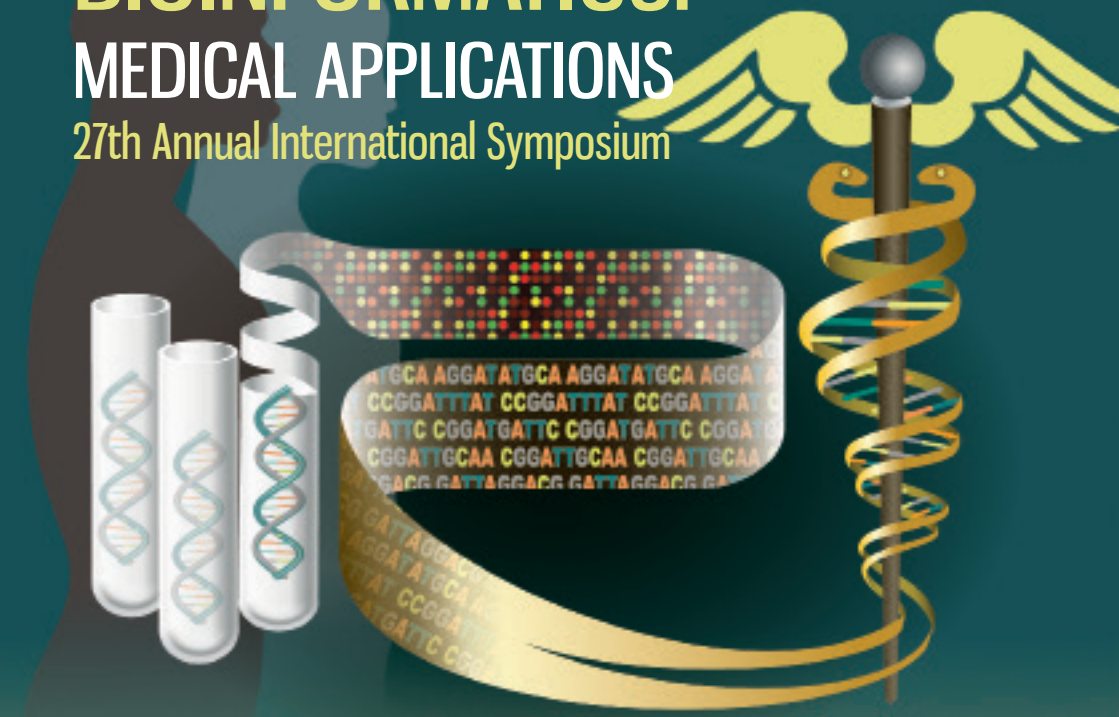
BIOINFORMATICS:
MEDICAL APPLICATIONS
27th Annual International Symposium



Hunter College of the City University of New York
695 Park Avenue (East 68th Street at Lexington Avenue)
Hunter West Building, Room 714 , New York, NY 10065

Thursday, May 29, 2014
8:30am-6:00pm

SYMPOSIUM PROGRAM

SYMPOSIUM SPONSORS:

Hunter College of the City University of New York, Center for Translational and Basic Research (CTBR)* & Clinical & Translational Science Center, Weill Cornell Medical College including its partners: Hunter College School of Nursing, Cornell University Cooperative Extension in NYC Memorial Sloan-Kettering Cancer Center, Hospital for Special Surgery

*(CTBR, formerly Center for Study of Gene Structure and Function/Gene Center)

http://ctbr.hunter.cuny.edu/bioinfo2014

27th Annual International Symposium

The 27th Annual International Symposium of the Center for Translational and Basic Research at Hunter College, with Weill Cornell Medical College Clinical and Translational Science Center, is supported by the National Institute on Minority Health and Health Disparities, National Institutes of Health (Grant #8 G12 MD007599-27) and the National Center for Advancing Translational Sciences (Grant #2UL1TR000457-06)

The Research Centers in Minority Institutions (RCMI) Program was transferred to the National Institute on Minority Health and Health Disparities (NIMHD) following the passing of the Consolidated Appropriations Act, 2012, which dissolved the National Center for Research Resources, the former home of the RCMI program.

The RCMI Program of the National Institutes of Health develops and strengthens the research infrastructure of minority institutions by expanding human and physical resources for conducting basic, clinical, and translational research. It provides grants to institutions that award doctoral degrees in the health professions or health-related sciences and have a significant enrollment of students from racial and ethnic minority groups that are underrepresented in biomedical sciences. The program also serves the dual purpose of bringing more racial and ethnic minority scientists into mainstream research and promoting minority health research because many of the investigators at RCMI institutions study diseases that disproportionately affect minority populations.

NIH National Institutes of Health
Turning Discovery Into Health

The NIH, a part of the U.S. Department of Health and Human Services, is the primary federal agency for conducting and supporting medical research. Composed of 27 Institutes and Centers, the NIH provides leadership and financial support to researchers in every state and throughout the world. Its mission is science in pursuit of fundamental knowledge about the nature and behavior of living systems and the application of that knowledge to extend healthy life and reduce the burdens of illness and disability. **http://www.nih.gov**

The mission of the **National Institute on Minority Health and Health Disparities (NIMHD)** is to promote minority health and to lead, coordinate, support, and assess the NIH effort to reduce and ultimately eliminate health disparities. In this effort NIMHD will conduct and support basic, clinical, social, and behavioral research, promote research infrastructure and training, foster emerging programs, disseminate information, and reach out to minority and other health disparity communities. **http://www.nimhd.nih.gov/default.html**

National Center for Advancing Translational Sciences is strategically positioned to facilitate interdisciplinary clinical and translational research. RTRN has established a solid technological foundation to support intellectual exchange, generate innovative inter- and multi-disciplinary research and facilitate the movement of scientific advances throughout the translational research spectrum. **http://www.ncats.nih.gov/**

Center for Translational and Basic Research

The RCMI Program at Hunter College of the City University of New York was established in 1985 through the vision of James Wyche, Harvey Ozer (former Program Coordinators for the Hunter RCMI Program) and Richard Mawe (former Program Director of the Hunter RCMI Program) with the support of the Research Centers in Minority Institutions (RCMI) Program.

