Report on the Cancer Biomarker Bioinformatics Workshop
April 29 – May 1, 2013 at Caltech, Pasadena, California

Executive Summary

A three-day workshop was held at the California Institute of Technology (Caltech) on April 29, 2013 – May 1, 2013 that highlighted informatics resources, needs, and challenges for cancer biomarker research with a specific emphasis on the identification of future priorities and plans. Significant progress has been made in informatics to improve the capture and analysis of cancer biomarker data across several programs at the National Cancer Institute (EDRN, TCGA, CPTAC, NCIP, etc). However, many of these techniques remain substantially ad hoc, with little standardization of methods for collecting and analyzing biomarker data. Improvements need to be explored in developing national, scalable data infrastructures that support data capture and integration. Integration of data pipelines may lead to increased data access for meta-analysis, authenticated provenance for data sets and methods, scalable and sustainable infrastructure for storage and analysis, and assured quality assessment. Taken together, these improvements could have a substantial impact on cancer research.

The workshop was organized by NASA's Jet Propulsion Laboratory (NASA/JPL) working in conjunction with the National Cancer Institute (NCI) and several of the Early Detection Research Network sites. The program committee consisted of the following people: Dan Crichton, NASA/JPL, David Elashoff, UCLA, Chris Kinsinger, NCI, Marc Lenburg, Boston University, Mervi Heiskanen, NCI, Christos Patriotis, NCI, Sudhir Srivastava, NCI, David Tabb, Vanderbilt University, Mark Thornquist, Fred Hutchinson Cancer Center, and Zhen Zhang, Johns Hopkins Medical Institute.

Keynotes and speakers were selected by the program committee. A multi-disciplinary group of researchers spanning biomedical research and informatics attended. "Bench" researchers provided case studies that reflected their informatics needs and the programs for which they are involved. Informatics personnel introduced new tools and data resources to the audience. Community dialogue occurred during discussion-panel sessions. Multiple researchers were selected to participate in a poster session.

The workshop produced several recommendations, which are expanded on in the full report, which include:

- Reproducibility: Captured data should enable reproducibility of results. Ensure reproducibility by through high quality capture of data and metadata. Capture multiple levels of data, processing and algorithms, scripts, and document steps to ensure an appropriate level of provenance is captured.

- Quality indicators should be captured with the data: Analyze and share information about data quality, recognizing that industry data is often higher quality than academic data. Quality indicators should be included as part of the capture and management of the data.

- Avoid common biases in studies examining the association between biomarkers and disease: Use PRoBE (Prospective specimen collection with Retrospective Blinded Evaluation) study design checklist for sample selection and development of the experimental design (JNCI 2008 Pepe et. al.).

- Data Curation Lifecycle: Increase emphasis on data curation within the data capture lifecycle. In many cases, manual curation must be included to ensure high quality capture of data and enable effective reuse.

- Automated pipelines: Automated pipelines should be developed to improve the quality and efficiency in the capture of data. These should include automated workflows that support the capture and provenance for reproducibility.

- Common Ontologies and Standards for Biomarker Research: Define a standard set of minimum information that should be published with biomarkers. The minimum information should form a baseline for biomarker databases and captured datasets. Define standards for data formats and structures to enable effective reuse of the data.

- Data Sharing and Interoperability: Ensure data and datasets can be shared between different applications/platforms. Integrate data from multiple resources including pathway and network databases, genomic and proteomic data, etc.

- Data Infrastructures: Provide robust data infrastructures and repositories to support collection and integration of wide variety of data types from biomarker research, ensuring longevity of data. These infrastructures should provide foundational building blocks to enable data analysis at a national scale following the model that the EDRN is establishing through the EDRN Knowledge System.

- Open source: Leverage and share open source software to enable and reuse common infrastructure, shared analysis tools, and promote common practices in biomarker research.

- Big data: Big data is challenging scalability of existing systems. Develop new approaches to scaling data management, distribution and analysis of cancer biomarker research results that integrate data infrastructures and data analysis.

