Click Here to Register Online!
Bio-ITWorldExpo.com

Cambridge Healthtech Institute's Fourteenth Annual

Bio-IT World CONFERENCE & EXPO '15

Enabling Technology. Leveraging Data. Transforming Medicine.

APRIL 21 – 23, 2015
SEAPORT WORLD TRADE CENTER
BOSTON, MA

CONFERENCE TRACKS:

1. IT Infrastructure – Hardware
2. Software Development
3. Cloud Computing
4. Bioinformatics
5. Next-Gen Sequencing Informatics
6. Clinical & Translational Informatics
7. Data Visualization & Exploration Tools
8. Pharmaceutical R&D Informatics
9. Clinical Genomics
10. Collaborations & Open Access Innovations
11. Cancer Informatics
12. Data Security

PLENARY SESSION SPEAKERS:

Philip E. Bourne, Ph.D.
Associate Director for Data Science (ADDS), National Institutes of Health

Chris Sander, Ph.D.
Computational and Systems Biology, Memorial Sloan Kettering Cancer Center

Benjamin Heywood
Co-Founder and President, PatientsLikeMe, Inc.

Andreas Kogelnik, M.D., Ph.D.
Founder, Open Medicine Institute

Katherine Wendelsdorf, Ph.D.
Field Application Scientist - Ingenuity Systems, QIAGEN Bioinformatics; Spokesperson, Empowered Genome Community

Register Early for Maximum Savings

EVENT FEATURES:

- **Access All 12 Tracks for One Price**
- **Network with 3,000+ Global Attendees**
- **Hear 150+ Technology and Scientific Presentations**
- **Attend Bio-IT World's Best Practices Awards**
- **Connect with Attendees Using CHI's Intro-Net**
- **Participate in the Poster Competition**
- **Choose from 17 Pre-Conference Workshops**
- **See the Winners of the following 2015 Awards:**
  - **Benjamin Franklin**
  - **Best of Show**
  - **Best Practices**
- **View Novel Technologies and Solutions in the Expansive Exhibit Hall**
- **And Much More!**

Platinum Sponsors:

aspera an IBM company · CYCLE COMPUTING · DDN STORAGE · ELSEVIER · EMC² · IBM · illumina · intel · okta · ORACLE HEALTH SCIENCES · SEAGATE · sgi · THINKMATE HIGH PERFORMANCE COMPUTING · THOMSON REUTERS

Organized by:
Cambridge Healthtech Institute

Bio-ITWorldExpo.com

Click Here to Register Online!
Bio-ITWorldExpo.com

Organized by:
Cambridge Healthtech Institute

SCHEDULE-AT-A-GLANCE

Tuesday, April 21, 2015

8:00am – 4:00pm	Pre-Conference Workshops
4:00 – 5:00pm	Plenary Session Presentation
5:00 – 7:00pm	Exhibit Hall Open
5:00 – 7:00pm	Welcome Reception in the Exhibit Hall with Poster Viewing

Wednesday, April 22, 2015

8:00 – 9:45am	Plenary Session, Benjamin Franklin Award Presentation, and Best Practices Awards Program
9:45am – 6:30pm	Exhibit Hall Open
9:45 – 10:50am	Coffee Break in the Exhibit Hall with Poster Viewing
10:50am – 12:30pm	Tracks 1-12
12:40 – 1:40pm	Luncheon Presentations (Sponsorship Opportunities Available)
1:50 – 3:25pm	Tracks 1-12
3:25 – 4:00pm	Refreshment Break in the Exhibit Hall with Poster Viewing
4:00 – 5:30pm	Tracks 1-12
5:30 – 6:30pm	Best of Show Awards Reception in the Exhibit Hall

Thursday, April 23, 2015

7:00 –7:50am	Breakfast Presentations (Sponsorship Opportunities Available)
8:00 – 10:00am	Plenary Session Panel
10:00am – 1:55pm	Exhibit Hall Open
10:00 – 10:30am	Coffee Break in the Exhibit Hall and Poster Competition Winners Announced
10:30am – 12:10pm	Tracks 1-12
12:20 – 1:20pm	Luncheon Presentations (Sponsorship Opportunities Available)
1:20 – 1:55pm	Dessert Refreshment Break in the Exhibit Hall with Poster Viewing
1:55 – 4:00pm	Tracks 1-12

Gain Further Exposure: PRESENT A POSTER & SAVE $50

6 Reasons Why You Should Present Your Research Poster at Bio-IT World Conference & Expo:

- Available to over 3,000 global attendees
- Will be seen by leaders from top pharmaceutical, biotech, academic, government institutes, and technology vendors
- Automatically entered in the Poster Competition, where two winners will each receive an American Express Gift Card
- Receive $50 off your registration fee
- Displayed in the Exhibit Hall – the central meeting place of the event – for maximum exposure
- Dedicated poster hours

Please visit Bio-ITWorldExpo.com for poster instructions and deadlines.

CONNECT WITH US:

#BioIT15

CHI's INTRO-NET Networking at its Best

CHI'S INTRO-NET: NETWORKING AT ITS BEST!

Maximize Your Experience Onsite at the Bio-IT World Conference & Expo!

The Intro-Net offers you the opportunity to set up meetings with selected attendees before, during and after this conference, allowing you to connect to the key people that you want to meet. This online system was designed with your privacy in mind and is only available to registered session attendees of this event. Registered conference attendees will receive more information on how to access the Intro-Net in the weeks leading up to the event!

Click Here to Register Online!
Bio-ITWorldExpo.com

Organized by:
Cambridge Healthtech Institute

PRE-CONFERENCE WORKSHOPS*

TUESDAY, APRIL 21, 2015

MORNING WORKSHOPS
8:00 – 11:30 am

W1: Aligning Projects with Agile Approach
Gurpreet Kanwar, Senior Project Manager, Information Management, NAV Canada

W2: Intelligent Methods Optimization of Algorithms for NGS
Michele Busby, Ph.D., Computational Biologist, Broad Technology Labs, Broad Institute
Mark D. M. Leiserson, Research Scientist, Benjamin Raphael Laboratory, Department of Computer Science & Center for Computational Molecular Biology, Brown University
James Lyons-Weiler, Ph.D., Managing Director, Ebola Rapid Assay Development Consortium

W3: Genome Assembly and Annotation
Robert Kuhn, Ph.D., Associate Director, UCSC Genome Browser, Center for Biomolecular Science & Engineering, University of California, Santa Cruz
Valerie A. Schneider, Ph.D., Staff Scientist, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health

W4: An Embarrassment of Riches: Choosing and Implementing Cloud Infrastructure
R. Mark Adams, Ph.D., CIO, Good Start Genetics
Jonathan Bingham, Product Manager, Google Genomics
Benjamin Breton, Senior Data Scientist, Good Start Genetics
William Brockman, Ph.D., Staff Software Engineer, Google Genomics
Jason Freimark, IT Manager, Good Start Genetics
Steve Marshall, Analytics Lead, Good Start Genetics
Angel Pizarro, Technical Business Development Manager, Scientific Computing, Amazon Web Services, Inc.

W5: Integrative Visualization Strategies for Large-Scale Biological Data
Nils Gehlenborg, Ph.D., Research Associate, Center for Biomedical Informatics, Harvard Medical School

W6: Biologics, Bioassay, and Biospecimen Registration Systems
Beth Basham, IT Director, Client Services, Biologics & Vaccines Discovery, Merck
Monica Wang, Ph.D., Lead System Engineer, Project and Program Manager, R&D Systems, Takeda Boston
Martin Romacker, Senior Scientist, Data and Information Architecture, Roche Innovation Center Basel, F. Hoffmann-La Roche AG
Rudolf Kinder, Senior Scientist, Roche Innovation Center Penzberg
Clemens Wrzodek, Ph.D., Scientific Software Engineer, Technical Project Manager, Roche Diagnostics GmbH

W8: Gamification of Science
Anselmo DiFabio, CTO, Applied Dynamic Solutions LLC
William Hayes, Ph.D., Senior Vice President, Platform Development, IT/Informatics, Selventa
Daniel Perry, Ph.D. Candidate and Researcher, Human Centered Design & Engineering
Eleanor Howe, Ph.D., Computational Biologist and Data Scientist, Broad Institute
Melanie Stegman, Ph.D., Owner, Molecular Jig Games, LLC.; Director, Science Game Center

AFTERNOON WORKSHOPS
12:30 – 4:00 pm

W9: The Impact of Research Informatics on Laboratory Evolution
Javier Roa, Head, Technical Operations and Research Infrastructure, F. Hoffmann-LaRoche LTD

W10: How Patient-Facing Data Networks are Transforming Biomedical Research
Moderator: Marcia Kean, Chairman, Feinstein Kean Healthcare
Ken Buetow, Ph.D., Director, Computation & Informatics Core Program, Complex Adaptive Systems, Arizona State University
Joanne S. Buzaglo, Ph.D, Vice President, Research & Training, Cancer Support Community
Robert McBurney, Ph.D., CEO, Accelerated Cure Project for MS
Christopher Boone, Ph.D., MSHA, FACHE, CPHIMS, PMP, Executive Director, Health Data Consortium
Laura Kolaczkowski, Lead Patient Representative, iConquerMS™ Governing Board
Sara Loud, Chief Operating Officer, Accelerated Cure Project and iConquerMS™ Project Team
William Tulskie, CEO, Life Data Systems, Inc. and iConquerMS™ Project Team
Toral Patel, Account Director, Feinstein Kean Healthcare and iConquerMS™ Project Team
Jamie Bull, Social Media Specialist, Feinstein Kean Healthcare and iConquerMS™ Project Team

W11: Determining Genome Variation and Clinical Utility
Heather McLaughlin, Ph.D., MB(ASCP)^{CM}, Instructor of Pathology, Massachusetts General Hospital and Harvard Medical School; Assistant Laboratory Director, Laboratory for Molecular Medicine, Partners HealthCare Personalized Medicine
Janusz Dutkowski, Ph.D., Founder and CEO, Data4Cure, Inc.
Zhaohui (Steve) Qin, Ph.D., Associate Professor, Biostatistics and Bioinformatics, Emory University

W12: Customizing Your Digital Research Environment with Genome Browsers
Mary E. Mangan, Ph.D., Director, Product and Content, OpenHelix

W13: Finding Innovation in Collaboration Environments: Documentum, Sharepoint, Veeva, and Tigers, Oh My!
Martin Leach, Ph.D., Vice President, Global Data Office, Biogen
Jay Bergeron, Director, Translational & Bioinformatics, Pfizer, Inc.
Scott Wilkins, Ph.D.,Enterprise Collaboration Director, Information Technology, AstraZeneca
Robert J. Boland, Senior Manager, External Innovation R&D IT, Janssen, Pharmaceutical Companies of Johnson & Johnson
Tom Arneman, President, Ceiba Solutions

W14: Converged IT Infrastructure in Life Science
Ari Berman, Ph.D., Director, Government Services and Principal Investigator, The BioTeam, Inc.
Aaron Gardner, Senior Scientific Consultant, BioTeam, Inc.
Adam Kraut, Principal Investigator, BioTeam, Inc.
Bhanu Rekepalli, Ph.D., Senior Scientific Consultant and Principal Investigator, BioTeam, Inc.

W15: Predictive Analytics
Mark Burfoot, Executive Director, Novartis
David King, CEO, Exaptive
Ted Snyder, Senior Solution Architect, Tamr
Alex Greenfield, Ph.D., Senior Systems Biologist, GNS Healthcare

W16: Large Scale NGS Analysis Using Globus Genomics
Paul Davé, Director, User Services at Computation Institute, University of Chicago
Ravi Madduri, Fellow, Computation Institute, University of Chicago and Argonne National Lab
Alex Rodriguez, Bioinformatics Expert, Computation Institute, University of Chicago and Argonne National Lab
Dinanath Sulakhe, Engagement Manager and Solutions Architect, Computation Institute, University of Chicago and Argonne National Lab

W17: Capturing, Managing and Exploiting Pre-Clinical *in vivo* Data - Challenges and Solutions
James Hinchliffe, Ph.D., Consultant, Life Sciences, Tessella
Bill Steel, Senior Consultant, Tessella
Caroline Hellawell, Senior Informatics Analyst, AstraZeneca

Sponsored by Tessella

** Separate registration required; for more details on the workshops, please visit Bio-ITWorldExpo.com*

Click Here to Register Online!
Bio-ITWorldExpo.com

Organized by:
Cambridge Healthtech Institute

2015 SPONSORS

PLATINUM SPONSORS

aspera an IBM® company
CYCLE COMPUTING
DDN STORAGE
ELSEVIER
EMC²
IBM
illumina
intel
okta™
ORACLE HEALTH SCIENCES
SEAGATE
sgi
THINKMATE HIGH PERFORMANCE COMPUTING
THOMSON REUTERS

GOLD SPONSORS

ACD/Labs
amazon web services
bina TECHNOLOGIES
BIOVIA
CAMBRIDGE SEMANTICS
cleversafe
Content Analyst Company
ConvergeHEALTH by Deloitte.
CRAY
DELL
DNAnexus
dotmatics
eClinicalOS
FLOWJO, LLC
GENERAL ATOMICS
idbs
INTERNET 2
Knowledgent
LabAnswer LABORATORY INFORMATICS
Linguamatics
Maverix BIOMICS
MOLECULAR HEALTH
paradigm4 answer harder questions
PerkinElmer For the Better
Qumulo
SCHRÖDINGER
SILICON MECHANICS SERVERS • STORAGE • HPC
SINEQUA REAL-TIME BIG DATA SEARCH & ANALYTICS
SwiftStack
tamr
TERADATA

BRONZE SPONSORS

databiology
epam
EXOSTAR
lenovo
PRECISION FOR MEDICINE
SoluSoft Business Process Solutions
SPEARHEAD
SQREAM TECHNOLOGIES
Tessella

OFFICIAL MEDIA PARTNER

Bio·IT World

Click Here to Register Online!
Bio-ITWorldExpo.com

Organized by:
Cambridge Healthtech Institute

PLENARY SESSION PRESENTATIONS:

TUESDAY, APRIL 21, 2015 * 4:00 – 5:00 PM

4:00 Event Chairperson's Opening Remarks
Cindy Crowninshield, RDN, LDN, Senior Conference Director, Cambridge Healthtech Institute

4:05 Plenary Session Introduction
Sanjay Joshi, CTO – Life Sciences, Emerging Technologies Division, EMC

4:15 The Vision for Data at the NIH
Philip E. Bourne, Ph.D., Associate Director for Data Science (ADDS), National Institutes of Health; Founding Editor in Chief, PLOS

Biomedical research and resultant health outcomes are increasingly defined by how we effectively use an ever increasing amount of digital data. This has become a focus at NIH as part of what we term the digital enterprise. That enterprise is based on community engagement, policies that make sense and a workable infrastructure all of which embraces both the public and private sector, the need to train the next generation of data scientists and is motivated by new research possibilities using data at different scales. Our progress and how this community can engage will be discussed.

WEDNESDAY, APRIL 22, 2015 * 8:00 – 9:45 AM

8:00 Chairperson's Opening Remarks
Phillips Kuhl, Co-Founder and President, Cambridge Healthtech Institute

8:05 Plenary Session Introduction
Jason Stowe, CEO, Cycle Computing

8:15 Precision Combination Therapy: Discover • Design • Deliver
Chris Sander, Computational and Systems Biology, Memorial Sloan Kettering Cancer Center

9:00 Benjamin Franklin Award & Laureate Presentation

9:30 Best Practices Awards Program

THURSDAY, APRIL 23, 2015 * 8:00 – 10:00 AM

8:00 Chairperson's Opening Remarks
Allison Byrum Profitt, Editorial Director, Bio-IT World & Clinical Informatics News

8:15 Plenary Session Introduction
Ketan Paranjape, General Manager Life Sciences, Intel Corp.

8:25 Data Custodians, Patient Advocates

One of the greatest challenges for every branch of medical science is rounding up patients for studies, trials, and endless measurements. What if patients could offer up their own data for research that interests and engages them? These panelists are working where big data science meets patient empowerment, finding ways that modern health records, patient portals, new data sources like next gen sequencing and activity trackers, and other technology innovations can give everyone a stake in how data are used. They will discuss how they keep patients in the research loop, and how non-professionals can be a driving force behind the medical breakthroughs of the future — in sickness and in health.

Benjamin Heywood, Co-Founder and President, PatientsLikeMe, Inc.
Andreas Kogelnik, M.D., Ph.D., Founder, Open Medicine Institute
Katherine Wendelsdorf, Ph.D., Field Application Scientist - Ingenuity Systems, QIAGEN Bioinformatics; Spokesperson, Empowered Genome Community

AWARDS PROGRAMS

Cambridge Healthtech Institute and Bio-IT World will again be recognizing and celebrating leaders in innovation through the following Awards Programs.

Best of Show Awards
The Best of Show Awards offer exhibitors an opportunity to distinguish their products from the competition. Judged by a team of leading industry experts and Bio-IT World editors, this award identifies exceptional innovation in technologies used by life science professionals today. Judging and the announcement of winners is conducted live in the Exhibit Hall. Winners will be announced on Wednesday, April 22 at 5:30pm. The deadline for product submissions is February 27, 2015. To learn more about this program, contact Ryan Kirrane at 781-972-1354 or email rkirrane@healthtech.com.

Best Practices Awards - Call for Entries!
Add value to your Conference & Expo attendance, sponsorship or exhibit package, and further heighten your visibility with the creative positioning offered as a Best Practices participant. Winners will be selected by a peer review expert panel in early 2015. Bio-IT World will present the Awards in the Amphitheater at 9:30am on Wednesday, April 22 during the Plenary Session and Awards Program. Early bird deadline (no fee) for entry is December 12, 2014 and final deadline (fee) for entry is February 6, 2015. Full details including previous winners and entry forms are available at Bio-ITWorld.com/BestPractices

2015 Benjamin Franklin Award
The Benjamin Franklin Award for Open Access in the Life Sciences is a humanitarian/bioethics award presented annually by the Bioinformatics Organization to an individual who has, in his or her practice, promoted free and open access to the materials and methods used in the life sciences. Nominations are now being accepted! The winner will be announced in the Amphitheater at 9:00am on Wednesday, April 22 during the Plenary Session and Awards Program. Full details including previous laureates and entry forms are available at www.bioinformatics.org/franklin.

Organized by:
Cambridge Healthtech Institute

Track 1

IT Infrastructure – Hardware

Big Data Storage Capabilities and Solutions in the R&D Ecosystem

TUESDAY, APRIL 21

7:00 am Workshop Registration and Morning Coffee

8:00 – 11:30 Recommended Morning Pre-Conference Workshops*
Aligning Projects with Agile Approach

12:30 – 4:00 pm Recommended Afternoon Pre-Conference Workshops*
Converged IT Infrastructure in Life Science
** Separate registration required*

2:00 – 6:30 Main Conference Registration

» 4:00 PLENARY SESSION
Please see page 5 for details.

5:00 – 7:00 Welcome Reception in the Exhibit Hall with Poster Viewing
Sponsored by IBM

WEDNESDAY, APRIL 22

7:00 am Registration Open and Morning Coffee

» 8:00 PLENARY SESSION
Please see page 5 for details.

9:00 Benjamin Franklin Awards and Laureate Presentation

9:30 Best Practices Awards Program

9:45 Coffee Break in the Exhibit Hall with Poster Viewing
Sponsored by INTERNET 2

TRENDS IN THE TRENCHES 2015

10:50 Chairperson's Opening Remarks
Wanmei Ou, Ph.D., Director, Product Strategy in Translational and Precision Medicine, Health Sciences Global Business Unit, Oracle

» 11:00 FEATURED PRESENTATION: HPC TRENDS IN THE TRENCHES 2015
Chris Dagdigian, Founding Partner & Director, Technology, BioTeam, Inc.
In one of the most popular presentations of the Expo, Chris delivers a candid assessment of the best, the worthwhile, and the most overhyped information technologies (IT) for life sciences.

12:00 pm Introduction to EVO:RAIL by VMware
Sponsored by THINKMATE HIGH PERFORMANCE COMPUTING
Michael McDonough, Senior Director, EVO:RAIL, VMware
VMware EVO:RAIL™ combines compute, networking, and storage resources into a hyper-converged infrastructure appliance to create a simple, easy to deploy, all-in-one solution offered by Qualified EVO:RAIL Partners. EVO:RAIL is a scalable Software-Defined Data Center (SDDC) building block that delivers compute, networking, storage, and management to empower private and hybrid cloud, end-user computing, test/dev, and branch office environments.

12:30 Session Break

12:40 Luncheon Presentation I: Big Data for Genomics -- SCALE, SPEED and SMART
Sponsored by IBM
Frank Lee, Ph.D., Lead Architect, Genomics Solution, IBM
Explosive growth of big data is challenging researchers in genomics and life sciences around the world. Learn about some of the latest solutions, architecture and best practice to 1) acquire, store, access data in scale; 2) build a high-throughput computing infrastructure to process large genomic data set; 3) gain insights and knowledge from the data through translational research. Illustrated through real-life projects and case studies, join this session to learn of the latest approaches to tackle big data, the evolving ecosystem, success stories and lessons learned that highlight the potential for collaboration among genomic research communities. Share in a preview of the upcoming IBM genomics turn-key platform currently under development.

1:10 Luncheon Presentation II: Optimizing Genomic Sequence Searches to Next-Generation Intel Architectures
Sponsored by intel
Bhanu Rekepalli, Ph.D., Senior Scientific Consultant and Principal Investigator, BioTeam, Inc.
Upcoming bioinformatics, and biomedical, research requires fast processing and analytic tools due to the immense growth of genomic data added to the biological knowledge base with the advent of next generation sequencing technologies. The design of these tools should adhere efficiently to homogeneous and heterogeneous architectures while supporting scalability, accuracy, and reproducibility. The National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) for genomics sequence searches is re-designed to scale on hybrid parallel architectures composed of Intel Xeon processors and Intel Xeon Phi coprocessors, denoted here as Highly Scalable Parallel Hybrid BLAST (HSPH- BLAST). Functionality enhancements, such as cross- compilation, dynamic load scheduling, master-worker model, input/output management, and database distribution are discussed. A performance evaluation of HSPH-BLAST demonstrates reduction in execution time, high scalability, and balanced processor utilization. HSPH-BLAST and similar tools integrated into scientific workflows pipelines can allow biologists to easily perform systematic studies resulting in rapid and high-impact scientific discovery.

1:40 Session Break

STORAGE EFFICIENCIES USING HADOOP & IMPROVING DATA WORKFLOWS

1:50 Chairperson's Remarks

1:55 Comparisons of Storage Efficiencies through Hadoop
Martin Gollery, CEO, Tahoe Informatics
Hadoop is widely used in 'Big-Data' applications, so much so that most modern cluster installations are now installing some version of Hadoop rather than the old style clusters. This talk will compare and contrast the storage techniques and the costs that are associated with them.

2:25 Rapid Integration of Cancer Genomics Data Using Hadoop and Cloudera's Impala
Sittichoke Saisanit, Ph.D., Data Scientist, Informatics, Pharmaceutical Research and Early Development Informatics, Roche Innovation Center New York
We explored Cloudera Impala for analysis of cancer genomics data. Without data transformation and reformatting, Impala tables can be created quickly from files on Hadoop file system with a simple command. Such speed and

Organized by:
Cambridge Healthtech Institute

Track 1

IT Infrastructure – Hardware

Big Data Storage Capabilities and Solutions in the R&D Ecosystem

flexibility enable us to interrogate data without spending much time on schema design, index creation, query tuning and data cleaning. Impala can be accessed through Spotfire allowing flexibility of data visualization.

2:55 Accelerating Biomedical Research Discovery: The 100G Internet2 Network – Built and Engineered for the Most Demanding Big Data Science Collaborations

Sponsored by Internet2

Christian Todorov, Director, Network Services Management, Internet2

Genomic & biomedical researchers have been forced to exchange big data via physical drives as advanced network connectivity was previously unavailable or cost prohibitive. Hear how colleagues are improving big data workflows using the 100G Internet2 Network, which provides the highest data transport rates available, along with dynamic cloud and trust applications that are interconnecting research and accelerating discovery.

