation rates for normoxia etc. Given a set of parameter values, the PDE system is solved numerically to get the distribution of densities at various instants of time.

The natural history model models the relationship between the size of the primary tumour at detection,the disease stage and the survival time of the patient. The model is driven by the choice of growth curve for the primary tumour volume. For instance the growth of the metastatic burden is a delayed and scaled version of the growth of the primary. When the metastatic burden exceeds a lethal threshold the patient dies. The model is controlled via parameters such as the growth rate of the primary, the curability threshold, the lethal metastatic burden etc. These parameters are hypothesized to be drawn from suitable distributions. One of our contributions is to use the growth curve generated by the PDE model in conjunction with the natural history model.

We present a mathematical investigation of the qualitative behaviour of such a model.This gives us valuable insights into the strengths and weaknesses of such a model. And also leads to a computationally feasible estimation procedure. We derive expressions for and compute confidence intervals for the estimated parameters.

poster

3

Characterization of Structural Modularity of Metabolic Networks using Mass Spectrometry based Metabolomics

Dinesh Kumar Barupal
University of California, Davis

Gert Wohgemuth
Oliver Fiehn

Introduction:
Modularity is an intrinsic organizational principle of biological networks[1, 2]. Here, we represent characterization of structural modularity of the metabolic networks using metabolomics datasets.

Methods:
A time-of-flight mass spectrometer (Leco Pegasus IV) coupled of gas chromatograph was employed for profiling of small molecules in perfused lungs tissues of adults rats. An informatics infrastructure BinBase executed post-acquisition data processing. Biochemical and chemoinformatics relationships available from open access databases were utilized for network mapping of known and unknown metabolites. Networks was visualized in cytoscape software and was characterized using MCODE plug-in for detection of modules.

Results:
A total 417 metabolic signals were consistently detected across multiple samples. Up to 140 signals were identified as genuine metabolites which passed the sets of quality criteria in BinBase. We used a three-tiered approach combining biochemical, chemical and mass spectral similarity distances that successfully generated network graphs including all identified metabolites co-localized with signals of novel, structurally unknown signals. Use of Tanimoto chemical similarity algorithm elucidated modular organization within network graphs. Distinct metabolic modules of fatty acids, sugar, nucleotides, amino acids and aromatics resembled structural modularity of metabolic networks. We found modular network graphs superior in clarity and information content compared to direct mapping to biochemical reaction and pathway network tools. Furthermore, the module detection algorithm MCODE also identified several modules. A dynamic modular response in the lung metabolism was observed suggesting alterations in functional modules of nucleotide metabolism and reactive oxygen species metabolism.

Conclusions:
Metabolomic network have showed clusters of metabolites that resemble classic biochemical modules and pathways, without losing information on novel compounds. Such modular networks are ideally suited to highlight statistical differences in test/control comparisons and thus guide biologists to areas of large metabolic perturbations.

References:
1. Kitano, H., Systems biology: a brief overview.Science, 2002. p. 1662-1664.
2. Hartwell, L., et al., From molecular to modular cell biology. Nature, 1999. 402(6761): p. 47.
3. Fiehn, O., Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. Comparative and Functional Genomics, 2001. 2(3): p. 155-168.

poster

4

Transfer learning for modeling genetic influence on multiple related diseases

Alexis Battle
Stanford University

Daphne Koller

Purpose:
We explore the use of transfer learning in the domain of predicting disease susceptibility from genotype for a set of related human diseases.

Materials and Methods:
We have developed a flexible transfer learning framework, called PriorNet, based on structured priors which are then applied models such as logistic regression. Here, we apply PriorNet to the WTCCC genome-wide association data, which proves data for several human diseases, and including genotype information for a large set of individuals with each disease. With PriorNet, we utilize the relationships between diseases and the known interactions between genes to help learn more accurate, and more biologically meaningful predictive models.

Results:
Our preliminary results indicate that using PriorNet to jointly learn predictive models of disease results in higher accuracy, especially when the number of individuals in the training set is limited. Further, our results highlight similarities between the WTCCC autoimmune diseases, providing a set of SNPs suggested to be relevant to multiple diseases.

Conclusions:
The use of structured priors for transfer learning holds promise for learning biologically meaningful predictive models for related diseases.

References:
Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. WTCCC. Nature. 2007;447;661-78.

poster

5

Quantitative Drug Screening and Discovery with Phosphoflow Cytometry

Tiffany J. Chen
Stanford University

Matt Clutter
Garry P. Nolan
Serafim Batzoglou

Although common cancer therapeutics exist, the mechanisms by which they act are not well defined. Drugs display differential action in cells with various molecular states (dividing, activated, etc), but therapeutics are compared using metrics like survival, ignoring cellular function. Subsequently, patients receive fairly unpersonalized doses of therapeutics, quantities of which are often derived by trial and error. The interplay between drugs and intracellular signaling networks is essential for a deeper understanding of how medications affect different cell subpopulations. This interplay is also key to creating a standard metric of comparison between individual drugs.

As a drug discovery tool, phosphospecific flow cytometry (phosphoflow) emerges as a cell-based drug screening method--systematic perturbations yield quantitative, multiple-parameter (DNA, protein) information for complex disease states. Because we can measure states at the single-cell level, we anticipate that phosphoflow will aid in investigating drug mechanisms. To effectively utilize mass amounts of interrelated data, there arises a need to develop automated tools for data aggregation, interpretation, and analysis. Most analyses are restricted to a couple biochemical parameters, lacking insight into constraints and interaction information. We hypothesize that systematic categorization and analysis of parameter states will provide insight into drug dosing, drug function, and underlying cell biology.

One of the first steps towards building this automated tool requires the isolation and categorization of cell subpopulations in multidimensional protein space. In this study, we illustrate why automated subpopulation discovery is difficult, as well as how finding meaningful subpopulations can result in the reclassification of very well-characterized cancer drugs. Furthermore, we are working towards building a tool to discover relationships between drug dose, exposure time, and downstream effects on individual cells.

poster

6

Sparse space-time decompositions of ECoG signals

C. Cheung
University of California, Berkeley

C. Cadieu
L. Secundo
E. Chang
B.A. Olshausen
R.T. Knight

Purpose:
We sought a novel decomposition of electrocorticogram (ECoG) data that is determined by the statistical structure of the data, and not by prior assumptions about the characteristic frequency bands contained within the data. This work is part of our ongoing project to relate behavior to the dynamics of large neuronal populations during cortical processing and communication.

Materials and Methods:
Three refractory epileptic patients (18-35 years) had undergone a craniotomy for chronic (1-2 weeks) implantation of a subdural 8 × 8 platinum–iridium electrode array. ECoG signals were sent to a clinical monitoring system and a custom-recording system. Subjects used a stylus on a touch-screen connected to a laptop computer to perform center-out target-directed arm movements.

The activity of the ECoG signal was modeled through a linear generative model in which sparse hidden causes produce temporal patterns [Olshausen 2002]. In the model, the sparsely active causes are each convolved with a temporal-pattern, called a basis function. We used unsupervised learning to adapt the basis functions to the statistics of the ECoG signal. By imposing a sparseness penalty, the model is forced to use as few causes as possible to account for the ECoG signal.

Results:
A total of 64 basis functions were derived to represent the ECoG signal. The model is able to represent the signal with a small number of causes per time point, ~2-5.

Analyzing the causes with respect to arm movement shows that several basis functions are strongly active when the arm is moving. This strong correlation in behavior is seen in both averaged activity and in single trials.

In addition, a simple spectral analysis of each basis function indicates that many basis functions learn structure in high-gamma. Furthermore, a number of basis functions reveal coupling between two or three frequency bands (ie. between alpha and gamma).

Conclusion:
We have learned a novel decomposition of ECoG signals into a set of sparse basis functions and causes. The learned basis functions share some properties with standard space-time decompositions, but they have unique properties that are dictated by the underlying structure of the ECoG data. The activation patterns of the sparse causes are correlated with arm movement and may improve BMI techniques. The activity of the sparse causes may be indicative of the underlying dynamics of large neuronal populations during cortical processing and communication.

References:
Olshausen, BA . Sparse Codes and Spikes (2002). In: Probabilistic Models of the Brain: Perception and Neural Function. R. P. N. Rao, B. A. Olshausen, and M. S. Lewicki, Eds. MIT Press. pp. 257-272.

poster

7 Electrostatics of Complex I of the Electron Transport Chain

Vernon Couch
University of California, Davis

Emile Medvedev
Alexei Stuchebrukhov

Purpose:
Respiratory complex I is the largest and most elusive of the primary protein complexes of the electron transport chain in mitochondria and respiring bacteria. Complex I serves as an entry point for electrons into the electron transport chain oxidizing NADH from the mitochondrial matrix and transferring electrons over 90 Å to quinone supplied from the mitochondrial inner membrane. This reaction is in turn coupled to the translocation of 4 protons across the inner membrane generating, in part, the electrochemical gradient necessary for ATP production[1]. Much like a battery, complex I utilizes flavin mononucleotide and several FeS cofactors, which form a "wire" through the enzyme, to effect the reaction between physically separate reductant and oxidant species [2]. The characterization of the electrostatics and redox chemistry of the system of electron transfer cofactors present in respiratory complex I is the focus of this work and ongoing investigations.

Materials and Methods:
The solution of the Poisson-Boltzmann equation was employed to calculate the electrostatic interactions between FeS clusters and their response to external fields applied across the membrane. Both the "Macroscopic Electrostatics with Atomic Detail" (MEAD) program and in house codes were used.

Results:
We have calculated the electrostatic interaction energies between the FeS cofactors; a necessary ingredient in interpreting redox titration experiments and determining midpoint reduction potentials for the FeS clusters. Furthermore, we propose an experiment to locate the FeS cofactors with respect to the membrane based on our calculations of the response of the FeS clusters to an externally applied electric field.

Conclusion:
The FeS cluster interaction energies are highly modified by the protein dielectric boundary as compared to Coulomb's Law estimates. The effective dielectric was shown to be greater 2 times the protein dielectric and is distance dependant [3]. These findings indicate that the protein shape and cofactor placement modify the interactions to a large extent. Our calculations have also shown that the midpoint reduction potentials of the terminal FeS cluster can be shifted by as much as 20% of the applied membrane potential and decreases exponentially with the cluster-membrane distance [3].

References:
[1] Zickermann et al, BBA - Bioenergetics 1787 (2009) 574-583.
[2] Sazanov et al, Science 311 (2006) 1430-1436.
[3] Couch et al, BBA- Bioenergetics 1787 (2009) 1266-1271.

poster

8

Possible pathway between alpha helical and beta helical structures of the C-terminal in the mammalian prion protein

Shan Dai
University of California, Davis

Daniel L. Cox

Purpose:
The normal form of the prion protein(PrPC) has mostly alpha-helical (AH) secondary structure in the C-terminal region (residues 166-230), while the infectious form (PrPSc) has been proposed to have a left-handed beta helical (LHBH) structure(1). The mechanism of conformational change from PrPC to PrPSc is unknown, but recent electron microscope data(2) and computer modeling(3) of in vitro grown prion fibrils suggest a possible LHBH structure in the C-terminal region. Therefore, we try to find a pathway between alpha helical and beta helical structures of the C-terminal region.

Materials and Methods:
We use high temperature (500K) AMBER molecular dynamics over 10 ns runtimes to study the unfolding transitions commencing from both LHBH and AH C-terminal starting structures. We also use OpenMM Molecular Dynamics Simulation Software with GPU to speed up the simulation process and compare with normal simulation on AMBER with CPU. After that,we further analyze our data using stability, contact map and reduced dimensionality trajectory analysis.

Results:
We find that both structures unfold on the nanosecond time scale to very similar AH-like conformations. Based upon comparison with all-atom structure based models (J. Singh and D.L. Cox, this meeting), we believe that we may be accessing a transition state between the conformations on the full unfolding pathway.

Conclusion:
The kinetic pathway found between alpha helical and beta helical structures of the C-terminal in the mammalian prion protein supports the idea that the C-terminal plays a role in conformation change of the prion protein from the cellular form to the diseased form. That both configurations access what appears to be a common transition state conformation on the time scale of nanoseconds suggests that the kinetics of the conformational change are reasonably fast.

References 1) Govaerts, C., et al. (2004) PNAS 101, 8342–8347.
2) Tattum, M. H., et al. . (2006) J. Mol. Biol. 357, 975–985
3) Kunes, K. C., et al. (2008) Prion 2, 81-90.

poster

9

Looking For LBH

Youval Dar
University of California, Davis

Jonathan Lawton
Daniel Cox
Rajiv Singh

Purpose:
Accumulation of amyloid protein fibrils in the brain or body is associated with many diseases, and can play a functional role in some organisms. Some believe that in the case of misfolded Prion protein, the insoluble protein fibrils form a left hand beta helical (LβH) structure. There are only few known proteins with an LβH structure which renders the commonly used homology driven thread predicting programs inefficient. Finding a possible thread of protein sequence onto LβH structure can be a time consuming problem. We've developed software that produces a space of 'Best' LβH threads in time order of one minute and scans a protein sequence for regions that might be sensitive to LβH formation, in time of order 10 minutes.

Materials and Methods:
Our software is based on a modified dynamic programming (DP) with a scoring method that derived using physical ideas and optimized using the known LβH proteins. The DP method is modified in a way that does not totally retain the idea of an additive score. Scoring is based upon local pseudo-energetics (which includes hydrophobicity, polarity, charged residue and volume packing) and calculates scores with some non-local considerations (including side-chain-to-side-chain bonding). Parameters weights optimization is done by minimizing the score difference between the scores of known LβH threads drawn from the PDB to scores of predicted threads. Threads produced by our program are then put to the test of molecular dynamics (MD) simulations for structural stability testing.

Results:
The program is used in our group to investigate wide range of proteins related to neurodegenerative diseases and amyloids. As a proof of principle, we have examined in detail LβH threads we produced for the known amyloidogenic yeast prion protein Ure2. The resultant threads were tested with AMBER all atom explicit solvent MD and via implicit solvent GROMACS MD via OpenMM. Both stability runs showed ~2.5 Å RMSD nearly constant over 10 ns of simulation time, with a visibly cohesive structure, which is an indication of good structure stability. We are investigating many other known, and large, amyloidogenic proteins. In particular, we have threaded the C-terminal of the Prion protein (Scores well), Curli (Scores poorly) and the CPEB protein speculated to play, via amyloid conversion, a molecular role in memory formation (this also scores well).