- Limited training/expertise in the use of computational capabilities and techniques. Many scientists lack expertise and training in the effective use of computing technologies to support biomarker research. This can affect the adoption of emerging informatics capabilities and limit their use in biomarker discovery and validation. Efforts are needed to better train scientists in learning and using informatics.

Detailed Report

Session 1: Informatics in Biomarker Discovery

The application of multiple biotechnologies to the development of cancer biomarkers is inextricably linked to the development of algorithms to support the analysis of complex datasets. These technologies and algorithms play a key role in biomarker discovery. Informatics capabilities to support the capture, integration and analysis of these datasets are critical to supporting biomarker discovery and enabling reproducibility of the results. This session included contributions from researchers who identified key needs for supporting experimental methodologies and from researchers who are developing needed algorithms for biomarker data capture, integration, and analysis.

Summary of Discussion

- Biomarker Data Curation: Manual biocuration requires infrastructure and expertise, and is time consuming, particularly for big data in biomarker research. It is, however, critical for ensuring reproducibility of the results and enabling effective reuse of captured datasets.
- Data Sharing: There are several curation efforts underway. These could be better coordinated to ensure curated datasets could be shared between complimentary applications. However, this requires definition of common APIs, shared data formats, and common data standards.
- Capture of high quality, primary datasets: The capture of primary (raw) data is critical to enable reuse of data and reproducibility of results. This is often a problem for data repositories that lack standards and proper curation procedures. Data is often difficult to use with different data formats, limited metadata, and limited documentation. Quite often, manual data curation and clean up is required, but that can also be severely constrained by the data that is captured.

Session 2: Informatics in Biomarker Validation

As new candidate biomarkers are reported in journals each week, the need to test and validate these biomarkers has become critical in the field. Speakers from this session involved in biomarker assessment presented case that highlighted the need for informatics improvements. The session also included researchers who have developed new algorithms to assist in biomarker data evaluation.

Summary of Discussion

- Reproducibility of results is a huge problem. This was a common theme throughout the entire workshop and is highly dependent on the proper information capabilities. In particular, proper capture of the data is essential. This includes documentation of pre-determined protocol, raw data, code, scripts, data provenance, and analysis workflows. The capture of data, metadata, and supporting documentation should ensure proper reuse of the data, particularly for validating biomarker results.
- Transparency. There is a direct need to share data and analytical methods to support reproducibility. Well curated data repositories that are maintained are needed that can support proper sharing of

data and methods. Journals should require publication of raw and processed data and metadata in these repositories that can be directly cited.

- Define/standardize minimum information that should be published with biomarkers. A minimum set of biomarker information should be included with each biomarker publication. This should also be directly linked and adopted by repositories that are capturing biomarker data. This minimal information should be used to directly annotate information about the quality and results of the biomarker and the associated study.
- Share information about data quality. Common approaches should be developed to share information about data quality in biomarker research. Currently, there are significant differences between industrial and academic data analysis. Industry is often higher quality than academia. Reuse of results is often limited due to data quality concerns and effective reproducibility of results.
- Sample heterogeneity impacts data analysis. Differences in sample collection can have a direct impact on data analysis. Standards such as what EDRN has developed for its Specimen Reference Sets, are critical to validating biomarker discovery results. The biomarker research community should place more emphasis on developing common reference sets for use and validation to improve the quality of data analysis.

Session 3: Quality-by-Design in Biomarker Research

The use of quality-by-design methodologies can make biomarker discovery and validation more rigorous, increasing the reproducibility of significant biomarker findings. This session discussed case studies, documentation tools, standards for minimum data to annotate, and algorithms that are being used to validate biomarker research results. This included discussions on supporting variability assessment for experimental methodology, the characterization of performance over time for methods, and the confidence associated with experimental results.

Summary of Discussion

- Recommendation to use PRoBE checklist for sample selection and experimental design: The Prospective enrollment, Retrospective random selection of cases and controls, Blinded specimen handling, Evaluation (JNCI 2008 Pepe et al) provides a checklist for the design and control of a biomarker study.

