August 13-15, 2012
The Westin Providence
Providence, RI

NGX

Cambridge Healthtech Institute's Seventh Annual

Applying Next-Generation Sequencing

Delivering Genomes at Exponential Rates

NGX

Cambridge Healthtech Institute's Fifth Annual

Next-Generation Sequencing Data Analysis

Progress in the Pipeline

Premier Sponsor

genomatix

Corporate Sponsors

DNAnexus

DNASTAR® Software for life scientists

EdgeBio

EMC® ISILON

life technologies™

Corporate Support Sponsors

geneious

Quantum

Organized by:
Cambridge Healthtech Institute

healthtech.com/**Sequencing**

Short Courses*

Pre-Conference Short Courses
MONDAY, AUGUST 13

Short Course One

Sequencing 101

In less than five years next-generation sequencing has gone from radical to routine. Naturally, in a field moving so fast, newbies can feel overwhelmed. Sequencing 101 is designed to provide definitions in terminology, types of platforms, workflow strategies, and current applications, to what's next? Learn about the genomic data revolution from data generation to management to analysis and interpretation. This course is perfect for those needing an introduction to next-generation sequencing technology and applications.

Instructors:

Kevin Davies, Ph.D., Editor-in-Chief, Bio-IT World

Shawn C. Baker, Ph.D., CSO, BlueSEQ

Tom Schwei, Vice President and General Manager, DNASTAR, Inc.

Danielle Dionne, Senior Process Development Associate, Genomics Platform, The Broad Institute

Additional Instructor to be Announced

Short Course Two

A Practical Guide to Choosing the Right SNP Detection Tool

There is an urgent need for the NGS community to be able to make the right choice from available SNP detection tools. We have used exome-seq data to evaluate the quality of SNP calls from selected discovery tools from both academic and commercial products. We compared the quality of SNP calls from selected tools side by side and assessed their global quality by checking for Mendelian inheritance inconsistencies within the SNP calls of family members and by using the SNP array data from the same samples. This course will show the comparison results of these selected tools and provide a practical guide to choosing and using the right SNP detection tool.

Instructor: Ming Yi, Ph.D., IT Manager, Functional Genomics/Bioinformatics Support Group, Advanced Biomedical Computing Center, SAIC-Frederick, FNLCR

Dinner Short Courses
TUESDAY, AUGUST 14

Short Course Three

NGS, an Overview of Technologies, Applications, and Analysis Strategies

If you are confused by the evolving, rapidly changing, and dynamic nature of the NGS market, then this is the course for you. This course will provide you with a comprehensive, clear outline of the various sequencing technologies currently available, their best suited applications, and which analysis strategies go with what applications in order to get the best answer from the data.

Instructor: Matthias Scherf, Ph.D., Managing Director & Head, Scientific Development, Genomatix Software GmbH

**** Separate Registration Required***

NGX Applying Next-Generation Sequencing and Next-Generation Sequencing Data Analysis

Plenary Session

MONDAY, AUGUST 13

8:00 am Short Course Registration

8:30-11:30 Short Courses

1:00 pm Conference Registration

Plenary Session: Successful Sequencing Centers

2:00 Chairperson's Opening Remarks
Kevin Davies, Ph.D., Editor-in-Chief, Bio-IT World

2:10 New York Genome Center Making the Technology Work: A Collaborative Ecosystem for Improved Patient Outcomes
Nancy J. Kelley, J.D., M.P.P., Founding Executive Director, New York Genome Center

In order to achieve our ultimate goals of improved patient outcomes and the delivery of personalized medicine, we must create a collaborative framework around innovative sequencing technology that will drive success. Drawing on her experience with the creation and launch of the New York Genome Center, Ms. Kelley will share insights on creating the framework necessary for unleashing innovative technology to deliver on the promise of personalized medicine.

2:40 Washington University Dealing with Terabase-Scale Sequencing Projects
George Weinstock, Ph.D., Associate Director, The Genome Institute, Washington University

As human genetics, cancer genomics, and metagenomics all move toward use of next-generation sequencing as a primary data source, project scales have rapidly reached the terabase level. This increased data density involves new challenges at the bench and the computer. Examples of such projects will be given as well as innovations in methodology to take advantage of the expanded sequencing capability.

3:10 The Broad Institute High-Throughput Sample Preparation and Sequencing at the Broad Institute: Challenges & Opportunities
Sheila Fisher, Assistant Director, Technology Development, The Broad Institute

The Broad Institute processes more than 40,000 samples per year for sequencing across a variety of technology platforms. Here, we present an overview of the pipelines and process improvements as well as the infrastructure developed to manage the unique challenges of different sequencing applications at scale.

3:40 Refreshment Break

4:00 Cold Spring Harbor Next-Generation Sequencing in Case Control and Family Studies of Disease
W. Richard McCombie, Ph.D., Resource Head, Cancer Center; Professor, Cold Spring Harbor Laboratory

I will explain the various concepts of NGS and how the change in the technology has changed the way we can do biology and soon change the way medicine is carried out. In particular, I will discuss how we use both exome sequencing and whole genome sequencing on different populations in order to understand the genetic variation that underlies major psychiatric disorders. The exome sequencing approaches let us study large case control populations where we need to look at thousands of individuals to determine the genetic variation that makes them susceptible to psychiatric disorders. In addition we are extensively involved in whole genome sequencing where we query every nucleotide of the genome of multiple members of affected families.

