National Institute of
Environmental Health Sciences

Workshop on Developing a Data Science Competent EHS Workforce

August 14-15, 2018

National Institutes of Health • U.S. Department of Health and Human Services

Table of Contents

Background and Scientific Need ---------- 2
Workshop Objectives and Framework ---------- 2
Agenda---------- 3
Biographies ---------- 6
Participant List ----------21

National Institute of Environmental Health Sciences Workshop on Developing a Data Science Competent Environmental Health Science Workforce

August 14 – 15, 2018
NIEHS Building 101, Rodbell Auditorium
111 TW Alexander Drive, Research Triangle Park, NC
NIEHS/DERT Planning Committee: Carol Shreffler, Jenny Collins, Danielle Carlin, Mike Humble, Chris Duncan

Background and Scientific Need:
With rapidly developing technology and more efficient data collection procedures, environmental health scientists are now collecting vast amounts of data. These data sets, termed "Big Data", can be large, complex, multidimensional, diverse, and are often generated using new technologies. They are associated with basic, translational, clinical, social, behavioral, environmental, or informatics research questions. Such data types may include imaging, phenotypic, genotypic, molecular, clinical, behavioral, environmental, and many other types of biological and biomedical data. Data science has emerged from its roots in applied statistics, analytics, and bioinformatics as a new area of research to meet the challenges in sharing, accessing, analyzing, and interpreting big data. "Data science", defined as the extraction of useful knowledge directly from data through a process of discovery, or of hypothesis formulation and hypothesis testing (https://bigdatawg.nist.gov/), refers to the management and execution of the end-to end data processes.

National Institutes of Health (NIH) has made early efforts to address the gap between the needed and existing biomedical data science skills through investments in training and education as part of the Big Data to Knowledge (BD2K) Initiative. The programs and Funding Opportunity Announcements released had two main, and somewhat separable, goals: 1) improving big data skills of biomedical scientists; and 2) increasing the number biomedical data scientists. These NIH-wide efforts were not domain-specific and were intended to develop resources which could benefit all NIH institutes.

This workshop will bring together experts from relevant research disciplines to examine existing data science and environmental health science (EHS) resources (trainee pipelines, mentors, research), identify how these resources can address EHS-specific training goals in data science, and make recommendations for National Institute of Environmental Health Sciences (NIEHS) in data science training.

Workshop Objectives and Framework:
The overarching goal of the workshop is to develop an overall strategy to build a data science competent EHS workforce. The workshop will be organized into three major sessions. The first session will be designed to understand the current state of data science in the EHS domain as it relates to training, and through the evaluation of representative scientific 'use cases' (nominated by The Division of Extramural Research and Training (DERT) program branches), current limitations for data science training in EHS will be identified. The second session will examine existing training resources relevant to the intersection of EHS and data science and will relate EHS training goals to the accomplishments of BD2K. The final session will formulate how to build EHS training in data science and will be a discussion with participant input into questions formulated before the workshop and during the planning.

Agenda

Workshop on Developing a Data Science Competent EHS Workforce

August 14-15, 2018

NIEHS Building 101, Rodbell Auditorium, Research Triangle Park, N.C.

AGENDA

Tuesday, August 14

8:30 – 8:45 a.m. **Opening Remarks and Purpose of the Workshop**
Carol Shreffler, Ph.D., NIEHS
Linda Birnbaum, Ph.D., Director, NIEHS and NTP

SESSION ONE: Understanding the Current State of Data Science in the EHS Domain As It Relates to Training

Goal: Through the evaluation of representative use cases, identify current limitations and rate-limiting steps for data science training in EHS. Presentations should include a high-level description of the persons or things involved, a sequence of actions and process flow, key competencies, and key obstacles. Discussion will entail a deep dive into use cases to identify relevant issues and challenges involved in solving data-intensive EHS research questions.

Moderator: *Charles Schmitt, Ph.D., NIEHS Office of Data Science*

8:45 – 8:50 a.m. **Introduction to Session One**
Charles Schmitt, Ph.D., NIEHS Office of Data Science

8:50 – 9:20 a.m. **Use Case 1: Tools, Skills, and Training for Quantitative Trait Locus Mapping**
Daniel Gatti, Ph.D., The Jackson Laboratory

9:20 – 9:50 a.m. **Use Case 2: Targeted Learning for Exposome Science**
Alan Hubbard, Ph.D., University of California, Berkeley

9:50 – 10:05 a.m. **Break**

10:05 – 10:35 a.m. **Use Case 3: Mapping What Matters: The Promise of Geospatial Health Informatics**
Marie Lynn Miranda, Ph.D., Rice University

10:35 – 11:05 a.m. **Use Case 4: Maximizing the Impact of Exposure Data in Children's Environmental Health Research**
Susan Teitelbaum, Ph.D., Icahn School of Medicine at Mount Sinai

11:05 a.m. – noon **Panel Discussion**

Noon – 1:00 p.m. **Lunch**

1:00 – 1:15 p.m. **Overview of Big Data Science Training**
Carol Shreffler, Ph.D., NIEHS

SESSION TWO: Relating BD2K Accomplishments to EHS Training Goals

Goal: Examine existing training resources relevant to the intersection of EHS and data science, including existing data science training resources (e.g., BD2K).