Conclusion:
We have developed a fast, specialized program for predicting potential LβH structure from sequence which works very well in a test case proof-of-principle (Ure2). Tests are ongoing for many other amyloidogenic proteins.

poster

10

Unbiased Reconstruction of a Mammalian Transcriptional Network Mediating Pathogen Responses

Yoni Donner
Stanford University

Purpose:
To develop a systematic strategy to reconstruct a mammalian transcriptional network and apply it on the network mediating response to pathogens

Materials and methods:
We profiled gene expression in dendritic cells (DCs) after stimulation with five pathogen-derived components and chose candidate regulators whose expression changed in response to at least one stimulus as well as a set of gene which captures most of the entropy in the gene expression patterns. We knocked down each of the regulators using shRNA and profiled the expression of the gene set. We used these measurements to construct a transcriptional network.

Results:
Our approach revealed the regulatory functions of 125 transcription factors, chromatin modifiers, and RNA binding proteins, which enabled the construction of a network model consisting of 24 core regulators and 76 fine-tuners that help to explain how pathogen-sensing pathways achieve specificity.

Conclusion:
Our work establishes a broadly applicable, comprehensive, and unbiased approach to reveal the wiring and functions of a regulatory network controlling a major transcriptional response in primary mammalian cells.

References:
Ido Amit et al. Unbiased Reconstruction of a Mammalian Transcriptional Network Mediating Pathogen Responses. Science 326, 257 (2009); DOI: 10.1126/science.1179050

poster

11

In Silico Functional Profiling of Human Disease-Associated and Polymorphic Amino Acid Substitutions

Uday S Evani
Buck Institute for Age Research

Matthew Mort
Vidhya Krishnan
Peter H. Baenziger
Predrag Radivojac
Sean D. Mooney

In Silico Functional Profiling of Human Disease-Associated and Polymorphic Amino Acid Substitutions

Matthew Mort, Uday S. Evani, Vidhya G. Krishnan, Peter H. Baenziger, Brandon Peters, Kishore Kamati, Rakesh Sayeth, Yanan Sun, Bin Xue, Eunseog Youn, Nigam Shah, Maricel Kann, David N. Cooper, Predrag Radivojac, Sean D. Mooney

Buck institute for Age Research, Novato, California and School of Informatics, Indiana University, Bloomington, Indiana. smooney@buckinstitute.org

Purpose:
An important challenge in translational bioinformatics is to understand how genetic variation gives rise to molecular changes at the protein level that can precipitate both monogenic and complex disease. We have co-opted a range of bioinformatic tools, designed to predict structural and functional sites in protein sequences, to the task of ascertaining whether intrinsic biases exist in terms of the distribution of different types of human amino acid substitutions (AAS) with respect to their structural, functional and pathological features.

Materials and Methods:
We applied these tools to compiled datasets of human disease-associated AAS in the contexts of inherited monogenic disease, complex disease, functional polymorphisms with no known disease association, somatic mutations in cancer, and neutral
polymorphic AAS.

Results:
The analysis revealed marked similarities in terms of the distribution of structural and functional sites between monogenic disease mutations and functional polymorphisms, with a bias toward those variants that impact protein function via structural disruption (P=3.8X10-24). Putative causative variants in both complex disease and cancer were significantly over-represented in intrinsically disordered regions (P=8.83X10-56) whilst cancer-associated mutations were enriched at certain molecular recognition sites (P=1.6X10-3).

Conclusions:
We postulate that missense mutations in complex disease and cancer are more likely than monogenic disease to impact on protein function directly through disruption of functional sites (e.g. protein interaction) rather than indirectly via structural disruption. Further analysis of subtypes of inherited disease (e.g. cardiovascular disease) served to identify several disease entities that differed significantly in terms of the distribution of specific causative molecular changes. For example, blood coagulation disorders were found to exhibit a 19-fold depletion in AAS at O-linked glycosylation sites. In overall terms, however, the disruption of a specific molecular function does not constitute a disease-specific phenomenon.

References:
Stenson PD. The Human Gene Mutation Database: 2008 update. Genome Med. 2009 Jan 22;1(1):13

poster

12

Developing an Accurate Gold Standard for Visual Similarity in a CBIR System for Radiological Images

Jessica Faruque
Stanford University

Daniel Korenblum
Christopher Beaulieu
Daniel Rubin
Sandy Napel

Purpose:
Content-based image retrieval (CBIR) has a wide range of applications to decision support systems for clinical diagnosis and personalized medicine. Evaluation of CBIR requires an accurate "gold standard" of image similarity for each clinical imaging application of interest. We approach developing these gold standards by obtaining information from readers about visual appearance of specific image features, thereby generating image feature similarity matrices.

Methods and Materials:
We have developed a web-based tool that allows users to rate a list of features for a set of images, where the feature list can be custom tailored to any clinical application (Fig. 1*). We applied this to a set of CT portal venous phase images of liver lesions, acquiring 5-point ratings for these features: (1) number of separable components, (2) average density, (3) margin definition (4) margin contour, and (5) rim density (Fig. 2*). We generated pairwise similarity matrices for each feature and reader by subtracting the absolute difference in ratings from 5. We also created a function allowing users to create gold standard matrices by selecting different methods and weights for combining the feature ratings. Finally, we trained and tested a set of weights to map the similarity matrices to a matrix evaluating overall lesion similarity, created independently of our 5-feature data.

Results:
We obtained data from 5 readers (3 radiologists, 2 non-radiologists), and compared differences between the ratings and the resulting similarity matrices (Fig. 3*). The standard deviations (averaged over all images) between readers' ratings for features (1), (2), (3), (4), and (5) were .33, .32, .79, .66, and .18, with maximum values 1.52, 1.34, 1.87, 1.52, and .89 points, respectively. The standard deviations in similarity matrices for features (1), (2), (3), (4), and (5) were .52, .47, .82, .77, and .31, with maximum values 1.79, 1.64, 1.95, 1.82, and 1.73 points, respectively. In training to an independent gold standard, the percentages of image pairs that differed in 0-1, 1-2, 2-3, 3-4, and 4-5 points were 48%, 34%, 12%, 5%, and 1%, respectively.

Conclusion:
These preliminary results show reasonable agreement with an independent gold standard, and moderate inter-reader variability. Variability may be reduced by including more readers, refining the feature list, and user training. Since viewers see each image only once, image similarity matrices can be created in time proportional to the number of images, allowing construction of large databases. With these efforts and further verification, our method will create accurate "gold standards" for content-based image retrieval systems for a variety of clinical applications.

*See Project Website.

poster

13

PharmReduce: A Symmetric Framework to Rank Pharmacogenes, Phenotypic Effects, and Small Molecules

Guy Haskin Fernald
Stanford University

Nicholas P. Tatonetti
Russ B. Altman

Computational prediction of pharmacogenes, drug effects, and drug uses are critical tasks which promise to speed the speed the pace of translational research. Previous attempts to make these predictions have been limited in scope and ability to make predictions accurately. Recently, data sources of gene networks, chemical effects, and pharmacogenes have grown sufficiently robust to allow latent patterns to be revealed using automated predictive methods. Here we present PharmReduce, a symmetric framework that builds on the relationships between genes, small molecules, and phenotypic effects to make rank list predictions of pharmacogenes for a drug, effects caused by a drug, or drugs to treat a disease.

poster

14

Modeling RNA structure from experimental results

Samuel C Flores
Stanford University

Samuel Flores
Yaqi Wan
Chris Bruns
Peter Eastman
Russ Altman

RNA structure is key to understanding function. But RNA is difficult to determine by crystallography and NMR, due largely to its charge and flexibility. Currently available computational methods only work well for small RNAs, due to poor scaling and the inability to incorporate available experimental information. RNABuilder constructs molecules using base pairing contacts which can be obtained by sequence alignment or relatively inexpensive experiments, as well as structure fragments from one or more related molecules. It scales well, promising applicability to larger molecules.

poster

15

Contributions from muscles and passive dynamics to swing initiation at different walking speeds

Melanie D. Fox
Stanford University

Scott L. Delp

Purpose :
Many children suffering from cerebral palsy, a neuromuscular condition resulting in impaired motor control and development, are unable to flex their knees sufficiently while walking. This common ambulatory disorder, known as stiff-knee gait, can cause falls, limit activity, and lead to compensatory motions that degrade health. Even after surgical treatment, many of these children remain "stiff" or even sustain further immobility [1], partly because one specific treatment does not address all of the possible causes of the condition in different individuals. To develop a scientific framework for determining the cause of an individual's stiff-knee gait, it is first necessary to understand how unimpaired children successfully coordinate muscles and passive dynamics to accelerate the knee into flexion during double support, the crucial period of gait before swing phase when both feet are on the ground. The goals of this study were to quantify the contributions of individual muscles, gravity, and velocity-related forces to pre-swing knee flexion acceleration and to compare these contributions across multiple walking speeds, since many children with stiff-knee gait walk at slow speeds.

Methods:
We analyzed muscle-actuated simulations of eight unimpaired children, each walking at four speeds [2]. We quantified how much each contributor (muscles, gravity, or velocity-related forces) flexed or extended the pre-swing knee during double support [3]. A repeated measures analysis of variance identified whether individual contributions were significantly affected by walking speed.

Results:
In the simulations, contributions from passive dynamics and muscles varied systematically with walking speed. Pre-swing knee flexion acceleration was achieved primarily by hip flexor muscles on the pre-swing leg with assistance from biceps femoris short head (the uni-articular hamstrings muscle) and gravity. The hip extensors and abductors on the pre-stance leg and velocity-related forces opposed knee flexion during double support.

Conclusion:
The study reveals how muscles and passive dynamics coordinate to produce knee flexion acceleration in double support, a key requirement for achieving sufficient knee flexion in swing. This provides a scientific framework for determining the cause of an individual's stiff-knee gait, enabling the potential for improved treatment outcomes.

References:
1. Yngve D, et al., 2002 J Ped Ortho 22, 672-676.
2. Liu M, et al., 2008 J Biomech 41, 3243-3252.
3. Liu M, et al., 2006 J Biomech 39, 2623-2630.

poster

16

Cell Frequency Deconvolution: A Novel Method for Generating Cell-Specific Gene Expression Profiles of Disease

Alexander A. Gaidarski
Stanford University

Shai Shen-Orr
Atul Butte

Background:
Despite recent advances in genomics, the etiology of many debilitating conditions, such as Chronic Fatigue Syndrome and organ transplant rejection, continues to elude scientists and doctors. Analysis of gene expression using DNA microarrays provides the ability to characterize disease at the cellular level, but the heterogeneous mixture of cell types that comprise a given tissue generates variation that makes it difficult to resolve abnormal states and distinguish causative genes. Ideally, differential expression analysis should be carried out for each cell type in a tissue; however the sorting of cells is costly and often impractical.

Purpose:
Here I present a statistical methodology that uses public data to estimate the relative frequencies of each cell type in a tissue, from which the cell-specific gene expression profile of a disease can be generated using a recently established method called Expression Deconvolution1,2. I apply my method to whole blood samples from pediatric kidney transplant patients in order to characterize the differences between the stable and acute rejection states.

Materials and Methods:
(1) I draw representative gene expression profiles for each white blood cell type from publicly available data. (2) Given the whole blood gene expression in query samples, I model the proportion of each cell type using linear regression. (3) Using these cell frequency estimates, I estimate gene expression of each cell type in each sample using a second linear regression.

Results: I validate my cell frequency predictions by showing high correlation (0.888) to the cell frequency estimates measured by Complete Blood Count (CBC) and compare the genes I identified as differentially expressed to those identified when the cell counts were known.

Conclusion: The feasibility of Expression Deconvolution has been demonstrated for many diseases of the blood when the proportions of different cell types are known. My method allows Expression Deconvolution to be expanded to blood samples for which the cell frequency is unknown and eventually to other tissues, which are considerably more difficult to sort than blood. By expanding both the depth and the scope of disease profiling, these methods carry great potential to push our understanding of disease mechanisms to an unprecedented resolution.

References:

1. Abbas, AR., et al. Plos One. 2009.
2. Shen-Orr, S., et al. (In review) 2009.

poster

17

Image Segmentation Using Gabor and Prolate Spheroidal Functions

Maria C. Gonzalez
University of California, Davis

Purpose:
The segmentation or partition of an image into regions is important in many fields: medical imaging, remote sensing, etc. The goal in all such applications is to delineate similar or dissimilar regions. There are many mathematical techniques used to identify a visual transition. Some are based on detecting a discontinuity, and others are based on regional homogeneity.

Connected with this problem is the understanding of how the visual system works. From experimental work we know that part of the mammalian visual system responds strongly to different spatial frequencies and orientations. It behaves as a spatial frequency analyzer with limited bandwidth. To model this characteristic, many functions have been proposed –i.e, the Gabor function [1], and recently, wavelets functions.

Materials and Methods:
We are using wavelets based on log-Gabor transform and Prolate Spheroidal functions. For the log-Gabor, we have used the Matlab code from [2] to do the projections. For the prolate Spheroide functions [3],[4], I re-used the orientation strategy of the log-Gabor wavelets, but developed my own approach for the frequency scale given the mathematical differences. The filter bank for both cases uses 4 scales and 6 orientations. These two sets are designed in 1D and are extended to 2D.

Results:
Images for the output of each filter and the progressive sum of the outputs to reconstruct the image are presented for both Gabor and Prolate wavelet functions.

Conclusion:
Segmentation of images by two different sets of wavelets is presented. Mathematically, these functions are well localized in time and frequency domains, so there is great interest in applying them to fields where these characteristics play a role. The Gabor set is not independent and there is redundancy in the coverage. For the wavelet Prolate set, we have developed low pass and band filters with increasing cutoff and center frequency. The set has similar redundancy as the Gabor case.