3:10 Managing Genomic Data at Scale! - Rules Based Intelligent Data Management

Sponsored by SEAGATE

Jose L. Alvarez, Principal Engineer, WW Director, Healthcare and Life Sciences, Seagate Cloud and Systems Solutions

The explosion of Genomic data due to new instrument chemistry and more powerful analysis tool sets has created a complex and manual data management problem for high-throughput NGS centers. We will discuss how an intelligent data management solution can address this problem. iRODS (Integrated Rules-Oriented Data System) enables this intelligent data orchestration and can even help with pipeline and workflow automation.

3:25 Refreshment Break in the Exhibit Hall with Poster Viewing

PERSPECTIVES OF LIFE SCIENCES SUPERCOMPUTING CENTERS

4:00 PANEL PRESENTATION/DISCUSSION: ICTBioMed: International Consortium for Technology in Biomedicine

Moderator: Anil Srivastava, President, Open Health Systems Laboratory

Panelists:

Rolf A. Heckemann M.D., Ph.D., Professor of Medical Imaging and Image Analysis, MedTech West, Sahlgrenska University Hospital

Cezary Mazurek, Ph.D., Director, Network Services Department, Poznan Supercomputing and Networking Center (PSNC)

Prof Asoke K Talukder, Ph.D., Adjunct Professor, Computer Science & Engineering, NIT Warangal; Co-founder & Chief Scientific Officer, InterpretOmics, Bangalore' Ex DaimlerChrysler Chair Professor, IIIT Bangalore

Panelists to be Announced

Open Health Systems Laboratory has brought together several life sciences supercomputing centers to form the International Consortium for Technology in Biomedicine (ICTBioMed). ICTBioMed members have been working together for almost two years to create a shared global cyberinfrastructure as a seamless and friction-free platform for the researchers worldwide for their collaborative research in consistent with the tenets of team science. ICTBioMed leadership team will present in this session both the shared resources and the research use cases that they have been supporting to validate and further develop the value added cloud services. The panelists will speak to a narrative framework of possible science using, what NSF describes, as International Research Network Connection, pursuing the Big Data to Knowledge goals of NIH.

5:00 Beyond Parallel Filesystems: NVMe Storage for Genomics Workflows

Sponsored by SGI

James Reaney, Ph.D., Senior Director, Research Markets, SGI

Network-attached storage. Clustered storage. Distributed parallel filesystem storage. Storage infrastructure for genomics workflows has always been about faster, easier, and especially more scalable storage solutions to keep pace with the data tsunami in next-gen sequencing. SGI and Intel present a new concept in storage architecture for these workflows, one with disruptive potential for the marketplace. Not only faster and very scalable, but drop-dead simple to use too.

5:15 The Expanding Face of Meta Data

Sponsored by EMC²

Steve Worth, Director of Engineering, EMC

Groups maintaining data repositories at the petabyte-scale are discovering that cataloguing associated metadata is necessary to properly access, recall and analyze data. Capturing and maintaining metadata long term is becoming as critical as the data itself. All the more when you consider the rapid cycling of underlying hardware technologies. We will discuss the evolving nature of metadata along with recent advancements and approaches

5:30 Best of Show Awards Reception in the Exhibit Hall with Poster Viewing

6:30 Close of Day

THURSDAY, APRIL 23

7:00 am Registration Open

7:00 Breakfast Presentation: Enabling Technology. Leveraging Data. Transforming Personalized Medicine

Sponsored by intel

Moderator: Ketan Paranjape, General Manager Life Sciences, Intel Corp.

Panelists: Sanjay Joshi, CTO, Life Sciences, EMC2 Isilon

Johannes Karten, CTO & Founder, Genalice

Walt Gall, Vice President, Healthcare & Strategic Partnerships, Saffron Technology

Pieter van Rooyen, CEO & President, Edico Genome

Shawn Dolley, Health & Life Science Big Data Expert, Cloudera

The $1000 genome is here, and the fundamental problems have shifted... it is no longer about shrinking the cost of sequencing but the explosive growth of big data: the downstream analytics with rapidly evolving parameters, data sources and formats; the storage, movement and management of massive datasets and workloads, and the challenge of articulating the results and translating the latest findings directly into improving patient outcomes. A panel will discuss these issues and more as we work to achieve the vision of personalized medicine.

» 8:00 PLENARY SESSION
Please see page 5 for details.

10:00 Coffee Break in the Exhibit Hall and Poster Competition Winners Announced

INFRASTRUCTURE AND PLATFORMS FOR BIG DATA: CAPABILITIES AND SOLUTIONS

10:30 Chairperson's Remarks

Peter Godman, Co-Founder & CEO, Qumulo

Organized by:
Cambridge Healthtech Institute

Track 1

IT Infrastructure – Hardware

Big Data Storage Capabilities and Solutions in the R&D Ecosystem

10:40 Intelligent Infrastructure Approaches for Emerging Life Sciences Data Management Issues at Scale

Sponsored by DDN STORAGE

George Vacek, Ph.D., Global Business Director, Life Sciences, DataDirect Networks

Dr. Vacek will deliver several in-depth case studies of global leaders HPC applications in Life Science. Case studies will focus on infrastructure approaches to solve the emerging issues of data at scale, including best practices in supporting high performance local workflows, collaborative and research communities, and life sciences clouds and hybrid cloud solutions.

10:55 How Next Generation Scale-Out Storage Fuels Breakthroughs in Life Sciences

Sponsored by Qumulo

Peter Godman, Co-Founder & CEO, Qumulo

Technology advances in DNA sequencing and other research data capture instruments are creating data at an unprecedented rate. As storage footprints grow further into petabyte scale, storage teams increasingly struggle to manage the massive amount of data stored. Next-generation scale-out storage provides instant insight into data at scale, abstracts away the underlying infrastructure, and achieves breakthrough price/performance using intelligent software and commodity hardware.

11:10 Infrastructure, Architecture, and Organization: Data Engineering at Scale at the Broad

Chris Dwan, Assistant Director, Research Computing and Data Services, Broad Institute of MIT and Harvard

As the Broad Institute enters its second decade, we are adapting to genomic research at a global scale. Among other things, this requires adopting hybrid cloud technologies, moving to object models for data storage, and embracing federated solutions for identity and authorization. The social and organizational aspects of these transitions are at least as challenging as the technical. This talk describes the interplay between the human and technical aspects of these changes, as well as specific lessons learned along the way.

11:40 Start Small, Collaborate Often, Grow Big – Scaling NGS Compute and Storage Solutions for Personalized Medicine

Sponsored by DELL

James Lowey, Vice President, Technology, Translational Genomics Research Institute (TGen)

Scaling an NGS IT solution doesn't have to be overwhelming. Collaborating with experienced Clinicians, Researchers, Vendors, and Partners, all using best practices - enables incremental success and effective development for high utilization and impactful results.

11:55 Out of the Trenches and Into the Future: Mixing File and Object Storage Architectures

Sponsored by EMC²

Patrick Combes, Principal Solution Architect, Life Science & HPC, EMC

Managing genomics and biomedical data across file and object storage architectures currently dominate the conversation within research IT groups. We will share insights and best practices to design and implement on premise, public cloud, and hybrid architectures. These architectures mix file and object approaches to achieve an optimal balance between performance, archive, and data governance & protection requirements.

12:10 pm Session Break

12:20 Luncheon Presentation I: Breaking the $1,000 Genome Sequencing Barrier with Object Storage

Sponsored by SwiftStack

Brandon Kruse, Senior Systems Engineer, HudsonAlpha Institute for Biotechnology

Joe Arnold, President and Chief Product Officer, SwiftStack

Peyton McNully, Technology Director, HudsonAlpha Institute for Biotechnology

Andrew Crouse, Ph.D., Intellectual Property and Industry Partnership Manager, HudsonAlpha Institute for Biotechnology

The next generation of human genome sequencers are a revolution in rapid disease diagnostics and custom gene therapy. In this talk, the HudsonAlpha Institute for Biotechnology will present how we are using OpenStack Swift as a component in our genomics-as-a-service. The recently purchased Illumnia X Ten's automatically place the institute in the petabyte-scale. OpenStack Swift enables long-term sustainable storage using commodity storage hardware and open source software. The talk will present our deployment and configuration of OpenStack Swift. Specifically, how OpenStack Swift storage policies are configured so that we can offer varying levels of storage durability and availability.

12:50 Luncheon Presentation II *(Sponsorship Opportunity Available)* or Lunch on Your Own

1:20 Dessert Refreshment Break in the Exhibit Hall with Poster Viewing

MANAGING BIG DATA AND SECURITY STRATEGIES

1:55 Chairperson's Remarks

John M. Conley, J.D., Ph.D., William Rand Kenan, Jr. Professor of Law, University of North Carolina, Chapel Hill ; Counsel, Robinson Bradshaw & Hinson

» 2:00 FEATURED PRESENTATION: IT AND INFORMATICS INNOVATION AT FDA

Roselie A. Bright, Sc.D., MS, PMP, Program Manager, Office of Information Management and Technology, Office of Informatics Technology and Innovation, Office of Operations, Office of the Commissioner, U.S. Food And Drug Administration (FDA)

OpenFDA was the first innovation created by Taha Kass-Hout, M.D., MS, upon joining FDA as the first Chief Health Information Officer in March 2013. OpenFDA was launched on June 2, 2014, allowing software developers, researchers and the public to tap into adverse events for drugs and medical devices; recalls, for drugs, devices and foods; and labeling for products on the market.

2:30 Global Developments in Privacy and Data Security Law

John M. Conley, J.D., Ph.D., William Rand Kenan, Jr. Professor of Law, University of North Carolina, Chapel Hill; Counsel, Robinson Bradshaw & Hinson

The international legal climate governing privacy and data security is changing. The European Union is in the midst of a fundamental shift in its approach. The U.S. still lacks a national data law, so the states and individual federal agencies are groping toward a strategy. This presentation focuses on the impact of these ongoing changes on genomics, bioinformatics and health research.

Track 1

IT Infrastructure – Hardware

Big Data Storage Capabilities and Solutions in the R&D Ecosystem

3:00 PANEL DISCUSSION: Achieving Much-Needed Innovation while Hurdling the Barriers of Stringent Regulation

Moderator: John M. Conley, J.D., Ph.D., William Rand Kenan, Jr. Professor of Law, University of North Carolina, Chapel Hill; Counsel, Robinson Bradshaw & Hinson

Panelists:

Roselie A. Bright, Sc.D., MS, PMP, Program Manager, Office of Information Management and Technology, Office of Informatics Technology and Innovation, Office of Operations, Office of the Commissioner, U.S. Food And Drug Administration (FDA)

Dana Caulder, Senior Software Engineer, Bioinformatics and Computational Biology, Genentech

Chris Dwan, Assistant Director, Research Computing and Data Services, Broad Institute of MIT and Harvard

Sanjay Joshi, CTO – Life Sciences, Emerging Technologies Division, EMC

Dave Peterson, Executive Director, Vendor & Third Party Assurance, National IT Compliance, Kaiser Permanente Information Technology

Vas Vasiliadis, Director, Products, Computation Institute, University of Chicago and Argonne National Laboratory

The growth in patient healthcare and life sciences innovations can be attributed to technology enhancements like cloud computing, big data analytics and mobile applications, but may conflict with increasing regulatory compliance demands to ensure protection of healthcare life and quality as well as patient data privacy and security. This panel of esteemed technology solution providers and regulators debates real-world challenges and how regulation must also innovate at technology's pace.

4:00 Conference Adjourns

Organized by:
Cambridge Healthtech Institute

Track 2

Software Development

Harnessing Data for Scientific Decision Making

TUESDAY, APRIL 21

7:00 am Workshop Registration and Morning Coffee

8:00 – 11:30 Recommended Morning Pre-Conference Workshops*
Aligning Projects with Agile Approach
Gamification of Science

12:30 – 4:00 pm Recommended Afternoon Pre-Conference Workshops*
Predictive Analytics
Large Scale NGS Analysis Using Globus Genomics
** Separate registration required*

2:00 – 6:30 Main Conference Registration

» 4:00 PLENARY SESSION
Please see page 5 for details.

5:00 – 7:00 Welcome Reception in the Exhibit Hall with Poster Viewing
Sponsored by IBM

WEDNESDAY, APRIL 22

7:00 am Registration Open and Morning Coffee

» 8:00 PLENARY SESSION
Please see page 5 for details.

9:00 Benjamin Franklin Awards and Laureate Presentation

9:30 Best Practices Awards Program

9:45 Coffee Break in the Exhibit Hall with Poster Viewing
Sponsored by INTERNET 2

TRENDS IN THE TRENCHES 2015

10:50 Chairperson's Opening Remarks
Wanmei Ou, Ph.D., Director, Product Strategy in Translational and Precision Medicine, Health Sciences Global Business Unit, Oracle

» 11:00 FEATURED PRESENTATION: HPC TRENDS IN THE TRENCHES 2015
Chris Dagdigian, Founding Partner & Director, Technology, BioTeam, Inc.
In one of the most popular presentations of the Expo, Chris delivers a candid assessment of the best, the worthwhile, and the most overhyped information technologies (IT) for life sciences.

12:00 pm Introduction to EVO:RAIL by VMware
Sponsored by THINKMATE HIGH PERFORMANCE COMPUTING

Michael McDonough, Senior Director, EVO:RAIL, VMware
VMware EVO:RAIL™ combines compute, networking, and storage resources into a hyper-converged infrastructure appliance to create a simple, easy to deploy, all-in-one solution offered by Qualified EVO:RAIL Partners. EVO:RAIL is a scalable Software-Defined Data Center (SDDC) building block that delivers compute, networking, storage, and management to empower private and hybrid cloud, end-user computing, test/dev, and branch office environments.

12:30 Session Break

12:40 Luncheon Presentation I: Big Data for Genomics -- SCALE, SPEED and SMART
Sponsored by IBM

Frank Lee, Ph.D., Lead Architect, Genomics Solution, IBM

Explosive growth of big data is challenging researchers in genomics and life sciences around the world. Learn about some of the latest solutions, architecture and best practice to 1) acquire, store, access data in scale; 2) build a high-throughput computing infrastructure to process large genomic data set; 3) gain insights and knowledge from the data through translational research. Illustrated through real-life projects and case studies, join this session to learn of the latest approaches to tackle big data, the evolving ecosystem, success stories and lessons learned that highlight the potential for collaboration among genomic research communities. Share in a preview of the upcoming IBM genomics turn-key platform currently under development.

1:10 Luncheon Presentation II: Optimizing Genomic Sequence Searches to Next-Generation Intel Architectures
Sponsored by intel

Bhanu Rekepalli, Ph.D., Senior Scientific Consultant and Principal Investigator, BioTeam, Inc.

Upcoming bioinformatics, and biomedical, research requires fast processing and analytic tools due to the immense growth of genomic data added to the biological knowledge base with the advent of next generation sequencing technologies. The design of these tools should adhere efficiently to homogeneous and heterogeneous architectures while supporting scalability, accuracy, and reproducibility. The National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) for genomics sequence searches is re-designed to scale on hybrid parallel architectures composed of Intel Xeon processors and Intel Xeon Phi coprocessors, denoted here as Highly Scalable Parallel Hybrid BLAST (HSPH- BLAST). Functionality enhancements, such as cross- compilation, dynamic load scheduling, master-worker model, input/output management, and database distribution are discussed. A performance evaluation of HSPH-BLAST demonstrates reduction in execution time, high scalability, and balanced processor utilization. HSPH-BLAST and similar tools integrated into scientific workflows pipelines can allow biologists to easily perform systematic studies resulting in rapid and high-impact scientific discovery.

1:40 Session Break

USING DATA TO DRIVE DECISIONS

1:50 Chairperson's Remarks
Brian Bissett, Senior Member, Institute of Electrical and Electronics Engineers

1:55 Lies, Damn Lies, and Big Data: How to Best Utilize Data to Drive Decisions
Brian Bissett, Senior Member, Institute of Electrical and Electronics Engineers

The audience will gain an appreciation for how to best utilize data to drive decisions. Common fallacies will be addressed, including the notion that Big Data sets are always superior to smaller data sets. The limitations of big data sets, the importance of quality data, effective display of quantitative information, boundary conditions, and the evaluation of quantitative and qualitative factors will all be discussed.

Organized by:
CHI Cambridge Healthtech Institute

Track 2

Software Development

Harnessing Data for Scientific Decision Making

2:25 Data Publication and Discovery Using Globus Research Data Management Software-as-a-Service

Vas Vasiliadis, Director, Products, Computation Institute, University of Chicago and Argonne National Laboratory

Globus is software-as-a-service for research data management, used at dozens of institutions and national facilities for moving, sharing, and publishing big data. This presentation will give an overview and demonstration of the Globus services, as well as case studies that illustrate how Globus is increasing researcher productivity and facilitating enhanced collaboration among researchers.

2:55 Leveraging Hadoop Mapreduce in Building Patient Timelines & Analyzing Health Resource Utilization

Sponsored by Knowledgent

Saar Golde Ph.D., Informationist, Knowledgent

During this presentation we will introduce methodological innovations in analyzing real world evidence and observational data in health outcomes research. Attendees will learn how we leveraged Hadoop data lake to transform the transaction-level data into a patient-centric data model and to run large scale analysis in an efficient manner, yielding robust results in a timely and cost-effective manner.

3:25 Refreshment Break in the Exhibit Hall with Poster Viewing

4:00 Semantic Integration of Unstructured Safety Study Data: Experiences and Outlook

Alain Nanzer, Ph.D., Global Head Safety & Development Workflows, Pharma Research and Early Development Informatics, Roche Innovation Center Basel

The presentation will share our experiences implementing a platform using semantic integration technologies to provide scientists search, evaluation, and advanced visualization capabilities for safety *in vivo* study data. Furthermore we will show how the platform has been extended providing fast access to real-time study data, and then evolved to a data turntable for external study data and submissions to regulatory authorities.

4:30 DIVOS: A Platform for Effective *in vivo* Study Knowledge Management at Genentech

Dana Caulder, Senior Software Engineer, Bioinformatics and Computational Biology, Genentech

Animal study data management provides unique challenges that often are not well addressed in a pharmaceutical research setting. More often than not, much of the *in vivo* workflow and process lives in email and spreadsheets. This is clearly not an effective way to manage some of the most valuable preclinical data on our therapeutics. We will present a unique success story in the realm of *in vivo* data management in an effort to share our knowledge with others in the industry.

5:00 Accelerate Life Sciences Data Processing in a Secure, HIPAA-compliant Cloud Platform

Ben Butler, Vice President, Business Development & Solutions Architecture, REAN Cloud Solutions

REAN Cloud has partnered with several leading life sciences organizations to deploy and manage genomics and personalized medicine research data processing pipelines on the Amazon Web Services cloud. In this session, learn about win-win design patterns that leverage the benefits of high-scale, low-cost compute and storage of the cloud while also being highly secure and meeting stringent compliance standards, specifically the requirements of the the U.S. Health Insurance Portability and Accountability Act (HIPAA). We will provide insights into several customer case studies which showcase how REAN Cloud accelerates data processing genomics research, while reducing the time required to meet compliance requirements. REAN offers an innovative solution to meet analytical challenges such as accommodating peak compute demand, coordinating secure access for teams of scientists and analysts, and securely sharing validated tools and results. Attendees will receive our blueprint for implementing a robust, defense-in-depth architecture that directly addresses working with processing data that contains protected health information (PHI)

5:30 Best of Show Awards Reception in the Exhibit Hall with Poster Viewing

6:30 Close of Day

THURSDAY, APRIL 23

7:00 am Registration Open and Morning Coffee

» 8:00 PLENARY SESSION PANEL
Please see page 5 for details.

10:00 Coffee Break in the Exhibit Hall and Poster Competition Winners Announced

INTEGRATING AND IMPLEMENTING DATA PLATFORMS AND WORKFLOWS

10:30 Chairperson's Remarks

Noel Southall, Ph.D., Informatics, National Center for Advancing Translational Sciences, NIH

10:40 A Case Study in Building a Clinical Research Database in a Translational Research Environment

Charlie Quinn, Director, Data Management & Software Development, Benaroya Research Institute

We have developed a database that integrates public and private clinical and experimental data in a translational research environment. We will discuss some of the challenges and solutions that we encountered in developing the database. In addition, we will discuss our new open source spreadsheet wrangling tool which is instrumental in allowing us to capture, integrate, and manage data.

11:10 Sciencescape - An Innovative Research Discovery Platform that Connects Users to Breaking Research As It Happens, Around the World, and Throughout History

Sam Molyneux, CEO & Co-Founder, Sciencescape

Sciencescape is an innovative research discovery platform that connects users to breaking research as it happens, around the world, and throughout history, enabling them to make discoveries and become leaders in their field. Showcasing Sciencescape at the Bio-IT Expo will allow industry leaders to revolutionize their workflow and join the Sciencescape network to make incredible discoveries and network within their field.

11:40 Building a Global Framework for the Exchange of Drug Substance Information

Noel Southall, Ph.D., Informatics, National Center for Advancing Translational Sciences, NIH

FDA needs a knowledge management system that can handle the enormous variety of substances found in commerce in a scientifically rigorous way. NIH's National Center for Advancing Translational Sciences (NCATS) is working with FDA, global regulators and stakeholders to build this software and enhance the cooperation between agencies. NCATS' charge is to develop, demonstrate and broadly disseminate tools for translational research that impact health care delivery, the proper use of medications, and their risk management. This project serves these goals and provides an example of how vision and innovation can come together within government to better serve public health.

12:10 pm Session Break

12:20 Luncheon Presentation *(Sponsorship Opportunity Available)* **or Lunch on Your Own**

1:20 Dessert Refreshment Break in the Exhibit Hall with Poster Viewing

Track 2

Software Development

Harnessing Data for Scientific Decision Making

MANAGING BIG DATA AND SECURITY STRATEGIES

1:55 Chairperson's Remarks

John M. Conley, J.D., Ph.D., William Rand Kenan, Jr. Professor of Law, University of North Carolina, Chapel Hill ; Counsel, Robinson Bradshaw & Hinson

» 2:00 FEATURED PRESENTATION: IT AND INFORMATICS INNOVATION AT FDA

Roselie A. Bright, Sc.D., MS, PMP, Program Manager, Office of Information Management and Technology, Office of Informatics Technology and Innovation, Office of Operations, Office of the Commissioner, U.S. Food And Drug Administration (FDA)

OpenFDA was the first innovation created by Taha Kass-Hout, M.D., MS, upon joining FDA as the first Chief Health Information Officer in March 2013. OpenFDA was launched on June 2, 2014, allowing software developers, researchers and the public to tap into adverse events for drugs and medical devices; recalls, for drugs, devices and foods; and labeling for products on the market.

2:30 Global Developments in Privacy and Data Security Law

John M. Conley, J.D., Ph.D., William Rand Kenan, Jr. Professor of Law, University of North Carolina, Chapel Hill; Counsel, Robinson Bradshaw & Hinson

The international legal climate governing privacy and data security is changing. The European Union is in the midst of a fundamental shift in its approach. The U.S. still lacks a national data law, so the states and individual federal agencies are groping toward a strategy. This presentation focuses on the impact of these ongoing changes on genomics, bioinformatics and health research.

3:00 PANEL DISCUSSION: Achieving Much-Needed Innovation while Hurdling the Barriers of Stringent Regulation

Moderator: John M. Conley, J.D., Ph.D., William Rand Kenan, Jr. Professor of Law, University of North Carolina, Chapel Hill; Counsel, Robinson Bradshaw & Hinson

Panelists:

Roselie A. Bright, Sc.D., MS, PMP, Program Manager, Office of Information Management and Technology, Office of Informatics Technology and Innovation, Office of Operations, Office of the Commissioner, U.S. Food And Drug Administration (FDA)

Dana Caulder, Senior Software Engineer, Bioinformatics and Computational Biology, Genentech

Chris Dwan, Assistant Director, Research Computing and Data Services, Broad Institute of MIT and Harvard

Sanjay Joshi, CTO – Life Sciences, Emerging Technologies Division, EMC

Dave Peterson, Executive Director, Vendor & Third Party Assurance, National IT Compliance, Kaiser Permanente Information Technology

Vas Vasiliadis, Director, Products, Computation Institute, University of Chicago and Argonne National Laboratory

The growth in patient healthcare and life sciences innovations can be attributed to technology enhancements like cloud computing, big data analytics and mobile applications, but may conflict with increasing regulatory compliance demands to ensure protection of healthcare life and quality as well as patient data privacy and security. This panel of esteemed technology solution providers and regulators debates real-world challenges and how regulation must also innovate at technology's pace.