Moderator: *Amy Herring, Sc.D., Duke University*

1:15 – 1:20 p.m.	**Introduction to Session Two** *Amy Herring, Sc.D., Duke University*
1:20 – 1:50 p.m.	**Applying the FAIR Principles to Online Data Science Training Resources** *John Van Horn, Ph.D., University of Southern California*
1:50 – 2:20 p.m.	**BD2K Training at the Harvard T.H. Chan School of Public Health** *John Quackenbush, Ph.D., Harvard T.H. Chan School of Public Health and Dana-Farber Cancer Institute*
2:20 – 2:35 p.m.	**Break**
2:35 – 3:05 p.m.	**The Physician Data Scientist: An Unexpected Journey** *Jonathan Chen, M.D., Ph.D., Stanford University*
3:05 – 3:35 p.m.	**The Training of Next Generation Data Scientists** *Lana Garmire, Ph.D., University of Hawaii*
3:35 – 4:05 p.m.	**NLM Training in Bioinformatics – Biomedical Data Science Training for Diverse Backgrounds and Career Paths** *Mark Craven, Ph.D., and Chris Bradfield, Ph.D., University of Wisconsin-Madison*
4:05 – 5:00 p.m.	**Panel Discussion**
5:00 p.m.	**Adjourn Day One**

Wednesday, August 15

8:30 – 8:35 a.m. **Day Two Opening Remarks: Chris Duncan, Ph.D., NIEHS**

SESSION THREE: Building EHS Training in Big Data Science

Goal: Examine existing NIEHS training resources (e.g., trainee pipelines) relevant to data science and discuss the next steps to increase training emphasis in big data. This may include both training for quantitative big data scientist career paths and needed big data skills for biomedical scientists in other environmental health programs. Identify the limitations and rate-limiting steps in data science training in EHS and make recommendations for priority areas.

Moderator: *Cheryl Walker, Ph.D., Baylor College of Medicine*

8:35 – 8:40 a.m.	**Introduction to Session Three** *Cheryl Walker, Ph.D., Baylor College of Medicine*
8:40 – 9:10 a.m.	**Experiences With Environmental Bioinformatics Training** *Fred Wright, Ph.D., North Carolina State University*
9:10 – 9:40 a.m.	**Data Science Training in Epidemiology** *Jim Gauderman, Ph.D., University of Southern California*
9:40 – 9:55 a.m.	**Break**
9:55 – 10:25 a.m.	**Building Multidisciplinary Environmental Health Data Science Training Resources** *Chirag Patel, Ph.D., Harvard University*
10:25 a.m. – 12:20 p.m.	**Panel Discussion on the Training Intersection of Data Science and Environmental Health Sciences** **Additional Panelists:** *Wesley Gray, Ph.D., Southern University and A&M College* *Ronald Hines, Ph.D., U.S. Environmental Protection Agency* *Jeanette Stingone, Ph.D., Icahn School of Medicine at Mount Sinai*
12:20 – 12:30 p.m.	**Meeting Wrap-Up and Closing Remarks** *Gwen Collman, Ph.D., Director, NIEHS/DERT*
12:30 p.m.	**Adjourn Meeting**

Biographies

Linda Birnbaum
National Institute of Environmental Health Sciences

Linda Birnbaum, Ph.D., is director of the National Institute of Environmental Health Sciences (NIEHS) of the National Institutes of Health (NIH), and the National Toxicology Program (NTP). A board-certified toxicologist, Birnbaum has served as a federal scientist for over 37 years. She has received many awards and recognitions, including the North Carolina Award in Science, the Women in Toxicology Elsevier Mentoring Award, the Society of Toxicology Public Communications Award, the U.S. Environmental Protection Agency's (EPA) Health Science Achievement Award and Diversity Leadership Award, the National Center for Women's 2012 Health Policy Hero Award, the Breast Cancer Fund Heroes Award, and 14 Science and Technology Achievement Awards, which reflect the recommendations of EPA's external Science Advisory Board, for specific publications. Birnbaum was also elected to the Institute of Medicine of the National Academies and received an honorary degree from Ben-Gurion University in Israel.