- Availability of large sample collections: The availability of high quality, specimens is critical to supporting validation and analysis of biomarker research results. The Women's Health Initiative (WHI) study, developed EDRN reference sets that were used as a baseline EDRN as a whole has found the development of specimen reference sets to be critical to improving the quality in designing biomarker research studies.
- Integration of data and knowledge. Biomarker research needs to move towards integrating all existing knowledge in data analysis. This requires ensuring that data is both accessible and of high quality.
- Measuring effectiveness. Quality by design requires the ability to measure effectiveness of the techniques that are used. The capture and analysis of these techniques are important for improving the quality of biomarker research and analyzing the research results.

Session 4: Data Integration and Biological Networks

The use of multiple biotechnologies introduces the challenge of integrating across multiple data types. Frequently, the incorporation of existing knowledge in the interpretation of an experiment implies organizing and evaluating results in accordance with existing biological network data. This session included case studies and algorithmic developments addressing these two challenges; it also identified areas of research that are most likely to pay dividends in cancer biomarker research.

Summary of Discussion

- Data and knowledge integration: Data needs to be brought together to integrate knowledge from multiple, disparate resources. This requires the development of standards, computing infrastructures, and methodologies that will enable data integration for biomarker research.
- Pathway and network integration. Pathway and network integration helps to discover driver mutations; mutations in signaling networks may be clinically informative.
- Combined analysis. Combined analysis of genomic and proteomics data to discover which variants are actually translated to proteins or alter protein function.
- Challenge: data in protein – protein interaction databases are variable quality.

Session 5: Biomarker Informatics Infrastructure

This session focused on discussing progress and challenges in developing large-scale informatics systems for cancer research and other efforts. This included large-scale data management, sharing, and discovery systems that support researchers for biomarker discovery, analysis and validation. It also included the development of data and metadata standards to enable data capture, sharing and analysis. Finally, it discussed the future in terms of developing common repositories and shared analysis architectures and infrastructures to address upcoming challenges biomarker research.

Summary of Discussion

- Biomarker Data Infrastructures: Data infrastructures are needed to support collection and integration of a wide collection of data types, ensuring longevity and sharing of data from biomarker research. The EDRN has invested in developing the EDRN Knowledge System that provides the capture and management of data from EDRN discovery and validation centers.
- High quality biomarker databases. Well curated, high quality databases should be captured within biomarker data infrastructures. These databases should work to employ standards in the capture of biomarker data.
- Common ontologies and standards. Use information models, data elements, controlled terminologies to capture data is critical to improving the annotation, search, and usability of the data.
- Open source. Open source software is important to promoting standards and moving towards shared and reusable data infrastructures. The National Cancer Institute Informatics Program (NCIP) is promoting new models of collaboration, open development (NCIP GitHub) and making data accessible. The EDRN and NASA/JPL have a strong connection to the Apache Software Foundation where they are sharing common data infrastructures. Open source communities explored include GitHub, Apache, and HubZero.
- Cloud data repositories. The NCIP is exploring cancer genomics clouds as an environment for data storage and computation, preloaded with public data (e.g., TCGA). These clouds serve as data repositories and can be integrated into a biomarker data infrastructure.
- Semantic workflows. Semantic workflows are critical to the generation and use of scientific data. They provide support to

track metadata and data provenance including how different variables (algorithms) affect the reproducibility and reuse of data.

- Integration of data and computing to support biomarker research. In addition to developing the biomarker data infrastructure, there are challenges with bringing together the tools, data and infrastructure. Visualization, for example, is critical to supporting analysis, particularly for massive data.
- Limited statistical methods. Current statistical methods are not adequate to deal with the increase of data. New methods that can be used to improve scientific inferences by reducing the uncertainty in analysis are needed as the data increases.
- Limited training/expertise in the use of computational capabilities and techniques. Many scientists lack expertise and training in the effective use of computing technologies to support biomarker research. This can affect the adoption of emerging informatics capabilities and limit their use in biomarker discovery and validation. Efforts are needed to better train scientists in learning and using informatics.
- Data quality, integration & reproducibility. A major theme across the entire workshop was the quality, integration and reproducibility of the data. Continued investments in standards, improved statistical methods, and computing infrastructures are needed to support biomarker research at a national scale.