4:00 Conference Adjourns

Organized by:
Cambridge Healthtech Institute

Track 3

Cloud Computing

Riding Cloud to Next-Generation Computing

TUESDAY, APRIL 21

7:00 am Workshop Registration and Morning Coffee

8:00 – 11:30 Recommended Morning Pre-Conference Workshops*

An Embarrassment of Riches: Choosing and Implementing Cloud Infrastructure

12:30 – 4:00 pm Recommended Afternoon Pre-Conference Workshops*

Large-Scale NGS Analysis Using Globus Genomics

** Separate registration required*

2:00 – 6:30 Main Conference Registration

» 4:00 PLENARY SESSION
Please see page 5 for details.

5:00 – 7:00 Welcome Reception in the Exhibit Hall with Poster Viewing

Sponsored by IBM

WEDNESDAY, APRIL 22

7:00 am Registration Open and Morning Coffee

» 8:00 PLENARY SESSION
Please see page 5 for details.

9:00 Benjamin Franklin Awards and Laureate Presentation

9:30 Best Practices Awards Program

9:45 Coffee Break in the Exhibit Hall with Poster Viewing

Sponsored by INTERNET 2

SECURITY: IT RISK MANAGEMENT

10:50 Chairperson's Opening Remarks

Dave Peterson, Executive Director, Vendor & Third Party Assurance, National IT Compliance, Kaiser Permanente Information Technology

» 11:00 FEATURED PRESENTATION: COMPLIANT CLOUD COMPUTING

Krista Woodley, Director, Information Technology, Biogen

We provide insight on how to best manage SaaS-based projects in a regulated world, by discussing best practices for Lifecycle management, change control, security management and IT risk management. IT and business project teams will have a clear understanding of how to optimize their IT deployments in this new cloud-based environment.

11:30 Rethinking Cloud Security: You Can't Control What You Can't See

Kevin Gilpin, CTO, Conjur, Inc.

As more companies adopt DevOps programs and build new infrastructure, the quantity and sensitivity of data being processed outside of the traditional IT stack are growing. Few organizations know where the access points into this information are, or how to secure them. We outline best practices for establishing visibility and control in this new space, drawing real-world examples from environments large and small.

12:00 pm Security in the Cloud: How AMAG Protects Company Data with Multi-factor Authentication

Sponsored by okta

Nathan McBride, Vice President, IA & Chief Cloud Architect, AMAG Pharmaceuticals

To stay competitive and deliver world-class care, organizations such as yours are increasingly adopting cloud and mobile-first IT strategies. These trends come with significant security and access management challenges. In this presentation, Nate McBride, VP of IT and Chief Cloud Architect at AMAG Pharmaceuticals will discuss AMAG's move to the cloud and their deployment strategy for securing data with multi-factor authentication.

12:30 Session Break

12:40 Luncheon Co-Presentation I: Are Your Researchers Paying Too Much for Their Cloud-Based Data Backups?

Sponsored by Silicon Mechanics

Dirk Petersen, Scientific Computing Manager, Fred Hutchinson Cancer Research Center (FHCRC)

Joe Arnold, President and Co-Founder, SwiftStack

Considering deploying a multi-petabyte storage-as-a-service offering in your research environment? Learn how an industry-leading software-defined object storage solution, architected by SwiftStack and Silicon Mechanics, helped shift hundreds of users to an object-based workflow for their archival data. With an emphasis on cost efficiencies, scalability, and manageability, see how this implementation at Fred Hutchinson Cancer Research Center (FHCRC) is continually evolving across new use cases and access methods.

1:10 Luncheon Co-Presentation II: Running Scalable and Cost Effective High-Throughput Sequencing Data Analysis on Amazon Web Services

Sponsored by amazon web services

Cory Funk, Ph.D., Research Scientist, Institute for Systems Biology

Dmitry Pushkarev, Ph.D., CEO and Founder, ClusterK

Here we present work by the Institute for Systems Biology, in collaboration with ClusterK and AWS, to run large cohort RNA-Seq comparative data analysis on the AWS Spot Market. We will showcase the SNAPR algorithm for transcriptome analysis, as well as highlight the advanced features of the ClusterK products that make full use of AWS Spot instances that resulted in significant cost savings over on-demand pricing.

1:40 Session Break

FLEXIBILITY: IT INFRASTRUCTURE

1:50 Chairperson's Remarks

Jonas S. Almeida, Ph.D., Professor and CTO, Biomedical Informatics Department, SUNY Stony Brook

1:55 Web Computing as Commodity Supercomputing for User-Facing Genomics Applications

Jonas S. Almeida, Ph.D., Professor and CTO, Biomedical Informatics Department, SUNY Stony Brook

Recently, web computing (computing distributed to web clients, typically web browsers) has been the object of both academic and commercial attention as an extreme low-cost, high-distribution, high-availability model for supercomputing. Genomics applications are particularly suitable for this model of cloud computing. There is, as always, a price to pay: some core sequence analysis algorithms need to be re-identified.

Organized by:
Cambridge Healthtech Institute

Track 3

Cloud Computing

Riding Cloud to Next-Generation Computing

2:25 Chameleon: A Large-Scale, Reconfigurable Experimental Environment for Cloud Research

Kate Keahey, Senior Fellow, Computation Institute, University of Chicago and Argonne National Laboratory; Principal Investigator, Chameleon

Chameleon is a large-scale, reconfigurable testbed for next-generation cloud computing research, established under the NSFCloud program. This talk describes the types of experiments it will support, the exciting hardware and software capabilities we will provide for cloud computing research, as well as the timeline in which these capabilities will be provided.

2:55 Leveraging the Cloud to Safeguard Genomic Data and Ensure Its Availability

Sponsored by SEAGATE

Tyna Callahan, Senior Manager, Healthcare Products & Strategy, E-Vault - Seagate Systems

Michael Leonard, Director, Product Management, Healthcare IT, Iron Mountain

The sheer volume of data that must be retained today is massive—as is the responsibility to keep it intact, and to keep private information out of the wrong hands. Storage experts from Iron Mountain and Seagate CSS will discuss the key considerations for developing a storage strategy that leverages object storage in the cloud to ensure data availability, data integrity, data security and privacy.

3:10 Web-Scale: The Genomic Data Commons Project

Sponsored by cleversafe

Piers D. Nash, Director, Business Development and Outreach, University of Chicago

Learn how using a web-scale data hub dramatically speeds up the pace of medical research by housing cancer genomic data. The Genomic Data Commons, a first of its kind facility established by the University of Chicago will not only centralize genomic data, but also harmonize it, enabling collaboration and engagement between researchers.

3:25 Refreshment Break in the Exhibit Hall with Poster Viewing

4:00 Simulating the Behavior of a Living Human Heart

Sponsored by BIOVIA

Karl D'Souza, SIMULIA Virtual Human Modeling

Given the prevalence of cardiovascular disease, major efforts are underway to understand cardiac function, design effective treatments, and accelerate the approval process. However, the lack of realistic simulation models and adequate computational resources has limited the use of in-silico methods to predict in-vivo drug or device performance. In response, Dassault Systemes is developing high fidelity multiphysics and multiscale models of the human heart and leveraging cloud computing. We present progress on this project to date.

4:15 Selected Oral Poster Presentation: Accurate HLA Genotyping to 3-field Level from Whole-Genome Sequencing Data Analyzed by Omixon Target HLA in G3's GLOBAL Clinical Study

Robert Pollok, Ph.D., Field Application Scientist, Omixon, Inc.

Successful characterization of HLA genes allows for more positive organ transplant outcomes and disease associations. Omixon, working with the Global Genomics Group, is validating NGS methods on HLA genes using Whole Genomic Sequencing, targeted amplifications and non-sequencing-based HLA typing methods to build confidence in using NGS with HLA genes.

4:30 Using Cloud Computing to Improve the Accuracy and Probability of Success of Drug Discovery

Ed Addison, Ph.D., CEO, Corporate, Cloud Pharmaceuticals, Inc.

Cloud computing combined with Moore's Law has provided an unprecedented opportunity for "*in silico*" drug discovery. The presentation explores why this approach got a bum rap in the past, how this has changed, accurate binding prediction, machine learning, use of QSAR, efficient search and property filtering.

5:00 Creating Customized Research Computing Environments on Cloud, While Addressing Needs for Faster Data Transfer, and a High Performance Parallel File System

Sponsored by CYCLE COMPUTING

Jason Stowe, CEO, Cycle Computing

Cloud provides researchers the ability to create customized computing environments for drug design and life sciences. But with that flexibility, come challenges. This session will review successful enterprise & start-up use cases to highlight how people are using cloud - in production - today. It will also offer a vision into how to address other needs like faster data transfer speeds, a high performance parallel file system (Lustre), and encryption/security.

5:30 Best of Show Awards Reception in the Exhibit Hall with Poster Viewing

6:30 Close of Day

THURSDAY, APRIL 23

7:00 am Registration Open and Morning Coffee

» 8:00 PLENARY SESSION PANEL
Please see page 5 for details.

10:00 Coffee Break in the Exhibit Hall and Poster Competition Winners Announced

APPLICATIONS: LARGE-SCALE TO SMALL-SCALE

10:30 Chairperson's Remarks

Jason Tetrault, Associate Director, Business and Information Architect, R&D IT, Biogen

10:40 Next-Generation Sequencing and Cloud Scale: A Journey to Large-Scale Flexible Infrastructures in AWS

Jason Tetrault, Associate Director, Business and Information Architect, R&D IT, Biogen

Biogen has built burst capabilities for large-scale NGS processing and collaboration with our partners. This extension of our infrastructure capability allows us to be more nimble, process more data and scale as needed. It also gives us unique options as we work with collaborators at scale. Of course, because it is NGS data, doing it securely is important.

11:10 Data Communications in BSL-3 and BSL-4 Containment: Safety, Compliance and Security

John McCall, Director, Information Technology and Telecommunications, National Emerging Infectious Diseases Laboratories, Boston University

Innovative solutions for BSL-3 and BSL-4 facilities address the asset tracking, personnel monitoring and worker communication problems associated with personal protective equipment and physical environment design. I scope out what it takes to plan and roll out a wireless networking and voice-over-IP system that meets safety, security and compliance requirements at Boston University's National Emerging Infectious Disease Laboratory.

11:40 Breaking the Classical Barriers to Collaboration and Scientific Discovery - Distance and Data Size

Sponsored by aspera an IBM® company

Serban Simu, Vice President, Engineering & Co-Founder

Life sciences organizations need to dramatically reduce analytics time and speed up clinical interventions, but most still rely on shipping physical disks due to inherent problems with existing networks and transfer protocol

Track 3

Cloud Computing

Riding Cloud to Next-Generation Computing

inefficiencies. Spending days to transport data is not a viable option, this session will explore technology infrastructure for file transfer that will catalyze the transition from 1GbE to 10GbE and beyond.

12:10 pm Session Break

12:20 Luncheon Presentation *(Sponsorship Opportunity Available)* **or Lunch on Your Own**

1:20 Dessert Refreshment Break in the Exhibit Hall with Poster Viewing

REGULATIONS: DATA PRIVACY AND SECURITY

1:55 Chairperson's Remarks
John M. Conley, J.D., Ph.D., William Rand Kenan, Jr. Professor of Law, University of North Carolina, Chapel Hill ; Counsel, Robinson Bradshaw & Hinson

» 2:00 FEATURED PRESENTATION: IT AND INFORMATICS INNOVATION AT FDA
Roselie A. Bright, Sc.D., MS, PMP, Program Manager, Office of Information Management and Technology, Office of Informatics Technology and Innovation, Office of Operations, Office of the Commissioner, U.S. Food And Drug Administration (FDA)

OpenFDA was the first innovation created by Taha Kass-Hout, M.D., MS, upon joining FDA as the first Chief Health Information Officer in March 2013. OpenFDA was launched on June 2, 2014, allowing software developers, researchers and the public to tap into adverse events for drugs and medical devices; recalls, for drugs, devices and foods; and labeling for products on the market.

2:30 Global Developments in Privacy and Data Security Law

John M. Conley, J.D., Ph.D., William Rand Kenan, Jr. Professor of Law, University of North Carolina, Chapel Hill ; Counsel, Robinson Bradshaw & Hinson

The international legal climate governing privacy and data security is changing. The European Union is in the midst of a fundamental shift in its approach. The U.S. still lacks a national data law, so the states and individual federal agencies are groping toward a strategy. This presentation focuses on the impact of these ongoing changes on genomics, bioinformatics and health research.

3:00 PANEL DISCUSSION: Achieving Much-Needed Innovation while Hurdling the Barriers of Stringent Regulation

Moderator: John M. Conley, J.D., Ph.D., William Rand Kenan, Jr. Professor of Law, University of North Carolina, Chapel Hill; Counsel, Robinson Bradshaw & Hinson

Panelists:

Roselie A. Bright, Sc.D., MS, PMP, Program Manager, Office of Information Management and Technology, Office of Informatics Technology and Innovation, Office of Operations, Office of the Commissioner, U.S. Food And Drug Administration (FDA)

Dana Caulder, Senior Software Engineer, Bioinformatics and Computational Biology, Genentech

Chris Dwan, Assistant Director, Research Computing and Data Services, Broad Institute of MIT and Harvard

Sanjay Joshi, CTO – Life Sciences, Emerging Technologies Division, EMC

Dave Peterson, Executive Director, Vendor & Third Party Assurance, National IT Compliance, Kaiser Permanente Information Technology

Vas Vasiliadis, Director, Products, Computation Institute, University of Chicago and Argonne National Laboratory

The growth in patient healthcare and life sciences innovations can be attributed to technology enhancements like cloud computing, big data analytics and mobile applications, but may conflict with increasing regulatory compliance demands to ensure protection of healthcare life and quality as well as patient data privacy and security. This panel of esteemed technology solution providers and regulators debates real-world challenges and how regulation must also innovate at technology's pace.

4:00 Conference Adjourns

Organized by:
Cambridge Healthtech Institute

Track 4

Bioinformatics

Developments and Applications for Big Data

TUESDAY, APRIL 21

7:00 am Workshop Registration and Morning Coffee

8:00 – 11:30 Recommended Morning Pre-Conference Workshops*

Data Visualization in Biology: From Basics to Big Data

12:30 – 4:00 pm Recommended Afternoon Pre-Conference Workshops*

How Data-Driven Patient Networks are Transforming Biomedical Research

The Impact of Research Informatics on Laboratory Evolutions

** Separate registration required*

2:00 – 6:30 Main Conference Registration

» 4:00 PLENARY SESSION
Please see page 5 for details.

5:00 – 7:00 Welcome Reception in the Exhibit Hall with Poster Viewing

Sponsored by IBM

WEDNESDAY, APRIL 22

7:00 am Registration Open and Morning Coffee

» 8:00 PLENARY SESSION
Please see page 5 for details.

9:00 Benjamin Franklin Awards and Laureate Presentation

9:30 Best Practices Awards Program

9:45 Coffee Break in the Exhibit Hall with Poster Viewing

Sponsored by INTERNET 2

BIG DATA, DIGITAL TOOLS AND BIOINFORMATICS ACROSS MULTIPLE RESEARCH INITIATIVES

10:50 Chairperson's Opening Remarks

11:00 An Algorithmic Rationale for the Irreversibility of Biological Ageing

Simon Berkovich, Professor, Computer Science, The George Washington University

Maryam Yammahi, Ph.D., Computer Science, The George Washington University

The presentation describes how the process of ageing is related to the specifics of the big data organization of biological information processing.

11:30 Role of Data and Digital Tools in Autoimmune Disorders

Bonnie Feldman, D.D.S., Digital Health Analyst, DrBonnie360

Turning data into useable information is challenging for complex chronic diseases like autoimmune disease. Tools now exist to begin building more personalized data sets from the ground up, while using this information to learn how to ask the right questions. This talk discusses innovations in personal data, new approaches to microbiome research around autoimmune disease, and bigger picture issues related to data sharing and data donation.

12:00 pm IBM Watson Cognitive Computing Applications in Healthcare and Life Sciences

Sponsored by IBM

Philip G. Abrahamson, Ph.D., Research Staff, IBM Watson

Information is being created faster than it can be consumed. This talk will share experiences applying IBM Watson Cognitive Computing to help researchers explore huge volumes of unstructured and structured content to discover insights and information. Examples include accelerating the understanding of the underlying biology of diseases; identifying, evaluating, and selecting drug targets and candidates, including leveraging safety and toxicity information; improving drug comparative effective studies; and competitive intelligence.

12:30 Session Break

12:40 Luncheon Co-Presentation I: How Revolutionary Machine Learning Advancements Improve Drug Research Productivity and Drive Discovery of Valuable Insights across Disparate Content Repositories

Sponsored by Content Analyst Company

Melissa Chapman, Principal, The Riverhead Group

Phillip Clary, Vice President, Content Analyst Company

Ensuring product quality, efficacy and safety by searching for correlations across disparate collections of eCTDs, articles, reports, and regulatory intelligence can be incredibly time-consuming. Boolean keyword searches can produce false positives and omit relevant results, and laborious taxonomies can be a burden to build and maintain. Using a live demonstration, attendees will see how the latest advances in machine learning technology can dramatically improve productivity and reveal key insights within large collections of unstructured content.

1:40 Session Break

BIG DATA, DIGITAL TOOLS AND BIOINFORMATICS ACROSS MULTIPLE RESEARCH INITIATIVES

1:50 Chairperson's Remarks

Michael Liebman, Ph.D., Managing Director, IPQ Analytics, LLC

1:55 Metabolic Biomarkers in Duchenne Muscular Dystrophy

Subha Madhavan, Ph.D., Director, Innovation Center for Biomedical Informatics, Georgetown University Medical Center; Director, Clinical Informatics, Lombardi Comprehensive Cancer Center; Director, Biomedical Informatics, Georgetown-Howard Universities CTSA; Associate Professor, Department of Oncology, Georgetown University

Duchenne Muscular Dystrophy (DMD) is a devastating degenerative X-linked disorder which affects approximately 1 in 5,000 newborn males and results in muscle degeneration, eventual loss of ambulation around the age of 9, and a life expectance of around 25 years of age. A bioinformatics platform for metabolic data interpretation has been developed and tested to identify DMD-associated biomarkers and will be made available on GitHub once validation is complete. This platform will be presented along with another use case from a breast cancer metabolomics study.

Organized by:
Cambridge Healthtech Institute

Track 4

Bioinformatics

Developments and Applications for Big Data

2:25 Personalized Medicine: Moving from Correlation to Causality in Breast Cancer

Michael Liebman, Ph.D., Managing Director, IPQ Analytics, LLC

Sabrina Molinaro, Ph.D., Institute for Clinical Physiology, National Research Council, Italy

We have developed a fundamental model of the disease process for breast cancer, from pre-disease through early detection, treatment and outcome, and apply a multi-scalar approach across the risk assessment-enhanced diagnosis-therapeutic decision axis and will present the modeling methodologies.

2:55 Streamline R&D and Catalyze Drug Repositioning by Identifying Expert Networks and Expertise

Sponsored by SINEQUA

Xavier Pornain, Vice President, Sales & Alliances, Sinequa

Finding networks of experts with similar or complementary expertise on a given subject helps avoid costly redundant research, shed light on a complex research problem from different angles, foster cooperation, facilitate drug repositioning, and accelerate time to market. This session will delve into the benefits pharmaceutical companies are seeing by employing Search & Analytics technology to: "link" researchers and teams with one another, create internal "journals of science" to share internal results and snippets, access "breaking science", with alerts and spotting trends across all scientific information. We show solutions for dealing with scientific vocabulary, detecting "synonyms" as well as "similar" and "complementary" notions, e.g. brand names for drugs, scientific names for the active ingredients, and even descriptions of molecules using a standard description language. In addition, we analyze vast quantities (200 to 500 million) of highly technical documents and data (billions of records), such as internal and external publications, patent filings, lab reports, clinical test reports, trade databases, etc.

3:10 Cloud-Based Solutions for Population-Scale, Whole Human Genome and Exome Analysis

Sponsored by DNAnexus

George Asimenos, Ph.D., Director, Science & Clinical Solutions, DNAnexus

Thanks to advances in sequencing technology, the size and scope of DNA sequencing projects is rapidly moving towards an era of thousands of whole genomes and tens of thousands of exomes per year. Learn how certain field-leading institutes are using a cloud-based bioinformatics platform to manage their big data deluge across multiple initiatives.

3:25 Refreshment Break in the Exhibit Hall with Poster Viewing

4:00 Using Games as Data Analytical Tools

Melanie Stegman, Ph.D., Owner, Molecular Jig Games, LLC.; Director, Science Game Center

Immune Defense is a video game, but it is also a molecular level simulation of the immune system. Individual data points tell us very specific details about cells, and a large database of these details should tell us a more complete story. But do we have enough data yet to tell the story of one cell, facing one bacterium? It has been a challenge gathering the knowledge to create this small story. Part of Immune Defense game development is the creation of a "game level editor." We can make new molecules, give them new binding partners, assign their affinities for each partner, increase or decrease their relative concentrations and give our enzymes activity... We have created a "medium data" analysis chamber--that is, not Big Data, but more data than one person can hold in their head. We are planning to build up our level editor as a tool for biochemists to analyze their data with much more perspective than ever before. We will also have a tool for scientists, students, public and game developers to use to create realistic scenarios for various purposes, from science fairs to testing to video game development. Play Immune Defense at www.MolecularJig.com/demo.

4:30 A Rigorous Methodology for Non-Randomized & Observational Study in Healthcare Testing

Gil Weigand, Ph.D., Director, Strategic Projects, Oak Ridge National Laboratory

We present an advanced rigorous science-based evaluation methodology for evaluation in healthcare testing. The methodology extends today's general practice, rapid cycle evaluation by introducing *in silico* methods of big data and modeling & simulation and tightly integrating the methods within a knowledge discovery infrastructure. We call this approach IDAMS-HC—integrated data, analytics, modeling, and simulation for healthcare.

5:00 Service-Oriented Bioinformatics – the CDC Influenza Sequence Data Management System

John M. Greene, Ph.D., CSM, Senior Director, Bioinformatics, Bioinformatics Solutions and Support, SRA International, Inc.

Next-Generation Sequencing technologies have opened enormous opportunities for improvements in the surveillance of infectious diseases such as influenza. However, effective use of such sequencing information depends on a robust system to store, manage, analyze, and interpret sequence data. The Influenza Sequence Data Management System (ISDMS) at the Centers for Disease Control and Prevention (CDC)'s Influenza Division in Atlanta fills this role using a service-based approach developed by SRA International that we refer to as 'service-oriented bioinformatics'. Services are small programs that are coordinated by an enterprise service bus, in this case Apache ServiceMix, based on the service-oriented architecture (SOA) model. Services can be written in different languages and act as modular components of the system, providing individual functionality, such as searching, annotation display, and location standardization. These services underpin data loading, data annotation, and data display, and services can be combined to implement new features and reused to speed development.

5:30 Best of Show Awards Reception in the Exhibit Hall with Poster Viewing

6:30 Close of Day

THURSDAY, APRIL 23

7:00 am Registration Open and Morning Coffee

» 8:00 PLENARY SESSION
Please see page 5 for details.

10:00 Coffee Break in the Exhibit Hall and Poster Competition Winners Announced

DATA CAPTURE, ANALYSIS, MODELING & SIMULATION

10:30 Chairperson's Remarks

Michael D. Stadnisky, Ph.D., CEO, FlowJo, LLC

10:40 Structure-Based Algorithms to Predict Drug-Mediated Toxicity

Khaled Barakat, Ph.D., Assistant Professor, Katz Group-Rexall Centre for Pharmacy & Health Research, University of Alberta

Drugs must physically fit into the binding site(s) within their targets. While reaching their targets they interact with many cellular components and may bind to undesired critical off-targets, leading to severe toxicity. This talk presents how state-of-the-art high performance computing and cutting-edge molecular dynamics simulations were used to predict drug-mediated toxicity and characterize these events at the atomic level.