Birnbaum is a former president of the Society of Toxicology, the largest professional organization of toxicologists in the world; former chair of the Division of Toxicology at the American Society of Pharmacology and Therapeutics; and former vice president of the American Aging Association. She is the author of more than 800 peer-reviewed publications, book chapters, and reports. She is also an adjunct professor at several universities, including the University of North Carolina at Chapel Hill and Duke University.

A native of New Jersey, Birnbaum received her master's and doctoral degrees in microbiology from the University of Illinois at Urbana-Champaign.

Christopher Bradfield
University of Wisconsin-Madison

Christopher Bradfield, Ph.D., is a professor of oncology and the director of the Molecular and Environmental Toxicology Center at the University of Wisconsin (UW)-Madison. He received his doctorate in toxicology/nutrition from the University of California, Berkeley and conducted postdoctoral research at UW-Madison. His laboratory has a long-standing interest in understanding how the chemical and physical environments influence human health; specifically, understanding the toxicology of 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) and related pollutants, as well as studies into signal transduction pathways. He has trained over a dozen graduate students and postdocs, many of whom are tenure-track faculty or leaders in government and industry.

Danielle Carlin
National Institute of Environmental Health Sciences

Danielle Carlin, Ph.D., is a program administrator with the Superfund Research Program (SRP). She provides guidance and advice to grantees applying for SRP P42 Center grants and serves as the lead liaison between SRP trainees and the various training opportunities offered by SRP. She also oversees the xenobiotic metabolism and asbestos grant portfolios (e.g., R01s). Her current research interests include chemical mixtures, combined exposures, metals, asbestos, and xenobiotic metabolism.

Prior to her current position, Carlin was a postdoctoral researcher for four years at the University of North Carolina: two years within the Eshelman School of Pharmacy Division of Molecular Pharmaceutics studying aerosolized drugs/vaccines for treatment and prevention of tuberculosis, and two years within the Curriculum in Toxicology conducting her research at the EPA in Research Triangle Park, North Carolina, where she studied the toxicological effects of exposure to Libby amphibole asbestos in the rat model. Her areas of expertise include cardiopulmonary/reproductive physiology and inhalation toxicology/pharmacology. She received her doctorate from the Kansas State University College of Veterinary Medicine, Department of Anatomy and Physiology. She also has a Bachelor of Science and Master of Science in animal science from New Mexico State University.

Jonathan Chen
Stanford Department of Medicine

Jonathan H. Chen, M.D., Ph.D., practices medicine for the concrete rewards of caring for real people and to inspire research focused on discovering and distributing the latent knowledge embedded in clinical data.

Chen co-founded a company to translate his computer science graduate work into an expert system for organic chemistry, with applications from drug discovery to an education tool for students around the world. To gain perspective tackling societal problems in health care, he completed training in internal medicine and a research fellowship in medical informatics. He has published influential work in the New England Journal of Medicine, JAMA, JAMA Internal Medicine, Bioinformatics, Journal of Chemical Information and Modeling, and the Journal of the American Medical Informatics Associations, and has received awards and recognition from the NIH Big Data 2 Knowledge (BD2K) initiative, the National Library of Medicine (NLM), the American Medical Informatics Association, Yearbook of Medical Informatics, and the American College of Physicians, among others.

In the face of ever escalating complexity in medicine, informatics solutions are the only credible approach to systematically address challenges in health care. Tapping into real-world clinical data like electronic medical records with machine learning and data analytics will reveal the community's latent knowledge in a reproducible form. Delivering this back to clinicians, patients, and health care systems as clinical decision support will uniquely close the loop on a

continuously learning health system. Chen's group seeks to empower individuals with the collective experience of the many, combining human and artificial intelligence approaches that will deliver better care than either can do alone.

Jennifer Collins
National Institute of Environmental Health Sciences

Jennifer Collins, M.R., is a health specialist in the NIEHS Division of Extramural Research and Training. She is responsible for coordinating activities related to programmatic oversight functions for the Nanotechnology Environmental Health and Safety (Nano EHS) program, the Children's Health Exposure Analysis Resource (CHEAR), and the Outstanding New Environmental Health Sciences (ONES) program. In addition, she is the program coordinator for the NIEHS Revolutionizing Innovative, Visionary Environmental Health Research (RIVER) program. This program is designed to support the majority of an individual's independent research program in the environmental health sciences by providing intellectual and administrative freedom, as well as sustained support for the pursuit of research in novel directions to achieve greater impacts. Collins is also a working group member for the NIH Common Fund High-Risk, High-Reward program which supports exceptionally creative scientists pursuing highly innovative research with the potential for broad impact in biomedical or behavioral science.

Prior to joining the Division of Extramural Research and Training in 2008, Collins spent 10 years in the Division of Intramural Research at NIEHS, primarily in the Microarray Core facility. As a biologist in this group, she was responsible for coordinating collaborations between the core facility and intramural researchers and for conducting the initial analyses of all gene expression data generated in the lab. Collins has a Bachelor of Science in biological sciences and a master's degree in functional genomics from North Carolina State University.