In 2012, the RCMI Program was transferred to the National Institute on Minority Health and Health Disparities (NIMHD) of the National Institutes of Health (NIH) after the closing of the National Center for Research Resources.

The program funds the Center for Translational and Basic Research (CTBR) for Addressing Health Disparities and Improving Health Outcomes, formerly the Center for Study of Gene Structure and Function (Gene Center), a consortium of researchers from the Hunter College departments of biology, chemistry, psychology, physics, anthropology, and urban public health, as well as from the Hunter-Bellevue School of Nursing. Since the CTBR's inception, the growing number of papers published in peer-reviewed journals and the number and amount of grants obtained by the faculty have been the most visible hallmarks of its success. The CTBR provides a vibrant research environment marked by workshops on cutting-edge research techniques; frequent research colloquia and seminars by guest scientists; an annual international symposium, which is a major event on the New York scientific calendar; and a strong emphasis on collaborative translational research. In addition, the CTBR encourages bright undergraduates, especially minorities, to develop a career in biomedical, drug abuse/neuroscience research by hosting a Summer Program for Undergraduate Research and supports the professional development of minority scientists through the JustGarciaHill science web site.

The CTBR is a key partner in the Clinical and Translational Science Center, an enterprise that also includes the Weill Cornell Medical College, Memorial Sloan-Kettering Cancer Center, the Hospital for Special Surgery and the Hunter-Bellevue School of Nursing. The Clinical Translational Science Center was established in 2007 with the aim of accelerating translational research. The overall goal is to facilitate the transition of laboratory work into state-of-the-art clinical research (T1 research), provide research that improves patient care (T2) and health outcomes in the general community (T3). The CTBR also participates in a national consortium, the Research Centers in Minority Institutions Translational Research Network (RTRN), which facilitates collaboration, large-scale projects, and sharing of facilities among Research Centers in Minority Institutions.

For more information about the CTSC, please visit **http://ctbr.hunter.cuny.edu/**

Clinical and Translational Science Center

The Clinical and Translational Science Center (CTSC) is a unique collaboration between renowned biomedical and community organizations centered on Manhattan's east side. Weill Cornell Medical College and the Graduate School of Medical Sciences is home to the administrative core of the CTSC, led by CTSC Program Director Julianne Imperato-McGinley, MD, Associate Dean of Translational Research and Education at Weill Cornell Medical College (WCMC).

In addition to WCMC, the CTSC partner institutions include:

- Hunter College, Center for Translational and Basic Research
- Hunter College, School of Nursing
- Hospital for Special Surgery
- Memorial Sloan-Kettering Cancer Center
- Cornell University Co-operative Extension in New York City

Affiliated hospitals include New York-Presbyterian Hospital, Lincoln Medical Center, Methodist Hospital, New York Downtown Hospital, New York Queens Hospital, Wyckoff Heights Medical Center, and Brooklyn Hospital.

The CTSC is designed to bring together the resources of all partner and affiliate institutions to facilitate novel translational research. Separately, these institutions include superb academic centers of excellence, a diverse patient base, and a unique community-engagement program designed to foster collaboration between community groups and translational research scientists. Each partner and affiliate has an unmistakable character that enhances multi-disciplinary interaction. Integration of these unique resources and intellectual assets will facilitate translation of research findings in the laboratory to clinical research at the bedside and ultimately to best practices within underserved communities.

This center is funded through the Clinical and Translational Science Awards (CTSAs), a national consortium that is transforming how clinical and translational research is conducted.

For more information about the CTSC, please visit http://www.med.cornell.edu/ctsc.

BIOINFORMATICS: MEDICAL APPLICATIONS

27th Annual International Symposium
May 29, 2014

Bioinformatic scientists develop computational and statistical methods to understand disease mechanisms and design disease interventions. Spurred by an avalanche of genomic and other omics data, "the biomedical research enterprise is increasingly becoming data-intensive and data-driven" *(Mission Statement, NIH Big Data to Knowledge [BD2K] Initiative).* Accordingly, bioinformatics has become an essential component of all biomedical projects and a core competency for all biomedical scientists. This symposium features distinguished bioinformatics scientists from the United States and around the world, who will present some of the latest applications of bioinformatics in medicine and health care.

The morning session will begin with a keynote presentation by **Søren Brunak** *(Professor, Director of the Center for Biological Sequence analysis, Technical University of Denmark),* who will share his experience of leveraging national electronic patient records to improve health care in Denmark. **Nicola Mulder** *(Associate Professor, University of Cape Town, South Africa; President, African Society for Bioinformatics and Computational Biology)* will describe H3ABioNet, a bioinformatics network supporting the Human Heredity and Health in Africa (H3Africa) project. **Timothy Thornton** *(Assistant Professor, Department of Biostatistics, University of Washington)* will present statistical methods for identifying disease-associated genes in Hispanic populations. **Tuuli Lappalainen** *(Assistant Professor, Department of Systems Biology, Columbia University; Junior Investigator and Core Member, New York Genome Center)* will offer insights into human genome functions based on integrated analyses of genomic and gene-expression data from thousands of individuals. **Brian Athey** *(Michael A. Savageau Collegiate Professor & Chair, Department of Computational Medicine and Bioinformatics, University of Michigan)* will demonstrate how bioinformatics helps the design of personalized treatments of psychiatric diseases.

In the afternoon, **Owen White** *(Professor & Director, Bioinformatics Department, Institute for Genome Sciences, University of Maryland School of Medicine)* will deliver a keynote speech on a bioinformatics infrastructure supporting the Human Microbiome Project. **Christine Eng** *(Professor, Medical Director of DNA Diagnostic Laboratory, Director of Storage Disorders Clinic, Baylor College of Medicine)* will describe the development of an exome sequencing facility for precise diagnosis of congenital diseases. **Joao Xavier** *(Assistant Faculty Member, Computational Biology, Memorial Sloan-Kettering Cancer Center)* will present how to develop probiotic treatments of bacterial infections based on mathematical modeling of microbial populations in human gut. **Weigang Qiu** *(Associate Professor, Department of Biological Sciences, Hunter College, City University of New York)* will showcase a bioinformatics pipeline for using whole-genome sequences to monitor emerging infectious diseases.