4:30 Yale University Recent Advancements at the Yale Center for Genome Analysis
Shrikant Mane, Ph.D., Senior Research Scientist, Genetics; Senior Deputy Director, MBB Keck Biotech Services; Director, Microarray Resource, Yale University

5:00 J. Craig Venter Institute Human Genomes – Lessons from Africa
Vanessa M. Hayes, Ph.D., Professor, Genomic Medicine, J. Craig Venter Institute

Current genome information is heavily biased towards recently diverged populations outside of Africa, where Africans are largely excluded from studies based on large scale DNA analysis. Our research focus has been to obtain a better understanding of the indigenous peoples of Southern Africa. The ability to define the evolutionary history of modern humans, and in parallel the evolution of human adaptation and disease, has been dramatically advanced by the ability to contrast and compare patterns of DNA sequence variation between different contemporary populations.

5:30 Welcome Reception in the Exhibit Hall with Poster Viewing

6:30 Close of Day

Applying Next-Generation Sequencing

Delivering Genomes at Exponential Rates

August 13-15, 2012
The Westin Providence | Providence, RI

Next-generation sequencing (NGS) has expanded our fundamental understanding of genomics, mutation, gene function and regulation, by delivering fast, precise, and large volumes of data. As the cost of sequencing declines, use in academic, biotech, medical, pharmaceutical, and ag-bio labs expands, paving the way for the creation of innovative new areas for bioresearch. Still, the question remains of how best to balance cost, quantity and quality when choosing a sequencing platform or service. Cambridge Healthtech Institute's **NGx: Applying Next-Generation Sequencing** investigates the expanding applications of NGS. Learn from experienced researchers from large sequencing centers, core laboratories, and specialized groups as they share their practical knowledge, real-world experiences and solutions.

MONDAY, AUGUST 13

1:00 pm Conference Registration

2:00 PLENARY SESSION **(See Page 2)**

TUESDAY, AUGUST 14

7:30 am Breakfast Presentation
(Sponsorship Opportunity Available)

8:30 Java and Jive Discussion Groups

Tips 'n Tricks

9:30 Chairperson's Remarks
Anthony Shuber, Ph.D., CTO, Applied Science, Predictive Biosciences

9:35 A Catalog of Loss-of-Function Variants in Healthy Human Genomes: Implications for Clinical Diagnostics
Suganthi Balasubramanian, Ph.D., Associate Research Scientist, Molecular Biophysics and Biochemistry, Yale University
We have obtained a comprehensive catalog of loss-of-function variations in protein-coding genes based on the analysis of 1000 genomes data. This set of variants have been thoroughly examined for sequence artifacts, mapping errors and validated experimentally. This will provide a useful resource for clinical sequencing studies to identify and differentiate potential disease-causing mutations from benign variations. Given that LoF variants are highly likely to be deleterious, this is an invaluable resource.

10:05 Whole-Genome Amplification, Sequencing, and Haplotyping of Single Human Cells
Sijia Lu, Research Scientist in Xie Lab, Department of Chemistry and Chemical Biology, Harvard University
Existing whole genome amplification methods are hindered by non-uniformity across the genome and introduction of spurious sequences. Here we report a new amplification method, which circumvents problems by performing linear pre-amplification with Multiple Annealing and Looping Based Amplification Cycles (MALBAC) before exponential amplification with PCR. We demonstrate sequencing single human cells with ~93% coverage at ~30x depth. We further perform digital whole genome amplification by MALBAC on single human cells from a normal donor, which enables us to determine the haplotype of these cells. Finally, simultaneously determining the genome and transcriptome is possible for single cells.

10:35 Coffee Break in the Exhibit Hall with Poster Viewing

11:15 Sequencing Isolated Chromosomes from Single Cells
Marek Malecki, M.D., Ph.D., Associate Professor of Genetics, Genomics, and Gene Therapy, Western University of Health Sciences
The ultimate goal of our work is to develop targeted, personalized therapy. Therefore, we developed synthetic biotags, targeting specific molecules on the surfaces of cancer cells, which allow us to capture the cancer and cancer stem cells and their clonal expansion. We further developed synthetic probes targeting specific chromosomes, which allow us to isolate spectra of chromosomes from isolated cells. Using liquid FISH and NGS, in addition to native immunoblots and RT qPCR, we characterized molecular profiles of these cells. This novel approach opens new routes for molecular profiling of cancer and cancer stem cells.

11:45 A Cost Effective Assay for Multiplexing Protein and DNA Biomarkers on a Single Analytical Platform
Anthony Shuber, Ph.D., CTO, Applied Science, Predictive Biosciences
To eliminate the need to perform separate procedures for protein and DNA, and to decrease assay complexity and cost, we are developing a method for simultaneous detection of protein (Prot-Seq) and DNA (DNA-Seq) on a NGS platform. We have identified a DNA aptamer that binds Matrix Metalloprteinase-2 (MMP-2) and can be used as template for quantitative protein detection and multiplexed with DNA marker detection using the same NGS platform. Our goal is to simultaneously detect protein and DNA using a single analytical platform (MADR-Seq).