Gwen Collman
National Institute of Environmental Health Sciences

Gwen Collman, Ph.D., is director of the NIEHS Division of Extramural Research and Training where she leads approximately 60 professional staff in areas of scientific program administration, peer review, and the management and administration of about 1,500 active grants each year. She directs scientific activities across the field of environmental health sciences, including basic sciences (i.e., DNA repair, epigenetics, environmental genomics), organ-specific toxicology (i.e., reproductive, neurotoxicology, respiratory), public health-related programs (i.e., environmental epidemiology, environmental public health), and training and career development. She also oversees the implementation of the Superfund Research Program and the Worker Training Program.

Prior to her current role, Collman served in program development and management, beginning in 1992 as a member, then as chief of the Susceptibility and Population Health Branch. During this time, she directed research on the role of genetic and environmental factors on the

development of human disease, from animal models of genetic susceptibility to population studies focusing on etiology and intervention. She was responsible for building the NIEHS grant portfolio in environmental and molecular epidemiology and developed several complex multidisciplinary research programs. These include the NIEHS Breast Cancer and the Environment Research Centers Program, the NIEHS/EPA Centers for Children's Environmental Health and Disease Prevention, and the Genes, Environment, and Health Initiative. Also, under her guidance, a team created a vision for the Partnerships for Environmental Public Health program for the next decade.

In recognition of her achievements, she is the recipient of numerous NIEHS Merit Awards, two NIH Director's Awards, and the HHS Secretary's Award for Distinguished Service. Collman received a doctorate in environmental epidemiology from the University of North Carolina School of Public Health where she was awarded the 2009 H.A. Tyroler Distinguished Alumni Award.

Mark Craven
University of Wisconsin-Madison

Mark Craven, Ph.D., is a professor in the department of biostatistics and medical informatics at the University of Wisconsin, and an affiliate faculty member in the department of computer sciences. He is the director of the Center for Predictive Computational Phenotyping, one of the NIH Centers of Excellence for Big Data Computing. He is also the director of the NIH/NLM-funded Computation and Informatics in Biology and Medicine (CIBM) Training Program, and a member of the Institute for Clinical and Translational Research, the Carbone Cancer Center, and the Genome Center of Wisconsin.

The focus of his research program is on developing and applying machine-learning methods to the problems of inferring models of, and reasoning about, networks of interactions among genes, proteins, clinical and environmental factors, and phenotypes of interest. Current projects in his group are focused on (i) learning models to assess risk for clinical events such as asthma exacerbations and post-hospitalization venothromboembolisms (VTEs) using electronic health records and genetic data; (ii) uncovering the intracellular networks involved in host-virus interactions, breast cancer susceptibility, and stress responses; (iii) learning models of viral genotype-phenotype associations; (iv) modeling, aligning, and classifying complex biomedical time series; and (v) automatically extracting molecular interactions and events from the biomedical literature.

Christopher Duncan
National Institute of Environmental Health Sciences

Christopher Duncan, Ph.D., is a program administrator in the Genes, Environment, and Health Branch at NIEHS. He is responsible for coordinating activities related to data science in the NIEHS Division of Extramural Research and Training. Duncan is involved in developing NIEHS extramural initiatives in data science and facilitating data management and sharing within

NIEHS-funded collaborative research programs. His scientific portfolio includes data resources, such as the Comparative Toxicogenomics Database (CTD) and the Toxicant Exposures and Responses by Genomic and Epigenomic Regulators of Transcription (TaRGET) Data Coordination Center. He has also contributed to cross-disciplinary initiatives including the NIEHS Mouse Methylome Project.

Duncan received his doctorate in pathology from Duke University for his work characterizing genomic alterations in human brain tumors. He conducted postdoctoral research at the University of Southern California, investigating the roles of DNA methylation and chromatin structure in transcriptional regulation. Prior to joining the Division of Extramural Research and Training, Duncan completed a cross-divisional postdoctoral fellowship at NIEHS with research focusing on the characterization of epigenomic landscapes through integrative data analysis.

Lana Garmire
University of Hawaii

Lana Garmire, Ph.D., is an associate professor in translational bioinformatics at the University of Hawaii Cancer Center, leading a multidisciplinary team of computational and experimental human genomics. Beginning with an NIH/BD2K K01 award (2014-2019), she went on to obtain several competitive federal grant awards of over $7.8 million as the principal investigator, including an NIH/National Institute of General Medical Sciences (NIGMS) P20 COBRE (2014-2019) and two concurrent NIH R01 grant awards from the National Institute of Child Health and Human Development (NICHD) (2016-2021) and the NLM (2016-2020).