The symposium will conclude with a keynote presentation by **Feng Zhang** *(Assistant Professor, Biological Engineering and Brain and Cognitive Sciences, Massachusetts Institute of Technology)*, who will describe the use of bioinformatics to improve specificity of the new CRISPR genome-editing technology.

MORNING SESSION

9:00 **Jesus Angulo, Ph.D.,** Professor of Biological Sciences at Hunter College, CUNY and Principal Investigator/Program Director of the Center for Translational and Basic Research

Jennifer J. Raab, J.D., Hunter College, City University of New York

Julianne Imperato-McGinley, M.D., Associate Dean of Translational Research, Weill Cornell Medical College

9:15 **Søren Brunak, Ph.D., Keynote Speaker,** Professor, Director of the Center for Biological Sequence Analysis, Technical University of Denmark
Fine-Grained Phenotypes, Comorbidities and Disease Trajectories from Data Mining of Electronic Patient Records

10:00 **Nicola Mulder, Ph.D.,** Associate Professor, University of Cape Town; President of the African Society for Bioinformatics and Computational Biology; Coordinator of the H3Africa Bioinformatics Network
H3ABioNet, an African Bioinformatics Network for Genomic Research into Human Diseases

10:35 **Coffee Break/Poster Session**

10:50 **Timothy Thornton, Ph.D.,** Assistant Professor, Department of Biostatistics, University of Washington
Gene Mapping of Complex Traits in Genetically Admixed Populations

11:25 **Tuuli Lappalainen, Ph.D.,** Assistant Professor, Department of Systems Biology, Columbia University; Junior Investigator and Core Member, New York Genome Center
Transcriptome Sequencing Uncovers Functional Variation in Human Genomes

12:00 **Brian Athey, Ph.D.,** Michael A. Savageau Collegiate Professor, Chair of the Department of Computational Medicine and Bioinformatics, Professor of Psychiatry and Internal Medicine, University of Michigan
Moving Bioinformatics into Medicine: The Emergence of 'Psychiatric Pharmacogenomics"

12:35 Lunch for **pre-registered** participants, Poster Session
Faculty Dining Room, Hunter College, West Building, 8th floor

AFTERNOON SESSION

2:00 **Remarks by Sponsoring Agency**

2:20 **Owen White, Ph.D., Keynote Speaker,** Professor of Epidemiology and Public Health, Director of the Bioinformatics Department, Institute for Genome Sciences (IGS), University of Maryland School of Medicine
Human Microbiome Project: Large-Scale Data Management and Analysis

3:05 **Christine Eng, M.D.,** Professor of Molecular and Human Genetics, Medical Director of DNA Diagnostic Laboratory, Director of Storage Disorders Clinic, Baylor College of Medicine
Clinical Whole Exome Sequencing for the Diagnosis of Mendelian Disorders

3:40 Coffee Break/Poster Session

3:55 **Joao Xavier, Ph.D.,** Assistant Faculty Member, Computational Biology, Memorial Sloan-Kettering Cancer Center
Shifts in the Intestinal Microbiota During Antibiotic Treatment

4:30 **Weigang Qiu, Ph.D.,** Associate Professor of Biology, Hunter College, City University of New York
Genomic Surveillance of Emerging Infectious Diseases

Closing Speaker

5:05 **Feng Zhang, Ph.D., Keynote Speaker,** Assistant Professor of Biological Engineering and Brain and Cognitive Sciences, Massachusetts Institute of Technology
Genome Engineering Using CRISPR-Cas9

5:50 **Poster Awards Ceremony**

Concluding Remarks
Paul Feinstein, Ph.D., Associate Professor of Biology, Hunter College, City University of New York

Søren Brunak

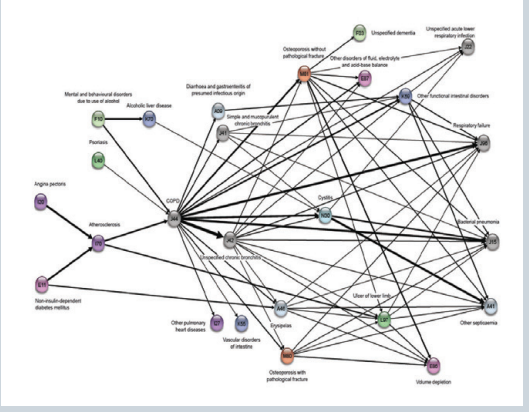
Keynote Speaker

Technical University of Denmark
& University of Copenhagen

Fine-Grained Phenotypes, Comorbidities and Disease Trajectories from Data Mining of Electronic Patient Records

Abstract: Electronic patient records remain a rather unexplored, but potentially rich data source for discovering correlations between diseases, drugs and genetic information in individual patients. Such data makes it possible to compute fine-grained disease co-occurrence statistics, and to link the comorbidities to the treatment history of the patients. A fundamental issue is to resolve whether specific adverse drug reaction stem from variation in the individual genome of a patient, from drug/environment cocktail effects, or both. Here it is essential to perform temporal analysis of the records for identification of ADRs directly from the free text narratives describing patient disease trajectories over time. ADR profiles of approved drugs can then be constructed using drug-ADR networks, or alternatively patients can be stratified from their ADR profiles and compared. Given the availability of longitudinal data covering long periods of time we can extend the temporal analysis to become more life-course oriented. We describe how the use of an unbiased, national registry covering 6.2 million people from Denmark can be used to construct disease trajectories which describe the relative risk of diseases following one another over time. We show how one can "condense" millions of trajectories into a smaller set which reflect the most frequent and most populated ones. This set of trajectories then represent a temporal diseaseome as opposed to a static one computed from non-directional comorbidities only.