11:10 An Informatics Solution for the Precise Registration and Visualization of Biological Molecules

Roxanne Kunz, Ph.D., Senior Scientist, Therapeutic Discovery, Amgen, Inc.

A custom bioinformatics software application for the registration and representation of biological molecules will be described. The system

Track 4

Bioinformatics

Developments and Applications for Big Data

includes a flexible, modality-independent editor to define biomolecules in a step-wise fashion, backed by a chemical structure-based database catalog to precisely capture atomic-level modifications of amino acids and other non-proteinaceous components. Data-driven visual representations based on canonical biological molecule structure reference types, such as IgG1 monoclonal antibodies and subtypes thereof, are dynamically constructed and interactive.

11:40 Man Versus Machine: Validating, Optimizing, and Predicting Outcomes in Single Cell Phenomics

Sponsored by FLOWJO, LLC

Michael D. Stadnisky, Ph.D., CEO, FlowJo, LLC

The exponential increase in the throughput and content of flow and mass cytometry assays has challenged the paradigm of DIY data management, manual analysis, and 2D visualization in single cell phenomics. We have developed and assessed the ability of an automated pipeline to direct analysis, statistical cluster comparison for iterative pipeline improvement, plug-and-play automated clustering algorithms, and predictive phenotype prediction.

12:10 pm Session Break

12:20 Luncheon Presentation I *(Sponsorship Opportunity Available)*

12:50 Luncheon Presentation II *(Sponsorship Opportunity Available)*

1:20 Dessert Refreshment Break in the Exhibit Hall with Poster Viewing

1:55 Chairperson's Remarks

2:00 Using Deep Learning Techniques for Word Vectors Generation for Insight into Poorly Structured Textual Data

Mark Pinches, Senior Scientist, Data Modeling and Bioinformatics, Drug Safety and Metabolism, AstraZeneca

Word vectors carry a number of interesting properties that can be applied to textual data in order to cluster/stratify data for additional analysis. This emerging approach has been used in other fields but here we apply it to biological data. This presentation illustrates a concrete example of this technique with clear and unique outcomes.

2:30 Examining the Health Effects of Multiple Environmental Exposures on Subpopulations Using a Big Data Platform

Chirag Patel, Ph.D., Research Associate, Center for Biomedical Informatics, Harvard Medical School/Pivotal, Inc.

The presentation will demonstrate how to apply these new computational paradigms to epidemiological datasets, enabling more complex analyses resulting in new insights. The resulting analysis of dataset underscores the utility of examining more complex relationships between multiple elements in the environment and attributes of individuals not commonly explored in traditional epidemiological association studies.

3:00 Streamlined Planning, Execution, Data Capture and Analysis of Peptide Preformulation Stability Studies

Roman Affentranger, Dr. sc. Nat, Head, Small Molecule Discovery Workflows, Roche

The presentation will illustrate what we have implemented for the peptide preformulation scientists in their electronic lab notebook to efficiently design peptide formulation stability studies. The study can cover a number of different formulations, and with the definition of time points, stress conditions and desired analytical methods the required number of vials as well as individual material amounts are automatically calculated.

3:30 Welcome to the Future: Data Analysis in a Language Workbench

Fabien Campagne, Ph.D., Assistant Professor and Laboratory Head, Institute for Computational Biomedicine, Weill Cornell Medical College

Our laboratory is developing innovative open-source, fully prototyped approaches to data analysis that can simplify the solution of many analysis problems (such as in high-throughput sequence data analysis, biomarker development and bioinformatics). The talk will focus on Language Workbench technology, practical applications of this technology to data analysis, and how both end-users and tool designers can benefit from its application.

4:00 Conference Adjourns

Organized by:
Cambridge Healthtech Institute

Track 5

Next-Gen Sequencing Informatics

Advances in Large-Scale Data Analysis and Interpretation

TUESDAY, APRIL 21

7:00 am Workshop Registration and Morning Coffee

8:00 – 11:30 Recommended Morning Pre-Conference Workshops*
Genome Assembly and Annotation
Intelligent Methods Optimization of Algorithms for NGS

12:30 – 4:00 pm Recommended Afternoon Pre-Conference Workshops*
Customizing Your Digital Research Environment with Genome Browsers
Large Scale NGS Analysis Using Globus Genomics
** Separate registration required*

2:00 – 6:30 Main Conference Registration

» 4:00 PLENARY SESSION
Please see page 4 for details.

5:00 – 7:00 Welcome Reception in the Exhibit Hall with Poster Viewing
Sponsored by IBM

WEDNESDAY, APRIL 22

7:00 am Registration Open and Morning Coffee

» 8:00 PLENARY SESSION
Please see page 4 for details.

9:00 Benjamin Franklin Awards and Laureate Presentation

9:30 Best Practices Awards Program

9:45 Coffee Break in the Exhibit Hall with Poster Viewing
Sponsored by INTERNET 2

EMERGING TRENDS AND PREDICTIONS OF NGS INFORMATICS

10:50 Chairperson's Opening Remarks
Narges Bani Asadi, Founder and CEO, Bina Technologies, Inc., a member of the Roche Group

11:00 Global Next Generation Sequencing Informatics Markets: Inflated Expectations in an Emerging Market
Greg Caressi, Senior Vice President, Healthcare and Life Sciences, Frost & Sullivan
This presentation evaluates the global next-generation sequencing (NGS) informatics markets from 2012 to 2018. Learn key market drivers and restraints, a detailed analysis of the changing competitive landscape, revenue forecasts, and important trends and predictions that affect market growth. Key highlights for many of the leading NGS informatics services providers, commercial primary and secondary data analysis tools vendors, commercial biological interpretation and clinical reporting tools vendors, and NGS LIMS vendors will be presented.

OPEN SOURCE AND LARGE-SCALE COMPUTING

11:30 Large-Scale NGS Analysis Using Globus Genomics: Challenges and User Success Stories
Ravi Madduri, Fellow, Computation Institute, University of Chicago and Argonne National Lab
Dinanath Sulakhe, Solutions Architect, Computation Institute, University of Chicago and Argonne National Lab
In this talk, we will present some of the challenges in scaling up NGS analysis on public cloud infrastructure and present user success stories where we have overcome them.

12:00 pm Turn-Key Variant Analysis for the Biologist Using the Maverix Analytic Platform
Sponsored by Maverix Biomics
Dan Kearns, Director, Software Development, Maverix Biomics, Inc.
Studies leveraging WGS, Exome, and Targeted sequencing data are commonly limited by the tools, infrastructure, and trained bioinformaticians necessary to process, interpret and manage the data. The Maverix Analytic Platform addresses these challenges through a unique environment designed for biologists. This cloud-based platform leverages best-in-class tools and methods, and provides an integrated environment to enable visualization and interpretation of results.

12:15 Developing and Provisioning Robust Automated Analytical Pipelines for Whole Genome-Based Public Health Microbiological Typing
Sponsored by DDN STORAGE
Anthony Underwood, Ph.D., Lead, Bioinformatics, Infectious Disease Informatics, Microbiology Services Division, Public Health England
Whole genome sequencing has great potential for microbial characterization in public health. Open source bioinformatics tools can generate necessary information, however converting these tools for usage in routine public health is challenging. They must be automated, auditable, timely, and robust, as well as record errors and log outputs. Dr Underwood will discuss the infrastructure, software architecture and algorithms used for this at Public Health England.

12:30 Session Break

12:40 Luncheon Presentation I: Sample Aggregation and Analytics in the Post-$1,000 Genome Era
Sponsored by illumina
John Shon, Vice President, Bioinformatics & Data Sciences, Illumina, Inc.
With the launch of the Illumina HiSeq X Ten system, the long-promised $1,000 genome became a reality. But as is often the case in science and engineering, the realization of one goal reveals new challenges to surmount. The economics of sequencing now make the sequencing of entire populations feasible, but aggregating, tracking, and analyzing whole human genome data cannot be done serially when it is produced in parallel. This presentation will discuss parallel sample processing approaches that enable multi-sample genome interpretation and analysis of large cohorts by employing cloud-scale computing.

1:10 Luncheon Presentation II *(Sponsorship Opportunity Available)*

1:40 Session Break

1:50 Chairperson's Remarks
Carlos P. Sosa, Ph.D., HPC Chemistry and Life Sciences Technical Lead, Biomedical Informatics and Computational Biology, Cray Inc, University of Minnesota Rochester

Organized by:
Cambridge Healthtech Institute

Track 5

Next-Gen Sequencing Informatics

Advances in Large-Scale Data Analysis and Interpretation

1:55 The Cloud Reigns: Enabling Scalable Analysis and Storage for High-Throughput Next-Gen Sequencing

John Penn, Associate Manager, NGS Data Analysis, Regeneron Genome Center

2:25 Data Intensive Academic Grid (DIAG): A Free Computational Cloud Infrastructure Designed for Bioinformatics Analysis

Anup Mahurkar, Executive Director, Software Engineering and IT, Institute for Genome Sciences, University of Maryland School of Medicine

2:55 Co-Presentation: The Challenges of Scaling Platforms for Translational Science: New Approaches and Case Studies

Sponsored by IBM

Houtan Aghili, Ph.D., Senior Technical Staff Member, Industry Solutions - Healthcare and Life Sciences; IBM Software Group

Janis Landry-Lane, Genomics Solutions, Software Defined Infrastructure, IBM World-Wide

As researchers build platforms for translational science, High Performance Data Centric Computing will be a key investment that must be considered in order to provide an integrated and scalable solution which fulfills the needs of multiple departments. In this session, we will cover: processing the NGS pipeline in order to bring omics data into a scalable information management platform, the role of natural language processing for integrating unstructured information, the integration of on-premise and cloud solutions, and effective data and content management at scale. IBM will present both a vision and potential solutions that have enabled our customers to build an effective architecture.

3:25 Refreshment Break in the Exhibit Hall with Poster Viewing

NGS VARIANTS & GENE MAPPNG AND EXPRESSION

4:00 Deep Sequencing Based Analysis of Ig repertoire in Humanized Mice

Stefan Klostermann, Ph.D., Expert Scientist, Bioinformatics, Data Science, Roche Innovation Center Penzberg

On our quest for human biotherapeutical antibodies we developed a novel methodology: Instead of replacing the mouse genomic immune loci by the human orthologs we reconstituted the humoral immune response in immunodeficient mice transplanted with human hematopoietic stem cells.

An in-depth characterization of the reconstituted immune system by data analysis of deep sequencing Ig repertoire validated the humanized mouse be immunological equivalent to human donors.

4:20 Development of Novel Algorithms for Assembly of RNA-seq Reads Into Transcriptomes

Guojun Li, Ph.D., Professor, Mathematics, Shandong University

We developed a more effective and efficient assembler to assemble RNA-seq reads into full-length transcripts encoded in a genome based on a new perception that the full-length transcripts would be better recovered from combinations of spliced junctions which can be detected by aligning RNA-seq reads against a reference genome using splice-awared aligner than from overlapped reads.

4:40 BLASTing with Chromatin Architecture: A Novel Method of Genomic Functional Element Identification and Annotation

Michael J. Buck, Ph.D., Associate Professor, Department of Biochemistry, SUNY at Buffalo; Director, Stem Cell Sequencing/ Epigenomics Center, The State University of New York at Buffalo; Co-Director, Next-Generation Sequencing & Expression Analysis Core, The State University of New York at Buffalo

In order to facilitate identification and characterization of new classes of genomic features, we developed and implemented a chromatin Architecture Basic Local Alignment Search Tool (ArchBLAST). We show ArchBLAST is capable of predicting both gene expression and genomic feature directionality as well as identifying cell-type specific enhancers using chromatin architecture and/or DNA-binding protein signatures.

5:00 High Performance Computing Technology and Methodology Applied to Next-Generation Sequencing Workflows

Sponsored by CRAY

Carlos P. Sosa, Ph.D., HPC Chemistry and Life Sciences Technical Lead, Biomedical Informatics and Computational Biology, Cray Inc, University of Minnesota Rochester

High Performance Computing (HPC) Technology and Methodology (profiling and optimizing) are enabling scientists in many disciplines to achieve progressively more demanding and valuable results. In this talk we will illustrate how the same technology and methodology can be used to dramatically accelerate next-generation sequencing (NGS) workflows.

5:30 Best of Show Awards Reception in the Exhibit Hall with Poster Viewing

6:30 Close of Day

THURSDAY, APRIL 23

7:00 am Registration Open and Morning Coffee

» 8:00 PLENARY SESSION PANEL
Please see page 5 for details.

10:00 Coffee Break in the Exhibit Hall and Poster Competition Winners Announced

NGS DATA MANAGEMENT, PROCESSING, AND ANALYSIS

10:30 Chairperson's Remarks

Alexander Wait Zaranek, Ph.D., Director of Informatics, Harvard Personal Genome Project; Chief Scientist, Curoverse, Inc.

10:40 Informatics Infrastructure for Secure Access, Visualization and Analysis of NGS data

Ted Kalbfleisch, Ph.D., Assistant Professor, Biochemistry and Molecular Biology, University of Louisville

This talk describes an augmentation to the Variant Call Format standard that will facilitate access to the source mapped dataset for inspection, or re-evaluation. We also describe an application programming interface that we have developed that allows access to these source NGS datasets to authorized users, that can function within a federated identity management environment.

11:10 NGS Data Management at Lilly: Progress and Challenges

Yuhao Lin, Consultant-Informatics Capabilities, Eli Lilly

11:40 Simplifying NGS Data Management with Metadata Centric Intelligent Storage

Sponsored by GENERAL ATOMICS

Robert Murphy, Big Data Program Manager, General Atomics

The rapid advance of NGS speed and cost reduction has opened the floodgates to staggering amounts of data. Managing overwhelming genomics data growth is critical to continued discovery. Adding workflow-specific NGS metadata is the key. With it, NGS constituents can find and access valuable data, share it world-wide for collaborative research, and make it available to support reproducibility mandates, while ensuring provenance, curation and

Track 5

Next-Gen Sequencing Informatics

Advances in Large-Scale Data Analysis and Interpretation

future data availability. This presentation will describe an easily deployable metadata-oriented data management system that can be used to simplify all aspects of NGS data management.

12:10 pm Session Break

12:20 Luncheon Presentation *(Sponsorship Opportunity Available)* **or Lunch on Your Own**

1:20 Dessert Refreshment Break in the Exhibit Hall with Poster Viewing

1:55 Chairperson's Remarks

Alexander Wait Zaranek, Ph.D., Director of Informatics, Harvard Personal Genome Project; Chief Scientist, Curoverse, Inc.

2:00 Technology and Data Analysis Methods for NGS Data

Yaoyu Wang, Ph.D., Associate Director, Center for Cancer Computational Biology, Dana Farber Cancer Institute

2:30 Talk Title to be Announced

Craig Pohl, Co-Director, Bioinformatics, The Genome Institute, Washington University

3:00 Reproducible NGS Research: Practical Approaches and Case Studies

Joseph Szustakowski, Ph.D., Group Director, Translational Bioinformatics, Bristol-Myers Squibb

3:30 An Open Source Precision Medicine Platform for Cloud Operating Systems

Alexander Wait Zaranek, Ph.D., Director of Informatics, Harvard Personal Genome Project; Chief Scientist, Curoverse, Inc.

The unique "big-data" requirements for precision medicine are best served by a common open-source platform developed collaboratively by and for the biomedical community. This platform can address the need to share the influx of human sequence data amongst various stakeholders (researchers, physicians, and the individuals themselves), stringent privacy and security guarantees that comply with government regulations, deep provenance for data reproducibility and analysis validation, and flexibility in efficiently compressing and searching of these data. We launched the Arvados project to meet community needs and are announcing its latest component, Lightning, an open-source, distributed query and translation engine.

4:00 Conference Adjourns

Organized by:
Cambridge Healthtech Institute

Track 6

Clinical & Translational Informatics

Transforming Biological Data to Clinical Development

TUESDAY, APRIL 21

7:00 am Workshop Registration and Morning Coffee

8:00 – 11:30 Morning Pre-Conference Workshops*

12:30 – 4:00 pm Recommended Afternoon Pre-Conference Workshops*

How Data-Driven Patient Networks are Transforming Biomedical Research

** Separate registration required*

2:00 – 6:30 Main Conference Registration

» 4:00 PLENARY SESSION
Please see page 5 for details.

5:00 – 7:00 Welcome Reception in the Exhibit Hall with Poster Viewing

Sponsored by IBM

WEDNESDAY, APRIL 22

7:00 am Registration Open and Morning Coffee

» 8:00 PLENARY SESSION
Please see page 5 for details.

9:00 Benjamin Franklin Awards and Laureate Presentation

9:30 Best Practices Awards Program

9:45 Coffee Break in the Exhibit Hall with Poster Viewing

Sponsored by INTERNET 2

TRANSLATIONAL AND CLINICAL TRIAL INFORMATICS INSIGHTS

10:50 Chairperson's Opening Remarks

Joy King, Principal Consultant & Practice Lead, Life Sciences, Teradata Corporation

11:00 Towards Patient-Centered Clinical Trial Eligibility Criteria Design

Chunhua Weng, Ph.D., Florence Irving Assistant Professor of Biomedical Informatics; Co-Director, Biomedical Informatics Core for CTSA, Columbia University

This talk will summarize the patterns in patient selection among > 170,000 clinical trials archived on Clinical Trials.gov and their association with recruitment outcomes. The need and opportunities for data-driven patient-centered eligibility criteria design will be described.

11:30 NIH/NCATS GRDRSM Program: A Model to Accelerate Rare Diseases Research

Barbara W. Brandom, M.D., Professor, Department of Anesthesiology, University of Pittsburgh; Director, North American Malignant Hyperthermia Registry of the Malignant Hyperthermia Association of the United States

NCATS has established the Global Rare Disease Patient Registry Data Repository NIH/NCATS GRDRSM Program. The aim is to develop a Web-based resource that aggregates, secures and stores de-identified patient information from many different registries for rare diseases, all in one place. The ultimate goal is to improve therapeutic development and quality of life for the many millions of people suffering with a rare disease.

12:00 pm Integrating Data is the Key to Translational Research and the Future of Personalized Medicine

Sponsored by PerkinElmer For the Better

Jens Hoefkens, Director, Research Strategic Marketing, PerkinElmer, Inc.

Emerging technologies are driving Translational Medicine research and PerkinElmer is developing tools, platforms, and algorithms to generate, analyze, visualize and store those data. This talk will describe how we integrate high-content data with clinical observations to enable our customers to derive and test unique hypotheses.

12:30 Session Break

12:40 Luncheon Presentation I: Implementing Continuous Improvement to Reduce Risks and Speed Clinical & Translational Informatics

Sponsored by TERADATA

Ed Acker, Ph.D., Principal Life Sciences Consultant, Teradata Corporation

Hear how Informatic organizations can innovate by implementing a continuous improvement strategy that reduces the risk of finding the right drug targets, the right treatment attributes for drugs and the right population that best responds to the treatment. Historically siloed research and clinical data repositories, along with today's large, continuously updated health data repositories, make these programs extremely difficult to implement … until now!

1:10 Luncheon Presentation II: Big Data in a Small World: Exercising Control in Global Clinical Trials

Sponsored by eClinicalOS

Don Turner, Senior Vice President, Business Strategy and Commercialization Global Sales, Marketing, and Partnerships with Merge eClinical

This presentation explores how advances in information technology and communications are strengthening researchers' ability to exercise the control needed to ensure successful and cost-efficient studies on a global stage. In addition, it will examine how digital data management is changing the dynamic of long-standing traditions that hamper global trials and enabling a wider array of research organizations to compete effectively regardless of size or location.

1:40 Session Break

COLLABORATIVE APPROACHES IN CLINICAL RESEARCH, BIOMEDICAL RESEARCH AND THE PHARMACEUTICAL SPACE

1:50 Chairperson's Remarks

Alex Sherman, Director, Strategic Development and Systems, Neurological Clinical Research Institute, Massachusetts General Hospital

1:55 Open Source National Network Facilitating Healthcare and Resource Data Sharing

Doug Macfadden, Chief Informatics Officer, Harvard Catalyst

Bhanu Bahl Director of Informatics, Harvard Catalyst

Accrual to Clinical trials (ACT) project supported by NCATS was launched with the goal of creating a network of 60 Clinical Translational Science Center Award (CTSA) sites. The network will facilitate investigators to query EHR data across all these sites for cohort exploration and subsequently engage and enroll identified patients into clinical trials. SHRINE (Shared Health Research Information Network)

Organized by:
Cambridge Healthtech Institute

Track 6

Clinical & Translational Informatics

Transforming Biological Data to Clinical Development

is a system developed by Harvard Catalyst for enabling clinical researchers to query across distributed hospital electronic medical record systems.

2:25 NeuroBANK™, Accelerated Research Environment as a Model for Collaboration and Cooperation in Clinical Research

Alex Sherman, Director, Strategic Development and Systems, Neurological Clinical Research Institute, Massachusetts General Hospital

NeuroBANK™, a patient-centric platform that allows clinicians and investigators to aggregate and cross-link clinical and research information from clinical visits, clinical studies, health records, and self-reported patient outcomes, and to connect it to biospecimen, images and genetic files. Will discuss how to find or create incentives for collaborations.

2:55 Data Access Models for Genetic Data Sharing – GSK SHARE and the GA4GH Beacon

Karen King, Head, Genetic Data Sciences, GlaxoSmithKline

3:25 Refreshment Break in the Exhibit Hall with Poster Viewing

VISUALIZATION TOOLS TO ADVANCE TRANSLATIONAL AND CLINICAL RESEARCH

4:00 Making Visualization and Exploration Tools Truly Useful in the Regulatory Setting

Timothy Kropp, Ph.D., Associate Director for Innovation, Office of Computational Science, US FDA/CDE

As FDA applies tools and technologies to regulatory data ("big" data as well as "little") a lot is being learned about what is truly useful and in what contexts (not what is pretty or simply interesting). This talk will provide an overview of what informatics approaches FDA/CDER is using for visualization and exploration of scientific/clinical review data, how we are modifying what we use for better usefulness, what our biggest challenges and opportunities are, and where we want to go.

4:30 Feeding the Analytics Engine: Targeting Optimal Clinical Trial Sites, a Case Study

James Gill, Ph.D., Director, Research Analytics and Visualization, Bristol-Myers Squibb

It is no surprise that as soon as an analytical approach is proposed, access to data becomes a hurdle. In this talk we review a successful approach to improving our clinical trials site selection process by leveraging unique data in a dashboard format. Our keys to success included a clear understanding of the impact of different factors on site performance, how we can find surrogates for non-existing data and using an exploratory process with our scientists.

5:00 Delivering Standardized Clinical and Preclinical Data to Scientists in Guided Analysis

Baisong Huang, Principal Statistical Analyst, Novartis Institutes for BioMedical Research, Inc.

As visualization tools evolve and become widely accepted in investigating and monitoring drug safety and efficacy, rapid access to standardized, interpretable data views is becoming essential. We will present some examples how we standardized and aggregated data in both translational and clinical settings and provided guided analysis to visualize the data in real-time.

5:30 Best of Show Awards Reception in the Exhibit Hall with Poster Viewing

6:30 Close of Day

THURSDAY, APRIL 23

7:00 am Registration Open and Morning Coffee

» 8:00 PLENARY SESSION PANEL
Please see page 5 for details.

10:00 Coffee Break in the Exhibit Hall and Poster Competition Winners Announced

ADVANCING TRANSLATIONAL RESEARCH WITH NEW ANALYTICS AND PLATFORMS

10:30 Chairperson's Remarks

Yuriy Gankin, Ph.D., Chief Life Science Officer, EPAM Systems

10:40 Translational R&D Analytics: Delivering 'Big Insights' to Drive Translational Research

Kaushal Desai, Associate Director, Translational R&D Analytics and Decision-Support, Research Informatics & Automation, Bristol-Myers Squibb

This session will explore case studies demonstrating how translational R&D analytics can inform patient stratification and trial design in early clinical and translational research. The talk will focus on the journey from a lack of discoverability for disjointed datasets to insights that drive key decisions in translational research. Challenges associated with delivering actionable information at the point of decision-making will be highlighted and opportunities to deliver business value will be outlined using real examples.