12:15 pm Close of Session

12:30 Luncheon Presentation
(Sponsorship Opportunity Available) **or Lunch on Your Own**

Conquering the Cancer Genome

2:00 Chairperson's Remarks

2:05 MI-ONCOSEQ - Experiences in Clinical Cancer Sequencing and the Path to Personalized Cancer Treatment
Dan Robinson, Ph.D., Director, Clinical Sequencing; Assistant Professor, Research, Michigan Center for Translational Pathology, Comprehensive Cancer Center, University of Michigan Health Systems
Presentation will discuss some of the hurdles in establishing next-generation sequencing protocols in a clinical setting. The utility of various sequencing strategies and the advantages of an integrated approach will be discussed. Examples of results impacting treatment decisions and illuminating tumor progression will be presented.

2:35 Cancer Characterization and Function
Adam Bass, M.D., Assistant Professor, Medicine, Harvard Medical School; Division of Cellular and Molecular Oncology, Dana-Farber Cancer Institute; Associate Member, The Broad Institute
The advent of NGS technologies has allowed us to perform comprehensive genome-wide surveys of the spectrum of mutations that accumulate in the cancer genome. I will discuss efforts to catalog and analyze the genomic alterations which exist across several types of gastrointestinal cancers. These studies have enabled us to identify novel genes which may contribute to these cancers. Additionally, they make us aware of the vast sea of genomic alterations present, creating the need to develop computational and experimental approaches to mine data to identify genes most important for cancer.

3:05 Incorporating NGS into a Gynecologic Cancer Program for Discovery of Novel Diagnostic Biomarkers and Therapeutic Targets - The MSSM Experience
John A. Martignetti, M.D., Ph.D., Associate Professor, Departments of Genetics and Genomic Sciences, Pediatrics, Oncological Sciences, Mount Sinai School of Medicine

3:35 Selected Poster Presentation: Alternative Transcripts: Perspectives on Associating Functions to Genes
Melanie Lehman, Ph.D., Research Fellow, Australian Prostate Cancer Research Centre, Queensland University of Technology

3:50 Refreshment Break in the Exhibit Hall with Poster Viewing

RNA-Seq

4:30 Using NGS Transcriptomic Tools in Non-Model Organisms
Suzy Strickler, Ph.D., Postdoctoral Researcher, Boyce Thompson Institute for Plant Research, Cornell University
Transcriptomic studies can benefit greatly from data generated using next-generation sequencing (NGS), even in non-model organisms. By applying tools developed for NGS to transcriptomic sequence data from non-model species, we can begin to elucidate gene expression in these organisms. There are various tools currently available for transcriptome assembly and analysis and some of these tools will be reviewed. Detecting differential expression, identifying SNPs, and performing annotation will be discussed using examples from the literature and data analysis performed in our lab.

FEATURED SPEAKER

5:00 Using RNA-Seq to Uncover Isoforms and Molecular Networks Involved in Nerve Regeneration
Vance Lemmon, Ph.D., Walter G. Ross Distinguished Chair in Developmental Neuroscience; Professor, Neurological Surgery, The Miami Project to Cure Paralysis, Center for Computational Sciences, University of Miami Miller School of Medicine
Most neurons in the central nervous system (CNS) are unable to re-grow axons that have been cut, yet neurons in the peripheral nervous system (PNS) have this ability. We have used RNA-seq to compare CNS and PNS neurons to identify differentially expressed mRNA isoforms and use this information to explore transcriptional and miRNA regulatory networks for subsequent experimental analysis using over expression and knock-down in neurons, combined with high content analysis.

5:30 Close of Day and Dinner Short Course Registration

5:45 – 8:45 Dinner Short Courses
(See Page 2)

WEDNESDAY, AUGUST 15

7:30 am Breakfast Presentation or Morning Coffee *(Sponsorship Opportunity Available)*

What's Next in Sequencing Platforms?

8:30 Chairperson's Remarks
Tim Harkins, Ph.D., Director of R&D, Advanced Applications & Collaborations, Life Technologies

Nanopore Sequencing

8:35 Nanopore Graphene-Based Electronic Devices
Marija Drndic, Ph.D., Department of Physics and Astronomy, University of Pennsylvania
Graphene is an exceptional material for high-speed electronics, as well as a revolutionary membrane material due to its strength and atomic thickness. Nanopores in suspended graphene membranes are currently regarded as candidates for ultrafast DNA sequencing. When a single DNA molecule passes through a nanopore, it blocks the field-driven ions passing through the pore and is detected by measuring the ion current reduction. Due to the thin nature of graphene membranes and reduced pore resistance, we observe larger current signals than in the case of traditional solid-state nanopores. Use of graphene as a membrane material opens the door to a new class of nanopore devices in which electronic sensing and control are performed directly at the pore.

9:05 Noise and Bandwidth Optimization for Nanopore Sensing at Sub-Microsecond Timescales
Jacob Rosenstein, Research Scientist, Department of Electrical Engineering, Columbia
We describe the integration of solid-state nanopores with custom low-noise complementary metal-oxide-semiconductor (CMOS) circuitry, in which the electronics are directly exposed to the electrolyte. As a result of this close integration and careful design, sub-microsecond temporal resolution has been demonstrated with nanopore recordings. Speeding up the electronics, complements recent efforts to control and "slow down" DNA translocation, leading to faster and more accurate reads.

9:35 Solving the Challenges of DNA Sequencing for Molecular Diagnostics
Stefan Roever, CEO & Founder, Genia Technologies
Genia has developed a versatile nanopore-based platform which allows for single molecule, electrical, real-time analysis without the need for complicated optics, labels, amplification, or fluidics. The sensor itself is truly transformative and allows very small electrical signals (~0.2 pA current levels) to be seen high above the noise floor. The data shows that with highly accurate analog electronics and clever data analysis techniques, single base discrimination is possible, and adequate SNR can be reached to perform DNA sequencing.