Bio: Søren Brunak, Ph.D., is professor of bioinformatics at the Technical University of Denmark. Prof. Brunak is the founding Director of the Center for Biological Sequence Analysis, which was formed in 1993 as a multi-disciplinary bioinformatics research group of molecular biologists, biochemists, medical doctors, physicists, and computer scientists. It has today 160 employees. In his work Søren Brunak combines molecular level systems biology data with the analysis of phenotypic data from the healthcare sector, such as electronic patient records, registry information and biobank questionnaires. A major aim is to discriminate between treatment related disease correlations and other comorbidities, thereby stratifying patients not only from their genotype, but also phenotypically based on the clinical descriptions in their medical records. His book with P. Baldi "Bioinformatics – The Machine Learning Approach" MIT Press, 1998/2001 was the first machine learning text book in the bioinformatics area and is highly cited. Prof. Brunak has published 250+ peer reviewed papers, and has around 30,000 WoK citations. Brunak has served on a large number of scientific advisory boards and funding bodies, including; the Bioinformatics Advisory Committee at the European Bioinformatics Institute; the "Molecules, Genes and Cells Funding Committee" at the Wellcome Trust, and the Scientific Council at the Institut Pasteur, Paris. Prof Brunak has been EMBO member since 2009. He was the chair of the interim board of the ELIXIR infrastructure for biological information in Europe.



Nicola Mulder

University of Cape Town and
H3Africa Bioinformatics Network

H3ABioNet, an African Bioinformatics Network for Genomic Research into Human Diseases

Abstract: The Human Heredity and Health in Africa (H3Africa) initiative funds research projects studying the genetic and environmental factors related to human diseases in Africa. The projects currently include a wide range of diseases, such as cardiometabolic disease, kidney disease, stroke, rheumatic heart disease, diabetes, etc., as well as the study of microbiomes and their impact on disease. Key to all these projects is a strong bioinformatics infrastructure to facilitate secure storage and management of sensitive genetic data from array and sequencing, and to enable African scientists to analyse their own data. H3ABioNet, a pan African bioinformatics network, was established to develop this infrastructure to support H3Africa projects, and consists of over 35 nodes in 15 African countries. The network is working on building computing infrastructure for managing and analysing large-scale genetics data, as well human capacity through a broad bioinformatics training program. Since the H3Africa projects have not yet generated data for analysis, the H3ABioNet nodes are preparing for the data generation by working on inter-node projects using already existing data at the nodes or public datasets. The presentation will describe the activities of H3ABioNet related to capacity development and then discuss in more detail an example project on the genetic basis for disease in an African population.

Bio: Associate Professor Nicola Mulder heads the Computational Biology Group at the University of Cape Town (UCT), which is located in the Institute for Infectious Disease and Molecular Medicine in the Health Science Faculty. After her PhD, she spent 8.5 years at the European Bioinformatics Institute as a Team Leader responsible for the InterPro and Gene Ontology Annotation projects. At UCT she works on bioinformatics of microbial pathogens, human genetics and infectious diseases, and manages a bioinformatics services team. Her research includes projects with local scientists and clinicians studying African population diversity and the genetic basis of disease, which includes development of new methods for studying admixed populations. She also works on analysis of functional interaction networks in *M. tuberculosis* and other mycobacteria, host pathogen interactions, and development of visualization and analysis tools for high-throughput biology. A/Prof Mulder coordinates H3ABioNet, a large pan African Bioinformatics Network for Human Heredity and Health in Africa (H3Africa), which aims to build capacity in Africa for genomics research. The H3Africa projects are aimed at unravelling the genetic and environmental basis for disease and H3ABioNet is building the bioinformatics infrastructure and human capacity to analyse the data on the continent.

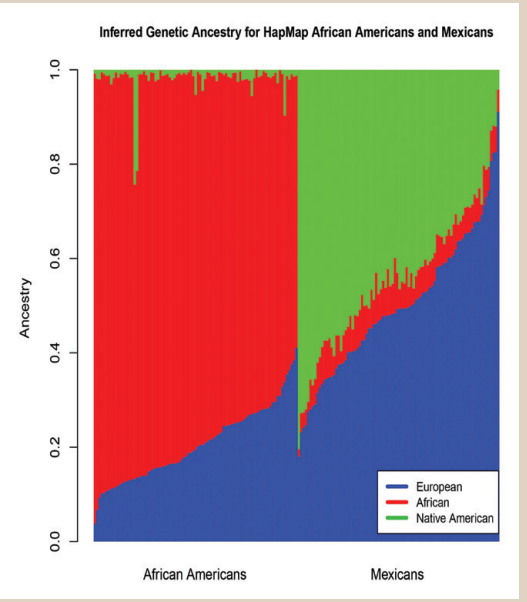
Timothy Thornton

University of Washington

Gene Mapping of Complex Traits in Genetically Admixed Populations

Abstract: To date, more than 1,500 large-scale genome-wide association studies (GWAS) have been conducted, leading to the discovery of numerous genetic variants that are associated with a variety of diseases. The vast majority of these genetic studies, however, have been conducted in populations of European descent, while very few studies have focused on the two largest minority groups in the United States, African Americans and Hispanics, despite these underrepresented populations bearing a disproportionately high burden for disease. African Americans and Hispanics are known to have admixed ancestry derived from multiple, previously isolated, continental progenitor groups. Genetic studies in ancestrally admixed populations offer exciting opportunities for the identification of novel genetic variants that underlie the diversity of traits both within and between populations. At the same time, the heterogeneous genetic background among admixed sample individuals pose special challenges for the mapping of genes. We will describe gene mapping approaches that exploit both genotype and ancestry information in genetically admixed populations. Illustrations and applications to clinical outcomes from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) and Hispanic women from the Women's Health Initiative (WHI) study will be presented.

Bio: Timothy Thornton is an Assistant Professor in the Department of Biostatistics and Institute for Public Health Genetics at the University of Washington. He is also an Affiliate Investigator at the Fred Hutchinson Cancer Research Center. The focus of his research is the development and application of statistical methods for the identification of genetic variants underpinning complex traits. His research lab also develops software for the statistical analysis of large-scale genotyping data. Prior to joining the faculty at the University of Washington, Dr. Thornton was a University of California President's Postdoctoral Fellow and he worked with Professor Neil Risch in the Institute for Human Genetic at UC-San Francisco. He earned a B.S. degree in mathematics from Hampton University and a Ph.D. in statistics from the University of Chicago. Dr. Thornton is currently a principal investigator (PI) of a National Cancer Institute funded Career Development Award (K01) and co-PI of a National Institute of General Medical Sciences funded Project Grant (P01).