References:
[1] Marcelja, S. “Mathematical description of the response of simple cortical cells”, J. Opt. Soc. Am., Vol. 70, No.11, November 1980.
[2] Kovesi, P.D., MATLAB and Octave Functions for Computer Vision and Image Processing., School of Computer Science & Software Engineering, The University of Western Australia. http://www.csse.uwa.edu.au/~pk/research/matlabfns.
[3] Slepian, D., “Prolate Spheroid Wave functions, Fourier Analysis, and Uncertainty-V: The discrete case”, The Bell System Technical Journal, Vol.57, No.5, 1978.
[4] Verma, T., S.Bilbao, and T.H.Y. Meng, “The digital prolate spheroidal window”, Proc. ICASSP-96, May 7-10, Atlanta, GA

poster

18

Optical deconstruction of parkinsonian neural circuitry

Viviana Gradinaru
Stanford University

Murtaza Mogri
Kim R. Thompson
Jaimie M. Henderson
Karl Deisseroth

Purpose:
Despite its known efficacy in the treatment of Parkinson's disease, the mechanisms of deep brain stimulation are still not well characterized. The development of new techniques for introduction of light-activated ion channels and pumps into excitable cells using gene therapy techniques ("optogenetics") has allowed unprecedented control of specific neuronal populations at millisecond timescales. We sought to examine the contribution of various circuit elements within the subthalamic nucleus (STN) to the behavioral effects of DBS using optogenetic techniques.

Materials and Methods:
The inhibitory rhodopsin eNpHR (a transmembrane Cl- pump activated by yellow light) was introduced into the subthalamic nucleus of hemiparkinsonian 6-OHDA rats using a lentiviral vector, targeting excitatory neurons with the CaMKIIκpromoter. An optical fiber linked with a recording electrode ("optrode") was then introduced and local inhibition was induced with 561 nm light. In a second set of experiments, the excitatory rhodopsin ChR2 (a Na+-K+ channel activated by blue light) was introduced to STN excitatory neurons and high frequency stimulation (HFS) with 473 nm light was carried out. In a third set of experiments, glial cells in the STN were targeted using GFAP::ChR2. In a final set of experiments, HFS of the STN was carried out in Thy1::ChR2 transgenic mice expressing ChR2 in projection neurons.

Results:
There were no demonstrable behavioral effects from either optical inhibition or HFS of neurons within the STN. Likewise, activation of glial cells produced inhibition of firing in the STN but no behavioral effects. HFS of both afferent fibers in the STN region and layer V cells in primary motor cortex in Thy1::ChR2 animals produced robust behavioral effects, equal or superior to those seen with electrical DBS.

Conclusions:
Excitation of cortical afferents to the STN reproduces the beneficial effects of electrical STN stimulation, ameliorating behavioral deficits in hemiparkinsonian rodents. Implications for understanding of basal ganglia function and future therapeutic options will be discussed.

References:
Gradinaru, V., Mogri, M., Thompson, K.R., Henderson, J.M., and Deisseroth, K. (2009). Optical deconstruction of parkinsonian neural circuitry. Science (New York, NY 324, 354-359.
Gradinaru, V., Thompson, K.R., and Deisseroth, K. (2008). eNpHR: a Natronomonas halorhodopsin enhanced for optogenetic applications. Brain cell biology 36, 129-139.
Zhang, F., Aravanis, A.M., Adamantidis, A., de Lecea, L., and Deisseroth, K. (2007a). Circuit-breakers: optical technologies for probing neural signals and systems. Nature reviews 8, 577-581.
Zhang, F., Wang, L.P., Brauner, M., Liewald, J.F., Kay, K., Watzke, N., Wood, P.G., Bamberg, E., Nagel, G., Gottschalk, A., et al. (2007b). Multimodal fast optical interrogation of neural circuitry. Nature 446, 633-639.

poster

19

Collaborative Targeted Maximum Likelihood Estimation, with an Application to Biomarker Discovery

Susan Gruber
University of California, Berkeley

Mark van der Laan

Purpose:
Medical researchers are often interested in identifying biomarkers associated with disease-related outcomes. Given a set of candidate biomarkers, we can ascertain the level of association with the outcome. We call these associations "variable importance parameters," and describe collaborative targeted maximum likelihood estimation (CTMLE), a novel methodology for obtaining unbiased, adjusted causal or variable importance estimates. (van der Laan & Gruber, 2009)

Materials and Methods:
Targeted maximum likelihood estimation (TMLE) (van der Laan & Rubin, 2006) is a general methodology for estimating many types of causal inference parameters and their variable importance analogs. The general procedure is to obtain an initial estimate of the data generating distribution, then fluctuate this initial estimate in a manner that reduces bias for the target parameter. Asymptotically TMLE achieves the Cramer-Rao lower bound when both the initial estimate and nuisance parameters related to the fluctuation are correctly specified. When only one of these is correct, target parameter estimates remain consistent and asymptotically normal.

CTMLE is a more sophisticated targeted estimation procedure that data-adaptively constructs a series of candidate TMLE estimators and uses cross-validation to select the best candidate with respect to the parameter of interest. CTMLE estimates are often more efficient than TMLE estimates, and more robust when estimating borderline identifiable parameters. We apply CTMLE to determine which among 26 candidate HIV mutations confer resistance to the anti-retroviral drug lopinavir. This question was considered in Bembom, Petersen, et.al (2009), using TMLE.

Results:
Here, CTMLE results are compared with previous estimates and with Stanford scores obtained from the Stanford HIV db scores database (http://hivdb.stanford.edu). CTMLE analysis correctly identifies the majority of mutations that confer resistance, and also identifies mutations believed to not confer resistance.

Conclusions:
CTMLE provides a data-adaptive machine learning approach to reliable, efficient estimation of variable importance and causal effect parameters.

References:
O. Bembom, M. Petersen, SY Rhee, W.J. Fessel, S.E. Sinisi, R.W.Shafter, M.J. van der Laan. Biomarker discovery using targeted maximum likeklhiood estimation: Application to the treatment of antiretroviral-restistant HIV infection. Statist. Med. 2009; 28:152-172.

M.J. van der Laan and D. Rubin. Targeted maximum likelihood learning. The International Journal of Biostatistics 2(1), 2006.

M.J. van der Laan and S. Gruber. Collaborative double robust penalized targeted maximum likelihood estimation. submitted to The International Journal of Biostatistics (July, 2009).

poster

20

Resequencing Human DNA Enriched for HAR Neighborhoods Reveals Evidence of Ongoing Biased Nucleotide Selection

Sol Katzman
University of California, Santa Cruz

Andrew Kern
Sofie Salama
David Haussler

Purpose:
The Human Accelerated Regions (HARs) of the genome are (~200bp) elements for which comparative genomic study has revealed evidence of conservation throughout vertebrate evolutionary history but that also have an unusual number of substitutions unique to the human lineage. To identify those HARs that have evolved to support some human-specific function we searched for evidence of a selective sweep by resequencing a 40kb neighborhood of each of 49 HARs in a set of 22 human chromosomal samples. To take advantage of the high throughput, multiplexed DNA sequencing technology of the ABI SOLiD platform while reducing costs, we enriched genomic DNA from barcoded samples by hybridization to a Nimblegen 385K Sequence Capture array.

Conclusion:
There is only modest evidence for selective sweeps in the 40kb neighborhoods of the top 49 HARs. This may be due to the decay of the signal since the time of any such sweep. By contrast, we find good evidence of a bias towards the selection of mutations from weak (A:T) to strong (G:C) basepairs in many of these regions, both as an ongoing as well as an historical phenomenon.

poster

21

A New Semi-Explicit Solvation Model: Fast Physics for Better Results

Charlie Kehoe
University of California, San Francisco

Christopher Fennell
Ken Dill

Purpose:
Computational physicists, chemists, and biologists have a critical need for better models of water and aqueous solutions. Detailed water models have been developed on the quantum, explicit (classical), and implicit (continuum) levels, depending on the levels of detail, accuracy, and performance needed. Various hybrid models have also been proposed, but we believe the current explicit/implicit combinations do not go far enough. Thus we present an exciting new solvation model called Semi-Explicit Assembly, which combines the speed of the fastest continuum models available with the strong physical basis and discrete water treatment afforded by explicit solvent simulations.

Materials and Methods:
We base our model on several simple physical properties of water as a solvent, collected directly from explicit solvent simulations (using both the TIP3P and SPC water models) for individual atomic solutes. We analyze these trajectories for water distances, orientational biases, and other important properties, and later use these statistics to construct putative configurations of water around arbitrary solutes. Our approach is purely physical and involves no fitting of parameters to data sets.

Results:
As a first test and application of our method, we compute solvation free energies based on dispersion and electrostatics, comparing our results to those of several popular methods. On a test set of 504 small molecules, our model executes as fast as the Generalized Born solvation model, but with substantially improved accuracy in prediction of experimental solvation free energies. Also, because of the structure of our model, improvements in simulation forcefields will improve our results as well.

Conclusion:
We present a fast, new solvation model with improved representation of the physics of water, resulting in improved accuracy without any increased computational cost. We emphasize that our results come without any artificial parameter adjustments; our model's properties are the same as those used directly in molecular dynamics. The energetic accuracy and detailed structural information we provide have wide-ranging implications for molecular modeling research.

References:
1. C. J. Fennell, A. Bizjak, V. Vlachy, and K. A. Dill. Ion pairing in molecular simulations of aqueous alkali halide solutions. J. Phys. Chem. B, 2009.
2. S. Rajamani, T. Ghosh, and S. Garde. Size dependent ion hydration, its asymmetry, and convergence to macroscopic behavior. J. Chem. Phys., 2004.
3. D. L. Mobley, C. I. Bayly, M. D. Cooper, M. R. Shirts, and K. A. Dill. Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations. J. Chem. Theory Comput., 2009.

poster

22

Ontology Based Annotation of Molecular Imaging and Contrast Agent Database

Kranthi Kode
Stanford University

Adrien Coulet
David Paik

Purpose:
The combinatorial nature of designing nanoparticles from their components created a huge space of possible nanoparticle compositions exceeding any individual's cognitive capacity [1-2]. Informatics tools are a natural fit for addressing this need by providing decision support in the analysis and design of nanoparticles. In particular, cancer nanotechnology, being a relatively new field, presents a unique opportunity for informatics-driven approaches to accelerate discovery and translation. We are indexing cancer nanotechnology informatics knowledge with Open Biomedical Resources (OBR) to create a powerful new toolset for information retrieval using biomedical ontologies.

Materials and Methods:
We are creating the new nanoinformation retreival toolset by (1) integrating new sources of nanoparticle data into Open Biomedical Resources (OBR) and (2) developing an interface for specifying nanoparticle structure to search amongst semantically-related nanoparticles. As part of the first step, we integrated Molecular Imaging and Contrast Agent Database (MICAD), a source of nanoparticle data with the OBR service since OBR contains only limited nanoparticle information needed by researchers in the cancer nanotechnology field. We used Ontology Based Annotator developed by National Center for Biomedical Ontology using NanoParticle Ontology to create sematic maps of the concepts extracted from free-text nanoparticle reports of MICAD database using Java based MicadAccessTool developed by us.

Results:
Currently, there are 725 free-text nanoparticle reports in the MICAD database. Using MicadAccessTool, we extracted the data from all the articles and categorized the data into 14 contexts allowing for the search results to point to the subsections of the corresponding articles. The extraction and annotatoin of the data from new articles that are added to MICAD is automated.

Conclusion:
The nanomedicine community has an immediate need for structured data on nanomaterials to understand their biological properties including nanomaterial toxicity and to design nanoparticles, nanodevices, and other materials with custom properties for specific biological applications. Ontology based annotation of databases like MICAD using ontology focusing on cancer nanotechnology allows us to create the structured data on nanoparticles that is specifically useful to biomedical community.

References:
[1] Ferrarri M, Cancer Nanotechnology: Opportunities and Challenges, Nat Rev Cancer 5:161-71, 2005.
[2] Ferrari M, The Mathematical Engines of Nanomedicine. Small 4(1):20-5, 2008.

poster

23

Using graphics processors for clustering of biological data sets

Kai J. Kohlhoff
Stanford University

Vijay S. Pande
Russ B. Altman

Purpose:
The acquisition and analysis of huge data sets is of central importance in modern day biology. Data analysis often requires the use of clustering algorithms to extract meaningful information from sources such as gene expression data and protein sequences and structures[1-4]. Sequential algorithms are often inadequate for large volumes of data as they can require substantial amounts of computer time. Hence, there are increasing efforts to harvest the parallel processing power of modern graphics processors (GPUs)[5, 6]. We have created a library of GPU-based clustering algorithms for the use in biological applications. Examples from the field of structural biology will be used to demonstrate how the achieved speed-up can improve data analysis.

Materials and Methods:
Clustering algorithms were implemented using the freely available programming API 'C for CUDA'[7]. Performance tests were carried out between sequential k-means and k-centers clustering on an Intel Xeon E5420 processor at 2.5 GHz, and parallel code executed on a single NVIDIA Tesla C1060 graphics multiprocessor at 1.3 GHz.

Two types of data were used for the tests: data vectors from the FEATURE analysis software for key biophysical and biochemical features of biological structures[8], and protein structures obtained by molecular dynamics runs on the simulation package Desmond[9].

Results:
We have implemented GPU versions of k-means, k-centers, and self-organizing maps clustering algorithms. While speed-ups are marginal for very small data sets, the algorithms achieve one to two orders of magnitude speed-ups for data sets involving in excess of 500,000 elements.

Conclusion:
The parallel processing power made available by the use of graphics processors can greatly enhance our abilities to analyze huge data sets. The speed-up seen in the algorithms presented here, for example, allows the computation of clusters in a fraction of the time, the use of much larger data sets, or the creation of analysis protocols that require frequent reclustering of the data.