Garmire obtained her doctorate from the University of California, Berkeley and completed postdoctoral training at the University of California, San Diego. She has published over 40 papers in high-quality journals including Cell and Nature, and has mentored over 30 M.D. fellows, postdocs, graduate students, and undergraduates of various academic backgrounds. She is an associate editor of BMC Bioinformatics and guest editor of PLoS Computational Biology.

Daniel Gatti
Jackson Laboratory

Daniel Gatti, Ph.D., obtained his doctorate in environmental sciences and engineering from the University of North Carolina at Chapel Hill, where he studied the effect of genetics on transcriptional regulation in the liver. After graduate school, he moved to The Jackson Laboratory as a bioinformatics analyst in the laboratory of Gary Churchill. His work focuses on finding associations between phenotypes and genotypes in complex mouse crosses.

Gatti is a co-author of over 35 publications. In the past several years, he has developed an interest in helping biologists learn computing skills and has become an instructor with Software Carpentry, a nonprofit organization that teaches programming and data sciences skills. In fall

2018, he will become the first professor of computer science at the College of the Atlantic in Bar Harbor, Maine.

W. James Gauderman
University of Southern California

W. James Gauderman, Ph.D., is professor and chief of the Biostatistics Division in the University of Southern California Department of Preventive Medicine. He is director of the Biostatistics Facility Core of the Southern California Environmental Health Sciences Center (SCEHSC), co-principal investigator of a program project grant to develop statistical methods for integrative genomics, and co-director of an NIEHS training grant in environmental genomics. He has conducted methodological research related to the design and analysis of genetic epidemiology studies for over 25 years, with particular focus on gene-environment (GxE) interaction. He has also developed user-friendly software that implements these methods, including Quanto for power calculations and GxEscan for genome-wide scans of GxE interaction.

Gauderman has applied novel statistical methods and software to address important substantive questions in the context of genetic- and environmental-epidemiology. As an example, he directed a genome-wide association study (GWAS) aimed at identifying novel genes and GxE interactions for asthma and lung function in the Children's Health Study (CHS). Based on CHS data, he has also published papers demonstrating a link between outdoor air pollution and lung function development and asthma in children, and related papers showing that cleaning up the air is associated with improvements in children's health. His current research related to air pollution focuses on the use of individual sensor monitors to collect exposure, time-activity, and biomarker data on asthmatic children, coupled with machine learning approaches to identify features in the data that may predict wheezing and other adverse events.

Wesley Gray
Southern University and A&M College

Wesley Gray, Ph.D., is the Walter Dumas Endowed Professor of Environmental Toxicology and professor of chemistry at Southern University Baton Rouge (SUBR), Louisiana. His professional career has been focused on establishing and maintaining a productive research and teaching program that advances the study of phytoestrogens and their role in steroid receptor function. Gray joined the faculty at SUBR as a tenure-track assistant professor in the newly established Ph.D. program in environmental toxicology with a dual appointment in the department of chemistry. He served as the program director for the Environmental Technology Consortium (2004-2008), established the Summer Research Institute for freshmen and sophomores (SRIFS) (2010), and is currently the campus coordinator for the Louisiana Biomedical Research Network (LBRN) program and the Science Research Education Program (SREP). In addition, Gray has served on the university sabbatical leave committee (2011-2012), as program director for the SUBR STEM Living Learning Community (2008-present), as chair for recruitment and retention for the department of environmental toxicology (2010-2016), and as a member of the advisory

council for the Center for Teaching and Learning Excellence. He is also a member of the Society of Toxicology, where he has served as regional chapter president (2012-2013) and on the education committee and K-12 committee.

Gray received his Bachelor of Science in biology from Wright State University in Dayton, Ohio, and earned his doctorate in biochemistry from The University of Maryland at Baltimore. He received postdoctoral training in molecular endocrinology as an NIH trainee in the biochemistry department at the University of Wisconsin-Madison.

Amy Herring
Duke University

Amy Herring, Sc.D., is the Sara and Charles Ayres Professor of Statistical Science and Global Health at Duke University. She received her doctorate in biostatistics from Harvard University after receiving bachelor's degrees in mathematics and English from the University of Mississippi. Her research has focused on statistical approaches for analyzing complex, high-dimensional data on individuals followed over time, and she has supported her research program as principal investigator (PI) of numerous grants from the NIH, the EPA, and the Maternal and Child Health Bureau. She has been PI of multiple T32 and R13 awards focused on training the next generation of data science researchers, including an NIEHS T32 interdisciplinary training program at the University of North Carolina at Chapel Hill.