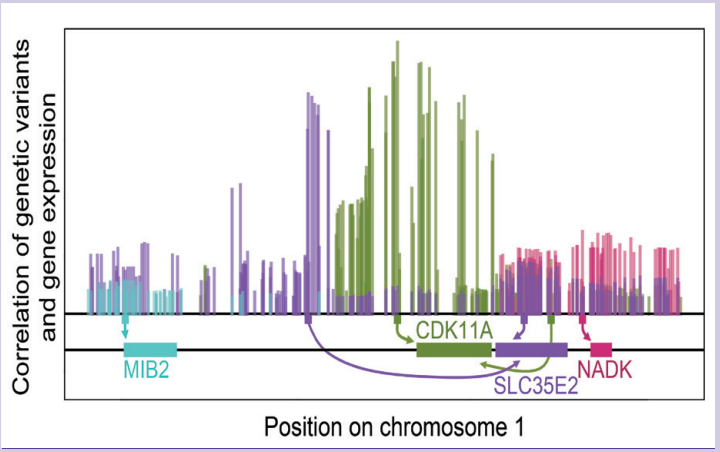
Tuuli Lappalainen

Columbia University and New York Genome Center

Transcriptome Sequencing Uncovers Functional Variation in Human Genomes

Abstract: Detailed characterization of cellular effects of genetic variants is essential for understanding biological processes that underlie genetic associations to disease. One approach to address this challenge is to combine genomic data to a functional readout of the cell, such as the transcriptome, measured by RNA-sequencing. In the Geuvadis project (Lappalainen et al. Nature 2013), we produced the first high-quality mRNA- and miRNA-seq data set from multiple human populations with high-quality genome sequences from 1000 Genomes. The tissue dimension of regulatory variation is being analyzed by Genotype-Tissue Expression project (GTEx), with pilot data from 1,642 RNA-seq samples and genome data from 174 donors across multiple tissues. In these projects, we have discovered extremely widespread genetic variation affecting both gene expression and splicing in different tissues, and characterized their impact in hundreds of disease-associated loci. We have further characterized how regulatory variants can affect the penetrance of deleterious coding variants, and systematically analyzed transcriptome effects of protein-coding loss-of-function variants. Altogether, our results shed light on the regulatory architecture of genome variation in humans, and demonstrate the power of integrating genome and transcriptome data not only to improve our general understanding of genetic variants, but also as a practical approach in future medical and clinical applications.

Bio: Tuuli Lappalainen is a junior investigator at the New York Genome Center and the first faculty member of the new institute, with a joint position as assistant professor at the Department of Systems Biology at Columbia University. Her work is focused on understanding functional variation in the human genome, and she has pioneered in integrating genome and transcriptome sequencing data to characterize regulatory variation. She is part of many major consortia in human genomics, including the 1000 Genomes project and the GTEx consortium. She did her postdoctoral training in the University of Geneva Medical School with Manolis Dermitzakis, and in Stanford University with Carlos Bustamante, and her PhD is from the University of Helsinki in 2009, supervised by Juha Kere and Päivi Lahermo, with a focus on genetic variation of North European populations.



Brian Athey

University of Michigan

Moving Bioinformatics into Medicine: The Emergence of 'Psychiatric Pharmacogenomics'

Abstract: Psychiatric Pharmacogenomics has been demonstrated to have value to patients suffering with Treatment Resistant Depression (TRD). This method utilizes Single Nucleotide Polymorphisms (SNPs) and Copy Number Variants (CNVs) in a defined base set of Pharmacokinetic and Pharmacodynamic genes. This 'combinatoric genotype' is then related to a patient's drug metabolizing phenotype, allowing personalization of drug treatment to be offered by the physician to the affected patient, many of whom have failed earlier treatment. Recent studies using a novel epigenomic annotation capability, applied to pharmacogenomic-based Genome Wide Association Studies (GWAS) and Psychiatric Association Studies, have yielded a family of new potential intragenic markers that are involved in regulating existing and newly suggested pharmacometabolic response elements. This appears to be a form of chromatin-based epigenomic gene regulation. Novel use of the NIH Epigenome Roadmap and the Allen Brain Atlas are featured in the analysis. This response is suggested to occur at the level of stratified populations of individuals, with a broad range of mapping to persons with mood disorders and also drug addiction behaviors. Experiments to validate these epigenomic "locus control region" markers at the microscopic and functional neuro-imaging levels are currently being planned, to be implemented on the tranSMART data integration and analysis platform. Prototypes of this translational bioinformatics system will be shown.

Bio: Brian Athey, Ph.D. is the Michael A. Savageau Collegiate Professor and Inaugural Chair of the Department of Computational Medicine and Bioinformatics at the University of Michigan Medical School. He is also a Professor of Psychiatry and of Internal Medicine. Brian serves as co-founder and Chief Scientific Officer of the tranSMART Foundation, a non-profit company founded to coordinate the development of the open source tranSMART communtiy and its code base. The tranSMART platform supports an integrated open data sharing and analytics platform used world-wide to accelerate clinical and translational research. Brian has led the National Library of Medicine (NLM) Next-Generation Internet (NGI) Visible Human Project and the DARPA Virtual Soldier Project. He is the founding Principal Investigator of the NIH National Center for Integrative Biomedical Informatics (NCIBI), one of eight NIH National Biomedical Computing Centers, funded by the National Institute on Drug Abuse (NIDA) and the NIH Common Fund. He has been a national leader in the NIH Clinical and Translational Scientists (CTSA) informatics Key Function Committee and U-M CTSA Informatics lead for 8 years. Brian has served as a special advisor to the Defense Scinces Office (DSO), DARPA (1994-1999) and to the Chief Informaton Officer (CIO) of the NIH (2007-2010). He has over 90 peer-reviewed scientific publications and proceedings, ranging from bioinformatics, epigenomics, metabolomics, chromatin structure, computational biology, clincal informatics, optical imaging, and grid computing. Brian serves as Chair of the Scientifc Advisory Board (SAB) of AssureRx Health, Inc. (Mason, Ohio). Brian is founding Chair of the One Mind for Research Technical Advisory Board, and is a standing member on its SAB.