References:
1. Andreopoulos B, et al, Brief Bioinform, 2009. 10(3): p. 297-314
2. Belacel N, Wang Q, Cuperlovic-Culf M, OMICS, 2006. 10(4): p. 507-31
3. Chodera JD, et al, J Chem Phys, 2007. 126(15): p. 155101
4. Zemla A, et al, Nucleic Acids Res, 2007. 35(22): p. e150
5. Shalom SAA, Dash M, Tue M, DaWaK 2008, 2008. LNCS 5182: p. 166-75
6. Wu R, Zhang B, Hsu M, 2009, ACM: Ischia, Italy
7. NVIDIA CUDA. Available from: http://www.nvidia.com/object/cuda_home.html
8. Liang MP, et al, Nucleic Acids Res, 2003. 31(13): p. 3324-7
9. Bowers KJ, et al, 2006, ACM: Tampa, Florida

poster

24

Automated Inference of Molecular Mechanisms of Disease from Amino Acid Substitutions

Vidhya G. Krishnan
Buck Institute for Age Research

Biao Li
Matthew E. Mort
Fuxiao Xin
Kishore K. Kamati
David N. Cooper, Sean D. Mooney, and
Predrag Radivojac

Purpose:
Due to the advancements in recent years in high-throughput sequencing, enormous volume of sequence and genetic variation data are available. This helps in analyzing huge datasets to understand the cause of diseases. Over the last decade, several computational methods and tools have been developed in order to classify amino acid substitutions that can affect protein function or disease causing or altering phenotypes. However, none of these methods help to understand the molecular mechanism underlying such effect or alterations.

Materials and Methods:
We collected five datasets of human amino acid substitutions from online databases and literature. These include substitutions data from cancer, kinase, HGMD and Swiss-Prot disease and putatively neutral. We generated broad range of attributes using protein sequence based on predicted protein structure, dynamics, functional properties, evolutionary information. To distinguish disease-causing substitutions from neutral we applied and compared Support Vector Machines (SVMs) and Random Forest (RF) classifiers.

Results:
We have developed a new computational model, MutPred, that is based on 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 a particular molecular mechanism responsible for the disease state. MutPred uses the established SIFT method but offers improved classification accuracy with respect to human disease mutations.

Conclusion:
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. 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.
The tool MutPred is available at http://mutdb.org/mutpred.

References:
1.Greenman,C. et al. (2007) Patterns of somatic mutation in human cancer genomes. Nature, 446, 153–158.
2.Mooney,S.D. (2005) Bioinformatics approaches and resources for single nucleotide polymorphism functional analysis. Brief Bioinformatics, 6, 44–56.
3.Radivojac,P. et al. (2008) Gain and loss of phosphorylation sites in human cancer. Bioinformatics, 24, i241–i247

poster

25

MD in Electronic Continuum: Making AMBER and CHARMM polarizable

Igor Leontyev
University of California, Davis

Alexei Stuchebrukhov

Purpose:
Molecular Mechanics, Molecular Dynamics, QM/MM, and other types of molecular simulations, have become an integral part of biological research. The success of such simulations depends critically on the quality of the molecular interaction modeling. It is well recognized that the electronic polarizability is an important factor in molecular interactions, and a significant effort is currently being made to develop electronically polarizable force fields. As fully polarizable models are being developed, there is also a growing interest in improving the existing empirical non-polarizable force fields to capture more accurately the effects of electronic polarization in MD simulations. Here we propose a simple empirical model which allows interpreting the concept of non-polarizable force fields and bridging the gap between completely non-polarizable and fully polarizable models.

Materials and Methods:
In our model, called MDEC, the electronic polarization is treated explicitly in terms of the electronic continuum (EC) approximation, while the nuclear dynamics is described by molecular dynamics (MD) simulations with a fixed-charge force field. In such a force field all atomic charges are scaled by the factor of about 0.7 to reflect the screening effect by the electronic continuum. A variety of MD simulations and quantum-mechanical calculations have been carried out to validate the model.

Results:
We demonstrate [1] that without accounting for the electronic polarization the electrostatic interactions between ionized groups, as given by AMBER or CHARMM force field, are exaggerated by a factor of about 2, while, dielectric constants and solvation free energies in low dielectric media are significantly underestimated. In contrast, MDEC model produces correct results, as shown in modeling of hydration free energies of ions [2], dielectric constants of neat alcohols, alkanes [3], protein interiors of Cytochrome c and Cytochrome c Oxidase [4] as well as non-equilibrium reorganization energies in water, dichloroethane, tetrahydrofuran and supercritical carbon-dioxide solvents [5].

Conclusions:
The mistreatment of the electronic polarization in the standard non-polarizable models, e.g. AMBER or CHARMM, is not noticeable in high-dielectric media, while, the approach completely fails in simulations of low-dielectric media. In contrast, the non-polarizable MD in the Electronic Continuum successfully models both high and low dielectric media [1-6] and can be considered as a low-cost alternative to fully polarizable MD algorithms.

References:
1. PNAS (2009) submitted.
2. JCP 119 (2003) 8038.
3. JCP 130 (2009) 085102.
4. JCP 130 (2009) 085103.
5. JPC B 110 (2006) 14950.
6. Chem. Phys. 319 (2005) 4.

poster

26

Dynamic DNA Damage Response Revealed by System Profiling of Cellular Signaling Network

Jia-Ren Lin
Stanford University

Karlene A. Cimprich

Purpose:
DNA damage response (DDR) simultaneously controls cell cycle progression, DNA damage repair, cell death and other cellular signaling pathway to protect cells from genomic instability and mutagenesis. Although DDR and its important regulators have been studied for years, the system overview on this complex response remains unclear. In addition, the integrated properties of DDR toward different types of DNA damage are largely unknown. Hence, we performed multiple-parameter probing of DNA damage signaling network by measuring protein phosphorylation, ubiquitination and stability to investigate the dynamic properties of DDR network.

Materials and Methods:
Human embryonic Kidney 293T (HEK293T) cells were used in this study. Cells are damaged by either UV-C (254nm) or methyl methanesulfonate (MMS) then lysed at indicated time point. The quantitative western blotting was performed on FluorChem HD2 (Alpha Innotech). The antibodies against specific proteins (cdc25a, cyclinD1, Rad18, RPA34, ATR and Rad1) or phosphorylation sites (pERK1/2, pc-Jun, pChk1, pRad17, pATM and pKAP1) were used in this study. The data analysis was done by Matlab and Microsoft Excel.

Results:
Time-course and dose-response of UV or MMS induced signaling pathways were captured by twelve different record outs, including cell-cycle regulators (cdc25a, cyclinD1), DNA repairs proteins (Rad18, Rad17, RPA, KAP1), checkpoint proteins (Chk1, ATR, ATM) and MAP kinase signaling (pc-Jun and pERK1/2). Over 800 readings were gathered and normalized. The principle component analysis was used to identify the primary responders for UV or MMS damage. In addition, the Bayesian network analysis reveals the functional connection between these DDR effectors. Finally, the time-series analysis was applied to find out the possible regulatory circuit of DDR signaling network.

Conclusion:
Simultaneous profiling of DDR network provides a novel, integrated view of system dynamics toward different DNA damage. This approach reveals how cells coordinate cell cycle regulation, DNA repair/checkpoint machinery and other cellular signaling pathways. Furthermore, this approach can be easily expanded to unexplored components of DDR system by incorporating more specific antibodies. Our future direction will be to apply specific perturbations (ex: siRNA and kinase/phosphatase inhibitors) to understand more about the connectivity and robustness of DDR network.

References:
The DNA damage response: ten years after. Harper JW, Elledge SJ. Mol Cell. 2007, 28(5):739-45.

poster

27

Crossover Breakpoint Detection with High Density SNP Markers in Three Generation Tri-Trio Pedigrees

Janet Y. Luo
University of California, Berkeley

Jack S. Chen
Mariel Vazquez
Christopher Lin
Allen M. Chen
Rainer K. Sachs

Purpose:
Recombination plays a crucial role in meiosis; locating recombination breakpoints is important in linking genetics to diseases. Because of different experimental designs and only sparse STS markers, previous experiments have detected crossover breakpoint regions with fairly low resolution.

We propose a combinatorial method using dense SNP markers in three-generation pedigrees to determine gene flow in grandchildren. By analyzing each SNP case by case, crossover breakpoints can be rapidly identified in the parental meioses with high accuracy. In addition, this program can correctly phase on average 94% of SNPs.

Materials & Methods:
We analyzed 5324 SNPs in twelve CEPH pedigrees on a 10 Mb region of chromosome 20[1], consisting of a paternal grandfather (F), paternal grandmother (M), maternal grandfather (f), maternal grandmother (m), two parents, and one child, creating a "tri-trio" structure.

Because SNPs are bi-allele, each SNP corresponds to one of the 2^10=1024 possible tri-trios. For each tri-trio, four combinations of gene flows are possible: (F,f), (F,m), (M,f), and (M,m). After deducing the possible gene flows for our given SNP, we minimized the number of crossovers, and detected the breakpoint regions to be bounded by the SNPs exhibiting a switch in gene flow from its previous informative marker.

Results:
Application of the algorithm on our data detected all eight recombination breakpoints previously found with microsatellite studies in 25 female meioses with ~100 times higher resolution, including one new, small double crossover. The median length of the crossover regions was 46 kb. No crossovers in the 25 male meioses were found. Furthermore, our algorithm, sensitive in detecting errors, identified a mislabeled grandson as a paternal grandfather in CEPH pedigree 1347 here and in published chromosome 22 data, dramatically reducing the unusually high rates of recombination and mutation previously reported in this pedigree[2].

Conclusion:
We have introduced a method that has empirically detected crossover regions at a higher resolution than in previous experiments. For future studies, this method can be applied to complex pedigrees, and other data, e.g. the entire human genome, or species such as mice or Drosophila.

References:
[1] Ke, X et al., The impact of SNP density on fine-scale patterns of linkage disequilibrium. Hum Mol Genet. 2004.
[2] Dawson, E. et al., A first-generation linkage disequilibrium map of chromosome 22. Nature. 2002.

poster

28

"In silico" mapping of liver iron levels in inbred mice

Stela Masle
University of California, Berkeley

Seung-Min Lee
Chris Vulpe

Purpose:
Both iron deficiency and iron excess are detrimental in multiple organisms including mammals. Previous studies in both mice and people suggest that genetic variation may influence iron status in mammals and susceptibility to these conditions. However, these genetic factors are not well defined.

Materials and Methods:
To address this issue, we measured basal liver iron levels and plasma diferric transferrin in 18 inbred strains of mice of both sexes on a defined iron diet. Large sets of publicly available genotyping data on number of inbred strains made performance of direct association studies on measured phenotypes by "in silico" mapping possible. Recent genome wide association study demonstrated the robustness of this method comparing the results with previously detected quantitative trait loci (QTL) from classical linkage analyses. We therefore carried out a genome-wide association mapping approach to identify haplotypes underlying differences in liver iron levels in subset of 14 inbred strains, for which genotype information was available. "In silico" haplotype association mapping (HAM) was preformed using web based program SNPster, based on method described by Pletcher et al., 2004 and McClurg et al., 2007. Briefly, inferred 3-window haplotype groups are used as independent factor in one-factor ANOVA. For each phenotype F-statistic is calculated. Significance level was estimated from background distribution which is simulated non-parametrically using 1 million bootstraps of phenotype. Log transformation was used to improve the normality of the data. Weighted bootstrapping is used to account for possible presence of population structure between the strains.

Results:
We found ~4 fold variation in liver iron in males (lowest 153.21 µg/g, highest 661.13 µg/g) and ~3 fold variation in females (lowest 222.12 µg/g, highest 658.66 µg/g). Using "in silico" approach we identified 14 regions which exceeded association score (AS) of 2.5 (median p-value ‹0.003) for iron levels; two resided in regions known to influence iron metabolism and 12 were novel. Most promising regions include 0.22 Mb region on chromosome 7, in which Adam12 is the only candidate gene, 0.32 Mb region on chromosome 11 with Gria1 gene and 0.74 Mb region on chromosome 15 with Trps1 gene.

Conclusion:
All identified quantitative trait loci (QTL) regions are good candidates for further refinement and subsequent functional studies.

References:
Pletcher et al., PLoS Biol, 2004; McClurg et al., Genetics, 2007

poster

29

Detection of allele-specific mRNA transcripts through an integrative analysis of genomic, Expressed Sequence Tags and Exon array data

V. Nembaware
University of Cape-town

B. Lupindo
K. Scheffler,
C. Seoighe

Background:
Accurate mRNA splicing depends on multiple regulatory signals encoded in the transcribed RNA sequence. Many examples of mutations within human splice regulatory regions that alter mRNA splicing qualitatively or quantitatively have been reported. Allelic differences in mRNA splicing are likely to be a common and important source of phenotypic diversity at the molecular level, in addition to their contribution to genetic disease and pharmacogenetics. However, because the impact of a mutation on the efficiency of mRNA splicing is often difficult to predict, many mutations that cause medically relevant phenotypes through an effect on splicing are likely to remain undiscovered.

Data and Methods:
We have combined a genome-wide scan for sequence polymorphisms likely to affect mRNA splicing with evidence from publicly available Expressed Sequence Tag (EST) and exon array data. The genome-wide scan uses published tools for SNPs (Single Nucleotide Polymorphisms) located within donor and acceptor splice sites, branch points and exon enhancer elements. We developed a novel probabilistic method to infer allele-specific splicing from EST data. The method makes use of SNPs and alternative mRNA isoforms that we mapped to EST sequences to model both regulated alternative splicing as well as allele-specific splicing.

Results:
From the genome scan, we identified 30,977 candidate splicing polymorphisms. For 1,085 candidate splicing polymorphisms, the difference in splicing between alternative alleles was corroborated by publicly available exon array data from 166 lymphoblastoid cell lines. We report a set of genes showing evidence of allele-specific splicing from the integrated analysis of SNPs, EST and exon array data including several examples for which there is experimental evidence of polymorphisms affecting splicing in the literature.

Conclusion:
Our results provide an extensive resource that can be used to assess the possible effect on splicing of human polymorphisms located in putative splice-regulatory sites.