Herring has been active in professional service as executive secretary of the International Society for Bayesian Analysis and a member of the National Academies' Committee on Applied and Theoretical Statistics, former president of the Eastern North American Region of the International Biometric Society, and former chair of the American Statistical Association Biometrics Section. She is associate editor of the Journal of the American Statistical Association. She was awarded the Mortimer Spiegelman Award as Outstanding Public Health Statistician under age 40 and is an elected fellow of the American Statistical Association and the International Statistical Institute.

Ronald Hines
U.S. Environmental Protection Agency

Ronald Hines, Ph.D., is the associate director for health at the National Health and Environmental Effects Research Laboratory in the Office of Research and Development at the EPA. In this position, he manages the research programs of the laboratory's health divisions, as well as the Research Core Unit.

Hines earned his doctorate in biochemistry from the University of Texas Southwestern Medical School. Following his postdoctoral fellowship at the University of Vermont College of Medicine, he became assistant professor (1983-1988) and, later (1988-1989), associate professor at the Eppley Institute for Research in Cancer and Allied Diseases and the department of biochemistry at the University of Nebraska Medical Center. Hines was recruited to the Wayne State

University School of Medicine as associate professor of pharmacology and in 1995 was promoted to professor of pharmacology. In 1999, he accepted a position as professor of pediatrics and pharmacology and toxicology at the Medical College of Wisconsin, where he also served as associate director of the Children's Research Institute (2005-2012) and co-section chief of clinical pharmacology, pharmacogenetics, and teratology in the department of pediatrics. Hines moved to his current position in 2012.

Hines is a member of the Drug Metabolism and Disposition and Current Topics in Toxicology editorial boards and an associate editor for Toxicological Sciences. He has more than 140 publications focused on mechanisms whereby exposures to environmental toxicants or drugs alters gene regulation and the genetic and/or epigenetic basis for population- and life-stage susceptibility to exposures.

Alan Hubbard
University of California, Berkeley

Alan Hubbard, Ph.D., is a professor of biostatistics at the University of California (UC), Berkeley. He the principal investigator of a study of statistical methods related to patient-centered outcomes research among acute trauma patients (PCORI), head of the computational biology Core E at the Superfund Center at UC Berkeley (NIH/EPA), and a consulting statistician on several federal and foundation projects.

Hubbard has published over 200 articles, including numerous peer-reviewed articles related to diarrheal disease. He has worked on projects ranging from molecular biology of aging to epidemiology to infectious disease modeling, but most of his work has focused on semi-parametric estimation in high-dimensional data. His current methods research focuses on precision medicine, variable importance, statistical inference for data-adaptive parameters, and statistical software implementing targeted learning methods. He is working in several areas of applied research, including early childhood development in developing countries, patient outcomes from acute trauma, environmental genomics, and comparative effectiveness research in diabetes care.

Michael Humble
National Institute of Environmental Health Sciences

Michael Humble, Ph.D., received his doctorate in toxicology from the University of North Carolina at Chapel Hill, performing his dissertation research in the intramural laboratories at NIEHS on a transgenic mouse model for skin cancer. Following a short postdoctoral training period, Humble began working in the Division of Extramural Research and Training at NIEHS as a program analyst. He is currently a program administrator. Humble oversees the extramural fellowship program, the Short Term Educational Experiences for Research (STEER) in the Environmental Health Sciences program, the Undergraduate Research Education Program (UP) to Enhance Diversity in the Environmental Health Sciences, as well as the research portfolios in

skin disease, immunotoxicology, and autoimmune disease. Humble was a high school chemistry teacher prior to returning to graduate school and receiving his doctorate.

Marie Lynn Miranda
Rice University

Marie Lynn Miranda, Ph.D., is the Howard R. Hughes Provost and Professor of Statistics at Rice University in Houston, Texas, and an adjunct professor of pediatrics at Duke University and Baylor College of Medicine.

Miranda specializes in research on environmental health, especially how the environment shapes health and well-being among children. She is the founding director of the Children's Environmental Health Initiative, a research, education, and outreach program committed to fostering environments where all people can prosper. Miranda's formal educational background is rooted in mathematical, statistical, and economic modeling; her professional experiences integrate environmental health sciences with sound social policies. She is a leader in the rapidly evolving field of geospatial health informatics and has applied spatial analytic approaches to a wide range of scientific issues.

Miranda maintains an active research portfolio, with a funding history that includes the EPA, the NIH, the Centers for Disease Control and Prevention, the National Association of Chronic Disease Directors, the U.S. Department of Agriculture, the state of North Carolina, the Robert Wood Johnson Foundation, the Wallace Genetics Foundation, the Mary Duke Biddle Foundation, and The Duke Endowment. She maintains a deep and abiding interest in environmental and social justice. Her research group received the 2008 EPA Environmental Justice Award.