Acknowledgements

The Program Committee:

Dan Crichton, NASA/JPL
David Elashoff, UCLA
Chris Kinsinger, NCI
Marc Lenburg, Boston University
Mervi Heiskanen, NCI
Christos Patriotis, NCI
Sudhir Srivastava, NCI
David Tabb, Vanderbilt University
Mark Thornquist, Fred Hutchinson Cancer Center
Zhen Zhang, Johns Hopkins Medical Institute

Local Organizing Committee:

Ferne Griffin, NASA/JPL
Heather Kincaid, NASA/JPL
Elizabeth Reyes, NASA/JPL

Appendix A – List of Registrants

Mulkah Ajagun-Ogunleye, B.S., M.S.
Kampala International University-Western Campus
Ishaka Uganda
mulka4real@yahoo.com

Michelle A. Anderson, M.D.
University of Michigan
maaander@med.umich.edu

Kristen Anton, M.S.
Geisel School of Medicine at Dartmouth
kristen.anton@dartmouth.edu

Keith Baggerly, Ph.D.
MD Anderson Cancer Center
kabagger@mdanderson.org

Kathleen M. Beckingham, B.A., M.A., Ph. D.
Rice University
kate@rice.edu

Dean E Brenner, M.D.
University of Michigan
dbrenner@umich.edu

C. Titus Brown, Ph.D.
Michigan State University
ctb@msu.edu

Andy Chang
Western University of Health Sciences
aychangwesternu@gmail.com

Maureen Colbert, M.S.
Geisel School of Medicine at Dartmouth
maureen.colbert@dartmouth.edu

Isabel Coutinho
iclcoutinho@gmail.com

Michael Craige, Ph.D.
Rutgers University and UMDNJ
craigemi@umdnj.edu

Dan Crichton, M.S.
NASA/JPL
dan.crichton@jpl.nasa.gov

George Djorgovski, Ph.D.
Caltech
george@astro.caltech.edu

Phil Douglass, Ph.D.
Industry
phil_douglass@agilent.com

Lewei Duan. M.S.
UCLA
lduan@mednet.ucla.edu

Rafael Dugarte
Integrated Analysis Inc
rdugarte@i-a-inc.com

David Elashoff, Ph.D.
UCLA
delashoff@mednet.ucla.edu

Michael Elashoff, Ph.D.
Celmatix
elashoff@gmail.com

Jeff Falk, Ph.D.
Industry
jeffrey.falk@mdxhealth.com

Gaddy Getz, Ph.D.
Broad Institue of MIT and Harvard, MGH
gadgetz@broadinstitute.org

Yolanda Gil, Ph.D.
University of Southern California/Information Sciences Institute
gil@isi.edu

Annalisa Gnoleba, M.S.A.
National Cancer Institute
annalisa.gnoleba@nih.gov

Tristan Grogan, M.S.
UCLA Department of Medicine
Tgrogan@mednet.ucla.edu

Daniel Gusenleitner, MSc. MSc.
Boston University
gusef@bu.edu

Andrew Hart, M.S.
NASA/JPL
andrew.f.hart@jpl.nasa.gov

Mervi Heiskanen, Ph.D.
NCI
heiskame@mail.nih.gov

Ying Huang, Ph.D.
Western University of Health Sciences
yhuang@westernu.edu

Trey Ideker, Ph.D.
UC San Diego School of Medicine
trey.ideker@gmail.com

Libin Jia, M.D.
NCI
libinj@mail.nih.gov

Yong Jiang, Ph.D.
AprexBio
yjiang@aprexbio.com

John Joern, Ph.D.
Meso Scale Discovery
jjoern@mesoscale.com

Evan Johnson, Ph.D.
Boston University School of Medicine
wej@bu.edu

Michael Joyce
JPL
michael.j.joyce@jpl.nasa.gov

Heather Kincaid, B.S.
NASA/JPL
heather.kincaid@jpl.nasa.gov

Christopher Kinsinger, Ph.D.
National Cancer Institute, NIH
kinsingc@mail.nih.gov