References:
Huang RS, Duan S, Bleibel WK, Kistner EO, Zhang W, Clark TA, Chen TX, Schweitzer AC, Blume JE, Cox NJ, Dolan ME (2007) A genome-wide approach to identify genetic variants that contribute to etoposide-induced cytotoxicity. Proc Natl Acad Sci U S A, 104: 9758-9763.
Nembaware V and Seoighe C (2009) Allele-specific splicing in human. Encyclopedia of Life Sciences (accepted).
Nembaware V, Lupindo B, Schouest K, Spillane C, Scheffler K and Seoighe C (2008) Genome-wide survey of allele-specific splicing in humans. BMC. Genomics 9: 265.
Wang GS, Cooper TA (2007) Splicing in disease: disruption of the splicing code and the decoding machinery. Nat Rev Genet, 8: 749-761

poster

30

Different Genomic signatures associated to Estrogen Receptor (ER) status in breast cancer patients

Alex Pankov
San Francisco State University

Daniel DeWoskin
Javier Arsuaga

Purpose:
Breast cancer is the second most common occurring form of cancer for women and is diagnosed in 1 in 8 women. Therefore, it is essential to properly diagnose the risk of cancer recurrence. However, because of a lack of indicators of patients' risk level, many women are unnecessarily exposed to aggressive chemotherapy treatment [1]. An assessment of various predictive indicators would ideally pave the way for a more personalized healthcare plan which would work to help each individual patient.

An approach to study this problem, that complements current expression profiling, is to look at the genomic alterations. These genomic aberrations can be detected through the use of array comparative genomic hybridization (aCGH) technology [1].

Materials and Methods:
We analyzed published aCGH data from 180 Patients that had been diagnosed with early stage, lymph-node negative breast cancer [2]. From these patients a total of 89 received anthracycline-based chemotherapy after surgery and 91 did not and 98 were estrogen receptor positive and 59 estrogen receptor negative. We compared the results obtained with the method Multidimensional Array CGH with other currently used methods such as combining circular binary segmentation and ANOVA. In order to use circular binary segmentation and ANOVA for performing association studies we first removed outliers, detected change-point locations, and segmented the data into regions of equal copy number [3].The data was then compared, probe by probe, between ER+ and ER- patients.

Results:
The preliminary results of ER + patients using MDACGH show that the regions 3q, 6p, and 13q are associated with ER status. Interestingly when we apply circular binary segmentation followed by ANOVA we find the strongest evidence this association is on chromosomes 4p, 15q, and 16p.
Conclusion

Different results found with MDACGH and other already established algorithms emphasizes the new features of MDACGH. By incorporating the results obtained here with those from other algorithms we aim at obtaining a more comprehensive view of which genes are associated with ER status and breast cancer.

Referrences:
[1] Climent J, Garcia JL, Mao JH, Arsuaga J, Perez-Losada J. Characterization of breast cancer by array comparative genomic hybridization. Biochem Cell Biol. 2007;85(4):497-508.
[2] Climent, J., Dimitrow, P., Fridlyand, J., Palacios, J. and Martinez-Climent, J.A. Deletion of chromosome 11q predicts response to anthracycline-based chemotherapy in early breast cancer. Cancer Res. 2007 67:818-826. PMID: 17234794.
[3] Olshen, A., Venkatraman, E., Lucito, R. and Wigler, M. (2004). Circular Binary Segmentation for the analysis of array-based DNA copy number data. Biostatistics 5 557-572.

poster

31

Warfarin Dose Prediction Incorporating Error Estimates and Warfarin Dose Interval Calculation

Jacob Stuart Porter
University of California, Davis

Purpose:
The anticoagulant Warfarin responds to genetic variability of the VKORC1 and the CYPC29 genes. A dosing algorithm that captures the main trend in a large dataset has been produced, but a dosing interval that establishes a reasonable interval of doses has not been computed. This interval can be used to establish a follow-up dose if the initial dose fails to stabilize. Additionally, a statistical test that incorporates parameter error estimates or that accounts for ambiguities in the definition of parameters may produce a more accurate dosing algorithm.

Materials and Methods:
The dosing interval and the statistical test were produced from a dataset of over 3500 people, and validated from a dataset of over 800 people. Validation tests were run on the intervals by checking if the real dose is within the interval, and by checking how far away from the ends of the interval the real dose is. The new statistical tests are compared to other statistical tests for accuracy.

Results:
The pharmacogenetic algorithm is more accurate than an algorithm based on clinical data alone. Incorporating errors in the measurements should be more accurate. A predictive dosing interval was computed that estimates the dose well and incorporates the main trend.

Conclusion:
An interval as well as a main trend-line provides a clinician with a range of doses to use if the initial dose fails to be stable for a patient. This is an improvement on both the fixed-dose approach and the simple trend-line pharamcogenetic approach. Incorporating error estimates gives an alternative dosing algorithm to existing algorithms.

References:
Estimation of the Warfarin Dose with Clinical and Pharmacogenetic Data by The International Warfarin Pharmacogenetics Consortium in The New England Journal of Medicine Vol. 360 No. 8 pp. 753 - 764
The element-wise weighted total least-squares problem by Markovsky et al. Computational Statistics and Data Analysis 50 (2006) 181-209
An Introduction to Total Least-Squares. P. de Groen. Nieuw Archief voor Wiskunde, Vierde serie, deel 14, 1996, pp. 237-253
Statistical Modeling, Analysis, and Management of Fuzzy Data edited by Carlo Bertoluzza, Maria A. Gil, and Dan A. Ralescu
Dosing Predictions for the Anticoagulant Warfarin by Ding et al to appear in the Fifteenth Industrial Mathematical and Statistical Modeling Workshop for Graduate Students published by the Center for Research in Scientific Computation of North Carolina State University

poster

32

Automatic detection of metastatic cells in the blood

Jesse M. Rodriguez
Stanford University

Ashley Powell
Serafim Batzoglou
Stefanie Jeffrey
David Paik

It is well-known that epithelial tumors metastasize by shedding cells into the bloodstream. These circulating tumor cells (CTCs) are of epithelial origin and can be easily distinguished from other cells in the blood stream from their surface markers. Studies have shown that the number of circulating tumor cells in patients' blood is correlated with patient survival, but the number alone is not a sufficiently accurate predictor. Current research is aimed at molecular characterization of these cells with the hope of both learning about their biological function and to improve prognostic accuracy.

Technology has been recently developed to isolate living CTCs using magnetic microbeads, but it currently limited by long hours of manual identification and extraction of cells from microscopy slides. This makes it difficult to acquire large samples necessary to study CTCs and reduces the technology's feasibility as a clinical diagnostic tool.

To address this, we have designed an automated machine learning method to identify cells in microscopy images against the background of non-specific magnetic microbeads. We achieve 96% precision and 91% recall in identifying cells with logistic regression using shape and size-based features of the cells and beads.

This method will improve throughput of isolating CTCs by allowing a fully automated robotic system to quickly identify and capture cells and will aid in the development of CTC-based diagnostic and monitoring protocols which will improve our understanding and treatment of metastatic cancer.

poster

33

Ovine Prion Polymorphisms Investigated by Threading to a Model Left Handed Beta Helical Structure Using Molecular Dynamics Simulation

Jamie F. Romnes
University of California, Davis

Daniel L. Cox
Rajiv R.P. Singh

We use AMBER all atom molecular dynamics (MD) to assess the stability of a model of the prion protein in its disease-causing conformation, PrP-sc. The model is based upon threading the ovine prion sequence onto a template left-handed beta-helical (LHBH) structure with 18 residues per turn. Five polymorphisms in the sheep prion protein, VRQ, ARQ, ARH, AHQ, and ARR, have been identified at residues 136, 154, and 171 respectively, which are roughly 18 amino acids apart which thus align approximately on the LHBH. Threading of the sequence was thus done with an emphasis on the locations of these special sites as a means to investigate their possible role in disease susceptibility as well as investigating the overall viability of the LHBH as a structural candidate for PrPsc. In comparison to known left handed beta-helical proteins, the resulting model for VRQ is shown in all atom MD to 10 ns to exhibit similar stability as indicated by a low root mean square deviation, the presence of substantial side-chain to side-chain hydrogen bonding, and volume packing fraction. Interestingly, and in corroboration with experimental data that it is a disease resistant variant, the same model for ARR exhibits much less stability. Each polymorphic site was also investigated individually by comparing results from models with only one site different and showed a good correlation to experimental data regarding the relation of the variants to disease susceptibility.

poster

34

An investigation of glutamic acid 242 as a proton pump valve in bovine Cytochrome c Oxidase using QM/MM Monte Carlo simulations

Benjamin M. Samudio
University of California, Davis

Vernon Couch
Alexei A. Stuchebrukhov

Cytochrome c Oxidase (CcO) is a mitochondrial inner membrane protein which catalyzes the reduction of oxygen to water and utilizes the free energy of this reaction to pump protons across the membrane from a lower concentration of protons (N-side) to a higher concentration of protons (P-side). This generates an electrochemical proton gradient which is ultimately used by ATP synthase to convert ADP to ATP. A key question is how CcO is able to maintain unidirectional translocation of protons across the membrane in the presence of this gradient. Glutamic acid 242 (bovine numbering) is a conserved residue in CcO which is found in the X-ray crystal structure to be a physical connection for protons from the N-side to the P-side of the membrane. It is hypothesized that Glu242 acts as a proton pump valve by delivering protons in one direction and preventing the backflow of these protons through protonation state dependent changes in its conformation. A model of CcO has been developed and the conformation space of Glu242 has been sampled using Monte Carlo simulations with energies calculated using the ONIOM QM/MM method. These calculations suggest a mechanism by which Glu242 facilitates unidirectional pumping and the prevention of proton leakage.

poster

35

A new SAEM algorithm for ordered-categorical and count data models: implementation and evaluation

Radojka Savic
Stanford University

Marc Lavielle, INRIA Saclay and University Paris 11

Background:
Analysis of categorical and count data from clinical trials using mixed effect analysis has recently become the method of choice. However, algorithms available for parameter estimation, including LAPLACE and Gaussian quadrature, are associated with limitations, including bias in parameter estimates as a consequence of approximations of the likelihood integral. This is amplified when the proportion of response categories are skewed- for ordinal data, and when models accounting for under- or over dispersion of individual variance compared to the mean are applied, in case of count data [1, 2]. The SAEM algorithm has proven to be a very efficient and powerful tool in the analysis of continuous data [3]. The aim of this study was to implement and investigate the performance of a new SAEM algorithm for discrete data.

Methods:
A new SAEM algorithm was implemented in MATLAB for estimation of both, parameters and the Fisher information matrix. Monte Carlo simulations were employed using scenarios as in previous studies with other algorithms [1, 2]. For ordered categorical data, the proportional –odds model was explored. For count data, six different probability distribution models, (i.e., Poisson, Zero-inflated Poisson, Generalized Poisson, Poisson with Markovian Features, Poisson with a mixture distribution for individual observations and Negative binomial models) were used. Performance of the algorithm was assessed by computing the relative bias (RB), root mean square error, and assessing the CPU time of the analysis. The accuracy of standard errors (SE) estimates was assessed as an absolute distance (AD) between actual and empirical relative SEs.

Results:
For proportional-odds model, RB was ‹ 8.13 % for all scenarios explored, including ones with skewed distributions of response categories. For count data models, AVRB was ‹ 4.13 % for all models studied including ones accounting for over- or under-dispersion. Estimates of standard errors were close to the empirical SEs, with AD ‹ 5.8% , for all explored scenarios. The longest CPU time was for the analysis of the Negative binomial model taking 40s for parameter estimation and 37s for SE estimation.

Conclusions:
The SAEM algorithm was extended for analysis of ordered categorical and count data with extensions to the Hidden Markov Model. It provides accurate estimates of parameters and standard errors. The estimation is significantly faster compared to other algorithms. The algorithm is implemented in Monolix 3.1.

References:
1. Jonsson S. et al, J Pharmacokinet Pharmacodyn, 2004, 31(4): p. 299-320.
2. Plan EL et al, J Pharmacokinet Pharmacodyn, 2009, 36(4): 353-66.
3. Kuhn E et al, Comput Stat Data Anal, 2005, 49 (4), 1020-38

poster

36

Isotropic MRI of the Healthy Shoulder with 3D-FSE-Cube: Preliminary Study

Lauren Shapiro
Stanford University

Ernesto Staroswiecki
Dr. Garry Gold

Purpose:
Conventional MRI is done with two dimensional fast-spin-echo (2D-FSE) resulting in anisotropic voxels, partial volume artifact, slice gaps and reformatting inability because it requires multiple acquisition planes. A newly devised sequence, three-dimensional fast-spin-echo (3D-FSE-Cube), allows for isotropic voxel acquisition which diminishes the drawbacks of 2D-FSE while reducing exam time 1-3. The shoulder, with its complex infrastructure and large range of motion is a valuable yet vulnerable joint, having a lifetime injury prevalence of up to 66.7% 4-5. This makes it an optimal region for MRI evaluation and research into technological advances. This study compares the clinical accuracy of 3D-FSE-Cube to 2D-FSE in evaluation of the healthy shoulder.

Materials and Methods:
Shoulders of 10 healthy subjects were imaged in the axial plane using an 8 channel shoulder coil. 3D-FSE-Cube images were acquired with TR/TE 2500/35ms, 384 x 288 matrix size, 0.6 mm sections, 20cm FOV, receiver bandwidth ±31.25kHz and ETL 60. 2D-FSE images were obtained in axial and oblique coronal planes with TR/TE 2500/35ms, 384 x 288 matrix, 3mm slices and 5mm gaps, 20cm FOV, ±31.25kHz receiver bandwidth and ETL 8. To allow noise measurements from identical noise images, both methods were also acquired with the RF pulse off. Regions of interest were placed in fluid, muscle and cartilage for both methods. ROIs placed in identical noise images calculated the standard deviation of noise. 3D-FSE-Cube SNRs were normalized for slice thickness. SNR and CNR were both compared with a non inferiority test. Images were paired, randomized and assessed on image quality, blurring and artifact by 2 radiologists. Ratings were analyzed with a non inferiority test.

(Expected) Results:
Cartilage, muscle and fluid SNR and fluid-cartilage CNR were similar between sequences. No significant differences will be seen in image quality, blurring and artifact comparisons. However, due to the reformation ability of 3D-FSE-Cube images, oblique and thin anatomy is able to be seen.