11:10 Integrated Genomics Platform: Putting Patients and Their Genomes into the Focus of Our Research

Nora Manstein, Ph.D., IT Project Manager, Bayer Business Services GmbH

We have established the Integrated Genomics Platform (IGP) as a central tool for genomics research in Cardiology, Oncology and Clinical Sciences. The platform supports advanced data analysis and is intended to simplify discovery processes. In this strategic project, we have overcome known bottlenecks and enabled true translational research by establishing a company-wide mandatory repository and toolbox for storage and analysis of genomics data as well as common standards for data annotation, privacy & security.

11:40 Building a Globally Distributed, Hybrid NGS Sequence Analysis and Integration Infrastructure for Oncology Discovery and Translational R&D

Sponsored by bina TECHNOLOGIES

Justin H. Johnson, Principal Scientist, AstraZeneca

Next-Generation Sequencing is changing the way pharmaceutical companies develop drugs, perform patient stratification, and evaluate treatment efficacy. However, managing the massive amounts of NGS data has introduced fundamental IT challenges. Here we discuss the implementation of a fast, flexible, scalable and validated IT infrastructure that can streamline the upkeep of the NGS analysis workflow and the distribution of genomic information throughout an organization for translational discovery.

12:10 pm Session Break

12:20 Luncheon Presentation *(Sponsorship Opportunity Available)* or Lunch on Your Own

Track 6

Clinical & Translational Informatics

Transforming Biological Data to Clinical Development

1:20 Dessert Refreshment Break in the Exhibit Hall with Poster Viewing

1:55 Chairperson's Remarks

Brenda Yanak, Ph.D., Director, Precision Medicine Leader, Clinical Innovation, Pfizer

2:00 GenISIS: Powering the Establishment of a Mega Scale National Resource for Biospecimens and Linkable Longitudinal Clinical and Genomic Data

Saiju Pyarajan, Scientific Director, MAVERIC, VA Boston Healthcare System

Veterans Administration (VA) embarked on the Million Veterans Program (MVP) in 2011 to collect and store consented biosamples from a million Veterans. GenISIS is the data integration and mining platform that powers MVP. GenISIS provides the framework for integration of longitudinal clinical & molecular data and the analytical platform & tools for performing genome-phenome analysis. Unlike many platforms targeted for specific diseases GenISIS allows for building analytical cohorts for a large number of diseases.

2:30 Technology Framework to Operationalize Biomarker-Focused Clinical Research

Brenda Yanak, Ph.D., Director, Precision Medicine Leader, Clinical Innovation, Pfizer

3:00 Optimizing Clinical Biomarker Data Collection for Translational Research

Al Wang, Associate Director, Exploratory Clinical & Translational Research IT, Bristol-Myers Squibb

Study teams are asked to plan and execute protocols that involve complex biological sample handling and distribution requirements, and significant effort is required to track those samples across multiple external partners to ensure high quality data. There are also real challenges in receiving the data back from these high-dimensional assays in a format that is consistent and ready for analysis. This talk will dig deeper into this problem, and offer some suggestions on possible solutions for real-time monitoring of sample collection and automation of data formatting.

3:30 PANEL DISCUSSION: How Has Translational Medicine Benefited from Big Data?

Moderator: Anastasia Christianson, Head, Translational R&D IT, Bristol-Myers Squibb

Panelists:

Justin H. Johnson, Principal Scientist, AstraZeneca

James Cai, Head, Data Science, Roche, Translational Clinical Research Center (TCRC)

Matthew V. St. Louis, Data Scientist, Predictive Informatics, R&D BT Business Insights, Pfizer

Heidi L. Rehm, Ph.D., FACMG, Chief Laboratory Director, Laboratory for Molecular Medicine, Partners HealthCare; Associate Professor, Pathology, Brigham & Women's Hospital and Harvard Medical School

4:00 Conference Adjourns

Organized by:
Cambridge Healthtech Institute

Track 7

Data Visualization and Exploration Tools

Genomics, Drug Discovery and Clinical Development

TUESDAY, APRIL 21

7:00 am Workshop Registration and Morning Coffee

8:00 – 11:30 Recommended Morning Pre-Conference Workshops*
Integrative Visualization Strategies for Large-Scale Biological Data

12:30 – 4:00 pm Recommended Afternoon Pre-Conference Workshops*
Customizing your Digital Research Environment with Genome Browsers
** Separate registration required*

2:00 – 6:30 Main Conference Registration

» 4:00 PLENARY SESSION
Please see page 5 for details.

5:00 – 7:00 Welcome Reception in the Exhibit Hall with Poster Viewing

Sponsored by IBM

WEDNESDAY, APRIL 22

7:00 am Registration Open and Morning Coffee

» 8:00 PLENARY SESSION
Please see page 5 for details.

9:00 Benjamin Franklin Awards and Laureate Presentation

9:30 Best Practices Awards Program

9:45 Coffee Break in the Exhibit Hall with Poster Viewing

Sponsored by INTERNET 2

VISUALIZATION TOOLS TO SUPPORT DRUG DISCOVERY, TRANSLATIONAL RESEARCH & CLINICAL DEVELOPMENT

10:50 Chairperson's Opening Remarks
Alexander Lex, Ph.D., Postdoctoral Fellow & Lecturer, Harvard School of Engineering & Applied Sciences

11:00 AIDEAS: An Integrated Cheminformatics Solution
Rishi Gupta, Senior Research Scientist, Platform Informatics and Knowledge Management, AbbVie, Inc.
AIDEAS is a novel concept that has brought together scientific tools and techniques under a unified platform that has enabled chemists and biologists to do their own data analysis and visualization. This presentation will be specifically directed towards a unique method called iSCORE that was developed as a probabilistic multi-parametric scoring methodology. iScore uses data based on AbbVie's proprietary *in vivo* and *in vitro* assay data as well as *in silico* ADMET models.

11:30 Bringing Proce ss, Chemical & Analytical Data Together: Data Mining & Visualization
Jean-Michel Adam, Ph.D., Senior Principal Scientist, Preclinical CMC Process Research, Roche Pharma Research & Early Development, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd.
Automated reactors, coupled with in-/off-line analytical tools, are routinely used in the chemical process R&D world. While these do help increase process knowledge and overall productivity, an increasing amount of data are is being generated, generally in a fragmented way. We would like to report a first approach aiming at integrating process data from automated reactors, analytical systems output as well as chemical information from Electronic Lab Notebook.

12:00 pm Collaborative Drug Design at Bristol-Myers Squibb

Sponsored by SCHRÖDINGER

Brian Claus, Senior Scientist, Bristol-Myers Squibb
Bristol-Myers Squibb has created an environment, based on the LiveDesign and Protein-Ligand Database (PLDB) products from Schrödinger that empowers its scientists to share ideas and modeling results, as well as to keep up-to-date on the latest data in their projects. The environment centralizes and connects computational tools with experimental data. Project team members can add idea compounds simultaneously or collaborate asynchronously. The presentation will discuss system design, customization, and lessons learned.

12:30 Session Break

12:40 Luncheon Presentation *(Sponsorship Opportunity Available)* **or Lunch on Your Own**

1:40 Session Break

1:50 Chairperson's Remarks
Hector Corrada Bravo, Ph.D., Assistant Professor, Center for Bioinformatics and Computational Biology, Department of Computer Science, University of Maryland, College Park

1:55 UpSet: Visualization of Intersecting Sets
Alexander Lex, Ph.D., Postdoctoral Fellow & Lecturer, Harvard School of Engineering & Applied Sciences
Understanding relationships between sets is an important analysis task in the life sciences. The major challenge for creating insightful visualizations is the combinatorial explosion of the number of set intersections if the number of sets exceeds a trivial threshold. To address this, we developed UpSet, a novel, interactive and web-based visualization technique for the quantitative analysis of sets, their intersections, and aggregates of intersections.

2:15 Presentation to be Announced
Heike Hofmann, Ph.D., Professor, Statistics, Iowa State University

Track 7

Data Visualization and Exploration Tools

Genomics, Drug Discovery and Clinical Development

2:35 Toward an Open Source Suite to Bridge the Gap between Plate-Based Screening and Results

Peter Henstock, Ph.D., Senior Principal Scientist, Research Business Technology Group, Pfizer, Inc.

Scientists in academic laboratories through large pharmaceutical companies have all encountered the challenges of efficiently extracting results from plate-based assay data. Issues from compound/reagent/plate management, assay format variability, instrumentation, output file formats, and analysis software invariably lead to a cumbersome process. To improve the efficiency, an open source suite of web-based tools is being developed that spans the key steps of plate editing, QC/QA calculation and visualization, and a user-driven non-coding approach to output file parsing. For results analysis, the suite includes visualization and computational approaches for interactively interpreting single-point, dose-response, and multivariate data.

Andrés Arslanian*, David Bonner*, Ivan Bugarinovic*, Mark Ford*, Alexander Galushka*, Cindy J. Liu*, Zachary Martin*, Frank O'Connor*, Alan A. Orcharton*, Gerson A. Rodrigues*, Sean M. Sinnott*, Timothy S. Stefanski*, Jaime A. Valencia*, Nikita E. Yaroshevsky*, Saaqib Zaman*, Robert Zupko*‡, Peter V. Henstock*†

*Harvard University, ‡Essen BioScience Inc., †Pfizer Inc.

2:55 Combining Machine & Human Intelligence to Successfully Integrate Biomedical Data

Sponsored by tamr

Timothy Danford, Ph.D., Field Engineer, Tamr

3:25 Refreshment Break in the Exhibit Hall with Poster Viewing

4:00 Making Visualization and Exploration Tools Truly Useful in the Regulatory Setting

Timothy Kropp, Ph.D., Associate Director for Innovation, Office of Computational Science, US FDA/CDE

As FDA applies tools and technologies to regulatory data ("big" data as well as "little") a lot is being learned about what is truly useful and in what contexts (not what is pretty or simply interesting). This talk will provide an overview of what informatics approaches FDA/CDER is using for visualization and exploration of scientific/clinical review data, how we are modifying what we use for better usefulness, what our biggest challenges and opportunities are, and where we want to go.

4:30 Feeding the Analytics Engine: Targeting Optimal Clinical Trial Sites, a Case Study

James Gill, Ph.D., Director, Research Analytics and Visualization, Bristol-Myers Squibb

It is no surprise that as soon as an analytical approach is proposed, access to data becomes a hurdle. In this talk we review a successful approach to improving our clinical trials site selection process by leveraging unique data in a dashboard format. Our keys to success included a clear understanding of the impact of different factors on site performance, how we can find surrogates for non-existing data and using an exploratory process with our scientists.

5:00 Delivering Standardized Clinical and Preclinical Data to Scientists in Guided Analysis

Baisong Huang, Principal Statistical Analyst, Novartis Institutes for BioMedical Research, Inc.

As visualization tools evolve and become widely accepted in investigating and monitoring drug safety and efficacy, rapid access to standardized, interpretable data views is becoming essential. We will present some examples how we standardized and aggregated data in both translational and clinical settings and provided guided analysis to visualize the data in real-time.

5:30 Best of Show Awards Reception in the Exhibit Hall with Poster Viewing

6:30 Close of Day

THURSDAY, APRIL 23

7:00 am Registration and Morning Coffee

» 8:00 PLENARY SESSION PANEL

Please see page 5 for details.

10:00 Coffee Break in the Exhibit Hall and Poster Competition Winners Announced

VISUALIZATION OF GENOMIC DATA

10:30 Chairperson's Remarks

William J.R. Longabaugh, MS, Senior Software Engineer, Institute for Systems Biology

» 10:40 KEYNOTE PRESENTATION: PEDIGREE VISUALIZATION IN GENOMICS

Jessie Kennedy, Dean of Research and Innovation, Edinburgh Napier University

Most visualizations that display pedigree structure for genetic research have been designed to deal with human family trees. However, due to the size and nature of plant and animal pedigree structures, human pedigree visualizations tools are unsuitable for use in studying animal and plant genotype data. We discuss two visualization tools, VIPER and Helium, and show how they support the work of biologists.

11:10 Visualization Tools for the Refinery Platform

Nils Gehlenborg, Ph.D., Research Associate, Center for Biomedical Informatics, Harvard Medical School

The Refinery Platform (http://www.refinery-platform.org) is a web-based data visualization and analysis system for epigenomic and genomic data designed to support reproducible biomedical research. The analysis backend employs the Galaxy Workbench and connects to a data repository based on the ISA-Tab data description format. In my talk I will discuss the exploratory visualization tools that we have integrated into Refinery.

11:40 Visualizing Genomic Variants and Annotations is Vital for Accurate Interpretation

Gabe Rudy, Vice President, Product & Engineering, Golden Helix, Inc.

In both the research and clinical context, the analytical steps to discover candidate variants of importance involves many transformations and cross-referencing of genomic datasets. Genomic visualization with tools like GenomeBrowse (http://genomebrowse.com) provide a genomic context critical for accurately interpreting function as well as detecting false-positive and false-negative calls and annotations. With visual case studies of variants, their alignments and genomic context, I will discuss the different representation of multi-nucleotide polymorphisms and other issues that impact public data annotations and functional classification of variants.

Organized by:
CHI Cambridge Healthtech Institute

Track 7

Data Visualization and Exploration Tools

Genomics, Drug Discovery and Clinical Development

12:10 pm Session Break

12:20 Luncheon Presentation *(Sponsorship Opportunity Available)* **or Lunch on Your Own**

1:20 Dessert Refreshment Break in the Exhibit Hall with Poster Viewing

1:55 Chairperson's Remarks

Nils Gehlenborg, Ph.D., Research Associate, Center for Biomedical Informatics, Harvard Medical School

2:00 Combing the Hairball: Network Visualization with BioTapestry and BioFabric

William J.R. Longabaugh, MS, Senior Software Engineer, Institute for Systems Biology

Networks models are crucial for understanding complex biological systems, yet traditional node-link diagrams of large networks provide very little visual intuition, and there is a need to develop scalable, unambiguous, and rational network visualization techniques. Our applications, BioTapestry (http://www.BioTapestry.org) and BioFabric (http://www.BioFabric.org), are designed to address this need, and I will discuss how they use novel approaches to avoid the "hairball" trap.

2:30 Visualization of Comparative Genomics Data: Results, Challenges, and Open Questions

Inna Dubchak, Ph.D., Senior Scientist, Lawrence Berkeley National Laboratory

As the rate of generating sequence data continues to increase, visualization tools for interactive exploration and interpretation of comparative data at the level of gene, genome, and ecosystem are of critical importance. We will talk about strengths and limitations of existing methods, and highlight new challenges in the visualization of huge volumes of complex comparative data.

3:00 Interactive and Exploratory Visualization of Epigenome-Wide Data

Hector Corrada Bravo, Ph.D., Assistant Professor, Center for Bioinformatics and Computational Biology, Department of Computer Science, University of Maryland, College Park

We will introduce epigenomics data visualization tools that provide tight-knit integration with computational and statistical modeling and data analysis: Epiviz (http://epiviz.cbcb.umd.edu), a web-based genome browser application, and the Epivizr Bioconductor package that provides interactive integration with R/Bioconductor sessions. This combination of technologies permits interactive visualization within a state-of-the-art functional genomics analysis platform. The web-based design of our tools facilitates the reproducible dissemination of interactive data analyses in a user-friendly platform.

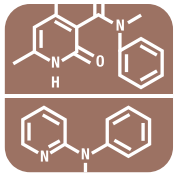
3:30 Visual-Analytic Systems for Integrative Genomic Analysis of Cancer Data

Raghu Machiraju, Ph.D., Professor, Ohio State University

Cancers are highly heterogeneous with different subtypes. Recently, integrative approaches were adopted that combined multiple types of omics data. In this talk, I present visual analytic solutions for the simultaneous and integrative exploration of multiple types genomics data including those from The Cancer Genome Atlas (TCGA) project. Using different combinations of mRNA and microRNA features we suggest potential combined markers for prediction of patient survival.

4:00 Conference Adjourns

Organized by:
Cambridge Healthtech Institute



Track 8

Pharmaceutical R&D Informatics

Collaboration, Data Science and Biologics

TUESDAY, APRIL 21

7:00 am Workshop Registration and Morning Coffee

8:00 – 11:30 Recommended Morning Pre-Conference Workshops*

Biologics, Bioassay, and Biospecimen Registration Systems

12:30 – 4:00 pm Recommended Afternoon Pre-Conference Workshops*

Finding Innovation in Collaboration Environments: Documentum, Sharepoint, Veeva, and Tigers, Oh My!

** Separate registration required*

2:00 – 6:30 Main Conference Registration

» 4:00 PLENARY SESSION
Please see page 5 for details.

5:00 – 7:00 Welcome Reception in the Exhibit Hall with Poster Viewing

Sponsored by IBM

WEDNESDAY, APRIL 22

7:00 am Registration Open and Morning Coffee

» 8:00 PLENARY SESSION
Please see page 5 for details.

9:00 Benjamin Franklin Awards and Laureate Presentation

9:30 Best Practices Awards Program

9:45 Coffee Break in the Exhibit Hall with Poster Viewing

Sponsored by INTERNET 2

DATA SCIENCE & ANALYTICS STRATEGIES

10:50 Chairperson's Opening Remarks

Jose L. Alvarez, Principal Engineer, WW Director, Healthcare and Life Sciences, Seagate Cloud and Systems Solutions

11:00 The Evolution of Data Science in Translational Medicine

Anastasia Christianson, Head, Translational R&D IT, Bristol-Myers Squibb

Eric Carleen, Director, Data Integration, Bristol-Myers Squibb

The role of the data scientist continues to evolve and the skills it requires continue to grow. This presentation will describe how the role has evolved in one Pharma company and how the collaboration between data scientists and related skills across organizational boundaries has delivered valuable insights to project teams.

11:30 Data Science in Translational Clinical Research

James Cai, Head, Data Science, Roche, Translational Clinical Research Center (TCRC)

In this talk I will outline a Data Science model that emphasizes mixed-capability teams and impact on science and business decisions. I will discuss how quantitative analytical skills, agile programming, novel technologies and business acumen all contribute to this model. I will illustrate with examples where Data Science was applied to clinical research resulting in new scientific insights and better business decisions.

12:00 pm Text Mining from Bench to Bedside - Where's the Value?

Sponsored by Linguamatics

Jane Reed, Ph.D., Head, Life Science Strategy, Linguamatics

Accessing the right information is critical to bench-to-bedside translational research. Much of the data is locked in textual format, such as scientific literature, clinical trial reports or electronic health records. This talk will demonstrate how advanced text analytics can provide a powerful solution to the challenges faced by researchers and clinicians, who need to extract the key facts rapidly and accurately to gain actionable insights for decision support.

12:15 Simplifying Analytical Knowledge Transfer in an Externalized World

Sponsored by ACD/Labs

Ryan Sasaki, Director, Global Strategy, ACD/Labs

The lion's share of chemical R&D today is being outsourced to external organizations. Subsequently, the potential of losing the 'proof of identity' for a sample, in the transfer of materials between a contractor and client, grows. As externalization and research virtualization continues to evolve, the task of mining these legacy analytical chemistry datasets and methods to help monitor and identify raw materials, impurities, and metabolites will be ever more difficult based on deficiencies in the knowledge exchange mechanisms. Fortunately, solutions are emerging. This session will present a use case for a new laboratory informatics external collaboration model.

12:30 Session Break

12:40 Luncheon Presentation I: Utilizing Big Data and Linked Data to Explore Relationships between Biological Entities for Drug Repurposing, Translational Medicine and Target Finding

Sponsored by THOMSON REUTERS

Tomasz Adamusiak, Ph.D., M.D., Senior Data Scientist, Technology Development, Thomson Reuters

With the advances in NGS technology and data generation and as traditional translational research is deemed inefficient and costly, pharmaceutical and biomedical industries are driven to seek new ways to better utilize their data to extract relevant biological information. Thomson Reuters Cortellis™ Data Fusion delivers a first-in-class Big Data solution to drive new scientific and strategic insights from all of the proprietary and public content.

1:10 Luncheon Presentation II: Where Science Intersects with Business – Creating Business Dashboards That Combine Data from Multiple Sources

Sponsored by ELSEVIER

Huijun Wang, Ph.D., Associate Principle Scientist, Cheminformatics, Merck & Co., Inc.

Eric Gifford, Ph.D., Principal Scientist, Systems Chemical Biology, Merck & Co., Inc.

Matthew Clark, Ph.D., Consultant, Life Science Services, Elsevier

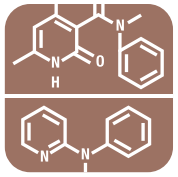
In today's highly competitive pharmaceutical environment it is imperative for project teams to monitor both business movements, and scientific developments that can affect the business proposition for the program. Elsevier is collaborating with Merck to develop a series of dashboards that can bring in information from multiple sources to create views with facets for drug, target, and disease related information. These dashboards will monitor scientific information gleaned from journals, patents & grant applications to provide a rich context for monitoring project status and competitive position.

1:40 Session Break

1:50 Chairperson's Remarks

Daniel H. Robertson, Ph.D., Senior Director, Research IT, Eli Lilly and Company

Organized by:
Cambridge Healthtech Institute



Track 8

Pharmaceutical R&D Informatics

Collaboration, Data Science and Biologics

1:55 Transforming IT and Informatics at Biogen to Drive Research

Hank Wu, Director, R&D IT, Biogen

Transforming IT and Informatics at Biogen is at the heart of the company's strategic commitment to use technology, data and analytics to inform the drug discovery process, unlock new insights, improve patient care and drive innovation. This presentation shares work in progress and lessons learned at Biogen.

2:25 PANEL DISCUSSION: Growing a Data Science Team

- Enabling Innovative Data-driven Approaches at the Intersection of Science, Medicine & Economics
- Assembly, Creation and Implementation of Data Science Groups for Pharma
- The Data Scientist - an Essential Component of Big Data Analytics – Difficult to Identify
- What are Data Sciences, Informatics and Bioinformatics?
- Should data scientists be centralized or embedded within other product/functional teams?
- How strong of a coder/programmer should members of a data science team be?
- How much domain knowledge does a data scientist need to have?

Moderator: Martin Leach, Ph.D., Vice President, Global Data Office, Biogen

Panelists:

Rainer Fuchs, CIO, Harvard Medical

Jason Johnson, Ph.D., Executive Vice President and Head of R&D, PatientsLikeMe

Jake Klamka, Founder, Insight Data Science Fellows Program

Daniel H. Robertson, Ph.D., Senior Director, Research IT, Eli Lilly and Company

Tom Plasterer, Ph.D., Director, US Cross-Science Lead, AstraZeneca

Sarah Aerni, Ph.D., Principal Data Scientist, Pivotal

Sponsored by idbs

2:55 Can Simplifying the Informatics Landscape Underpin Your Lab or the Future?

Paul Denny-Gouldson, Ph.D., Vice President, Strategic Solutions, IDBS

A core concept of the lab of the future is simplifying day to day tasks and providing easy access to information concerning materials, results and reports. To realize these aspirations, it is essential to modernize existing R&D data workflows but importantly not to just automate the current state. With an upgrade of infrastructure comes a great opportunity to reassess what is done, how it is done and how this can all be optimized. Removing paper and capturing IP can be drivers for a change – but don't miss the opportunity to get more out of the change. We will use case studies of the good and the bad to show what can be done and how it can be done.

Sponsored by BIOVIA

3:10 BIOVIA ScienceCloud: Automating Collaboration Workflows

Ton van Daelen, Ph.D., ScienceCloud Product Director, BIOVIA

The amount of R&D spending beyond company boundaries is approaching 50% of the overall R&D budget, yet informatics infrastructures are challenged to support this changing environment. We will present a comprehensive, cloud-based solution stack for externalized, collaborative research for pharma/biotech and CROs that addresses these challenges and we will discuss how developing customized business rules and synchronizing cloud with on-prem data are critical success factors.