Karl Krueger, Ph.D.
NIH
kruegerk@mail.nih.gov

Sudeep Kumar, Ph.D.
Industry
skumar@mesoscale.com

Chung Lee, Ph.D.
UC Irvine
c-lee7@northwestern.edu

Ming-Chou Lee, Ph.D.
MDxHealth
ming-chou.lee@mdxhealth.com

Marc Lenburg, Ph.D.
University
mlenburg@bu.edu

Danni Li, Ph.D.
University of Minnesota
dannili@umn.edu

Dave Li, M.D.
Wako Diagnostics
daveli@wakousa.com

Yaojun Li, Ph.D.
The Methodist Hospital Research Institute
yli4@tmhs.org

Mandy Liu, M.S.
Western University of Health Sciences
mmliu@westernu.edu

Dan Masys, M.D.
University of Washington
dmasys@u.washington.edu

Raja Mazumder, Ph.D.
George Washington University
mazumder@gwu.edu

Shannon McWeeney
OHSU Knight Cancer Institute

Aregay, Mengsteab Fantahun, M.Sc.
www.uhasselt.be
mycs.zab@gmail.com

Christine Micheel, Ph.D.
Vanderbilt-Ingram Cancer Center
christine.micheel@vanderbilt.edu

Vishnu Mishra, Ph.D.
MDxHealth
vishnu.mishra@mdxhealth.com

Akhilesh Pandey, M.D., Ph.D.
Johns Hopkins University
pandey@jhmi.edu

Christos Patriotis, Ph.D.
National Cancer Institute, NIH
patriotisc@mail.nih.gov

Samuel Payne, Ph. D.
Pacific Northwest National Lab
samuel.payne@pnnl.gov

Suzanna Reid, Ph.D.
Fred Hutchinson Cancer Research Center
sreid@fhcrc.org

Juri Reimand, Ph.D.
University of Toronto
Juri.Reimand@utoronto.ca

Robert Rheaume, B.S.
Geisal School of Medicine at Dartmouth
Robert.L.Rheaume@dartmouth.edu

Markus Riester, Ph.D.
Dana-Farber Cancer Institute
markus@jimmy.harvard.edu

Mark A Rubin, M.D.
Weill Cornell Medical College
rubinma@med.cornell.edu

Ashish Saini
Deakin University
asain@deakin.edu.au

Gautam Saxena
Integrated Analysis Inc
gsaxena888@gmail.com

Dejun Shen, M.D., Ph.D.
University of Alabama at Birmingham
shend@uab.edu

Leming Shi, Ph.D.
leming.shi@gmail.com

Steven S. Smith, Ph.D.
Beckman Res. inst. of the City of Hope
ssmith@coh.org

Bharat Somkuwar
Intitute of Bioresources and Sustainable Development,
Takyelpat Industrial Area, Imphal
bgs.ibsd@nic.in

Sudhir Srivastava, Ph.D.
NCI
srivasts@mail.nih.gov

Deanna Stelling, B.A.
EDRN, FHCRC
dstellin@fhcrc.org

David L. Tabb, Ph.D.
Vanderbilt University Medical School
david.l.tabb@vanderbilt.edu