10:05 Technology Showcase HLA Typing from Genomic Data Using Omixon Target
Attila Berces, Ph.D., CEO, Omixon Biocomputing
Next generation sequencing offers inexpensive determination of the HLA type at the highest possible resolution, but the analysis is challenging. I show results obtained with Omixon Target from targeted amplification, HLA-seq kit, exome kits and whole genome data from Illumina, Ion Torrent and Roche454 sequencers and validation on HapMap samples.

10:20 Coffee Break in the Exhibit Hall with Poster Viewing

Desk-Top Sequencers

11:00 Scientific Advancements using Ion Torrent's PGM *Sponsored by life technologies*
Tim Harkins, Ph.D., Director of R&D, Advanced Applications & Collaborations, Life Technologies
Dr. Timothy Harkins will discuss several emerging applications performed using the PGM. The PGM is a bench top, next generation sequencer that is able to scale in throughput to meet the individual needs of a project. This presentation will cover several advancements from low frequency variant detection in cancer to shotgun metagenomics to *de novo* sequencing using long read chemistries.

11:30 Towards Enterprise-Scale Genomics: The Gene Partnership Project at Boston Children's Hospital
David Margulies, M.D., Executive Director, The Gene Partnership, Children's Hospital Boston
This talk will present the approach, priorities, and early experiences as Boston Children's Hospital and the HMS Center for Biomedical Informatics seek to deploy genomic measurement and interpretive methods to the full array of clinical and research applications. Policies supporting large scale recruitment of patients and families will be discussed.

12:00 pm Luncheon Presentation *(Sponsorship Opportunity Available)* **or Lunch on Your Own**

Structural Variation

1:45 Chairperson's Remarks
Kevin Davies, Ph.D., Editor-in-Chief, Bio-IT World

1:50 Rapid Detection and Characterization of Structural Variation in Hundreds of Human Genomes
Ira M. Hall, Ph.D., Assistant Professor, Department of Biochemistry and Molecular Genetics; Center for Public Health Genomics, University of Virginia
Structural variation (SV) is a major source of genomic diversity in mammals, but accurate detection of SV from NGS data remains a challenge. Here, we present a multi-sample version of our breakpoint detection algorithm, HYDRA, that can simultaneously detect and genotype SV breakpoints in hundreds to thousands of whole genome sequence datasets while using minimal RAM. We further outline a general framework for the characterization

of both simple and complex SV in cancer genomes, and present results from an initial study of 64 tumor/normal pairs.

2:20 Discovery and Functional Impact of Structural Variation across 1000 Genomes

Ryan E. Mills, Ph.D., Assistant Professor, Department of Computational Medicine and Bioinformatics; Department of Human Genetics, University of Michigan Medical School

Genomic structural variants (SVs) are a poorly understood class of genetic variation. Through the 1000 Genomes Project, we have analyzed sequence data for 1,092 individuals across 14 populations and have precisely identified many variable regions that have been assessed for functional implications. We have additionally constructed integrated maps of genetic variation incorporating a subset of high quality SNP, INDEL, and deletion variants in these individuals onto phased haplotypes. These data sets will help further our understanding of the formation and prevalence of common and rare chromosomal rearrangements, informing future studies investigating the impact of such variation in human health and disease phenotypes.

2:50 Refreshment Break, Last Chance for Exhibit and Poster Viewing

3:15 Poster Awards

Sponsored by Quantum

3:30 Combining Effects from Rare and Common Genetic Variants in Exome-Wide Association Study of Sequence Data

Hugues Aschard, Ph.D., Research Fellow, Department of Epidemiology, School of Public Health, Harvard University

Next-generation sequencing has the potential to revolutionize complex trait genetics by directly measuring common and rare genetic variants within a genome-wide context. Strategies that model the effects of both common and rare variants could enhance the power of identifying disease-associated genes. We evaluated various strategies for association of rare, common, or a combination of both rare and common variants on quantitative phenotypes in unrelated individuals using the Genetic Analysis Workshop 17 data.

4:00 Taking NGS into the Clinic

Gholson J. Lyon, M.D., Ph.D., Assistant Professor, Human Genetics, Cold Spring Harbor Laboratory; Research Scientist, Utah Foundation for Biomedical Research; Adjunct Assistant Professor, Psychiatry, New York University Child Study Center

For the implementation of genomic analysis in the clinic, it will be critically important to optimize and standardize pipelines with high sensitivity and specificity for variant calling. This includes having sufficient sequencing depth as well as high quality of sequencing data. One way to reduce false positives could be to use variants called by two or more variant calling approaches on one set of sequencing data. We are using various tools such as ANNOVAR and VAAST to prioritize variants from various families for follow-up in case-control and biological studies.

FEATURED SPEAKER

4:30 Tools for Processing Next-Generation Sequencing Data

Mark Gerstein, Ph.D., Albert L. Williams Professor of Biomedical Informatics, Molecular Biophysics and Biochemistry, Computer Science, Yale University

A central problem for 21st century science is annotating the human genome and making this annotation useful for the interpretation of personal genomes. My talk will focus on annotating the bulk of the genome that does not code for canonical genes, concentrating on intergenic features such as TF binding sites and non-coding RNAs (ncRNAs) and structural variations and pseudogenes. Much of this work has been carried out within the ENCODE and 1000 Genomes projects.