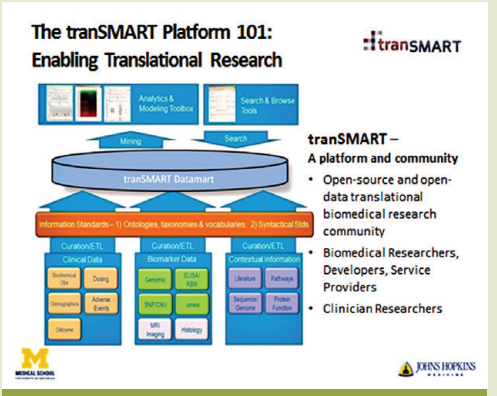


Figure: Overview of the tranSMART integration and analysis validation platform

Owen White

Keynote Speaker

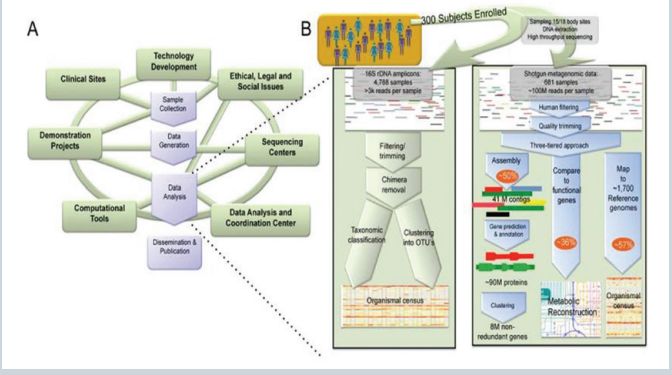
University of Maryland School of Medicine

Human Microbiome Project: Large-Scale Data Management and Analysis

Abstract: The HMP is designed to fuel research into the microbes that live in the various environments of the human body. HMP data sets now include over hundreds of reference genomes isolated from the human body, as well as 16S ribosomal RNA, and whole metagenome shotgun sequencing of samples collected from multiple body sites and individuals. Successful utilization of this data requires the use of controlled vocabularies, the application of quality control measures, and the development of large-scale data management procedures. As part of this initiative, the HMP Data Analysis and Coordination center (DACC) has provided a data management and analysis infrastructure to support the collection, integration and standardization at several levels (see: hmpdacc.org). As many people are aware the intersection of second generation sequencing technologies and the field of metagenomics is driving an explosion of data -- our effort to meet these unique informatics challenges will be presented. A major goal of the HMP is to define a data set of metagenomic samples derived from healthy individuals to serve in comparisons between each sample as well as comparisons of data derived from individuals with disease phenotypes. I will present the results of the HMP publications, as well as an overview for how users may access this data, and the analysis tools that are now as well as tools that will be made available in the near future.

Bio: Owen White is Director of Bioinformatics at the University of Maryland School of Medicine. In addition to these responsibilties, he is the Director of the Informatics Resource Center at IGS. White has overseen the annotation of hundreds of genomes sequenced using computer analyses such as pairwise searches, multiple sequence alignments, and numerous other methods in combination with systematic manual evaluation. This administration of analysis has served to generate highly uniform annotation that includes the genomes for Arabidopsis, the mosquito Aedes agypti, parasitic organisms such as Trypanosoma Brucei and Plasmodium falciparum, human ESTs and many Bacterial and Archaeal species. He has also developed automated annotation systems such as TIGRFams, Genome Properties, as well as the Annotation Engine.

Dr. White has also been at the forefront of creating web based genome analysis tools. He was the PI for a number of tools that include the TIGR Comprehensive Microbial Resource, the NIAID funded Bioinformatics Resource Center, and the BRC Portal. Dr. White is currently the PI for the NHGRI funded Data Analysis and Coordiantion Center (DACC) for the Human Microbiome Project and is associated with Gemina, a web based tool designed to identify infectious pathogens and their representative genomic sequences through selection of associated epidemiology metadata.



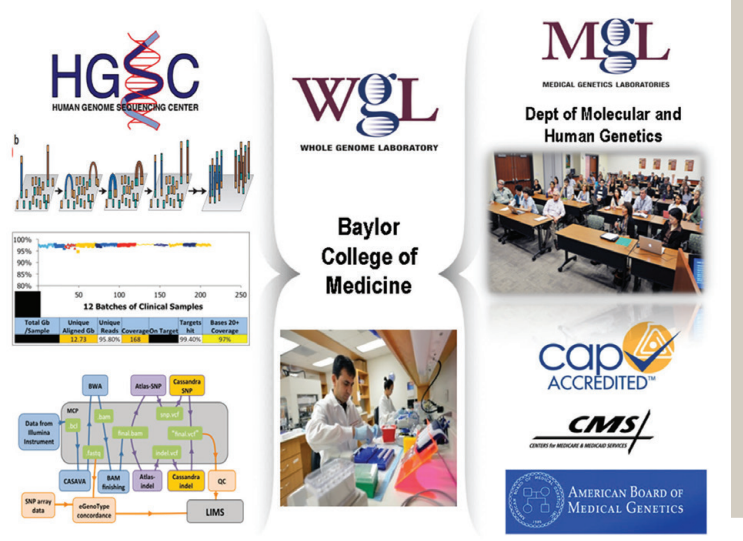
Christine Eng

Baylor College of Medicine

Clinical Whole Exome Sequencing for the Diagnosis of Mendelian Disorders

Abstract: Next generation sequencing technologies have experienced increased utilization in the clinical arena as a more efficient approach to the molecular characterization of challenging patient phenotypes. In October 2011, the Baylor Human Genome Sequencing Center (HGSC) and the Baylor Medical Genetics Laboratories developed and implemented a CAP/CLIA whole exome sequencing service. The whole exome sequencing test (WES) includes massively parallel sequencing of the human exome to a mean depth of >140X coverage (>95% target bases greater than 20X coverage), a SNP array for quality control (SNP concordance to genotype >99.8%), annotation of variants, clinical interpretation, and validation of significant findings by Sanger sequencing. To date, over 3000 samples have been submitted for clinical testing, of which 85% are from pediatric-age patients. The majority of probands have neurological phenotypes and have had extensive genetic testing that has not yielded a diagnosis. Of the first 2000 cases reported, 504 positive results (25%) consistent with the patient's described phenotype were detected. In this experience, there are examples of diagnoses that helped direct patient care, such as the identification of congenital myasthenic syndrome due to mutation in RAPSN. Medically actionable incidental results were reported, including mutations in genes causing Marfan syndrome, and arrthymogenic right ventricular dysplasia, which were relevant to both the proband and family members. The experience with exome sequencing in the clinical setting has demonstrated the clinical utility of this approach in which 25% of previously undiagnosed patients have definitive positive findings that can lead to improved clinical decision-making.