Miranda is a Phi Beta Kappa, summa cum laude graduate of Duke University, where she earned her bachelor's degree in mathematics and economics and was named a Truman Scholar. She has a doctorate and a Master of Arts, both in economics, from Harvard University, where she held a National Science Foundation Graduate Research Fellowship. She served on the faculty at Duke from 1990-2011, and then as dean of the School of Natural Resources and Environment at the University of Michigan for four years. Miranda became provost at Rice University in July 2015.

Chirag Patel
Harvard Medical School

Chirag Patel, Ph.D., is an assistant professor in the department of biomedical informatics at Harvard Medical School. Patel's long-term research goal is to address problems in human health and disease by developing computational and bioinformatics methods to reason over high-throughput information spanning molecules to populations. His group focuses on computational strategies to efficiently and reproducibly uncover the complex interaction between the exposome, genome, and phenome toward development of new tools for disease

diagnosis and therapy, supported by NIEHS (K99/R00), the National Institute of Allergy and Infectious Diseases (NIAID), and the National Science Foundation (NSF).

Patel trained in biomedical informatics and epidemiology at Stanford University. Prior to graduate work, he was a software engineer in the biotechnology industry. He teaches introductory courses in data science, is a mentor to three postdoctoral associates, and advises three Ph.D. students. Patel also co-leads a Harvard National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) T32 postdoctoral program in bioinformatics applied to diabetes, obesity, and metabolism.

John Quackenbush

Harvard T.H. Chan School of Public Health, Dana-Farber Cancer Institute

John Quackenbush, Ph.D., is a professor of biostatistics and computational biology at the Dana-Farber Cancer Institute and a professor of computational biology and bioinformatics at the Harvard T.H. Chan School of Public Health. Quackenbush holds a doctorate in theoretical physics. In 1992, he received a fellowship from the NIH to work on the Human Genome Project, which led him from the Salk Institute to Stanford University to The Institute for Genomic Research before moving to Harvard in 2005. He currently directs the Master of Science in Computational Biology and Quantitative Genetics Program and is principal investigator of the BD2K Training Grant at the Harvard Chan School.

Quackenbush's research uses massive data from DNA sequencing and other assays to model functional networks in human cells. By comparing networks between groups of individuals, he has found new drug targets, explored chemotherapy resistance, and investigated differences between the sexes. He has made pioneering discoveries about how the genetic variants work together to determine our traits. Quackenbush has published more than 280 papers; his work has been cited more than 65,000 times. He has received numerous awards for his work, including recognition in 2013 as a White House Open Science Champion of Change. He is also the co-founder of Genospace, a precision medicine software company that was purchased by the Hospital Corporation of America in 2017.

Charles Schmitt

National Institute of Environmental Health Sciences

Charles Schmitt, Ph.D., is a computer scientist with a background in data science, data mining, and software engineering. He joined NIEHS as a senior advisor in 2017 to help advance the application of data science tools and methods. Prior to joining NIEHS, he served as chief technology officer and director of data science at the Renaissance Computing Institute (RENCI), a research institution focused on advancing the use of computational and data-driven approaches in science. At RENCI, he led or participated in projects in the areas of data security, data mining and analysis, data management, genomic sequencing, bioinformatics, patient monitoring, cheminformatics, and medical decision support.

Schmitt has served as an investigator or co-principal investigator on multiple grants from the NIH, the NSF, and the Department of Homeland Security. He served as the acting science information officer at NIEHS in 2014 and led the formation of the international iRODS Consortium in 2013 to advance open-source data management solutions. Prior to joining RENCI, he worked in industry for over 10 years in the areas of bioinformatics, data mining, and software engineering at IBM, BD Technologies, and several small companies. He holds a Bachelor of Science in physics and a doctorate in computer science from the University of North Carolina.

Carol Shreffler
National Institute of Environmental Health Sciences

Carol Shreffler, Ph.D., is the program officer for the Training and Career Development Program at NIEHS. Shreffler is the NIEHS representative on the NIH Training Advisory Committee and the NIH Molecular Libraries and Imaging Roadmap Committee. She served for two years as the co-chair of the Training Subcommittee (now the Key Function Committee) of the Clinical and Translational Science Awards, and on the Re-Engineering the Clinical Research Enterprise Roadmap Committee. Her other training-related experience includes serving on the Interdisciplinary Workforce Roadmap Committee, the Trans-NIH Neurosciences Blueprint Training Subcommittee, and the K30 Clinical Curriculum Development Steering Committee. She is the NIEHS coordinator for the Diversity and Re-entry Supplements program, and is credited with developing and managing the Outstanding New Environmental Scientist Program, an R01 research grant aimed at outstanding new early stage investigators in the environmental health sciences.

Shreffler's scientific portfolio is the Liver, Kidney, and Digestive Systems Toxicology Research and she served on the NIH Liver Disease Subcommittee of the Digestive Diseases Interagency Coordinating Committee.