Mark Thornquist, Ph.D.
Fred Hutchinson Cancer Research Center
mthornqu@fhcrc.org

Abhilash Venugopalan, M.Sc
National Cancer Institute
venugopalana@mail.nih.gov

Rishi Verma. B.S.
NASA Jet Propulsion Laboratory
Rishi.Verma@jpl.nasa.gov

Garrick Wallstrom, Ph.D.
Arizona State University
Garrick.Wallstrom@asu.edu

Zhijun Wang, Ph.D.
Western University of Health Sciences
zwang@westernu.edu

Garba Yahaya
AMINU KANO TEACHING HOSPITAL KANO NIGERIA
garbayahaya@gmail.com

Hiro Yamada, M.S.
Wako Life Sciences, Inc.
h-yamada@wakousa.com

Tzu-Yi Yang
Purdue University
yang127@purdue.edu

Wei Yang
wei.yang@cshs.org

Xiaoqing Yang, M.D., M.P.H.
Industry
quinnie.yang@abbott.com

Zhen Zhang, Ph.D.
Johns Hopkins University
zzhang7@jhmi.edu

Appendix B – Workshop Agenda

Cancer Biomarker Bioinformatics Workshop
California Institute of Technology
Beckman Institute
Pasadena, California
Monday, April 29 – Wednesday, May 1, 2013

Monday, April 29, 2013

8:00 a.m. – 8:30 a.m.	**Registration and Light Refreshments**
8:30 a.m. – 8:45 a.m.	Welcome Dan Crichton, NASA Jet Propulsion Laboratory
8:45 a.m. – 9:15 a.m.	Bioinformatics Tailored to Meet Cancer Biomarker Research Needs Sudhir Srivastava, Ph.D., MPH, MS, National Cancer Institute
9:15 a.m. – 10:00 a.m.	**Keynote –** Blueprint or Parts List? The value of biomarkers for 21st century discovery science and for cancer prevention, diagnosis and treatment Dan Masys, M.D., University of Washington School of Medicine
10:00 a.m. – 10:15 a.m.	**Break**

Session 1: Informatics in Biomarker Discovery

Moderator:

- Christos Patriotis,Ph.D., National Cancer Institute

10:15 a.m. – 10:45 a.m.	Developing curated databases for cancer biomarkers Akhilesh Pandey, Ph.D., Johns Hopkins School of Medicine
10:45 a.m. – 11:15 a.m.	My Cancer Genome: Knowledge Resource to Support Genetically Informed Cancer Medicine Christine Micheel, Ph.D., Vanderbilt University
11:15 a.m. – 11:45 a.m.	A comparative meta-analysis of prognostic gene signatures for late-stage ovarian cancer Markus Riester, Ph.D., Dana-Farber Cancer Center/Harvard School of Public Health

11:45 a.m. – 12:15 p.m.	Panel Discussion
12:15 p.m. – 1:45 p.m.	**Lunch** (on your own)

Session 2: Informatics in Biomarker Validation

Moderators:

- Marc Lenburg, Ph.D.,
  Boston University School of Medicine
- David Elashoff, Ph.D.,
  University of California, Los Angeles

1:45 p.m. – 2:15 p.m.	Development and Validation of a Blood-Based Gene Expression Diagnostic for Coronary Disease Michael Elashoff, Ph.D., Celmatix
2:15 p.m .– 2:45 p.m.	Data preprocessing and integration strategies for reproducible multi-omic biomarker discovery and validation Evan Johnson, Ph.D., Boston University School of Medicine
2:45 p.m. – 3:00 p.m.	**Break**
3:00 p.m. – 3:30 p.m.	Lessons Learned from the FDA-led MAQC-II Project on the Development and Validation of Microarray-based Predictive Models Leming Shi, Ph.D., School of Pharmacy, Fudan University, Shanghai, China and National Center for Toxicological Research, US Food and Drug
3:30 p.m. – 4:00 p.m.	Biomarker Discovery for Heterogeneous Diseases Garrick Wallstrom, Ph.D., Arizona State University

4:00 p.m. – 4:30 p.m.	Panel Discussion
4:30 p.m. – 4:45 p.m.	**Break**
4:45 p.m. – 5:45 p.m.	**Poster Session**

Tuesday, April 30, 2013

8:00 a.m. – 8:30 a.m.	**Light Refreshments**
8:30 a.m. – 9:15 a.m.	**Keynote –** Cancer Genome Analysis: Mutational patterns, evolution and significance across cancer Gaddy Getz, Ph.D., Broad Institute
9:15 a.m. – 9:30 a.m.	**Break**