Bio: Christine M. Eng is Professor of Molecular and Human Genetics at Baylor College of Medicine. She received her B.A. from Yale University and M.D. from the Tulane University School of Medicine. She is a pediatrician, medical geneticist and senior director of the Baylor Medical Genetics Laboratories. She has directed DNA-based genetic testing laboratories for over 20 years. Most recently, her focus has been on clinical exome sequencing which the Baylor lab launched as a clinical test in 2011. The group recently published their experience with the first 250 clinical exome cases in the New England Journal of Medicine and they have now completed and reported over 2500 cases. She is particularly interested in examining how best to integrate complex genomic testing in clinical medicine.



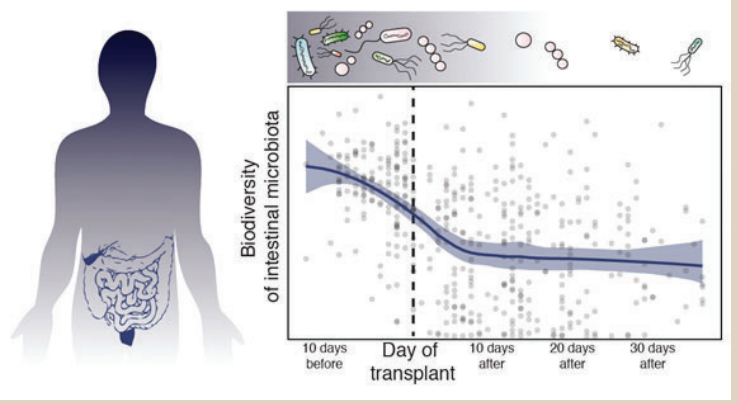
Joao Xavier

Memorial Sloan Kettering Cancer Center

Shifts in the Intestinal Microbiota During Antibiotic Treatment

Abstract: Understanding how the gut microbiota confers resistance against enteric pathogens and how antibiotics disrupt that resistance is key to the prevention and cure of intestinal infections. I will describe a novel bioinformatics method to infer microbial community ecology directly from time-resolved metagenomics data from next generation sequencing. The new method is inspired by the Lotka-Volterra mathematical model from ecological theory. Data from antibiotic-mediated *Clostridium difficile* infections are analyzed to quantify microbial interactions, commensal-pathogen interactions, and the effect of antibiotics on the community. Stability analysis reveals that the microbiota is intrinsically stable, explaining how antibiotic perturbations and *C. difficile* inoculation can produce catastrophic shifts that persist even after removal of the perturbations. Importantly, we identify a subnetwork of bacterial groups implicated in protection against *C. difficile*, which may be used in the rational development of probiotic treatments or the design of antibiotic treatments that minimize harm to microbiota biodiversity.

Bio: João Xavier was trained as a multidisciplinary scientist. He originally studied chemical engineering at Instituto Superior Técnico in Portugal, and later, during his PhD at Universidade Nova de Lisboa, developed an interest in emergent properties of biological systems. João Xavier subsequently moved to the Delft University of Technology in The Netherlands for his postdoctoral work, where he developed a computer program of bacterial biofilms, which he applied to problems in environmental biotechnology such as wastewater treatment. He then decided to shift his research focus to more fundamental questions and pursued a second postdoctoral fellowship at Harvard University where he studied the evolution of cooperation using bacteria as model systems. In late 2009, he joined the faculty at Memorial Sloan-Kettering Cancer Center, where he investigates how bacteria interact in infection and how cell-cell interactions are implicated in cancer development. In 2011 he received the NIH Director's New Innovator Award, a program that funds provocative research projects by early stage investigators



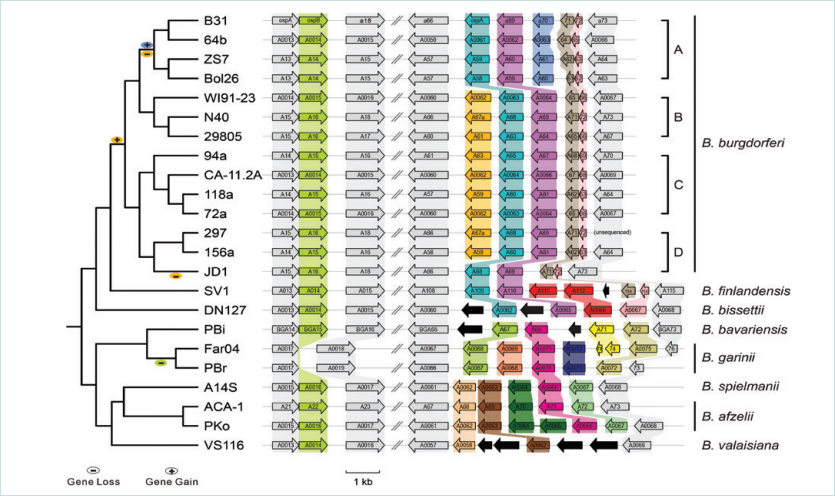
Weigang Qiu

Hunter College, City University of New York

Genomic Surveillance of Emerging Infectious Diseases

Abstract: Individual pathogen genomes carry unique genetic signatures consisting of single-nucleotide polymorphisms (SNPs) and DNA rearrangements. Improving upon traditional typing methods based on DNA sequences at single locus or multiple loci, whole-genome sequencing of bacterial and viral pathogens provides the ultimate resolution in tracing the origin and identifying the mechanisms of emerging infectious diseases. With the increasing availability of high-throughput genome sequencing technology, however, development of efficient bioinformatics tools has become a bottleneck for real-time genome-based surveillance of pathogens. Lyme disease, caused by the spirochetes *Borrelia burgdorferi* and transmitted by hard-bodied *Ixodes* ticks, is the most prevalent vector-borne disease in the US and Europe. The geographic range of Lyme disease continues to expand in association with climatic and ecological changes. We sequenced and compared the genomes of over twenty representative *Borrelia* strains from the US and Europe. Genome analyses revealed extensive genetic exchange among coexisting strains, genomic diversification driven by interactions with the host immunity, as well as cross-species genome hybridization. I will present bioinformatics tools we have been developing for comparative analyses of pathogen genomes including genome databases, genome-processing software tools, and a web-based genome browser. I will outline bioinformatics challenges and emphasize a phylogeny-based informatics framework for visualizing and analyzing pathogen genomes.