3:25 Refreshment Break in the Exhibit Hall with Poster Viewing

MODELING & ANALYTICS

4:00 The Construction of a Scientific Modeling Culture and Technology Platform at Merck

Chris L. Waller, Ph.D., Director and Head, Scientific Modeling Platforms, Merck Research Laboratories

Merck Research Laboratories is undergoing a transformation in the way that it prosecutes R&D programs. Through the adoption of a "model-driven" culture, enhanced R&D productivity is anticipated. To support this emerging culture, an ambitious IT program has been initiated to implement a harmonized platform to facilitate cross-domain workflows and decision-making through agile persona driven data and predictive model access.

4:30 Separating the Wheat from the Chaff: Using Proprietary and Public Genomic Information to Identify Biomarkers from Cancer Cell Line Profiling Studies

Yue Webster, Ph.D., Senior Research Scientist, LRL IT Informatics, Eli Lilly and Company

Like most companies, Lilly uses large panels of cancer cell lines to discover genes, transcripts, proteins and/or metabolites which influence response to treatment. The potential for generating false positive findings is significant, and low concordance was highlighted by recent publication (Nature 504, 389–393). The use of co-expression networks and integration across various resources helps identify higher quality relationships. Advanced visualization tools help biologists navigate through thousands of putative relationships.

Sponsored by LabAnswer LABORATORY INFORMATICS

5:00 Helping Our Clients Succeed in Their Distributed R&D Environments by Delivering Excellence in Scientific and Laboratory Informatics

John F. Conway, Global Director, R&D Strategy and Solutions, LabAnswer

Many organizations have chosen to distribute or externalize large portions of their R&D. Consequently, these same organizations are struggling to collaborate with their external partners. Sharing and capturing of data and information in these environments is requiring extra (inefficient) effort. Through discussion and case studies attendees will get to see firsthand how LabAnswer is helping our clients develop strategies, technologies and best practices that help solve some of the headaches associated with the distributed R&D business model.

Sponsored by CAMBRIDGE SEMANTICS

5:15 Co-Presentation: A Data Lake for Competitive and Clinical Trial Intelligence

Christine Blazynski, Ph.D., Chief Science Officer & Senior Vice President, New Product Development, Informa

Ben Szekely, Vice President, Solutions, Cambridge Semantics

Semantic Data Lakes combine rich, conceptual models with cloud storage and computing technologies to link multi-structured content. This paradigm enables user-friendly and intuitive search, analytics and visualization across wide and diverse data sets. In this talk, Cambridge Semantics and Informa will present the Semantic Data Lake they have created across Informa's rich content sources including Citeline and Sagient. We will walk through some interesting use cases that illustrate the value of developing a Semantic Data Lake.

5:30 Best of Show Awards Reception in the Exhibit Hall with Poster Viewing

6:30 Close of Day

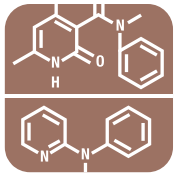
THURSDAY, APRIL 23

7:00 am Registration Open and Morning Coffee

» 8:00 PLENARY SESSION PANEL
Please see page 5 for details.

10:00 Coffee Break in the Exhibit Hall and Poster Competition Winners Announced

Organized by:
Cambridge Healthtech Institute



Track 8

Pharmaceutical R&D Informatics

Collaboration, Data Science and Biologics

MODELING & ANALYTICS

10:30 Chairperson's Remarks

Yuriy Gankin, Ph.D., Chief Life Science Officer, EPAM Systems

10:40 Translational R&D Analytics: Delivering 'Big Insights' to Drive Translational Research

Kaushal Desai, Associate Director, Translational R&D Analytics and Decision-Support, Research Informatics & Automation, Bristol-Myers Squibb

This session will explore case studies demonstrating how translational R&D analytics can inform patient stratification and trial design in early clinical and translational research. The talk will focus on the journey from a lack of discoverability for disjointed datasets to insights that drive key decisions in translational research. Challenges associated with delivering actionable information at the point of decision-making will be highlighted and opportunities to deliver business value will be outlined using real examples.

11:10 Integrated Genomics Platform: Putting Patients and Their Genomes into the Focus of Our Research

Nora Manstein, Ph.D., IT Project Manager, Bayer Business Services GmbH

We have established the Integrated Genomics Platform (IGP) as a central tool for genomics research in Cardiology, Oncology and Clinical Sciences. The platform supports advanced data analysis and is intended to simplify discovery processes. In this strategic project, we have overcome known bottlenecks and enabled true translational research by establishing a company-wide mandatory repository and toolbox for storage and analysis of genomics data as well as common standards for data annotation, privacy & security.

11:40 Building a Globally Distributed, Hybrid NGS Sequence Analysis and Integration Infrastructure for Oncology Discovery and Translational R&D

Sponsored by bina TECHNOLOGIES

Justin H. Johnson, Principal Scientist, AstraZeneca

Next-Generation Sequencing is changing the way pharmaceutical companies develop drugs, perform patient stratification, and evaluate treatment efficacy. However, managing the massive amounts of NGS data has introduced fundamental IT challenges. Here we discuss the implementation of a fast, flexible, scalable and validated IT infrastructure that can streamline the upkeep of the NGS analysis workflow and the distribution of genomic information throughout an organization for translational discovery.

12:10 pm Session Break

12:20 Luncheon Presentation *(Sponsorship Opportunity Available)* **or Lunch on Your Own**

1:20 Dessert Refreshment Break in the Exhibit Hall with Poster Viewing

INTEGRATION & INNOVATE CAPTURE OF INTERNAL & EXTERNAL DATA

1:55 Chairperson's Remarks

Dermot McCaul, Director, PreClinical Development and Biologics IT, Merck

2:00 Best Practices to Thrive Under the BioPharma Big Data Deluge

Tom Plasterer, Ph.D., Director, US Cross-Science Lead, AstraZeneca

Knowing how to ingest, harmonize and query information can be a tremendous advantage both internally and to the ecosystem of partners and providers we depend upon. Examples using linked data approaches illustrate how the information stream can be tamed, focusing first on getting data out of containers. Once this is established, terminologies can be applied to derive meaningful answers across otherwise-siloed content. The Open PHACTS project and Bio2RDF projects show how this approach has been used to solve real big data questions for BioPharma.

2:30 Beyond Data Integration – Consumable Expert Knowledge in Chemical Biology

Jeremy L. Jenkins, Ph.D., Senior Investigator II, High Throughput Biology, Developmental & Molecular Pathways, Novartis Institutes for BioMedical Research

An emerging challenge is how to create inferences from knowledge bases that enable automation of expert opinions at large scale. We present a system that creates summary-level assertions based on diverse chemical biology data sources to address the problem of ranking tool compounds for targets, and vice versa, quantifying target confidence for compounds. Overall this approach provides a data-driven opinions about compounds that reflect those of an informed chemical biologist.

3:00 Development and Implementation of a Nonclinical Data Warehouse

Gregory Woo, Principal IS Business Systems Analyst, Research & Development Informatics, Amgen, Inc.

Amgen has implemented an integrated data warehouse for nonclinical toxicology studies, including data from internal systems and at Contract Research Organizations. The goal of the system is to allow scientists to rapidly search, query, and visualize historical toxicology, pathology, and toxicogenomics data. This presentation will discuss the system's design, key features, challenges and lessons learned.

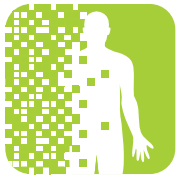
3:30 The Data Integration Challenge

Mark Davies, Technical Lead, Computational Chemical Biology, European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus

The UniChem resource allows users to quickly and dynamical integrate the chemical content from a growing number sources, which currently stands at 25 and contains more than 70 million compound structures. We use the example of the new and open SureChEMBL patent system to demonstrate how UniChem can assist with data integration. We also identity the new challenges we face and how we can embrace other technologies and methodologies, such as Linked Data, to help stay on top of the data integration challenge.

4:00 Conference Adjourns

Organized by:
Cambridge Healthtech Institute



Track 9

Clinical Genomics

Tools for Investigation, Integration and Implementation

TUESDAY, APRIL 21

7:00 am Workshop Registration and Morning Coffee

8:00 – 11:30 Recommended Morning Pre-Conference Workshops*
Genome Assembly and Annotation

12:30 – 4:00 pm Recommended Afternoon Pre-Conference Workshops*
Determining Genome Variation and Clinical Utility
** Separate registration required*

2:00 – 6:30 Main Conference Registration

» 4:00 PLENARY SESSION
Please see page 5 for details.

5:00 – 7:00 Welcome Reception in the Exhibit Hall with Poster Viewing

Sponsored by IBM

WEDNESDAY, APRIL 22

7:00 am Registration Open and Morning Coffee

» 8:00 PLENARY SESSION
Please see page 5 for details.

9:00 Benjamin Franklin Awards and Laureate Presentation

9:30 Best Practices Awards Program

9:45 Coffee Break in the Exhibit Hall with Poster Viewing

Sponsored by INTERNET 2

GENOMICS: CLINICAL CHALLENGES AND MEDICAL OPPORTUNITIES

10:50 Chairperson's Opening Remarks

Scott Kahn, Ph.D., Vice President, Commercial Enterprise Informatics, Illumina, Inc.

» 11:00 FEATURED PRESENTATION: THE PENETRANCE OF INCIDENTAL FINDINGS IN GENOMIC MEDICINE

Robert C. Green, M.D., MPH, Director, G2P Research Program; Associate Director, Research, Partners Personalized Medicine, Division of Genetics, Department of Medicine, Brigham & Women's Hospital and Harvard Medical School

Much of the controversy surrounding the implementation of incidental findings in clinical sequencing is due to uncertainty about the penetrance of such findings in persons unselected for clinical features or family history. This uncertainty also influences the question of genomic population screening, i.e., whether actionable sequence variants should be sought and reported in ostensibly healthy individuals. In this talk, new data will be presented estimating the penetrance of actionable incidental findings.

» 11:30 FEATURED PRESENTATION: CHALLENGES AND OPPORTUNITIES IN ESTABLISHING IT SUPPORT FOR CONTINUOUS LEARNING IN HEALTHCARE: THE POTENTIAL FOR APPLYING LESSONS LEARNED FROM CLINICAL GENOMIC IT SUPPORT TO BROADER CONTINUOUS LEARNING CHALLENGES

Samuel (Sandy) Aronson, Executive Director, IT, Partners HealthCare Center forPersonalized Genetic Medicine

Continuously updated knowledge bases will be required to enable a true continuous learning healthcare environment. However, modern healthcare pressures make their maintenance difficult. The clinical genomic IT community has been wrestling with this issue for some time. We present lessons learned from supporting clinical genomic IT processes that may be generalizable to broader development of IT support for continuous learning healthcare processes

12:00 pm Census of the Apoptosis Pathway

Sponsored by ELSEVIER

Philip L. Lorenzi, Ph.D., Department of Bioinformatics and Computational Biology & the Proteomics and Metabolomics Core Facility, MD Anderson Cancer Center

We recently compared several different "omic" approaches to constructing the autophagy pathway de novo, including siRNA screening, mass spectrometry-based proteomics, and three different pathway analysis software packages. Unexpectedly, although merging all of the validated data sets yielded 739 autophagy-modulating genes, each individual approach alone yielded sparse coverage of the autophagy pathway. The best individual siRNA screen, for example, yielded only 169 of the 739 (23%) genes. Nevertheless, text mining-based pathway analysis with Pathway Studio in conjunction with manual curation provided the most comprehensive coverage, yielding 417 targets (56% of the pathway). Here, we explored the generalizability of those findings by examining a more well-characterized pathway—apoptosis. We compiled apoptosis-modulating genes from 12 published siRNA screens and two pathway analysis software packages—Ingenuity Pathway Analysis (IPA) and Pathway Studio. The resulting inventory of 6,882 proteins consisted of 215 targets identified by siRNA screening, 3,378 targets by IPA, and 6,381 targets by Pathway Studio. The extensive coverage (93%) of the apoptosis pathway provided by text mining with Pathway Studio can likely be attributed to recent upgrades in the software, including an expanded database and collection of full-text articles. Together with our previous autophagy pathway analysis, the new apoptosis results support the generalizable conclusions that: 1) siRNA screening has a large false negative rate (i.e., fails to identify many true "hits"), and 2) text mining-based pathway analysis using Pathway Studio provides the most comprehensive pathway coverage.

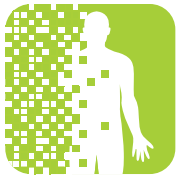
12:30 Session Break

12:40 Luncheon Presentation I: Computational Enablement of the Hippocratic Oath in a Clinical Oncology Setting

Sponsored by MOLECULAR HEALTH

David B. Jackson, Ph.D., Chief Innovation Officer, Molecular Health, Gmbh

The clinical response of cancer patients to oncolytic agents is influenced by three major classes of molecular determinant; tumor intrinsic factors (e.g. tumor biomarkers); patient intrinsic factors (e.g. polymorphisms) and patient extrinsic factors (e.g. co-medications). In my talk, I will present a novel computational technology and associated treatment decision support process that was designed to provide this knowledge-driven approach to clinical care in oncology.



Track 9

Clinical Genomics

Tools for Investigation, Integration and Implementation

1:10 Luncheon Presentation II: A High Performance Application Development Platform for Collaborative Genomics Research

Sponsored by illumina

Paul Flook, Ph.D., Senior Director, Enterprise Informatics, Illumina Inc.

Collaborative research among groups working with genomic data presents major logistical challenges. Transferring huge volumes of data can be prohibitively expensive for projects utilizing WGS data sets. Illumina has addressed this challenge by building a platform that enables collaborators to not only share data in a secure multitenant environment, but to develop and deploy their own applications close to the data.

1:40 Session Break

GENOMIC DATA SECURITY AND PRIVACY

1:50 Chairperson's Remarks

Nora Manstein, Ph.D., IT Project Manager, Bayer Business Services GmbH

1:55 Security vs. Freedom – It's Not a Matter of Philosophy

Nora Manstein, Ph.D., IT Project Manager, Bayer Business Services GmbH

We contribute to the debate on how patient's rights and wishes are respected and meaningful research with patient data can be done. In order to support this, we have developed an organizational process and a technical tool by which patients' informed consents are an integral part of the authorization process, allowing compliant access to and scientific analysis of patient data.

2:25 Privacy, Access Control and Security in Clinical Genomics Environments

Toby Bloom, Ph.D., Deputy Scientific Director, Informatics, New York Genome Center

The integration of clinical and genomic data introduces new, complex problems in privacy and security. These include protecting the anonymity of clinical data when it is linked to "self-identifying" genomic data; managing the fine-granularity access control required to share data from multiple projects; and overcoming the regulatory and legal hurdles associated with clinical genomic data. We discuss these and other access issues.

2:55 Blocking the Cyber Barbarians

Betsy Fallen, Global Head, Program and Business Development, SAFE-BioPharma Association

Identity trust is necessary for secure and regulated Internet communications. The presentation explains the issues associated with establishing online trust and the role of the industry-driven SAFE-BioPharma global identity management/digital signature standard in assuring that only authorized identities have access to protected information. Participants will learn about types and levels of identity credentials, government and industry organizations involved in establishing identity trust infrastructures, applicable standards, governance models and approaches to cloud-based identity management.

3:25 Refreshment Break in the Exhibit Hall with Poster Viewing

4:00 Data Integration, Privacy and Openness at PatientsLikeMe, a Social Network for Patients with Life-Altering Conditions

Marcia M. Nizzari, MS, Vice President, Engineering, PatientsLikeMe, Inc.

PatientsLikeMe provides a social network and research platform for capturing, curating and analyzing patient-reported data. With 300,000+ users, 2,300+ conditions represented and over 25 million health datapoints collected, it provides a new, rich source of data to integrate with EHR and genomic data to drive new insights about disease. We discuss trade-offs in privacy and openness when combining EHR and other sources of clinical and research data – such as -omics – with patient-reported data.

4:30 Differential Privacy: Future-Proof Protection for Sensitive Data

Ishaan Nerurkar, CEO & Founder, Shroudbase, Inc.

In the analysis of sensitive data, current methods of privacy protection severely compromise utility, access and opportunities for collaboration. Shroudbase is a cloud software that creates and manages permanently de-identified copies of high-dimensional data with strong, mathematically proven guarantees of statistical accuracy. Our patent-pending technology enables previously untouchable information to be open-sourced and analyzed while maintaining differential privacy, the gold standard of data privacy. This presentation discusses an instance of the Shroudbase platform optimized to handle the unique privacy challenges posed by genomics data.

5:00 PANEL DISCUSSION: Genomic Research: Utility vs. Patient Rights

Moderator:

Toby Bloom, Ph.D., Deputy Scientific Director, Informatics, New York Genome Center

Panelists:

Betsy Fallen, Global Head, Program and Business Development, SAFE-BioPharma Association

Nora Manstein, Ph.D., IT Project Manager, Bayer Business Services GmbH

Ishaan Nerurkar, CEO & Founder, Shroudbase, Inc.

Marcia M. Nizzari, MS, Vice President, Engineering, PatientsLikeMe, Inc.

Juhapekka Piiroinen, Head, Personalized Medicine Development, MediSapiens, Ltd.

What software tools, organizational processes and regulatory minefields must researchers and clinicians understand to not only improve drug development and personalized therapies, but also preserve the privacy of patient data? Experts share their perspectives on informed consent, security access, technical infrastructures, genomic and clinical data integration and more.

5:30 Best of Show Awards Reception in the Exhibit Hall with Poster Viewing

6:30 Close of Day

THURSDAY, APRIL 23

7:00 am Registration Open and Morning Coffee

» 8:00 PLENARY SESSION PANEL

Please see page 5 for details.

10:00 Coffee Break in the Exhibit Hall and Poster Competition Winners Announced

DATABASES, SHARING AND INTEGRATION

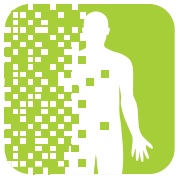
10:30 Chairperson's Remarks

Heidi L. Rehm, Ph.D., FACMG, Chief Laboratory Director, Laboratory for Molecular Medicine, Partners HealthCare; Associate Professor, Pathology, Brigham & Women's Hospital and Harvard Medical School

10:30 ClinGen: The Clinical Genome Resource

Heidi L. Rehm, Ph.D., FACMG, Chief Laboratory Director, Laboratory for Molecular Medicine, Partners HealthCare; Associate Professor, Pathology, Brigham & Women's Hospital and Harvard Medical School

ClinGen is an NIH-funded program developing resources to understand genomic variation and optimize its use in medicine. ClinGen interfaces research and clinical testing and includes the development of standards for variant and gene interpretation as well as broad data sharing. The ClinVar



Track 9

Clinical Genomics

Tools for Investigation, Integration and Implementation

database serves as the primary site for deposition and retrieval of variant interpretations as well as aggregation for expert curation.

10:55 Human Genome Analysis

Mark Gerstein, Ph.D., Albert L. Williams Professor, Biomedical Informatics, Yale University

Identification of noncoding cancer "drivers" from thousands of somatic alterations is an unsolved problem. Here, we developed a computational framework to annotate cancer regulatory mutations. The framework combines an adjustable data context summarizing large-scale genomics and cancer-relevant datasets with an efficient variant prioritization pipeline. To prioritize high-impact variants, we developed a weighted scoring scheme to score each mutation's impact.

11:20 Clinical Genomicist Workstation: Analyze, Interpret and Report Next-Gen-Based Molecular Diagnostic Studies

Rakesh Nagarajan, M.D., Ph.D., Associate Professor, Pathology & Immunology and Genetics, Washington University in St. Louis; Chief Biomedical Informatics Officer, PierianDx

Clinical NGS has been gaining traction over the past few years. The Clinical Genomicist Workstation was developed to address the informatics barriers that limit the more broad and rapid adoption of this technology broadly in the medical community. This presentation discusses adoption of the CGW by molecular diagnostic laboratories and approaches to data and information sharing.

11:40 Coordinated Care in the Age of Personalized Medicine

Sponsored by ORACLE HEALTH SCIENCES

Ketan Patel, Ph.D., Healthcare Solutions Consultant, Oracle Health Sciences

Advances in genome diagnostics are starting to make an impact on patient care. A key challenge is how to enable a multidisciplinary care team to collaborate using genomic data from an individual patient. We describe an architecture which enables researchers, clinicians, molecular pathologists and genetic counselors to interact with the data using role-based interfaces which are tuned to their task in the clinical workflow. Such a system can speed up adoption of genomic data into routine clinical care scenarios.

12:10 pm Session Break

12:20 Luncheon Presentation *(Sponsorship Opportunity Available)* or Lunch on Your Own

1:20 Dessert Refreshment Break in the Exhibit Hall with Poster Viewing

CLINICAL UTILITY OF GENOME VARIATION

2:25 Chairperson's Remarks

Louis Fiore, M.D., MPH, Executive Director, MAVERIC, Research, Veterans Affairs Boston Healthcare System

2:30 Epigenetic Profiling of DNA Methylation to Identify Breast Tumor Aggressiveness

Adam Marsh, Ph.D., Professor, Center for Bioinformatics and Computational Biology, University of Delaware; CSO, Genome Profiling, LLC

Women with triple-negative genotypes (i.e., normal for the three common marker mutations for breast cancer) are still at risk for developing aggressive breast tumors. We identify a suite of differentially methylated CpG sites between normal and tumor breast tissues using NGS that indicate a high degree of epigenetic conservation among different triple-negative patients who have developed advanced-stage breast tumors. Subtle epigenetic shifts in methylation status may provide a key line of evidence for assessing tumor risk and informing therapy decisions between surgery or versus noninvasive treatments.

3:00 Establishing Clinical-Grade RNA Sequencing

Sheng Li, Ph.D., Instructor, Bioinformatics, Neurological Surgery, Weill Cornell Medical College

High-throughput sequencing drastically expands the potential for large-scale whole transcriptome profiling of clinical samples for disease monitoring and diagnosis. Here we established standard approach and analysis methods and benchmark datasets for evaluation of RNA-seq performance of different platforms, protocols and various qualities of input materials.

3:30 The Department of Veterans Affairs Precision Oncology Program: The Crossroads of Clinical Care and Research

Louis Fiore, M.D., MPH, Executive Director, MAVERIC, Research, Veterans Affairs Boston Healthcare System

This presentation describes a model for creation of "Learning Healthcare Systems" through integration of a clinical precision oncology program with a tailored research program that leverages and augments the clinical investment. Databases and applications that support clinical trial matching, capture of patient reported outcomes, clinician collaboration and patient outcome prediction will be discussed.

4:00 Conference Adjourns

Organized by:
Cambridge Healthtech Institute

Track 10

Collaborations and Open Access Innovations

Using Collaborative Technologies and Methodologies to Accelerate Basic, Translational and Clinical Research

TUESDAY, APRIL 21

7:00 am Workshop Registration and Morning Coffee

8:00 – 11:30 Recommended Morning Pre-Conference Workshops*
Gamification of Science

12:30 – 4:00 pm Recommended Afternoon Pre-Conference Workshops*
How Data-Driven Patient Networks are Transforming Biomedical Research

IT & Informatics in Support of Collaboration and Externalization
** Separate registration required*

2:00 – 6:30 Main Conference Registration

» 4:00 PLENARY SESSION
Please see page 5 for details.

5:00 – 7:00 Welcome Reception in the Exhibit Hall with Poster Viewing

Sponsored by IBM

WEDNESDAY, APRIL 22

7:00 am Registration Open and Morning Coffee

» 8:00 PLENARY SESSION
Please see page 5 for details.

9:00 Benjamin Franklin Awards and Laureate Presentation

9:30 Best Practices Awards Program

9:45 Coffee Break in the Exhibit Hall with Poster Viewing

Sponsored by INTERNET 2

OPEN SOURCE SOFTWARE & STANDARDS

10:50 Chairperson's Opening Remarks
Janak Joshi, Product Strategist, ConvergeHEALTH by Deloitte, Deloitte Consulting LLP

11:00 Imitation and Disruption: Impact on Open Source Software Success in the Life Sciences
Jay Bergeron, Director, Translational and Bioinformatics, Pfizer, Inc.
There are many successful examples of open source software (OSS) both within and outside of the life sciences community. However, investigators sponsoring software-based efforts need to consider many factors, including resourcing, architecture fragmentation, maintenance and their customer community when selecting between commercial and open license alternatives. This presentation will review such factors.