Conclusion:
3D-FSE-Cube allows for more efficient isotropic imaging of the shoulder compared to 2D-FSE. With decreased imaging times, 3D-FSE-Cube is better suited for patients in pain as well as claustrophobic and pediatric patients. 3D-FSE-Cube, with its ability to rapidly acquire thorough isotropic data, is a promising high-resolution MR imaging technique which may improve visualization of complex shoulder anatomy.

References:
1 Gold et al., AJR 2007; 188: 1287-1293
2 Stevens et al., Radiology 2008; 249
3 Yao et al., AJR 2007; 188: W199-W201
4 Meyers et al., S Afr Med J 1982; 403-5
5 Lumine et al., Scand J Rheumatol 2004; 73-81

poster

37

Filtering RNA decoys with small angle x-ray scattering and clustering analysis

Adelene Sim
Stanford University

Michael Levitt

Purpose:
RNA molecules have previously been regarded as "boring" molecules which merely relay genetic information from DNA to proteins. However, they are now known to also exhibit a wide range of gene regulation functions. Like proteins, the function of an RNA depends on its three dimensional structure. Here, we discuss how we can incorporate low-resolution experimental data (namely, small angle x-ray scattering or SAXS) to score the RNA models (also known as decoys) generated using existing RNA structure prediction tools [1].

Part of the problem of making use of low-resolution scoring functions is the effects of outliers. In some instances, these outliers fortuitously give good scoring results, while in others they are detrimental in the overall scoring scheme. Furthermore, for SAXS data, three-dimensional structure information is flattened out into a one-dimensional scattering profile, markedly emphasizing the effects of degeneracy, or the result of having multiple different structures with similar scattering profiles. In such a case, having outliers can further mar the efficacy of the scoring function.

Materials and Methods:
In order to ameliorate the effects of outliers, we attempted to minimize the amount of noisy data (models that are not often sampled) in our RNA dataset. Removing these noisy data gives us information about the true sampling features, and will further reduce the effects of sampling outliers on our low-resolution scoring function.
To do this, we conducted multiple independent k-means clustering on our RNA models, and studied their similarities. Decoys belonging to a well-sampled region would naturally be clustered together more often than decoys that are rarely sampled. Hence by picking out models that are often found in these cluster intersections (defined as cluster-cluster similarities between different clustering runs), we can effectively remove the poorly sampled data.

Results:
The combination of SAXS scoring and cluster analysis allows for effective scoring of RNA decoys. The cluster analysis also gives us information about the sampling efficiency of current RNA structure prediction tools.

Conclusion:
Clustering analysis accentuates sampling features and allows the effective removal of outliers to more reliably assess the efficacy of low-resolution scoring functions. Since the cluster analysis approach is not limited to RNAs, we can use the same tools to assess protein structure prediction tools. This will give us a better idea of their strengths and weaknesses.

References:
[1] Parisien, M. and Major, F., Nature, 452, 51 (2008)

poster

38

A Faster Measurement of Volumetric Breast Density from Magnetic Resonance Imaging Data

Lisa Singer
University of California, San Francisco

Nola Hylton
Catherine Klifa

Purpose:
Breast cancer is the most common non-skin cancer in women and the second most common cause of cancer death in women (1). Mammographic breast density is an independent risk factor for breast cancer (2). Our group has been interested in measuring breast density from magnetic resonance imaging (MRI) data and has shown a correlation between MRI and mammographic density (3); however, current MRI methods are time and labor intensive. This project aims to automate volumetric breast density (VBD) calculation from MRI data with the goal of facilitating its clinical use in risk assessment and treatment monitoring. In previous work, we showed that VBD measurements from coronal and axial MRI volumes were comparable. In this study, we aimed to develop a faster method of VBD measurement using coronal MRI volumes. Reduced analysis time and costs could facilitate measurement of MRI-based breast density in clinical care and research.

Methods:
MRI noncontrast-enhanced volumes were obtained from healthy pre-menopausal women. VBD was calculated using a method based on coronal volumes in which breast volume delineation was circumvented using techniques in image processing. Within the subsequently identified breast volume, fibroglandular tissue was segmented from fat using an algorithm based on fuzzy-c-means, allowing for the automated calculation of the fibroglandular tissue volume. Finally, VBD was calculated as a volumetric ratio of the fibroglandular tissue volume to total breast volume.

Results:
We simplified the step in VBD calculation of extracting the complete breast volume from MRI data. The number of slices requiring manual extraction of the breast volume from background noise was reduced from approximately 20-50 (dependent on the MRI data and user technique) to a total of two slices per volume and the need for manual delineation of breast contours was eliminated, reducing time required for breast volume delineation.

Conclusions:
We present a new technique based on coronal data that expedites VBD calculation. Further integration of the programming interfaces used in our laboratory would help reach our long-term goal of fully-automated VBD measurement, from the step of importing an MRI exam to the final step of calculating density. By reducing the time and resources needed for VBD calculation, faster methods of VBD calculation could facilitate clinical and research applications of MRI-based breast density measurement.

References:
1. Kopans DB. Breast Imaging. Baltimore, MD: Lippincott Williams & Wilkins, 2007.
2. Boyd NF, et al. Methods Mol Biol 2009; 472:343-360.
3. Klifa C, et al. Magn Reson Imaging 2009.

poster

39

Structure-based models for alpha-helical to beta-helical conformation change in the C-terminal of the mammalian prion protein

Jesse P. Singh
University of California, Davis

Daniel L. Cox
Paul C. Whitford

Purpose:
We employ all atom structure-based models with mixed basis contact maps to explore where there are any significant geometric or energetic constraints limiting conjectured conformational transitions between the alpha-helical (κH) and the left handed beta helical (LHκH) conformations for the C-terminal (residues 166-230) of the mammalian prion protein. The LHκH structure has been proposed to describe infection oligomers1 and one class of in vitro grown fibrils2,3, as well as possibly self-templating the conversion of normal cellular prion protein to the infectious form.

Materials and Methods:
The structure-based model uses GROMACS based molecular dynamics with a two-dimensional weighted histogram analysis method (WHAM) being applied to study projected energy surfaces. This model uses harmonic potentials to represent bond lengths, bond angles, improper dihedrals, and planar dihedrals. The native contacts of each conformation are used in the same Hamiltonian, maintained by a Lennard-Jones potential. Every other non-local interaction is given a sole repulsive term.

Results and Conclusion:
Our preliminary results confirm that the kinetics of the conformation change are not strongly limited by the large scale geometry modification, and evidence exists for a pathway linking the two conformations with a common folding intermediate, also suggested by all atom unfolding simulations4.

References:
1Govaerts C., et. al. Evidence for assembly of prions with left-handed beta-helices into trimers. Proc Natl Acad Sci USA 2004; 101; 8342-8347
2Tattum M. H., et. al. Elongated oligomers assemble into mammalian PrP amyloid fibrils. J. Mol. Biol 2006; 357; 975-985
3Kunes K., et. al. Left handed κhelix models for mammalian prion fibrils. Prion 2008; 2; 81-90
4See S. Dai and D.L. Cox, abstract elsewhere for this meeting

*Research supported in part by the International Institute for Complex Adaptive Matter, NSF Grant DMR-0844115

poster

40

A Novel Method for Scoring Candidate Genes in Association Studies: Application to Warfarin Response

Nicholas P. Tatonetti
Stanford University

Nicholas P. Tatonetti
Joel T. Dudley
Hersh Sagreiya
Russ B. Altman

A key challenge in pharmacogenomics is the identification of genes whose variants contribute to drug response phenotypes, which can include severe adverse effects. Pharmacogenomics GWAS attempt to elucidate genotypes predictive of drug response. However, the size of the available samples has severely limited their power and potential application. Improved statistics can result from aggregation of information about SNPs at the gene level. Given an identified gene, we can probe individual genotypes more deeply. We propose a novel approach for identifying genes impacting drug response. Our method characterizes the degree to which uncommon alleles of a gene are associated with drug response. We first use pre-existing knowledge sources to rank genes based on their likelihood of affecting drug response. We then define a summary score for each gene based on their minor allele frequencies and train classifiers to use these to predict drug response phenotypes. We validate our method on a warfarin GWAS data set from 181 individuals. We find that our method can increase the power of the GWAS and identifies both VKORC1 and CYP2C9, where the original analysis had only identified VKORC1. Additionally, we find that our method discriminates between low-dose (AUROC=0.886) and high-dose (AUROC=0.764) responders. Our method offers a new route for candidate pharmacogene discovery.

poster

41

Systematic identification of pathologic DNA variants in human mitochondrial disorders

Sreedevi Thiyagarajan
Stanford University

Curt Scharfe

Purpose:
Mitochondrial diseases are a group of clinical disease phenotypes caused by breakdowns in the cells power plants - the mitochondria. Defects in hundreds of mitochondrial genes encoded in the nuclear genome may be associated with these conditions. The currently known gene defects are mostly rare, display clinically similar or different diseases, and are limiting the medical genetics counseling of the affected individuals. Here we prioritize pathologic variants from a larger list of DNA variants, which we identified through re-sequencing of candidate genes in phenotyped disease populations.

Methods:
We utilized publicly available prediction tools including SIFT, PMut, and PolyPhen to predict the potentially damaging effects of non-synonymous DNA variants. To evaluate the predictive power of each algorithm, we tested their performance using information on mitochondrial gene mutations with known pathologic effects. A set of 107 clinically studied variants in the mitochondrial polymerase κgene (POLG; http://tools.niehs.nih.gov/polg/) and 23 neutral POLG variants from dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/) were used to evaluate the predictive power of the different algorithms for each POLG amino acid substitution.

Results:
PolyPhen predicted approximately 80% of the confirmed disease-associated POLG mutations with a false discovery rate of 10%, and SIFT predicted 71% of the mutations with a false discovery rate of 14%, while PMut identified 69% of the mutations with a false discovery rate of 15%. The false negative rates were higher for PMut (31%) followed by SIFT (27%) and PolyPhen(20%). In addition, our comparative performance analysis of the three methods showed a higher predictive power when combining PolyPhen and Sift predictions. A combined threshold of Polyphen scores ›1.3 or SIFT ‹0.1 resulted in a lower false negative rate of 0.12 (sensitivity = 88%).

Conclusion:
These results suggest that an integrative analysis using a combination of different computational methods can improve the sensitivity of predicting the potential pathologic DNA variants. Our approach is useful in the analysis of DNA variants in mitochondrial candidate genes identified through high-throughput re-sequencing technologies.

poster

42

Uncertainty Quantification in Blood Flow Simulations of Glenn Patients

Guillaume A. Troianowski
Stanford University

Alexandre Birolleau
Charles A. Taylor
Jeffrey A. Feinstein
Irene E. Vignon-Clementel

Purpose:
Congenital heart defects (CHD) affect ~1 in 100 newborns. While most of them can be treated without significant long term sequelae, some require complex surgical palliation/repair and alter the normal circulatory physiology. The single-ventricle defect is one such defect and leaves the child with only one operational ventricle, requiring the systemic and the pulmonary circulations to be placed in series through several operations performed during young childhood. Numerical simulations may be used to investigate the nature of the flow and its connection to post-operative failures and defects but heavily rely on boundary condition prescription. The goal of this paper is to present a precise method to address this question and study the impact of uncertainties in the input data on the results. The sensitivity of commonly used hemodynamic indicators to compare patients is discussed in this context.

Methods:
We constructed a patient-specific 3D model of the pulmonary arteries and superior vena cava from MRI data on five "Glenn" patients, aged 3 - 5 years. Using flow split information extracted from PC-MR images, and transpulmonary pressure gradients (TPG) from catheterization, we assigned as outflow boundary conditions a 3-element morphometric-based 0D electric analog. From this configuration, we introduced uncertainties at two levels (central venous pressure (CVP), flow split) and some variability in the method solution parameters to analyze the impact on several possible indicators (Energy Loss, Efficiency, Mean Wall Shear Stress (MWSS), Oscillatory Shear Index (OSI)) of outcomes.

Results:
Perturbation on the CVP appeared to be strongly correlated to the efficiency, while it left the flow distribution unchanged. Changes in the flow distribution introduced non-linear variations in the energy-loss and the efficiency of the geometries and, to a smaller extent, in the MWSS and OSI. Finally, the choice parameter in the method had a strong impact on both energy-loss and MWSS.

Conclusion:

These results suggest that the energy based indicators as well as the efficiency, both widely used to measure the "performance" of a geometry, are very sensitive to the input parameters and the methods used to incorporate them into the models. This emphasizes the necessity of defining robust indicators and methods, while taking into account the uncertainties of the clinically measured data.

poster

43

Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1-39)

Vincent A. Voelz
Stanford University

Gregory R. Bowman
Kyle Beauchamp
Vijay S. Pande

Purpose:
To date, the slowest-folding proteins folded ab initio by all-atom molecular dynamics simulations have had folding times in the range of nanoseconds to microseconds (1). We report simulations of several folding trajectories, each from fully unfolded states, of the 39-residue protein NTL9(1-39), which has a folding time of ~1.5 milliseconds.

Methods:
Molecular dynamics simulations on GPU processors were performed on the Folding@Home distributed computing platform (2). An AMBER forcefield with GBSA implicit solvation were used to generate ensembles of trajectories out to ~40 μs for several temperatures and starting states.

Results:
At a temperature less than the melting point of the forcefield, we observe a small number of productive folding events, consistent with predictions from a model of parallel uncoupled two-state simulations. The posterior distribution of the folding rate predicted from the data agrees well with the experimental folding rate (~640/sec) (3). Markov State Models (MSMs) built from the data shows a gap in the implied time scale, indicative of two-state folding (4,5). Structural analysis and transition path theory analysis of a 2000-state MSM shows a compact unfolded state with residual structure, and a great heterogeneity in mesoscopic substates and pathways consistent with several mechanistic models of folding, with the rate-limiting role of non-local strand pairing clearly seen.

Conclusion:
We plan to use this data to seed adaptive resampling simulations of metastable transitions in explicit solvent, which we believe will be a promising new method for achieving statistically converged descriptions of folding landscapes at longer time scales than ever before.