Bio: Weigang Qiu is an Associate Professor in the Department of Biological Sciences at Hunter College of the City University of New York. Qiu received his Ph.D. in Ecology and Evolution from the State University of New York at Stony Brook and was a Postdoctoral fellow in Bioinformatics at the University of Maryland Biotechnology Institute in Baltimore, Maryland. Qiu's research interest is in the population genomics of microbial species. The focus of his research is the comparative analysis of multiple genomes of the Lyme disease pathogen. The goals include reconstructing the history of worldwide diversification of the Lyme disease bacteria, an understanding of the mechanisms of their genome evolution (e.g., the roles of recombination and natural selection), and inference of gene and genome functions. His research approach is in bioinformatics, and more specifically, the computational and statistical testing of evolutionary hypotheses.



Feng Zhang

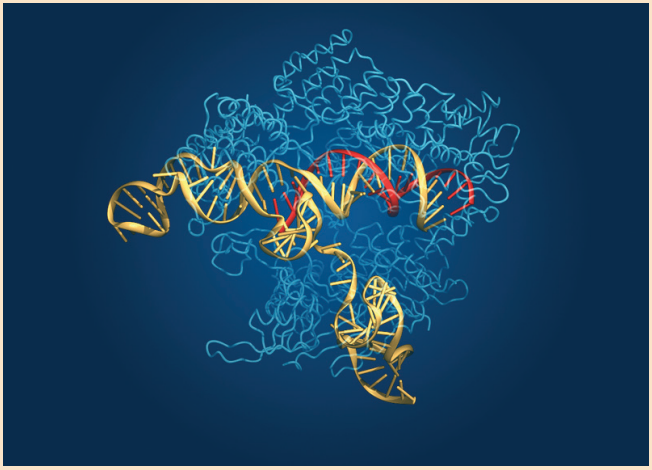
Keynote Speaker

Massachusetts Institute of Technology

Genome Engineering Using CRISPR-Cas9

Abstract: Functional elucidation of causal genetic variations and genetic elements requires precise genome manipulation technologies. We have recently developed a new class of eukaryotic genome engineering technology based on the bacterial CRISPR (clustered regularly interspaced short palindromic repeats) adaptive immune system. We reconstituted the CRISPR crRNA processing and interference system in mammalian cells and demonstrate that the Cas9 nuclease can be targeted to specific genomic loci by short crRNA guides to induce DNA double strand breaks. In a variety of cell types and species, Cas9 mediates editing of endogenous chromatin. Here we describe most recent advances for the Cas9 technology through interrogation and enhancement of targeting specificity, conversation of Cas9 into a modular DNA targeting domain, as well as application of the Cas9 system to probe gene function and genetic variations. Our results demonstrate the versatility of the RNA-guided CRISPR Cas9 nuclease system and open the possibility for efficient and multiplexed eukaryotic genome engineering for a variety of biomedical research and biotechnological applications.

Bio: Feng Zhang is a Core Member of the Broad Institute of MIT and Harvard and the W. M. Keck Career Development Professor of Biomedical Engineering at MIT. As a graduate student at Stanford University, Zhang worked with advisor Karl Deisseroth to invent a set of technologies for dissecting the functional organization of brain circuits. His lab works on developing and applying disruptive technologies including optogenetics and genome engineering (TALEs and CRISPR) to understand nervous system function and disease. Zhang's long-term goal is to develop novel therapeutic strategies for disease treatment. He obtained a bachelor's degree from Harvard University and a PhD in chemistry and bioengineering from Stanford University. Before joining the MIT faculty he was a junior fellow of the Harvard University Society of Fellows. He is a recipient of the Perl/UNC Prize in Neuroscience, the NIH Director's Pioneer award, and awards from the Searle Scholars Program, McKnight, Keck, and Damon Runyon foundations.



Symposium Planning Committee

SYMPOSIUM PLANNING COMMITTEE

Co-chair: **Paul Feinstein,** Hunter College of the City University of New York

Co-chair: **Weigang Qiu,** Hunter College of the City University of New York

Fabien Campagne, Weill Cornell Medical College

Hilary Monaco, Graduate Student, Memorial Sloan-Kettering Cancer Center

Joao Xavier, Memorial Sloan-Kettering Cancer Center

Denise Charles, Hunter College CTBR

Carlos Lijeron, Hunter College CTBR

PROGRAM LEADERSHIP

Jesus Angulo, PI/PD, Center for Translational and Basic Research, Professor of Biological Sciences, Hunter College, CUNY

Rodrigo Valles, Jr., Associate Program Director, Center for Translational and Basic Research, Hunter College, CUNY

Julianne Imperato-McGinley, Associate Dean of Translational Research, Weill Cornell Medical College

GENE CENTER STAFF

Coordinator: Denise Charles, Program Administrator for Communications and Outreach

Leah Abraha, Program Assistant and Evaluation Liaison

Megan Anderson, Collaborations and Development Coordinator

Richard Baldwin, Web Programmer

Vincent Cayenne, Web Programmer

Christine Gonzalez, Associate Program Manager

Carlos Lijeron, Assistant Director, Bioinformatics Facility

Jeanne Waxman, Program Manager

Shirley Yang, Program Coordinator

Notes

Notes

Notes

Notes

Notes

HUNTER
The City University of New York

Clinical & Translational
SCIENCE CENTER

Thank you for attending the 27th Annual International Symposium of the Hunter College Center for Translational and Basic Research.

This event is sponsored by Hunter College Center for Translational and Basic Research and Weill Cornell Medical College, Clinical & Translational Science Center.

Continue to follow us @
https://twitter.com/CTBRHunter

&

Like us on
https://www.facebook.com/CTBRatHunter

Designed by Stan Povelikin tel: 212.865.3759 web: www.spdesign.org