11:30 OpenBEL: Collaborative Knowledge Base and Tools for Biomedical Research
Natalie Catlett, Ph.D., Senior Computational Scientist, Engineering, Selventa
Anthony Bargnesi, Application Architect, Engineering, Selventa
OpenBEL is an open source platform for managing biological knowledge, comprised of the Biological Expression Language (BEL) and a knowledgebase platform. We will describe a next-generation Semantic Web RDF platform for harmonization, storage, and access of BEL knowledge; language expansion; and development of an exchange format for biological models derived from BEL knowledge networks.

12: 00 An Integrated Informatics Solution to Optimize Collaborative Research

Sponsored by dotmatics

Robert Brown, Ph.D., Vice President, Global Informatics, Dotmatics
Conducting research projects across multiple organizations presents a number of challenges which must be overcome for them to be successful. Using case studies from pharma, biotech and CROs, this talk will discuss how dedicated hosted informatics systems designed to support collaborative research can help enhance the success of such projects.

12:30 Session Break

12:40 Luncheon Presentation I: Accelerators to Collaboration between Pharma and Providers

Sponsored by ConvergeHEALTH by Deloitte.

Dan Housman, CTO, ConvergeHEALTH by Deloitte, Deloitte Consulting LLP
Janak Joshi, Product Strategist, ConvergeHEALTH by Deloitte, Deloitte Consulting LLP

1:10 Luncheon Presentation II *(Sponsorship Opportunity Available)* **or Lunch on Your Own**

1:40 Session Break

COLLABORATIVE APPROACHES IN CLINICAL RESEARCH, BIOMEDICAL RESEARCH AND THE PHARMACEUTICAL SPACE

1:50 Chairperson's Remarks

1:55 Open Source National Network Facilitating Healthcare and Resource Data Sharing
Doug Macfadden, Chief Informatics Officer, Harvard Catalyst
Bhanu Bahl Director of Informatics, Harvard Catalyst
Accrual to Clinical trials (ACT) project supported by NCATS was launched with the goal of creating a network of 60 Clinical Translational Science Center Award

Organized by:
Cambridge Healthtech Institute

Track 10

Collaborations and Open Access Innovations

Using Collaborative Technologies and Methodologies to Accelerate Basic, Translational and Clinical Research

(CTSA) sites. The network will facilitate investigators to query EHR data across all these sites for cohort exploration and subsequently engage and enroll identified patients into clinical trials. SHRINE (Shared Health Research Information Network) is a system developed by Harvard Catalyst for enabling clinical researchers to query across distributed hospital electronic medical record systems.

2:25 NeuroBANK™, Accelerated Research Environment as a Model for Collaboration and Cooperation in Clinical Research

Alex Sherman, Director, Strategic Development and Systems, Neurological Clinical Research Institute, Massachusetts General Hospital

NeuroBANK™, a patient-centric platform that allows clinicians and investigators to aggregate and cross-link clinical and research information from clinical visits, clinical studies, health records, and self-reported patient outcomes, and to connect it to biospecimen, images and genetic files. Will discuss how to find or create incentives for collaborations.

2:55 Data Access Models for Genetic Data Sharing – GSK SHARE and the GA4GH Beacon

Karen King, Head, Genetic Data Sciences, GlaxoSmithKline

3:25 Refreshment Break in the Exhibit Hall with Poster Viewing

COLLABORATING IN CHRONIC AND RARE DISEASES

4:00 CureAccelerator™: How a New Global Platform Will Help Propel Cures for the World's Unsolved Diseases

Bruce Bloom, D.D.S., J.D., President and Chief Science Officer, Cures Within Reach

More than 7,000 diseases have no fully effective treatment, affecting more than 500 million people worldwide. CureAccelerator™ is the world's first open-access, online platform dedicated to repurposing research – the quest to create new medical treatments from existing therapies, to drive more cures more quickly to more patients. Learn how this innovative IT tool will enable researchers, funders, the biomedical industry and patient groups to collaborate far more efficiently, to propel the pace of repurposing research.

4:30 Delivering on a Promise: Achieving a Patient-Centric Open Information Ecosystem

Walter Capone, President and CEO, The Multiple Myeloma Research Foundation

Patient-centric, open access research models are widely acknowledged catalysts of scientific and medical progress. Here we present a first-in-kind model for combining an observational clinical study with a participatory, community driven research program for Mutliple Myeloma, and in doing so, providing a powerful example that can lead the way forward in other disease areas.

5:00 There and Back Again: AstraZeneca's Collaboration Journey

Robert Albert, DocuSign Project Manager and Collaboration Specialist, AstraZeneca

AstraZeneca has always been on the cutting edge of innovation. Despite having over 50,000 employees worldwide, we know we don't have all of the answers. In the past 3 years, we have been transforming into a more collaborative company. The journey began with a culture event where we discussed how we could solve difficult challenges through the power of crowdsourcing. In 2012 we partnered with Innocentive to deliver iSolve/ @ work – a bulletin board where R&D challenges could be solved by colleagues from around the world. 2013 saw the AstraZeneca world-wide culture jam, which brought together 34,000 people from 100 countries together over a 3 day span. In 2014 we launched the R&D Open Innovation website. On the site we have 5 modules. They include: the clinical compound bank; the pharmacology toolbox; target innovation/ compound library; new molecule profiling, and the R&D module. My talk will focus specifically on the R&D challenges, and how we are crowdsourcing answers from Innocentive's solver network. Our aim is to deliver life-saving medicines as quickly as we can. One powerful option is through open collaboration with the greater scientific community.

5:30 Best of Show Awards Reception in the Exhibit Hall with Poster Viewing

6:30 Close of Day

THURSDAY, APRIL 23

7:00 am Registration and Morning Coffee

» 8:00 PLENARY SESSION PANEL
Please see page 5 for details.

10:00 Coffee Break in the Exhibit Hall and Poster Competition Winners Announced

CONSORTIA EFFORTS: UPDATES, OPPORTUNITIES AND DISCUSSION

10:30 Chairperson's Remarks

Robert Albert, DocuSign Project Manager and Collaboration Specialist, AstraZeneca

10:40 PANEL DISCUSSION: Multiple Sclerosis Conquered by the Data Science Revolution: Patients Win

Ken Buetow, Ph.D., Director, Computational Sciences and Informatics, Arizona State University

Robert McBurney, CEO, Office of the Chief Executive, Accelerated Cure Project

Marcia Kean, Chairman, Strategic Initiatives, Feinstein Kean Healthcare

Joseph LaFerrera, J.D., Partner, Gesmer Updegrove LLP

Funded by a grant from PCORI (Patient-Centered Outcomes Research Institute), the Accelerated Cure Project for MS is collaborating with all the key organizations in the MS community, gathering patient-reported and EHRs from 20,000 patients. It's a best-practice model for data-enabled research; patient-centricity; and public-private partnerships. The key players from life sciences, data sciences, medicine patient advocacy and communications will describe the winning formulas that are making it successful. Attendees will learn how to design and fund such an initiative; how to collect standards-based data including the horrendous challenges around EHRs; best tools for analytics and data visualization; handling research queries; and overcoming the traditional silos that prevent seamless data exchange and global big data-enable basic, clinical and comparative effectiveness research.

Track 10

Collaborations and Open Access Innovations

Using Collaborative Technologies and Methodologies to Accelerate Basic, Translational and Clinical Research

11:40 The Open PHACTS Foundation - Semantic Data Integration for Life Sciences

Bryn Williams-Jones, Founder & Chief Operating Officer, Open PHACTS, Connected Discovery

Building on the success of the Open PHACTS IMI project, the Open PHACTS Foundation is a not-for-profit membership organisation, supporting the Open PHACTS Discovery Platform. We will describe some of the capabilities of The Open PHACTS Discovery Platform, as well as show how commitment to pre-competitive and open innovation remains at the heart of The Open PHACTS Foundation with opportunities for all to get involved.

12:10 pm Session Break

12:20 Luncheon Presentation I: Accelerating Cancer Informatics at Foundation Medicine using SciDB

Sponsored by paradigm4 answer harder questions

Eric Neumann, Ph.D., Neurobiology and Developmental Genetics, Vice President, Knowledge Informatics, Foundation Medicine

Marilyn Matz, CEO, Paradigm4

Alex Poliakov, Solutions Architect, Paradigm4

Much can be learned from the proper analysis of large sets of genomic data. We will describe a few examples of scalable analytics applied to cancer genomics, and how SciDB enables this kind of analytics. Combining statistical analysis with other knowledge discovery tools can help accelerate this transformation of large data sets into biological insights.

1:20 Dessert Refreshment Break in the Exhibit Hall with Poster Viewing

1:55 Chairperson's Remarks

2:00 Incubating Open Innovation - the IMI2 Case Study

Anthony Rowe, Ph.D., Principal Research Scientist, External Innovation R&D IT, Janssen R&D, LLC

This presentation provides a better understanding of the Innovative Medicine Initiative (IMI) public-private consortia framework, what the scientific areas of interest are and the opportunities in IMI2 and what the process to propose and incubate new topics is. Learn from the cumulative experience of the EFPIA companies about the value of open innovation through the specific example of the IMI framework.

2:30 Allotrope Foundation: A Collaboration Making Real Progress Addressing the Data Management Problems Facing the Analytical Laboratory

Dana Vanderwall, Ph.D., Associate Director, Cheminformatics, Research Information Technology and Automation, Bristol-Myers Squibb

Allotrope Foundation is building a framework (a software toolkit) to embed a set of public, non-proprietary standards for analytical data in software utilized throughout the entire analytical chemistry data lifecycle. We will discuss why embedding standards addresses the fundamental, root causes of our data challenges, rather than merely treating symptoms of the problem.

3:00 Creating an Open Innovation Platform for the Promotion of Precompetitive Collaboration – The Pistoia Alliance's Interactive Project Portfolio Platform (IP3)

Carmen Nitsche, Executive Director Business Development North America, Pistoia Alliance

In order to promote the free exchange of ideas and increase the transparency of the portfolio development process, the Pistoia Alliance recently launched its Interactive Project Portfolio Platform (IP3). In this talk we will review the development and application of the platform as a key tool to advance the Pistoia Alliance's mission.

3:30 The New tranSMART Platform v1.2 Provides Unparalleled Functionality for Translational Medicine

Rudy Potenzone, Ph.D., Vice President, Marketing, tranSMART Foundation

The tranSMART Platform is in active use by over 50 organizations worldwide and the basis for a growing number of large data integration and analysis projects. It is becoming the premier platform for translational studies as the Community continues to expand this Open Source platform by their contributions. Learn about the Platform, see a large number of user examples and see the breadth of the user community contributing to the Platform.

4:00 Conference Adjourns

Organized by:
Cambridge Healthtech Institute

Track 11

Cancer Informatics

Applying Computational Biology to Cancer Research & Care

TUESDAY, APRIL 21

7:00 am Workshop Registration and Morning Coffee

8:00 – 11:30 Recommended Morning Pre-Conference Workshops*
Neuroinformatics Tools
Biologics, Bioassay and Biospeciment Registration Systems

12:30 – 4:00 pm Recommended Afternoon Pre-Conference Workshops*
How Data-Driven Patient Networks are Transforming Biomedical Research
The Impact of Research Informatics on Laboratory Evolutions
** Separate registration required*

2:00 – 6:30 Main Conference Registration

» 4:00 PLENARY SESSION
Please see page 5 for details.

5:00 – 7:00 Welcome Reception in the Exhibit Hall with Poster Viewing

Sponsored by IBM

WEDNESDAY, APRIL 22

7:00 am Registration Open and Morning Coffee

» 8:00 PLENARY SESSION
Please see page 5 for details.

9:00 Benjamin Franklin Awards and Laureate Presentation

9:30 Best Practices Awards Program

9:45 Coffee Break in the Exhibit Hall with Poster Viewing

Sponsored by INTERNET 2

GENOMICS: CLINICAL CHALLENGES AND MEDICAL OPPORTUNITIES

10:50 Chairperson's Opening Remarks

Scott Kahn, Ph.D., Vice President, Commercial Enterprise Informatics, Illumina, Inc.

» 11:00 FEATURED PRESENTATION: THE PENETRANCE OF INCIDENTAL FINDINGS IN GENOMIC MEDICINE

Robert C. Green, M.D., MPH, Director, G2P Research Program; Associate Director, Research, Partners Personalized Medicine, Division of Genetics, Department of Medicine, Brigham & Women's Hospital and Harvard Medical School

Much of the controversy surrounding the implementation of incidental findings in clinical sequencing is due to uncertainty about the penetrance of such findings in persons unselected for clinical features or family history. This uncertainty also influences the question of genomic population screening, i.e., whether actionable sequence variants should be sought and reported in ostensibly healthy individuals. In this talk, new data will be presented estimating the penetrance of actionable incidental findings.

» 11:30 FEATURED PRESENTATION: CHALLENGES AND OPPORTUNITIES IN ESTABLISHING IT SUPPORT FOR CONTINUOUS LEARNING IN HEALTHCARE: THE POTENTIAL FOR APPLYING LESSONS LEARNED FROM CLINICAL GENOMIC IT SUPPORT TO BROADER CONTINUOUS LEARNING CHALLENGES

Samuel (Sandy) Aronson, Executive Director, IT, Partners HealthCare Center forPersonalized Genetic Medicine

Continuously updated knowledge bases will be required to enable a true continuous learning healthcare environment. However, modern healthcare pressures make their maintenance difficult. The clinical genomic IT community has been wrestling with this issue for some time. We present lessons learned from supporting clinical genomic IT processes that may be generalizable to broader development of IT support for continuous learning healthcare processes.

12:00 pm Census of the Apoptosis Pathway

Sponsored by ELSEVIER

Philip L. Lorenzi, Ph.D., Department of Bioinformatics and Computational Biology & the Proteomics and Metabolomics Core Facility, MD Anderson Cancer Center

We recently compared several different "omic" approaches to constructing the autophagy pathway de novo, including siRNA screening, mass spectrometry-based proteomics, and three different pathway analysis software packages. Unexpectedly, although merging all of the validated data sets yielded 739 autophagy-modulating genes, each individual approach alone yielded sparse coverage of the autophagy pathway. The best individual siRNA screen, for example, yielded only 169 of the 739 (23%) genes. Nevertheless, text mining-based pathway analysis with Pathway Studio in conjunction with manual curation provided the most comprehensive coverage, yielding 417 targets (56% of the pathway). Here, we explored the generalizability of those findings by examining a more well-characterized pathway—apoptosis. We compiled apoptosis-modulating genes from 12 published siRNA screens and two pathway analysis software packages—Ingenuity Pathway Analysis (IPA) and Pathway Studio. The resulting inventory of 6,882 proteins consisted of 215 targets identified by siRNA screening, 3,378 targets by IPA, and 6,381 targets by Pathway Studio. The extensive coverage (93%) of the apoptosis pathway provided by text mining with Pathway Studio can likely be attributed to recent upgrades in the software, including an expanded database and collection of full-text articles. Together with our previous autophagy pathway analysis, the new apoptosis results support the generalizable conclusions that: 1) siRNA screening has a large false negative rate (i.e., fails to identify many true "hits"), and 2) text mining-based pathway analysis using Pathway Studio provides the most comprehensive pathway coverage.

12:30 Session Break

12:40 Luncheon Presentation I: Computational Enablement of the Hippocratic Oath in a Clinical Oncology Setting

Sponsored by MOLECULAR HEALTH

David B. Jackson, Ph.D., Chief Innovation Officer, Molecular Health, Gmbh

The clinical response of cancer patients to oncolytic agents is influenced by three major classes of molecular determinant; tumor intrinsic factors (e.g. tumor biomarkers); patient intrinsic factors (e.g. polymorphisms) and patient extrinsic factors (e.g. co-medications). In my talk, I will present a novel computational technology and associated treatment decision support process that was designed to provide this knowledge-driven approach to clinical care in oncology.

Track 11

Cancer Informatics

Applying Computational Biology to Cancer Research & Care

1:10 Luncheon Presentation II: A High Performance Application Development Platform for Collaborative Genomics Research

Sponsored by illumina

Paul Flook, Ph.D., Senior Director, Enterprise Informatics, Illumina Inc.

Collaborative research among groups working with genomic data presents major logistical challenges. Transferring huge volumes of data can be prohibitively expensive for projects utilizing WGS data sets. Illumina has addressed this challenge by building a platform that enables collaborators to not only share data in a secure multitenant environment, but to develop and deploy their own applications close to the data.

1:40 Session Break

BIG DATA, DIGITAL TOOLS AND BIOINFORMATICS ACROSS MULTIPLE RESEARCH INITIATIVES

1:50 Chairperson's Remarks

1:55 Metabolic Biomarkers in Duchenne Muscular Dystrophy

Subha Madhavan, Ph.D., Director, Innovation Center for Biomedical Informatics, Georgetown University Medical Center; Director, Clinical Informatics, Lombardi Comprehensive Cancer Center; Director, Biomedical Informatics, Georgetown-Howard Universities CTSA; Associate Professor, Department of Oncology, Georgetown University

Duchenne Muscular Dystrophy (DMD) is a devastating degenerative X-linked disorder which affects approximately 1 in 5,000 newborn males and results in muscle degeneration, eventual loss of ambulation around the age of 9, and a life expectance of around 25 years of age. A bioinformatics platform for metabolic data interpretation has been developed and tested to identify DMD-associated biomarkers and will be made available on GitHub once validation is complete. This platform will be presented along with another use case from a breast cancer metabolomics study.

2:25 Personalized Medicine: Moving from Correlation to Causality in Breast Cancer

Michael Liebman, Ph.D., Managing Director, Strategic Medicine, Inc.

Sabrina Molinaro, Ph.D., Institute for Clinical Physiology, National Research Council, Italy

We have developed a fundamental model of the disease process for breast cancer, from pre-disease through early detection, treatment and outcome, and apply a multi-scalar approach across the risk assessment-enhanced diagnosis-therapeutic decision axis and will present the modeling methodologies.

2:55 Streamline R&D and Catalyze Drug Repositioning by Identifying Expert Networks and Expertise

Sponsored by SINEQUA

Xavier Pornain, Vice President, Sales & Alliances, Sinequa

Finding networks of experts with similar or complementary expertise on a given subject helps avoid costly redundant research, shed light on a complex research problem from different angles, foster cooperation, facilitate drug repositioning, and accelerate time to market. This session will delve into the benefits pharmaceutical companies are seeing by employing Search & Analytics technology to: "link" researchers and teams with one another, create internal "journals of science" to share internal results and snippets, access "breaking science", with alerts and spotting trends across all scientific information. We show solutions for dealing with scientific vocabulary, detecting "synonyms" as well as "similar" and "complementary" notions, e.g. brand names for drugs, scientific names for the active ingredients, and even descriptions of molecules using a standard description language. In addition, we analyze vast quantities (200 to 500 million) of highly technical documents and data (billions of records), such as internal and external publications, patent filings, lab reports, clinical test reports, trade databases, etc.

3:10 Cloud-Based Solutions for Population-Scale, Whole Human Genome and Exome Analysis

Sponsored by DNAnexus

George Asimenos, Ph.D., Director, Science & Clinical Solutions, DNAnexus

Thanks to advances in sequencing technology, the size and scope of DNA sequencing projects is rapidly moving towards an era of thousands of whole genomes and tens of thousands of exomes per year. Learn how certain field-leading institutes are using a cloud-based bioinformatics platform to manage their big data deluge across multiple initiatives.

3:25 Refreshment Break in the Exhibit Hall with Poster Viewing

COLLABORATING IN CHRONIC AND RARE DISEASES

4:00 CureAccelerator™: How A New Global Platform Will Help Propel Cures for the World's Unsolved Diseases

Bruce Bloom, D.D.S., J.D., President and Chief Science Officer, Cures Within Reach

More than 7,000 diseases have no fully effective treatment, affecting more than 500 million people worldwide. CureAccelerator™ is the world's first open-access, online platform dedicated to repurposing research – the quest to create new medical treatments from existing therapies, to drive more cures more quickly to more patients. Learn how this innovative IT tool will enable researchers, funders, the biomedical industry and patient groups to collaborate far more efficiently, to propel the pace of repurposing research.

4:30 Delivering on a Promise: Achieving a Patient-Centric Open Information Ecosystem

Walter Capone, President and CEO, The Multiple Myeloma Research Foundation

Patient-centric, open access research models are widely acknowledged catalysts of scientific and medical progress. Here we present a first-in-kind model for combining an observational clinical study with a participatory, community driven research program for Mutliple Myeloma, and in doing so, providing a powerful example that can lead the way forward in other disease areas.

5:00 There and Back Again: AstraZeneca's Collaboration Journey

Robert Albert, DocuSign Project Manager and Collaboration Specialist, AstraZeneca

AstraZeneca has always been on the cutting edge of innovation. Despite having over 50,000 employees worldwide, we know we don't have all of the answers. In the past 3 years, we have been transforming into a more collaborative company. The journey began with a culture event where we discussed how we could solve difficult challenges through the power of crowdsourcing. In 2012 we partnered with Innocentive to deliver iSolve/ @ work – a bulletin board where R&D challenges could be solved by colleagues from around the world. 2013 saw the AstraZeneca world-wide culture jam, which brought together 34,000 people from 100 countries together over a 3 day span. In 2014 we launched the R&D Open Innovation website. On the site we have 5 modules. They include: the clinical compound bank; the pharmacology toolbox; target innovation/ compound library; new molecule profiling, and the R&D module. My talk will focus specifically on the R&D challenges, and how we are crowdsourcing answers from Innocentive's solver network. Our aim is to deliver life-saving medicines as quickly as we can. One powerful option is through open collaboration with the greater scientific community.

5:30 Best of Show Awards Reception in the Exhibit Hall with Poster Viewing

6:30 Close of Day

Organized by:
Cambridge Healthtech Institute

Track 11

Cancer Informatics

Applying Computational Biology to Cancer Research & Care

THURSDAY, APRIL 23

7:00 am Registration Open and Morning Coffee

» 8:00 PLENARY SESSION PANEL
Please see page 5 for details.

10:00 Coffee Break in the Exhibit Hall and Poster Competition Winners Announced

COLLABORATING IN CHRONIC AND RARE DISEASES

10:30 Chairperson's Remarks
Robert Albert, DocuSign Project Manager and Collaboration Specialist, AstraZeneca

10:40 PANEL DISCUSSION: Multiple Sclerosis Conquered by the Data Science Revolution: Patients Win
Ken Buetow, Ph.D., Director, Computational Sciences and Informatics, Arizona State University
Robert McBurney, CEO, Office of the Chief Executive, Accelerated Cure Project
Marcia Kean, Chairman, Strategic Initiatives, Feinstein Kean Healthcare
Joseph LaFerrera, J.D., Partner, Gesmer Updegrove LLP
Funded by a grant from PCORI (Patient-Centered Outcomes Research Institute), the Accelerated Cure Project for MS is collaborating with all the key organizations in the MS community, gathering patient-reported and EHRs from 20,000 patients. It's a best-practice model for data-enabled research; patient-centricity; and public-private partnerships. The key players from life sciences, data sciences, medicine patient advocacy and communications will describe the winning formulas that are making it successful. Attendees will learn how to design and fund such an initiative; how to collect standards-based data including the horrendous challenges around EHRs; best tools for analytics and data visualization; handling research queries; and overcoming the traditional silos that prevent seamless data exchange and global big data-enable basic, clinical and comparative effectiveness research.

11:40 Sponsored Presentation *(Opportunity Available)*

12:10 pm Session Break

12:20 Luncheon Presentation I: Accelerating Cancer Informatics at Foundation Medicine using SciDB
Sponsored by paradigm4 answer harder questions
Eric Neumann, Ph.D., Neurobiology and Developmental Genetics, Vice President, Knowledge Informatics, Foundation Medicine
Marilyn Matz, CEO, Paradigm4
Alex Poliakov, Solutions Architect, Paradigm4
Much can be learned from the proper analysis of large sets of genomic data. We will describe a few examples of scalable analytics applied to cancer genomics, and how SciDB enables this kind of analytics. Combining statistical analysis with other knowledge discovery tools can help accelerate this transformation of large data sets into biological insights.