5:15 Close of Conference

NGX

Cambridge Healthtech Institute's Fifth Annual

Next-Generation Sequencing Data Analysis

Progress in the Pipeline

August 13-15, 2012

The Westin Providence | Providence, RI

Sequencing a genome is only the beginning. Several layers of analysis are necessary to convert raw sequence data into understanding of functional biology. First, error sources in the original raw data from multiple platforms and diverse applications must be accounted for. Then, as computational methods for assembly, alignment, and variation detection continue to advance, a broad range of genetic analysis applications including comparative genomics, high-throughput polymorphism detection, analysis of coding and non-coding RNAs, and identifying mutant genes in disease pathways can be addressed. CHI's **Next-Generation Sequencing Data Analysis** conference combines unique perspectives from a variety of researchers, engineers, biostatisticians, and software developers involved in NGS data analysis

MONDAY, AUGUST 13

1:00 pm Conference Registration

2:00 PLENARY SESSION (See Page 2)

TUESDAY, AUGUST 14

7:30 am Breakfast Presentation *(Sponsorship Opportunity Available)*

8:30 Java and Jive Discussion Groups (See page 3 for details)

Assembly's Role in Gene Annotation

9:30 Chairperson's Remarks

Herve Tettelin, Ph.D., Associate Professor, Department of Microbiology and Immunology, Institute for Genome Sciences, University of Maryland

9:35 Improving Pan-Genome Annotation Using Whole Genome Multiple Alignment

Herve Tettelin, Ph.D., Associate Professor, Department of Microbiology and Immunology, Institute for Genome Sciences, University of Maryland

Rapid annotation and comparisons of genomes from multiple isolates (pan-genomes) is becoming commonplace due to advances in sequencing technology. Genome annotations can contain inconsistencies and errors that hinder comparative analysis even within a single species. We developed Mugsy-Annotator that identifies orthologs and evaluates annotation quality in prokaryotic genomes using whole genome multiple alignment. Mugsy-Annotator identifies anomalies in annotated gene structures, including inconsistently located translation initiation sites and disrupted genes due to draft genome sequencing or pseudogenes.

10:05 Written in the Genome: Assembly Consequences for Gene Annotation

Liliana Florea, Ph.D., Assistant Professor, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine

As a growing number of species are being sequenced and assembled to various degrees of completion, a key question is: how does the quality of the assembled sequence affect gene annotations? We found that even between successive high-quality draft assemblies there were significant changes in the annotations, due primarily to genome mis-assembly events and local sequence variation. We illustrate with a comparison of annotations in assemblies produced with different methods, from reads generated by diverse sequencing technologies, and undergoing various amounts of curation.

10:35 Coffee Break in the Exhibit Hall with Poster Viewing

11:15 Identifying Genomic Features by Chromatin Architecture

Michael J. Buck, Ph.D., Assistant Professor, Department of Biochemistry, The State University of New York at Buffalo

The accurate identification and annotation of genomic features is a key to understanding the human genome. However, finding these locations within the genome can be a laborious and expensive undertaking requiring site-specific assays. Even more difficult is identifying entirely new classes of genomic features. To facilitate identification and characterization of new classes of genomic features, we have developed chromatin architecture alignment and search tools. We have successfully applied these tools to various datasets from multiple genomes and will present our results for insulators, enhancers, transcriptional start sites, and origins of replication.

11:45 Evaluating NGS Factors Affecting Prediction Performance, and Exploring Feasibility of NGS Meta Analysis

May Dongmei Wang, Ph.D., Associate Professor, Biomedical Engineering, Electrical and Computer Engineering, Hematology and Oncology, Winship Cancer Institute, Georgia Institute of Technology

We have been developing computing algorithms and infrastructure to evaluate among various parameters involved in NGS data analysis, which factors may affect prediction performance. In addition, we have been exploring whether meta-analysis can be used in NGS data analysis.

12:15 pm A Community-Inspired Collaborative and Scalable Data Technology Platform

Sponsored by DNAnexus

Andreas Sundquist, Ph.D., CEO & Co-founder, DNAnexus

DNAnexus' instant genomics data and analysis center is a novel, open, and flexible cloud-based platform for DNA data management, analysis, and visualization. It allows collaboration in a secure environment and the integration of own and external tools into custom workflows to make any DNA data analysis a success.

12:30 Luncheon Presentation *(Sponsorship Opportunity Available)* **or Lunch on Your Own**

Assembly and Alignment's Role in Analysis

2:00 Chairperson's Remarks

Tom Schwei, Vice President and General Manager, DNASTAR, Inc.

2:05 Assessing and Improving the Reliability of Short Read Alignments

Matthew Ruffalo, Research Scientist, Department of Electrical Engineering and Computer Science, Case Western Reserve University

As the developers of alignment tool software optimize their algorithms with respect to various considerations, the relative merits of different software packages remain unclear. However, for researchers who generate and use NGS data for their specific projects, an important consideration is choosing software that is most suitable for their application. Here, we identify criteria to evaluate the performance of alignment algorithms, compare various algorithms in terms of performance with respect to criteria, and discuss how their output can be used to accurately assess the reliability of mappings.

2:35 Performance Metrics in the Cloud: A Storage Perspective

Sponsored by EMC® ISILON

Sanjay Joshi, Solutions Architect, Life Sciences, EMC Isilon Storage Division

As the process of assembling and interpreting genomic data moves to the Cloud, understanding its performance dictates operational costs, resources and planning for growth within a storage context. We present the most common use cases for NGS on the Cloud and their performance guidelines and best practices.