References:
1. Snow, C.; Sorin, E.; Rhee, Y.; Pande, V. Annual review of biophysics and biomolecular structure 2005, 34, 43-69.
2. Shirts, M.; Pande, V. Science 2000, 290, 1903-1904
3. Horng, J.-C.; Moroz, V.; Raleigh, D. P. Journal of Molecular Biology 2003, 326, 1261-1270.
4. Noé, F.; Fischer, S. Current Opinion in Structural Biology 2008, 18, 154-162.
5. Bowman, G. R.; Huang, X.; Pande, V. S. Methods 2009, in press.

poster

44

RNA expression patterns in *Coccidioides*: Using RNAseq to unravel a fungal pathogen

Emily Whiston
University of California, Berkeley

Ginger Jui
Thomas J. Sharpton
Chiung-Yu Hung
Garry T. Cole
John W. Taylor

Purpose:
Coccidioides spp., the causative agent of Coccidioidomycosis, is a dimorphic fungus, with a saprobic hyphal phase and a pathogenic spherule phase that infects mammals, including humans. There are two species of *Coccidioides*: *C. immitis* (found in California and Mexico) and *C. posadasii* (found in Arizona, Texas, Mexico and South America); both species have the same growth and disease phenotype. Interestingly, *Coccidioides*' closest relatives are non-pathogenic. Therefore, what is the genetic basis of adaptation, and specifically pathogenicity, in *Coccidioides*? Here, we present mRNA expression data to investigate this question.

Materials and Methods:
Recently, full genome sequences have been produced for *C. immitis* strain RS and *C. posadasii* strain c735. We have isolated mRNA from the hyphal and spherule growth forms of these 2 strains with 3 biological replicates for Solexa sequencing. We have compared transcriptional differences between growth forms and strains using an overdispersed log-linear model.

Results:
We have observed statistically significant increased expression in the hyphal growth form for both *C. immitis* RS and *C. posadasii* c735 for 269 genes. Of these, the genes with the most extreme differential expression include a fungal hydrophobin, and an acetyltransferase. We have also observed statistically significant increased expression in the spherule growth form for both *C. immitis* RS and *C. posadasii* c735 for 260 genes. Of these, the genes with the most extreme differential expression include RTM1, a putative major facilitator superfamily transporter, and a polysaccharide deacetylase. We have applied RNAseq data to a set of 119 previously identified vaccine candidate genes. Of these, 78 genes have detectable mRNA expression in all 3 bio-replicates of the spherule phase of both *C. immitis* and *C. posadasii*. Only one of the vaccine candidate genes has statistically significant higher expression in the spherule phase compared with the hyphal phase, and may be of particular interest in follow-up studies.

Conclusion:
As expected, we see major differences in gene expression between the hyphal and spherule growth forms of *Coccidioides*. These changes include many cell wall associated proteins, which is unsurprising considering the very different morphology and growth environments of the two *Coccidioides* growth forms. This data is also very helpful for the prioritization of vaccine candidates and drug targets for further studies in preventing and treating Coccidioidomycosis, which is potentially fatal in humans.

poster

45

Physics-based filtering of vessel wall motion from cardiac-gated 4D CT

Guanglei Xiong
Stanford University

Guanglei Xiong
Charles Taylor

Purpose:
Recent advances in cardiac-gated CT technique can provide useful dynamic information of vessel wall motion. However, the 4D (3D+t) measurement from this modality suffers from uncertainties due to limited spatial and temporal resolution as well as a variety of noises. On the other hand, the technique on physics-based simulation of vessel wall is able to generate motion data in enough spatial and temporal resolution. But it is difficult to obtain exact mechanical model and subject-specific wall mechanical properties for the model.
Our goal is to develop methods to let these two techniques benefit each other. (1) Physics-based correction of the 4D data can reduce noises that do not satisfy mechanical law. (2) Simulation with real measurements can estimate subject-specific wall properties that will make the simulation more reliable.

Material and Methods:
We proposed a physics-based state filter to achieve (1). The filter feeds the difference between the simulated and measured wall motion into the simulation of vessel wall. Therefore, the vessel wall motion is controlled by the influence of both measurements and the model. The weight of two influences depends on the relative uncertainty in them.

We proposed to employ ensemble Kalman filter to achieve (2). Again, the filter feeds the difference between the simulated and measured wall motion into the simulation of vessel wall. During the correction of the wall motion by the state filter, the Kalman filter gradually corrects the parameters of the model, i.e. wall properties. The promise of parameter correction is that the difference between the simulated and measured wall motion become smaller even without the state filter.

Results:
The two filters were tested on synthetic and real data. On the synthetic data, the state filter quickly corrects the wall motion and the Kalman filter is able to correctly estimate the wall properties even under relatively high noises. On the real data, although there is no ground truth, the state filter also fits the model to the measured motion. The wall properties by the Kalman filter are in the range of literature values. We also observe the trend of wall stiffening down the aorta.

Conclusion:
Our resutls demonstrate that the cardiac-gated CT and physics-based simulation can be combined to obtain more reliable vessel motion as well as estimate mechanical wall properties.

poster

46

Novel Shape Descriptors for Liver Lesions

Jiajing Xu
Stanford University

Chris F. Beaulieu
Daniel L. Rubin
Sandy Napel

We developed an efficient descriptor of the shape of an imaged object for use in content-based image retrieval of medical images containing lesions. We combined three distinct contour-based shape features: compactness, centroid distance signal, and multi-scale local area integral invariant (LAII). The distance between two images is defined as the weighted sum of the absolute difference between the three features. We tested this descriptor using a database of 72 portal-venous-phase CT liver images containing lesions classified by radiologists to be either round, lobulated or ovoid. A reference standard for similarity was established by setting pair-wise similarity to 3 if the classifications were identical and 1 if different. For each of the 72 query images, the remaining 71 images in the database were ranked according to distance from the query image in ascending order. Our computational shape descriptor is highly correlated with radiologists' annotation and could be used to retrieve images of lesions with similar boundaries, or combined with other features for more general types of similarity queries.

poster

47

Improving the prediction of regulatory SNPs using functional information

Yiqiang Zhao
Buck Institute for Age Research

Yiqiang Zhao
Matt Mort
David Cooper
Sean Mooney

Purpose:
Single nucleotide polymorphisms (SNPs) are the most common form of genetic variation and can effect gene expression, transcript processing and protein function. SNPs that are located in promoter regions might be functional by affecting regulation of gene transcription. However, identification of expression affecting SNPs is challenging for many reasons including our lack of quantitative models of cis acting regulation, linkage disequilibrium and an abundance of neutral variants. In this study, we are attempting to identify functional SNPs from non-functional SNPs in the putative transcription regulatory region (defined as 2500bp upstream the TSS and 500 bp downstream TSS) using supervised machine learning methods.

Materials and Methods:
We use a number of experimentally validated SNP features on set of 338 annotated functional SNPs from the Human Gene Mutation Database (HGMD) as well as random controls from dbSNP. We evaluated different machine learning schemes including Support Vector Machines (SVMs), Bayes Networks and Random Forest. The Random Forest algorithm, implemented in Waikato Environment for Knowledge Analysis (WEKA, was finally chosen with the generation of 100 trees on the basis of superior performance. Evaluation was performed using 10-fold cross-validation in all experiments.

Results:
We found the distance to transcription start site was of great importance. To test whether regulatory SNPs tend to occur within functionally important genes, we extend our approach by incorporating features of gene expression, codon usage and functional complexity. With the incorporation of this information, we achieved a receiver operator characteristics plot AUC of 89%, which is better than using SNP features alone.

Conclusion:
Our results suggested functional SNPs in the putative transcription regulatory region would be better predicted when both SNP and gene information were considered.

References:
Montgomery SB, Griffith OL, Schuetz JM, Brooks-Wilson A, Jones SJ. 2007. A survey of genomic properties for the detection of regulatory polymorphisms. PLoS Comput Biol 3(6):e106.
Torkamani A, Schork NJ. 2008. Predicting functional regulatory polymorphisms. Bioinformatics 24(16):1787-92.

poster

48

Multidimensional Analysis of Glioblastoma aCGH data using computational homology

Wenjing Zheng
University of California, Berkeley

Javier Arsuaga

Purpose:
Brain cancer is among the cancer types with worst prognosis outlook. It is estimated that 22,000 people will be diagnosed with and 13,000 will die of brain cancer in 2009 [1]. Glioblastoma is the most aggressive and common form of brain cancer, with median survival time of 12 months with standard medical treatment. It is most common in adults over 50, and it has been observed that increased age of diagnosis is associated with increased prognostic risk. Genomic deletions and amplifications affect transcriptional programs and are found to be associated to the genesis and development of cancer. Array CGH is a microarray technology that can measure DNA copy number changes on the genome. We use a novel method, called Mutidimensional Analysis of CGH (MDACGH) , to identify regions of the genome that are associated with the development of the disease.

Materials and Methods:
The MDACGH method consists of a mapping of the CGH profile to a high dimensional point cloud followed by the application of computational topology tools for its analysis. This approach allows us to perform association analysis of chromosome arms and clinical parameters in a given group of samples. We apply this method to CGH dataset from TCGA's glioblastoma pilot project.

Results:
In our first analysis, we compared CGH profiles between tumor samples and controls and identified copy number changes associated with the disease. MDACGH identified some of the chromosome arms found to have broad aberrant regions in [3], such as chromosome 6q, 7, 9p, 10, 13, 14, 15q, 22, as well as arms that have found to have smaller aberrant regions, such as chromosome 1 and 12q. We have also found regions that were not identified before (such as 2q, 5q).

Conclusions:
We believe that MDACGH is a promising method for analyzing aCGH that will complement well the existing statistical methods that are more focused on the linearity of the aCGH profile.

References:
[1] National Cancer Institute, "SEER Cancer Statistics Review"
[2] American Brain Tumor Association
[3] TCGA, "Comprehensive genomic characterization defines human glioblastoma genes and core pathways", Nature (455), 2008, 1061-1068.
[4] DeWoskin, "Using computational homology for prediction of breast cancer response to chemotherapy", Topology and Its Applications(In press).

bcats2009

Invited Poster Abstracts

poster

49

SOCR Motion Charts: An Efficient, Interactive and Dynamic Applet for Visualizing Multivariate Data

Jameel Al-Aziz
University of California, Los Angeles

Nicolas Christou
Ivo D. Dinov

Purpose:
The amount, complexity and provenance of data have dramatically increased in the past five years. Visualization of observed and simulated data is a critical component of any social, environmental, biomedical or scientific quest. Dynamic, exploratory and interactive visualization of multivariate data, without preprocessing by dimensionality reduction, remains an insurmountable challenge. The goals of this study is to utilize the Statistics Online Computational Resources (www.SOCR.ucla.edu) and address this challenge in the context of probability, statistics and bioinformatics education, technology based instruction and statistical computing.

Materials and Methods:
We have developed a new Java-based infrastructure, SOCR Motion Charts, for discovery-based exploratory analysis of multivariate data. This interactive data visualization tool enables the visualization of high-dimensional longitudinal data. SOCR Motion Charts allows mapping of ordinal, nominal and quantitative variables onto time, axes, size, colors, glyphs and appearance characteristics, which facilitates the interactive display of multidimensional data.

Results:
We validated this new visualization paradigm using several publicly available multivariate datasets including Ice-Thickness, Housing Prices, Consumer Price Index, and California Ozone Data. SOCR Motion Charts is designed using object oriented programming, implemented as a Java Web-applet and provided to the entire community on the web at http://www.SOCR.ucla.edu/SOCR_MotionCharts. It can be used as instructional tool for rendering and interrogating high-dimensional data in the classroom, as well as a research tool for exploratory data analysis.

Conclusion:
SOCR Motion Charts is an open-source, anonymously-accessible Java applet that facilitates the visualization of complex multimodal data and meta-data. It may be used directly online, downloaded for local execution, extended to fit specific new research or educational needs, and employed for classroom training or research informatics projects.

References:
Che, A., Cui, J., and Dinov, I. (2009), "Socr Analyses: Implementation and Demonstration of a New Graphical Statistics Educational Toolkit," JSS, 1-19.
Dinov, I. (2006a), "Statistics Online Computational Resource," Journal of Statistical Software, 16, 1-16.
Dinov, I. D., Sanchez, J., and Christou, N. (2008), "Pedagogical Utilization and Assessment of the Statistic Online Computational Resource in Introductory Probability and Statistics Courses," Journal of Computers & Education, 50, 284–300.

poster

50

Exploratory analysis of a genomic segmentation with segtools

Orion J. Buske
University of Washington

Michael M. Hoffman
William Stafford Noble

Purpose:
Novel computational methods help discover patterns in large-scale functional genomics data, but create a new challenge in understanding and interpreting these patterns. ChIP-seq and other assays are generating whole-genome, single-base-resolution readouts of many genomic properties, including histone modification, open chromatin, RNA expression, and transcription factor (TF) binding. Automatic segmentation methods can label genomic regions that exhibit consistent patterns across such diverse data, but it is difficult to determine if these segmentations are biologically meaningful. We developed a software package called segtools to investigate the properties of a segmentation in a genomic context and to suggest biological interpretations of the segment labels.

Methods:
The software we developed takes a genomic segmentation and annotations as input and visualizes segmentation properties and relationships between the segment labels and the annotations. Segtools currently displays, for each segment label, the length, base, and segment distributions, the mono- and dinuclotide frequencies, the distribution of the underlying signal data, and the co-occurrence with annotated features in terms of both aggregation around features and predictive power of segment labels.

Results:
Segtools was provided with a 25-label whole-genome segmentation. This segmentation was produced using Segway[1] and was based on 31 ChIP-seq and DNase-seq signal tracks (histone modification, open chromatin, and TF binding data) generated by the ENCODE Consortium[2]. Segway was trained on the subset of these data found in nine of the 30 ENCODE pilot regions (0.15% of the human genome).