Session 3: Quality by Design in Biomarker Research

Moderators:

- Mark Thornquist, Ph.D.,
  Fred Hutchinson Cancer Research Center
- Zhen Zhang, Ph.D.,
  Johns Hopkins Medicine

9:30 a.m. – 10:00 a.m.	Selecting Clinical Specimens for Biomarker Evaluation Margaret Pepe, Ph.D., Fred Hutchinson Cancer Research Center
10:00 a.m. – 10:30 a.m.	When is Reproducibility an Ethical Issue? Forensic Bioinformatics in High-Throughput Biology Keith Baggerly, Ph.D., MD Anderson Cancer Center
10:30 a.m. – 10:45 a.m.	**Break**

10:45 a.m. – 11:15 a.m.	Design of Clinical Proteomic Studies – Bridge the Gap from Discovery to Clinical Applications Zhen Zhang, Ph.D., Johns Hopkins Medicine
11:15 a.m. – 11:45 a.m.	QuaMeter: Assessing longitudinal variability in LC-MS/MS proteomics David Tabb, Ph.D., Vanderbilt University
11:45 a.m. – 12:15 p.m.	Panel Discussion
12:15 p.m.– 1:45 p.m.	**Lunch** (on your own)

Session 4: Data Integration and Biological Networks

Moderators:

- David Tabb, Ph.D.,
  Vanderbilt University
- Chris Kinsinger, Ph.D.,
  National Cancer Institute

1:45 p.m. – 2:15 p.m.	Biomarkers as networks, not individual genes and proteins Trey Ideker, Ph.D., UC San Diego School of Medicine
2:15 p.m. – 2:45 p.m.	Discovering Cancer Driver Mutations in Cell Signaling Networks Jüri Reimand, Ph.D., University of Toronto
2:45 p.m. – 3:00 p.m.	**Break**
3:00 p.m. – 3:30 p.m.	Using personalized genome information in proteomics analyses Sam Payne, Ph.D., Pacific Northwest National Laboratory

3:30 p.m. – 4:00 p.m.	Analysis and curation of short-read data from CGHub/TCGA using HIVE-XLD: Evaluation of non-synonymous variation in cancer cases and controls Raja Mazumder, Ph.D., George Washington University
4:00 p.m. – 4:30 p.m.	Panel Discussion

Wednesday, May 1, 2013

8:00 a.m. – 8:30 a.m.	**Light Refreshments**
8:30 a.m. – 9:15 a.m.	**Keynote –** Virtual Observatory, Astroinformatics, and Data-Rich Astronomy in the 21st Century George Djorgovski, Ph.D., Caltech
9:15 a.m. – 9:30 a.m.	**Break**

Session 5: Biomarker Informatics Infrastructure

Moderators:

- Dan Crichton, NASA Jet Propulsion Laboratory
- Mervi Heiskanen, Ph.D., National Cancer Institute

9:30 a.m. – 10:00 a.m.	Informatics Support for Precision Medicine George Komatsoulis, Ph.D., National Cancer Institute
10:00 a.m. – 10:30 a.m.	The EDRN Knowledge System: A Big Data Infrastructure for Cancer Biomarker Research Dan Crichton, NASA Jet Propulsion Laboratory
10:30 a.m. – 10:45 a.m.	**Break**
10:45 a.m. – 11:15 a.m.	Fast and Efficient Streaming Variant Calling in Resequencing Data C. Titus Brown, Ph.D., Michigan State University
11:15 a.m. – 11:45 a.m.	Intelligent Assistance to Disseminate Best Practices and Accelerate Discoveries in Cancer Omics Shannon McWeeney, Ph.D., OHSU Knight Cancer Institute
11:45 a.m. – 12:15 p.m.	Panel Discussion
12:15 p.m. – 12:45 p.m.	Recommendations and Other Needs Moderator: Mervi Heiskanen, Ph.D., National Cancer Institute