3:05 Improving Structural Variation Analysis with Next-Generation DNA Sequencing Data

Anna Ritz, Research Scientist, Department of Computer Science & Center for Computational Molecular Biology, Brown University

Detecting structural variants (SVs) in mammalian genomes using next-generation DNA sequencing data is complicated by the fact that many SVs lie in repetitive regions. Sequence reads from such regions have numerous possible alignments to the reference genome, and determining whether ambiguous alignments result from an SV is challenging. We describe probabilistic algorithms for SV detection that combine multiple signals indicating the presence of a variant. We demonstrate the results of our algorithms on Illumina and Pacific Biosciences sequencing data.

3:35 Towards a Human Clinical Grade Genome – EdgeBio and The Archon Genomics X PRIZE presented by Medco

Sponsored by EdgeBio

Justin Johnson, Director, Bioinformatics, EdgeBio

The Archon Genomics X PRIZE presented by Express Scripts created a "Validation Protocol" that is helping to define for the first time what it means to have a complete and accurate "medical grade" whole human genome sequence. This presentation will describe the Validation protocol in detail.

3:50 Refreshment Break in the Exhibit Hall with Poster Viewing

4:30 How Basic Assembly Parameters Affect *de novo* and Reference-Guided Next-Gen Sequence Assembly

Matthew R. Keyser, Next-Gen Applications Scientist, DNASTAR, Inc.

Most sequence assembly software provide adjustable parameters that can dramatically affect assembly quality and commonly used benchmarks like assembly time and contig size. However, assemblies that are optimized for speed and contig size often show a decrease in assembly quality that may impede downstream analysis. This discussion will investigate how adjusting basic parameters affects assembly quality and performance in both *de novo* and reference-guided projects, and the need to use assembly software that provides a proper balance.

5:00 Assembly and Alignment Panel Discussion with Afternoon Speakers

Moderator: Tom Schwei, Vice President and General Manager, DNASTAR, Inc.

5:30 Close of Day and Dinner Short Course Registration

5:45-8:45 Dinner Short Courses

WEDNESDAY, AUGUST 15

7:30 am Breakfast Presentation or Morning Coffee *(Sponsorship Opportunity Available)*

Analysis: RNA-Seq

8:30 Chairperson's Remarks

8:35 Global Analysis of RNA Alternative Splicing by Deep Sequencing

Yi Xing, Ph.D., Assistant Professor, Department of Internal Medicine; Department of Biomedical Engineering, University of Iowa

Ultra-deep RNA sequencing (RNA-Seq) has become a powerful approach for genome-wide analysis of pre-mRNA alternative splicing. We developed MATS (multivariate analysis of transcript splicing), a flexible statistical framework for detecting differential alternative splicing patterns from RNA-Seq data. Using RNA-Seq, we investigated the complete spectrum of alternative splicing events that are regulated by the epithelium-specific splicing regulatory proteins ESRP1 and ESRP2. Our study provides a comprehensive picture of an extensive epithelial alternative splicing program in development and disease.

9:05 Design Considerations for Measuring Differential Gene Expression Using RNA Sequencing Data

Michelle Busby, Research Scientist in Marth Lab, Biology, Boston College

An RNA-Seq experiment's ability to detect differential gene expression is based on the ability to distinguish true, biological differences between conditions from the variability that occurs in repeated measurements of the same condition. We will demonstrate how to quantify these sources of variability in existing data sets and show how this can be used to design experiments with adequate number of samples sequenced to a sufficient depth, both cost-effectively, and with reproducible results.

9:35 Selected Poster Presentations: RNA-Seq and MicroRNA Expression Profiling Reveal Networks of RNA Interactions in Regenerating Dorsal Root Ganglion Neurons

Dario Motti, Ph.D., Postdoctoral Fellow, The Miami Project to Cure Paralysis, University of Miami, Miller School of Medicine

9:50 AssemEval: Illumina RNA-Seq Assembly Evaluation; Pitfalls and Opportunities for Non-Model Organisms

Shaadi F. Pooyaei Mehr, Research Scientist, American Museum of Natural History

10:05 NGS Data Analysis and Interpretation – From Raw Sequences to Clinical Relevance

Sponsored by genomatix

Martin Seifert, Ph.D., CEO, Genomatix

While NGS wet-lab protocols have become straightforward, data analysis remains difficult, with BioIT specialists often sacrificing ease-of-use for efficiency. Quality NGS data interpretations require intuitive tools and extensive background information. We will present an example of a Genomatix integrated solution for gene-fusion analysis that led to a clinically relevant result.

10:20 Coffee Break in the Exhibit Hall with Poster Viewing

11:00 RNA-Seq Analysis of Citrus Mature Leaves and Young Seedlings

Sponsored by DNASTAR

Chunxian Chen, Ph.D., Associate, Citrus Genomics, University of Florida, IFAS, Citrus Research and Education Center

RNA-seq analysis of 454 data from citrus mature leaves and young seedlings were performed using DNASTAR's QSeq. Among the genes aligned with at least one read in both experiments (R2 = 0.6377), 6515 genes showed expression change with 2-fold, 1308 with 4-fold, and 254 with 8-fold. Selected validation is underway.