Of these 25 labels, segtools found 13 that correspond to gene organization, including five associated with transcription initiation and elongation, four associated with transcription termination, one associated with insulation, and two possibly associated with poising. Another nine labels correspond to repressed regions and broad “dead zone” regions with little signal activity. Additionally, two labels were found to be differentially enriched between exonic and intronic regions.

Conclusion:
By presenting the segment labels in a variety of genomic contexts, the quality and character of a segmentation was able to be quickly evaluated. Segtools automated the exploratory analyses that were necessary to begin understanding a segmentation and identifying areas of additional investigation.

References:
[1] Hoffman MM, Buske OJ, Bilmes JA, Noble WS. Segway: a dynamic Bayesian network for genomic segmentation. In preparation.
[2] ENCODE Project Consortium. 2007. Identification and analysis of functional elements[...]. Nature. 447(7146):782-3.

poster

51

Multivariate Analysis of Functionally-Annotated Metagenomes from Multiple Environments

Katherine Isaacs
San Jose State University

Rob Edwards
Barbara A. Bailey
Peter Salamon
Imre Tuba
Elizabeth Dinsdale

Purpose:
A metagenome is a sampling of genetic sequences from the entire microbial community within an environment or community. Recent studies have shown the viability and reliability of functional metagenomic analysis in describing and identifying environmental microbial and viral samples [1]. Most of the focus in metagenomics has been on single environments and the few comparative analyses have been based on small fractions of available metagenomic data. We develop statistical tools to identify functional differences between multiple environments.

Material and Methods:
All publicly available microbial metagenomic sequence sets were acquired from the SEED database [2]. Sequences were aligned and categorized into 27 functional hierarchies. We normalized the hierarchy counts for each metagenomic sample to a percent composition by function. Each metagenome was initially classified by the environment from which it was sampled. Unsupervised grouping was also done by K-means silhouettes. Random Forests were used to determine which functional hierarchies grouped the samples most strongly. A combination of canonical discriminant analysis and linear discriminant analysis was then used to create an environment predictor for metagenome samples.

Results:
We were able to differentiate and predict among the seven largest environments using only eight of the functional hierarchies with an error rate of 12.5%, verifying the utility of those functional hierarchies in profiling metagenomic environments. The technique gave similar insight when restricting the focus to further specified subgroups within a single environment. We also observed differentiation of environment classes when the technique was applied to a second tier of functional hierarchies.

Conclusions:
Our findings demonstrate the ability to differentiate and predict environments based on only a subset of key functional hierarchies. The functional profile of the metagenomes accurately distinguish the activity of microbial communities from different environments.

References:
[1] EA. Dinsdale et al. Nature, 452:629–632, 2008.
[2] F Meyer et al. BMC Bioinformatics, 9(1):386, 2008.

poster

52

Computational prediction of drug side effects by identifying human antitargets

Adrian Laurenzi
University of Washington

Jeremy Horst
Ram Samudrala

Purpose:
A tremendous cost of developing new drugs results from pursuing candidate compounds that fail due to toxicological and clinical safety issues, accounting for 30% of failures in human clinical trials (1). To increase drug development efficiency computational methods can be applied early on in development to identify the potential compounds likely to fail. We propose a computational method to predict the propensity for a candidate compound to cause side effects.

Materials and Methods:
Our method outputs a list of human proteins predicted to bind to a potential inhibitor that targets a given input protein. The functional sites of the input protein and each human protein from GenBank are predicted using an algorithm to generate meta-functional signatures (2). Predicted functional residues are superimposed onto sequence alignments that compare the input protein to each human protein. An inhibitor interaction potential (IIP) is calculated for each alignment according to the similarity of the residues within the predicted functional regions. The higher the IIP, the more likely an inhibitor targeting the input protein would bind the human protein and potentially cause side effects. IIPs are used to calculate the overall inhibitor interaction potential (OIIP) reflecting the overall side effect potential. We evaluate the method using data from the BindingDB to create groups of proteins known to bind the same ligand and control groups of random proteins. The proteins within each ligand or control group are compared amongst each other to generate IIPs.

Results:
Preliminary evaluation was performed on a set of 25 proteins making up 5 ligand groups and a control group containing 5 random proteins. The alignments of the proteins within each ligand group had an average of 10 out of a possible 60 overlaps amongst predicted functional residues and there were no overlaps in the control group alignment. The method produced more overlaps amongst proteins that bind the same ligand demonstrating it would be useful in identifying human proteins that bind the same inhibitor as an input target protein.

Conclusion:
Drug developers can use our method to select drug targets least likely to produce side effects when targeted by an inhibitor by inputting a set of target proteins from a pathogen with sequence data available and comparing the OIIPs output for each target. We will apply our method to the entire BindingDB for more extensive evaluation.

References:
1. Kola I et al. (2004) Can the pharmaceutical industry reduce attrition rates? Nature Rev. Drug Discov.
2. Wang K et al. (2008) Protein Meta-Functional Signatures from Combining Sequence, Structure, Evolution, and Amino Acid Property Information. PLoS Comput Biol.

poster

53

Virtual Screening for Specific Inhibitors of the Dual-Specificity Phosphate SSH-2

Matt Mui
University of California, San Diego

Marshall J. Levesque
Jason H. Haga

Purpose:
Cellular functions, in particular cell growth and movement, are controlled in part through the activation of the dual specificity phosphatase (DSP) called Slingshot-2 (SSH-2). SSH-2 is known to contribute to the progression of cancer and Alzheimer's disease. Finding a specific inhibitor for SSH-2 may have a profound impact in clinical treatments for different diseases. In contrast to the traditional wet-bench method of directly determining binding affinities of different chemical compounds, we have employed a computational approach to screen for potential inhibitors for SSH-2 using grid computer technologies.

Materials and Methods:
Proteins in the DSP family are screened against the ZINC "drug-like" library in parallel over clusters of processors in the PRAGMA grid. First, SSH-2 is screened against the entire library to determine the binding affinity of each compound to the protein. Amongst the 2 million compounds in the library, only the top 1% binding compounds to SSH-2 are retained. These compounds are then rescreened against other DSP family members. The differences in binding scores between SSH-2 and other family members will determine the binding specificity to the compound. A large difference suggests high specificity and a small difference indicates low specificity toward the proteins of interest.

Results:
The results from five DSP screenings, specifically SSH-2, VHR, VH3, PTEN and KAP, suggest that 3-[(4,5-dimethoxy-3-oxo-1H-isobenzofuran-1-yl)amino]-4-methyl-benzoic acid shows the highest affinity for SSH-2, but the lowest affinity for the other DSPs, among the best 100 SSH-2 binding compounds. These results suggest that this compound has high specificity toward SSH-2.

Conclusions:
This study successfully identified a chemical compound that is specific for SSH-2. Further computational screenings with the remaining DSP family members are necessary to ensure that it does not bind to other DSPs. This study reduces the cost associated with manually testing a very large chemical compound space by employing grid computer technologies in virtual molecular docking experiments to identify promising potential SSH-2 inhibitors. Wet bench verification will be necessary to further confirm results from the identified potential inhibitors list and to verify in-vitro activity. These results will help better understand the molecular interactions between SSH-2 and cofilin and possibly the drug discovery process of cancer and Alzheimer's disease.

References:
Levesque MJ, et al. "Design of a grid service-based platform for in silico protein-ligand screenings", Comput Methods Programs Biomed. 2009, pp. 73-82.

poster

54

Determining the Optimal Pacing Sites for Biventricular Pacing the Failing Heart with Left Bundle Branch Block

Scott Revelli
University of California, San Diego

Dr. Roy C. P. Kerckhoffs

Purpose:
The goal of this project is to determine the location of the leads, for biventricular pacing a failing heart with left bundle branch block, which will produce the optimal cardiac performance.

Methods:
This task was pursued though the use of Continuity 6, a finite element modeling program designed by UCSD's Cardiac Mechanics Research Group for modeling biological systems. In Continuity 6, an anatomically simplified biomechanical (BM) and an electrophysiology (EP) model of a rabbit heart were created. In order to solve a problem of this nature which involves both mechanical and electrical properties, a fully coupled electromechanical model was employed. By employing Nimrod, a parametric modeling system, designed and devolved by Dr. Abramson and his colleagues at the Message Lab, at the University of Monash in Melbourne Australia, a parameter sweep will be conducted on the fully coupled model. To produce results as accurate as possible, the existing finite element rabbit heart model was heavily refined, producing one of the most computationally demanding problems ever created in Continuity 6.

Results:
The Puglisi-Bers EP model used in this project was validated in 2004 by Dr. Saucerman. However, because the EP model was used to simulate the entire heart, in this project, convergence had to first be reconfirmed. The accuracy of the refined BM model was checked by passively inflating the left ventricle. To accomplish this task, a pressure was applied to the left ventricle while no active tension parameters were implemented. While the results produced by the refined model follow the same trend as the results produced by the original biomechanical model, created by Dr. Vetter and Dr. McCulloch's model, and the right ventricular volumes are just about the same, the left ventricle in the refined model did not inflate as much as Dr. Vetter and Dr. McCulloch's model. The lack of inflation indicates that the elements in the refined mesh are slightly stiffer than those present in Dr. Vetter and Dr. McCulloch's model. Work is currently being conducted to resolve this issue, as well as on coupling the EP and BM models so that the parameter variations can be started.

References:
UCSD Cardiac Mechanics Laboratory
Dr. Roy C. P. Kerckhoffs, Dr. Jazmin Aguado-Sierra, Fred Lionetti, Stewart G. Campbell
University of Monash MeSsAGE Lab/ DSSE
Dr. David Abramson, Blair Bethwaite, Tom Peachey, Colin Enticott
UCSD PRIME
Dr. Peter Arzberger, Dr. Gabriele Wienhausen, Teri Simas, Tricia Taylor

poster

55

Oscillatory driving of an engineered mevalonate network to increase biofuel yields

Catherine Shi
University of California, San Diego

Tal Danino
Howard Chou
Jeff Hasty

Previously a mevalonate pathway in E. coli was engineered for high production of terpenoids, which are valuable compounds of numerous commercial uses such as anti-malarial drugs and possibly biofuel candidates. However, non-native synthetic pathways often inhibit cell growth due to an unbalanced production of toxic intermediates or cause a high metabolic burden by taking up resources that would feed normal cell growth and function. The goal of this project is to use computational modeling & experimental biocircuits to determine whether oscillatory production of enzymes in this network can mitigate the metabolic load, relieve cell toxicity and thus lead to higher yields of the desired products. Using a previous synthetic oscillator [1] we constructed several plasmids that drive either the MevB, MevT, or nudF genes responsible for producing isopenteneol at a particular frequency. We simulated this network with a discrete time-delay model for the enzymes and growth of cells involved. Calibrating our model to experimental results, we were able to show that downstream production of isopentenol can be increased by a factor of 5 at high frequencies as compared to a control with average level of production.

References

1. Jesse Stricker, Scott Cookson, Matthew R. Bennett, William H. Mather, Lev S. Tsimring & Jeff Hasty. A fast, robust and tunable synthetic gene oscillator. Nature 456, 516-519 (27 November 2008)

poster

56

Trypanosoma cruzi Proline Racemase: A promising new drug target for Chagas' disease

Michelle Zhou
University of California, San Diego

Cesar Augusto Fernandes de Oliveira, PhD
Barry Grant, PhD
Prof. James Andrew McCammon

Purpose:
Chagas' disease, caused by Trypanosoma cruzi infection, currently affects 16-18 million people, mostly in Central and South America. Here we describe a computational study on a promising new drug target, the T. cruzi secreted proline racemase (TcPRAC) [1], believed to be essential for parasite virulence. TcPRAC has been shown to be a potent mitogen for host B cells, preventing the development of an effective parasite-directed immune response [3]. The availability of a recent high-resolution crystal structure in complex with a modified proline inhibitor affords the opportunity for structure based computational design of novel inhibitors with important therapeutic potential.

Materials and Methods:
We utilized Virtual Screening Workflow with ligand docking program GLIDE to dock available ligand structures to our protein crystal structure. We also applied statistical computing program R to undergo conservation analysis of residues, as well as cross correlation and principal component analysis. The molecular dynamics simulations were performed using AMBER10.

Results:
High-scoring ligand structures were identified through docking. Our docking results suggest the existence of a second potential binding site on the open form of the enzyme. Conservation analysis showed that many of the residues around the secondary pocket were well-conserved, and that several of these residues interacted directly with the docked ligands. Cross-correlation and principal components analyses revealed that the enzyme can undergo large conformational changes in the apo state around the newly identified binding pocket. Molecular dynamics simulations also displayed significant conformational changes, namely the distinct opening and closing movements of the κchain.

Conclusion:
We identified several promising drug structures through docking, as well as a possible secondary binding site. We hypothesize that mitogenic properties of TcPRAC may depend on the exposure of epitopes located around the newly identified pocket. Conservation analysis further indicates that this secondary pocket may play a significant role in the structure and function of the protein. We believe that the strategy adopted in this work is useful for rationalizing the molecular basis of inhibition and representing a starting point for designing more potent inhibitors.

References:
[1] Chamond, N. et al. (2005). Molecular Microbiology 58(1):46-60.
[2] Reina-San-Martin, B. et al. (2000). Nature Medicine 6(8):890-897.

BCATS 2009 thank you

Guidance and Help

Blanca Pineda
Russ Altman
Vijay Pande
Larry Fagan

Previous Organizers

Tiffany Chen
Adam Grossman
Xuhui Huang
Hedi Razavi
Jesse Rodriguez
Katherine Steele
Rebecca Taylor

2009 Organizing Committee

David Chen (Chair)
Sarah Aerni (Co-Chair)
Robert Bruggner
Samuel Hamner
Jonathan Karr
Linda Liu
Daniel Newburger
Chirag Patel

Platinum Sponsors

National Science Foundation
Stanford Biomedical Informatics Training Program
Sandia National Laboratories
Simbios
Bio-X
Applied Biosystems

Gold Sponsors

Genentech
Agilent

Silver Sponsors

Butte Lab

Other Sponsors

Ingenuity Systems
Solid Spark Creative
Stanford BMIR
Stanford Bookstore

University Liaisons

Angela Brooks
Jeremy Phillips
Josh Stuart