1:20 Dessert Refreshment Break in the Exhibit Hall with Poster Viewing

DATA CAPTURE, ANALYSIS, MODELING & SIMULATION

1:55 Chairperson's Remarks
Nils Gehlenborg, Ph.D., Research Associate, Center for Biomedical Informatics, Harvard Medical School

2:00 Combing the Hairball: Network Visualization with BioTapestry and BioFabric
William J.R. Longabaugh, MS, Senior Software Engineer, Institute for Systems Biology
Networks models are crucial for understanding complex biological systems, yet traditional node-link diagrams of large networks provide very little visual intuition, and there is a need to develop scalable, unambiguous, and rational network visualization techniques. Our applications, BioTapestry (http://www.BioTapestry.org) and BioFabric (http://www.BioFabric.org), are designed to address this need, and I will discuss how they use novel approaches to avoid the "hairball" trap.

2:30 Visualization of Comparative Genomics Data: Results, Challenges, and Open Questions
Inna Dubchak, Ph.D., Senior Scientist, Lawrence Berkeley National Laboratory
As the rate of generating sequence data continues to increase, visualization tools for interactive exploration and interpretation of comparative data at the level of gene, genome, and ecosystem are of critical importance. We will talk about strengths and limitations of existing methods, and highlight new challenges in the visualization of huge volumes of complex comparative data.

3:00 Interactive and Exploratory Visualization of Epigenome-Wide Data
Hector Corrada Bravo, Ph.D., Assistant Professor, Center for Bioinformatics and Computational Biology, Department of Computer Science, University ofMaryland, College Park
We will introduce epigenomics data visualization tools that provide tight-knit integration with computational and statistical modeling and data analysis: Epiviz (http://epiviz.cbcb.umd.edu), a web-based genome browser application, and the Epivizr Bioconductor package that provides interactive integration with R/ Bioconductor sessions. This combination of technologies permits interactive visualization within a state-of-the-art functional genomics analysis platform. The web-based design of our tools facilitates the reproducible dissemination of interactive data analyses in a user-friendly platform.

3:30 Visual-Analytic Systems for Integrative Genomic Analysis of Cancer Data
Raghu Machiraju, Ph.D., Professor, Ohio State University
Cancers are highly heterogeneous with different subtypes. Recently, integrative approaches were adopted that combined multiple types of omics data. In this talk, I present visual analytic solutions for the simultaneous and integrative exploration of multiple types genomics data including those from The Cancer Genome Atlas (TCGA) project. Using different combinations of mRNA and microRNA features we suggest potential combined markers for prediction of patient survival.

4:00 Conference Adjourns

Organized by:
Cambridge Healthtech Institute

Track 12

Data Security

Meeting the Challenge in a Data-Centric World

TUESDAY, APRIL 21

7:00 am Workshop Registration and Morning Coffee

8:00 – 11:30 Recommended Morning Pre-Conference Workshops*

An Embarrassment of Riches: Choosing and Implementing Cloud Infrastructure

12:30 – 4:00 pm Recommended Afternoon Pre-Conference Workshops*

Large-Scale NGS Analysis Using Globus Genomics

** Separate registration required*

2:00 – 6:30 Main Conference Registration

» 4:00 PLENARY SESSION
Please see page 5 for details.

5:00 – 7:00 Welcome Reception in the Exhibit Hall with Poster Viewing

Sponsored by IBM

WEDNESDAY, APRIL 22

7:00 am Registration Open and Morning Coffee

» 8:00 PLENARY SESSION
Please see page 5 for details.

9:00 Benjamin Franklin Awards and Laureate Presentation

9:30 Best Practices Awards Program

9:45 Coffee Break in the Exhibit Hall with Poster Viewing

Sponsored by INTERNET 2

DESIGNING A SECURE CLOUD

10:50 Chairperson's Opening Remarks
Dave Peterson, Executive Director, Vendor & Third Party Assurance, National IT Compliance, Kaiser Permanente Information Technology

» 11:00 FEATURED PRESENTATION: COMPLIANT CLOUD COMPUTING
Krista Woodley, Director, Information Technology, Biogen
We provide insight on how to best manage SaaS-based projects in a regulated world, by discussing best practices for Lifecycle management, change control, security management and IT risk management. IT and business project teams will have a clear understanding of how to optimize their IT deployments in this new cloud-based environment.

11:30 Rethinking Cloud Security: You Can't Control What You Can't See
Kevin Gilpin, CTO, Conjur, Inc.
As more companies adopt DevOps programs and build new infrastructure, the quantity and sensitivity of data being processed outside of the traditional IT stack are growing. Few organizations know where the access points into this information are, or how to secure them. We outline best practices for establishing visibility and control in this new space, drawing real-world examples from environments large and small.

12:00 pm Security in the Cloud: How AMAG Protects Company Data with Multi-factor Authentication

Sponsored by okta

Nathan McBride, Vice President, IA & Chief Cloud Architect, AMAG Pharmaceuticals
To stay competitive and deliver world-class care, organizations such as yours are increasingly adopting cloud and mobile-first IT strategies. These trends come with significant security and access management challenges. In this presentation, Nate McBride, VP of IT and Chief Cloud Architect at AMAG Pharmaceuticals will discuss AMAG's move to the cloud and their deployment strategy for securing data with multi-factor authentication.

12:30 Session Break

12:40 Luncheon Co-Presentation I: Are Your Researchers Paying Too Much for Their Cloud-Based Data Backups?

Sponsored by Silicon Mechanics

Dirk Petersen, Scientific Computing Manager, Fred Hutchinson Cancer Research Center (FHCRC)
Joe Arnold, President and Co-Founder, SwiftStack
Considering deploying a multi-petabyte storage-as-a-service offering in your research environment? Learn how an industry-leading software-defined object storage solution, architected by SwiftStack and Silicon Mechanics, helped shift hundreds of users to an object-based workflow for their archival data. With an emphasis on cost efficiencies, scalability, and manageability, see how this implementation at Fred Hutchinson Cancer Research Center (FHCRC) is continually evolving across new use cases and access methods.

1:10 Luncheon Co-Presentation II: Running Scalable and Cost Effective High-Throughput Sequencing Data Analysis on Amazon Web Services

Sponsored by amazon web services

Cory Funk, Ph.D., Research Scientist, Institute for Systems Biology
Dmitry Pushkarev, Ph.D., CEO and Founder, ClusterK
Here we present work by the Institute for Systems Biology, in collaboration with ClusterK and AWS, to run large cohort RNA-Seq comparative data analysis on the AWS Spot Market. We will showcase the SNAPR algorithm for transcriptome analysis, as well as highlight the advanced features of the ClusterK products that make full use of AWS Spot instances that resulted in significant cost savings over on-demand pricing.

1:40 Session Break

DATA SECURITY VS. DATA PRIVACY IN HEALTHCARE

1:50 Chairperson's Remarks
Nora Manstein, Ph.D., IT Project Manager, Bayer Business Services GmbH

1:55 Security vs. Freedom – It's Not a Matter of Philosophy
Nora Manstein, Ph.D., IT Project Manager, Bayer Business Services GmbH
We contribute to the debate on how patient's rights and wishes are respected and meaningful research with patient data can be done. In order to support this, we have developed an organizational process and a technical tool by which patients' informed consents are an integral part of the authorization process, allowing compliant access to and scientific analysis of patient data.

Organized by:
Cambridge Healthtech Institute

Track 12

Data Security

Meeting the Challenge in a Data-Centric World

2:25 Privacy, Access Control and Security in Clinical Genomics Environments

Toby Bloom, Ph.D., Deputy Scientific Director, Informatics, New York Genome Center

The integration of clinical and genomic data introduces new, complex problems in privacy and security. These include protecting the anonymity of clinical data when it is linked to "self-identifying" genomic data; managing the fine-granularity access control required to share data from multiple projects; and overcoming the regulatory and legal hurdles associated with clinical genomic data. We discuss these and other access issues.

2:55 Blocking the Cyber Barbarians

Betsy Fallen, Global Head, Program and Business Development, SAFE-BioPharma Association

Identity trust is necessary for secure and regulated Internet communications. The presentation explains the issues associated with establishing online trust and the role of the industry-driven SAFE-BioPharma global identity management/digital signature standard in assuring that only authorized identities have access to protected information. Participants will learn about types and levels of identity credentials, government and industry organizations involved in establishing identity trust infrastructures, applicable standards, governance models and approaches to cloud-based identity management.

3:25 Refreshment Break in the Exhibit Hall with Poster Viewing

4:00 Data Integration, Privacy and Openness at PatientsLikeMe, a Social Network for Patients with Life-Altering Conditions

Marcia M. Nizzari, MS, Vice President, Engineering, PatientsLikeMe, Inc.

PatientsLikeMe provides a social network and research platform for capturing, curating and analyzing patient-reported data. With 300,000+ users, 2,300+ conditions represented and over 25 million health datapoints collected, it provides a new, rich source of data to integrate with EHR and genomic data to drive new insights about disease. We discuss trade-offs in privacy and openness when combining EHR and other sources of clinical and research data – such as -omics – with patient-reported data.

4:30 Differential Privacy: Future-Proof Protection for Sensitive Data

Ishaan Nerurkar, CEO & Founder, Shroudbase, Inc.

In the analysis of sensitive data, current methods of privacy protection severely compromise utility, access and opportunities for collaboration. Shroudbase is a cloud software that creates and manages permanently de-identified copies of high-dimensional data with strong, mathematically proven guarantees of statistical accuracy. Our patent-pending technology enables previously untouchable information to be open-sourced and analyzed while maintaining differential privacy, the gold standard of data privacy. This presentation discusses an instance of the Shroudbase platform optimized to handle the unique privacy challenges posed by genomics data.

5:00 PANEL DISCUSSION: Genomic Research: Utility vs. Patient Rights

Moderator:

Toby Bloom, Ph.D., Deputy Scientific Director, Informatics, New York Genome Center

Panelists:

Betsy Fallen, Global Head, Program and Business Development, SAFE-BioPharma Association

Nora Manstein, Ph.D., IT Project Manager, Bayer Business Services GmbH

Ishaan Nerurkar, CEO & Founder, Shroudbase, Inc.

Marcia M. Nizzari, MS, Vice President, Engineering, PatientsLikeMe, Inc.

Juhapekka Piiroinen, Head, Personalized Medicine Development, MediSapiens, Ltd.

What software tools, organizational processes and regulatory minefields must researchers and clinicians understand to not only improve drug development and personalized therapies, but also preserve the privacy of patient data? Experts share their perspectives on informed consent, security access, technical infrastructures, genomic and clinical data integration and more.

5:30 Best of Show Awards Reception in the Exhibit Hall with Poster Viewing

6:30 Close of Day

THURSDAY, APRIL 23

7:00 am Registration Open and Morning Coffee

» 8:00 PLENARY SESSION PANEL
Please see page 5 for details.

10:00 Coffee Break in the Exhibit Hall and Poster Competition Winners Announced

FROM COLLABORATION TO COMPLIANCE

10:30 Chairperson's Remarks

Jason Tetrault, Associate Director, Business and Information Architect, R&D IT, Biogen

10:40 Next-Generation Sequencing and Cloud Scale: A Journey to Large-Scale Flexible Infrastructures in AWS

Jason Tetrault, Associate Director, Business and Information Architect, R&D IT, Biogen

Biogen has built burst capabilities for large-scale NGS processing and collaboration with our partners. This extension of our infrastructure capability allows us to be more nimble, process more data and scale as needed. It also gives us unique options as we work with collaborators at scale. Of course, because it is NGS data, doing it securely is important.

11:10 Data Communications in BSL-3 and BSL-4 Containment: Safety, Compliance and Security

John McCall, Director, Information Technology and Telecommunications, National Emerging Infectious Diseases Laboratories, Boston University

Innovative solutions for BSL-3 and BSL-4 facilities address the asset tracking, personnel monitoring and worker communication problems associated with personal protective equipment and physical environment design. I scope out what it takes to plan and roll out a wireless networking and voice-over-IP system that meets safety, security and compliance requirements at Boston University's National Emerging Infectious Disease Laboratory.

11:40 Breaking the Classical Barriers to Collaboration and Scientific Discovery - Distance and Data Size

Sponsored by aspera an IBM® company

Michelle Munson, President and CEO, Aspera, an IBM Company

Life sciences organizations need to dramatically reduce analytics time and speed up clinical interventions, but most still rely on shipping physical disks due to inherent problems with existing networks and transfer protocol inefficiencies. Spending days to transport data is not a viable option, this session will explore technology infrastructure for file transfer that will catalyze the transition from 1GbE to 10GbE and beyond.

12:10 pm Session Break

12:20 Luncheon Presentation *(Sponsorship Opportunity Available)* or Lunch on Your Own

1:20 Dessert Refreshment Break in the Exhibit Hall with Poster Viewing

Track 12

Data Security

Meeting the Challenge in a Data-Centric World

REGULATIONS: DATA PRIVACY AND SECURITY

1:55 Chairperson's Remarks

John M. Conley, J.D., Ph.D., William Rand Kenan, Jr. Professor of Law, University of North Carolina, Chapel Hill ; Counsel, Robinson Bradshaw & Hinson

» 2:00 FEATURED PRESENTATION: IT AND INFORMATICS INNOVATION AT FDA

Roselie A. Bright, Sc.D., MS, PMP, Program Manager, Office of Information Management and Technology, Office of Informatics Technology and Innovation, Office of Operations, Office of the Commissioner, U.S. Food And Drug Administration (FDA)

OpenFDA was the first innovation created by Taha Kass-Hout, M.D., MS, upon joining FDA as the first Chief Health Information Officer in March 2013. OpenFDA was launched on June 2, 2014, allowing software developers, researchers and the public to tap into adverse events for drugs and medical devices; recalls, for drugs, devices and foods; and labeling for products on the market.

2:30 Global Developments in Privacy and Data Security Law

John M. Conley, J.D., Ph.D., William Rand Kenan, Jr. Professor of Law, University of North Carolina, Chapel Hill; Counsel, Robinson Bradshaw & Hinson

The international legal climate governing privacy and data security is changing. The European Union is in the midst of a fundamental shift in its approach. The U.S. still lacks a national data law, so the states and individual federal agencies are groping toward a strategy. This presentation focuses on the impact of these ongoing changes on genomics, bioinformatics and health research.

3:00 PANEL DISCUSSION: Achieving Much-Needed Innovation while Hurdling the Barriers of Stringent Regulation

Moderator: John M. Conley, J.D., Ph.D., William Rand Kenan, Jr. Professor of Law, University of North Carolina, Chapel Hill; Counsel, Robinson Bradshaw & Hinson

Panelists:

Roselie A. Bright, Sc.D., MS, PMP, Program Manager, Office of Information Management and Technology, Office of Informatics Technology and Innovation, Office of Operations, Office of the Commissioner, U.S. Food And Drug Administration (FDA)

Dana Caulder, Senior Software Engineer, Bioinformatics and Computational Biology, Genentech

Chris Dwan, Assistant Director, Research Computing and Data Services, Broad Institute of MIT and Harvard

Sanjay Joshi, CTO – Life Sciences, Emerging Technologies Division, EMC

Dave Peterson, Executive Director, Vendor & Third Party Assurance, National IT Compliance, Kaiser Permanente Information Technology

Vas Vasiliadis, Director, Products, Computation Institute, University of Chicago and Argonne National Laboratory

The growth in patient healthcare and life sciences innovations can be attributed to technology enhancements like cloud computing, big data analytics and mobile applications, but may conflict with increasing regulatory compliance demands to ensure protection of healthcare life and quality as well as patient data privacy and security. This panel of esteemed technology solution providers and regulators debates real-world challenges and how regulation must also innovate at technology's pace.

4:00 Conference Adjourns

Click Here to Register Online!
Bio-ITWorldExpo.com

Organized by:
Cambridge Healthtech Institute

HOTEL & TRAVEL

Conference Venue:
Seaport World Trade Center
200 Seaport Boulevard
Boston, MA 02210

Host Hotel:
Seaport Hotel *(Located directly across the street)*
One Seaport Lane
Boston, MA 02210
T: 617-385-4514
Reservations: Please visit the travel page of Bio-ITWorldExpo.com

Discounted Room Rate: $259 s/d
Discounted Room Rate Cut-off Date: March 19, 2015

Please visit Bio-ITWorldExpo.com for additional hotels and travel information

MEDIA PARTNERS

Official Media Partner:
Bio·IT World

Lead Sponsoring Publications:
CLINICAL INFORMATICS NEWS
GEN Genetic Engineering & Biotechnology News
Science AAAS
TheScientist EXPLORING LIFE, INSPIRING INNOVATION

Sponsoring Publications:
ASSAY and Drug Development Technologies
BIG DATA
CanBiotech
DDNEWS
INSIGHT PHARMA REPORTS
Pharmaceutical-Tech.com
PHARMA FOCUS ASIA
PharmaVOICE
Scientific Computing
SelectScience

Official PR Partner:
PAN COMMUNICATIONS

Sponsoring Organizations:
Bioinformatics Organization
ISCB AFFILIATED CONFERENCE
MASSACHUSETTS HIGH TECHNOLOGY COUNCIL
MASSACHUSETTS LIFE SCIENCES CENTER
Investing in the State of Innovation

Web Partners:
Biocompare The Buyer's Guide for Life Scientists
Big3Bio
BioSpace.com
FierceBiotech THE BIOTECH INDUSTRY'S DAILY MONITOR
FierceBiotech IT
FiercePharma
genomeweb
labroots
mendelspod.com
Pharmaceutical Outsourcing™

Click Here to Register Online!
Bio-ITWorldExpo.com

Organized by:
Cambridge Healthtech Institute

SPONSORSHIP, EXHIBIT, AND LEAD GENERATION OPPORTUNITIES

CHI offers comprehensive sponsorship packages which include presentation opportunities, exhibit space, branding and networking with specific prospects. 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 early will allow you to maximize exposure to qualified decision-makers.

Podium Presentations – Within the Main Agenda!

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

Luncheon Podium 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!

Exhibit

Exhibitors will enjoy facilitated networking opportunities with qualified delegates. Speak face-to-face with prospective clients and showcase your latest product, service, or solution.

Additional branding and promotional opportunities are available, including:

- Mobile App (SOLD)
- Hotel Room Keys (SOLD)
- Footprint Trails
- Staircase Ads
- Conference Tote Bags (SOLD)
- Badge Lanyards (SOLD)
- Program Guide Advertisement

Looking for additional ways to drive leads to your sales team?

CHI's Lead Generation Programs will help you obtain more targeted, quality leads throughout the year. We will mine our database of 800,000+ life science professionals to your specific needs. We guarantee a minimum of 100 leads per program! Opportunities include:

- Whitepapers
- Web Symposia
- Custom Market Research Surveys
- Podcasts

For sponsorship and exhibit information, please contact:

Companies A-K:
Katelin Fitzgerald
Sr Business Development Manager
781-972-5458 | kfitzgerald@healthtech.com

Companies L-Z:
Elizabeth Lemelin
Business Development Manager
781-972-1342 | elemelin@healthtech.com

CURRENT SPONSORS & EXHIBITORS

(November 25, 2014)

Accelrys
Accunet Solutions
Aequor Technologies
Alpha Clinical Systems, Inc.
Annai Systems, Inc.
Arista Networks
Arxspan
Aspera, Inc.
Avere Systems
Ayasdi
Bina Technologies
BioFortis, Inc.
Bioinformatics Organization, Inc.
Biomatters Geneious Software
Biomax Informatics AG
BioTeam
BSI
BT Global
Cambridge Computer
Cambridge Semantics
Certara
ChemAxon
Chemical Computing Group
Cleversafe
Collaborative Consulting
ConvergeHEALTH by Deloitte
Core Informatics
Cray, Inc
Cycle Computing
DataDirect Networks
Dell
DeltaSoft, Inc.
DNAnexus
Dotmatics
Eagle Genomics
Elsevier
EMC Isilon
Exostar LLC
Freezerworks
GENALICE
Genedata
GenoSpace
GGA Software Services
Globus
IBM
Illumina, Inc.
INFINIDAT
Intel
Internet2
inviCRO, LLC
ISCB
LabAnswer
LabVantage Solutions, Inc.
Liaison Healthcare Informatics
Linguamatics
Lumenogix
Maverix Biomics, Inc.
MediSapiens, Inc.
Metrum Research Group
Nebula
NextCODE Health
NNIT
Okta
OpenEye Scientific Software
Oracle Health Sciences
OSTHUS GmbH
Partek
Parthys Reverse Informatics
PerkinElmer
Qlucore
Qumulo
RAID Incorporated
RCH Solutions
SAS Institute Inc., JMP Division
Schrödinger
Scigilian Software, Inc.
Scilligence
Seagate
Seven Bridges Genomics
SGI
Sidus BioData
Silicon Mechanics
Simulations Plus, Inc.
SINEQUA
Sterne Kessler Goldstein & Fox
Studylog Animal Study Software
Tessella
Thinkmate
Thomson Reuters
TimeLogic
Titian Software
TopQuadrant
tranSMART Foundation
Univa Corporation
Waters Corporation
XTechnology Global LLC
Zifo Technologies

Click Here to Register Online!
Bio-ITWorldExpo.com

Organized by:
Cambridge Healthtech Institute

How to Register: Bio-ITWorldExpo.com

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

Please use keycode 1520 F when registering

Pricing and Registration Information

WORKSHOP PRICING

	Commercial	Academic, Government, Hospital-affiliated	Student*
One Half-Day Workshop	$599	$299	$149
Two Half-Day Workshops	$899	$499	$249

Please refer to Workshop list on page 3.

MAIN CONFERENCE PRICING (excludes workshops)

Registrations after March 13, 2015, and on-site	$2,099	$979	$329

Conference Tracks

Track 1: IT Infrastructure - Hardware	Track 7: Data Visualization & Exploration Tools
Track 2: Software Development	Track 8: Pharmaceutical R&D Informatics
Track 3: Cloud Computing	Track 9: Clinical Genomics
Track 4: Bioinformatics	Track 10: Collaborations & Open Access Innovations
Track 5: Next-Gen Sequencing Informatics	Track 11: Cancer Informatics
Track 6: Clinical & Translational Informatics	Track 12: Data Security

** Full-time graduate students and PhD candidates can attend Bio-IT World Conference & Expo at a special Student Rate. Students are encouraged to present a research poster and receive an additional $50 off their registration fee. Research posters will be seen by leaders from top pharmaceutical, biotech, academic, government institutes, and technology vendors. Posters will be automatically entered in the Poster Competition, where two winners will each receive an American Express Gift Card. Poster abstracts are due by March 6, 2015.*

Student rate cannot be combined with any other discount offers, except poster discount. Students must present a valid/current student ID to qualify for the student rate. Limited to the first 100 students that apply.

CONFERENCE DISCOUNTS

Exclusive Offer to Attend Medical Informatics World and Clinical Research Informatics World Conferences*

Cambridge Healthtech Institute presents a series of informatics programs in Boston this spring with the goal of bridging the healthcare and life science worlds. Paid attendees of Bio-IT World Conference & Expo can attend the **Medical Informatics World Conference** (May 4-5) and **Clinical Research Informatics World Conference** (May 6-7) for a special discounted rate (20% discount off the registration fee for the main conference).

To receive this exclusive 20% discount, mention keycode **1520BITXP** when registering for Medical Informatics World and/or Clinical Research Informatics World. *Please note: Our records must indicate you are a paid attendee of Bio-IT World Conference & Expo 2015 to qualify.*

** Discount applies to paid attendees of Bio-IT World Conference & Expo 2015 only. Applies to new registrations only and cannot be combined with other discount offers, except poster discount. Discount does not apply to workshops. Discount taken off lowest priced item(s).*

Poster Submission-Discount ($50 Off)
Poster abstracts are due by March 6, 2015. 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.

International Society for Computational Biology (ISCB) Member-Discount (10% Off)

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

If you are unable to attend but would like to purchase the Bio-IT World Conference & Expo 2015 conference CD for $750 (plus shipping),please visit Bio-ITWorldExpo.com. Massachusetts delivery will include sales tax.

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 www.healthtech.com/regdetails

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