11:30 Self-Training Algorithm for Splice Junction Detection Using RNA-seq

Jian Ma, Ph.D., Assistant Professor of Bioengineering, Assistant Professor of Biophysics and Computational Biology Affiliate, Assistant Professor of Computer Science Institute for Genomic Biology, University of Illinois at Urbana-Champaign

Accurately mapping RNA-seq reads to splice junctions is critically important for the analysis of alternative splicing and isoform construction. We developed a novel method, called TrueSight, utilizing semi-supervised learning to precisely identify splice junctions. Both simulation and real data evaluations showed that TrueSight achieved higher sensitivity and specificity than other tools. We believe this new tool will be highly useful to comprehensively study splice variants based on RNA-seq.

12:00 pm Luncheon Presentation *(Sponsorship Opportunity Available)* **or Lunch on Your Own**

Analysis: Structural Variation

1:45 Chairperson's Remarks

Kevin Davies, Ph.D., Editor-in-Chief, Bio-IT World

1:50 Rapid Detection and Characterization of Structural Variation in Hundreds of Human Genomes

Ira M. Hall, Ph.D., Assistant Professor, Department of Biochemistry and Molecular Genetics; Center for Public Health Genomics, University of Virginia

Structural variation (SV) is a major source of genomic diversity in mammals, but accurate detection of SV from NGS data remains a challenge. Here, we present a multi-sample version of our breakpoint detection algorithm, HYDRA, that can simultaneously detect and genotype SV breakpoints in hundreds to thousands of whole genome sequence datasets while using minimal RAM. We further outline a general framework for the characterization of both simple and complex SV in cancer genomes, and present results from an initial study of 64 tumor/normal pairs.

2:20 Discovery and Functional Impact of Structural Variation across 1000 Genomes

Ryan E. Mills, Ph.D., Assistant Professor, Department of Computational Medicine and Bioinformatics; Department of Human Genetics, University of Michigan Medical School

Genomic structural variants (SVs) are a poorly understood class of genetic variation. Through the 1000 Genomes Project, we have analyzed sequence data for 1,092 individuals across 14 populations and have precisely identified many variable regions that have been assessed for functional implications. We have additionally constructed integrated maps of genetic variation incorporating a subset of high quality SNP, INDEL, and deletion variants in these individuals onto phased haplotypes. These data sets will help further our understanding of the formation and prevalence of common and rare chromosomal rearrangements, informing future studies investigating the impact of such variation in human health and disease phenotypes.

2:50 Refreshment Break, Last Chance for Exhibit and Poster Viewing

3:15 Poster Awards

Sponsored by Quantum

3:30 Combining Effects from Rare and Common Genetic Variants in Exome-Wide Association Study of Sequence Data

Hugues Aschard, Ph.D., Research Fellow, Department of Epidemiology, School of Public Health, Harvard University

Next-generation sequencing has the potential to revolutionize complex trait genetics by directly measuring common and rare genetic variants within a genome-wide context. Strategies that model the effects of both common and rare variants could enhance the power of identifying disease-associated genes. We evaluated various strategies for association of rare, common, or a combination of both rare and common variants on quantitative phenotypes in unrelated individuals using the Genetic Analysis Workshop 17 data.

4:00 Taking NGS into the Clinic

Gholson J. Lyon, M.D., Ph.D., Assistant Professor, Human Genetics, Cold Spring Harbor Laboratory; Research Scientist, Utah Foundation for Biomedical Research; Adjunct Assistant Professor, Psychiatry, New York University Child Study Center

For the implementation of genomic analysis in the clinic, it will be critically important to optimize and standardize pipelines with high sensitivity and specificity for variant calling. This includes having sufficient sequencing depth as well as high quality of sequencing data. One way to reduce false positives could be to use variants called by two or more variant calling approaches on one set of sequencing data. We are using various tools such as ANNOVAR and VAAST to prioritize variants from various families for follow-up in case-control and biological studies.

FEATURED SPEAKER

4:30 Tools for Processing Next-Generation Sequencing Data

Mark Gerstein, Ph.D., Albert L. Williams Professor of Biomedical Informatics, Molecular Biophysics and Biochemistry, Computer Science, Yale University

A central problem for 21st century science is annotating the human genome and making this annotation useful for the interpretation of personal genomes. My talk will focus on annotating the bulk of the genome that does not code for canonical genes, concentrating on intergenic features such as TF binding sites and non-coding RNAs (ncRNAs) and structural variations and pseudogenes. Much of this work has been carried out within the ENCODE and 1000 Genomes projects.

5:15 Close of Conference

Sponsorship & Exhibit Opportunities

CHI offers comprehensive sponsorship packages which include presentation opportunities, exhibit space and branding, as well as the use of the pre and post show delegate lists. Sponsorship allows you to achieve your objectives before, during, and long after the event. Any sponsorship can be customized to meet your company's needs and budget. Signing on earlier will allow you to maximize exposure to hard-to-reach decision makers.

Agenda Presentations

Showcase your solutions to a guaranteed, highly-targeted audience. Package includes a 15 or 30-minute podium presentation within the scientific agenda, exhibit space, on-site branding and access to cooperative marketing efforts by CHI.

Breakfast & Luncheon Presentations

Opportunity includes a 30-minute podium presentation. Boxed lunches are delivered into the main session room, which guarantees audience attendance and participation. A limited number of presentations are available for sponsorship and they will sell out quickly. Sign on early to secure your talk!

Invitation-Only VIP Dinner/Hospitality Suite

Sponsors will hand-pick their top prospects from the conference pre-registration list for an evening of networking at the hotel or at a choice local venue. CHI will extend invitations and deliver prospects. Evening will be customized according to sponsor's objectives (i.e. purely social, focus group, reception style or plated dinner, plated dinner with specific conversation focus).

CHI Lead Generation

CHI can help you with lead generation throughout the year. Our internal database includes over 800,000 prospects in the life sciences. By leveraging the database and mining for your specific requirements, we can produce multiple custom projects which will deliver your prospective buyers:

- Web Symposia
- Podcasts
- White Papers
- Custom Market Research Surveys
- And More!

Exhibit Information

Exhibitors will enjoy facilitated networking opportunities with qualified decision makers and end-users, making it the perfect platform to launch a new product, collect feedback and generate new leads. Exhibit space sells out quickly, so reserve yours today!

Additional Networking & Promotional Opportunities Are Available!

For sponsorship & exhibit information, please contact:
Jay Mulhern
Manager, Business Development
781-972-1359 | jmulhern@healthtech.com

Hotel & Travel Information

Conference Hotel:

The Westin Providence
One West Exchange Street
Providence, RI 02903
www.westin.com/Providence
T: 401-598-8000

Discounted Room Rate: $159 s/d
Discounted Room Rate Cut-off Date: July 20, 2012

Please visit our conference website to make your reservations online or call the hotel directly to reserve your sleeping accommodations. You will need to identify yourself as a Cambridge Healthtech Institute conference attendee to receive the discounted room rate with the host hotel. **Reservations made after the cut-off date or after the group room block has been filled (whichever comes first) will be accepted on a space- and rate-availability basis. Rooms are limited, so please book early.**

Flight Discounts:
Special discounts have been established with American Airlines for this conference.

- Call American Airlines 1-800-433-1790 and use Conference code 9482BG.
- Go to www.aa.com/group and enter Conference code 9482BG in promotion discount box.
- Contact our dedicated travel agent, Wendy Levine at 1-877-559-5549 or wendy.levine@protravelinc.com.

Car Rental Discounts:
Special rental discounts have been established with Hertz for this conference.

- Call Hertz 1-800-654-3131 and use our Hertz Convention Number (CV): 04KL0003
- Go to www.hertz.com and use our Hertz Convention Number (CV): 04KL0003

NGX

Applying Next-Generation Sequencing

and

Next-Generation Sequencing Data Analysis

Pricing and Registration Information

SHORT COURSE PRICING

	Commercial	Academic, Government, Hospital-affiliated
One short course	$695	$395
Two short courses	$995	$695

Monday, August 13th	Tuesday, August 14th
Sequencing 101	NGS, an Overview of Technologies, Applications, and Analysis Strategies
A Practical Guide to Choosing the Right SNP Detection Tool	

CONFERENCE PRICING

(Includes access to 1 conference, excludes short courses)

Registrations after July 13, 2012, and on-site	$2195	$945

Applying Next-Generation Sequencing **OR** **Next-Generation Sequencing Data Analysis**

CONFERENCE DISCOUNTS

Poster Submission-Discount ($50 Off)
Poster abstracts are due by July 13, 2012. Once your registration has been fully processed, we will send an email containing a unique link allowing you to submit your poster abstract. If you do not receive your link within 5 business days, please contact jring@healthtech.com. *CHI reserves the right to publish your poster title and abstract in various marketing materials and products.

REGISTER 3 - 4th IS FREE: Individuals must register for the same conference or conference combination and submit completed registration form together for discount to apply.

Additional discounts are available for multiple attendees from the same organization. For more information on group rates contact David Cunningham at +1-781-972-5472

CAMBRIDGE HEALTHTECH MEDIA GROUP

Receive a FREE eNewsletter by signing up at chimediagroup.com

Weekly Update
The latest industry news, commentary and highlights from Bio-IT World

eCliniqua
Innovative management in clinical trials

INSIGHT PHARMA REPORTS

A series of diverse reports designed to keep life science professionals informed of the salient trends in pharmaceutical technology, business, clinical development, and therapeutic disease markets. For a detailed list of reports, visit InsightPharmaReports.com, or contact Rose LaRaia, rlaraia@healthtech.com, +1-781-972-5444.

BARNETT EDUCATIONAL SERVICES

Barnett is a recognized leader in clinical education, training, and reference guides for life science professionals involved in the drug development process. For more information, visit barnettinternational.com.

ADDITIONAL REGISTRATION DETAILS

Each registration includes all conference sessions, posters and exhibits, food functions, and access to the conference proceedings link.

Handicapped Equal Access: In accordance with the ADA, Cambridge Healthtech Institute is pleased to arrange special accommodations for attendees with special needs. All requests for such assistance must be submitted in writing to CHI at least 30 days prior to the start of the meeting.

To view our Substitutions/ Cancellations Policy, go to http://www.healthtech.com/regdetails

Video and or audio recording of any kind is prohibited onsite at all CHI events.

If you are unable to attend but would like to purchase the NGx Sequencing CD for $750 (plus shipping), please visit healthtech.com/Sequencing. Massachusetts delivery will include sales tax.

How to Register: **healthtech.com/Sequencing**

reg@healthtech.com • P: 781.972.5400 or Toll-free in the U.S. 888.999.6288

Please use keycode **DTD F** when registering!

Please refer to the Registration Code below:

Cambridge Healthtech Institute
250 First Avenue, Suite 300, Needham, MA 02494
www.healthtech.com • Fax: 781-972-5425