Shreffler received a Bachelor of Science in mathematics from Texas Tech University. Following experience as an aerospace engineer at Johnson Spacecraft Center, she returned to graduate school and received a doctorate in genetics from Oregon State University. She spent time in the Intramural Research Program at NIEHS before joining the Division of Extramural Research and Training as a scientific review administrator. In 1999, she assumed her current position as director of the Training and Career Development Program.

Jeanette Stingone
Icahn School of Medicine at Mount Sinai

Jeanette Stingone, Ph.D., is a postdoctoral fellow in the department of environmental medicine and public health at the Icahn School of Medicine at Mount Sinai. She conducts research that couples data science techniques with epidemiologic methods to investigate how prenatal and early-life environmental exposures affect children's health and development. Currently, Stingone's research is supported by a K99/R00 Career Development Award from NIEHS. Using a

two-stage approach of hypothesis generation followed by targeted analysis, she will apply data science techniques to investigate the relationship between prenatal exposure to multiple air toxics and subsequent academic outcomes in childhood within an administrative data linkage of 200,000 New York City children. Stingone also has a strong interest in the use of collective science initiatives to advance environmental health research, and currently works as part of the CHEAR Data Center to develop methods and approaches for data harmonization across diverse studies of children's health. She will be transitioning to a faculty position in the department of epidemiology at the Mailman School of Public Health at Columbia University this fall.

Stingone received an undergraduate degree in biomedical engineering from Boston University, a master's in public health from the Mount Sinai School of Medicine, and a doctorate in epidemiology from the University of North Carolina at Chapel Hill.

Susan Teitelbaum
Icahn School of Medicine at Mount Sinai

Susan Teitelbaum, Ph.D., is a classically trained environmental and cancer epidemiologist with extensive experience conducting epidemiologic field studies, assessing complex environmental exposures, and performing advanced statistical analysis of large epidemiologic data sets. Her collaboration on studies with a wide array of designs, including case-control studies (population-based and international, multi-site), prospective cohort studies, and methodological development studies, has provided her with a range of epidemiology experience, including study design, data collection protocol design and oversight, field operation management, data quality assurance and management, and data analysis.

Teitelbaum's research projects include investigations of environmental endocrine disruptors in childhood health, growth, and development, as well as breast cancer. She is also the principal investigator of the CHEAR Data Center. CHEAR will provide the extramural research community access to laboratory and data analyses to add or expand the inclusion of environmental exposures in their children's health research.

John Van Horn
University of Southern California

John Van Horn, Ph.D., is an associate professor of neurology with additional appointments in neuroscience and in electrical engineering at the University of Southern California (USC) in Los Angeles, California. He is based in the USC Mark and Mary Stevens Neuroimaging and Informatics Institute. He received his bachelor's degree in psychology from Eastern Washington University, a master's in electrical engineering and computer science from the University of Maryland, College Park, and his doctorate from the University of London in the United Kingdom. He conducted a postdoctoral fellowship at the National Institute of Mental Health on the NIH main campus in Bethesda, Maryland, specializing in human neuroimaging investigation of brain function. He has held faculty positions at Dartmouth College, the University of California, Los Angeles, and USC. He is an accomplished author (over 150 publications) and

university-level educator, and is known internationally as an expert in neuroinformatics and data sharing.

Among Van Horn's research articles are many publications on multimodal neuroimaging of the brain and the characterization of mild and severe traumatic brain injury (TBI). This includes using MRI and diffusion tensor imaging to model the morphological effects of brain injury as well as the effect on white matter fiber pathways. He is the past education chair and program chair for the Organization for Human Brain Mapping, and is the current vice president of the Society for Claustrum Research. He directs the Master of Science in Neuroimaging and Informatics Program at USC – a one-of-its-kind program covering the spectrum of human neuroimaging research and practice – and contributes to other USC graduate programs.

Van Horn is also the principal investigator of the NIH BD2K Training Coordinating Center, an effort to synthesize data science educational content from around the Internet, index it into a common database framework, and make the information searchable, sortable, and openly available for users to organize into personalized training plans. He directs a unique series of five-day mentored and facilitated Data Science Innovation Lab events on specific topics such as mobile health, the microbiome, and single cell dynamics.

Van Horn also oversees a data science scholarly rotation program which pairs junior biomedical researchers with more senior data scientists to work on projects of mutual interest.

Cheryl Walker
Baylor College of Medicine

Cheryl Walker, Ph.D., is the director of the Center for Precision Environmental Health at Baylor College of Medicine, where she holds the Alkek Presidential Chair in Environmental Health and is a professor in the departments of molecular and cell biology and medicine. Walker is internationally recognized for her work in environmental health and elucidation of molecular mechanisms of disease. Her studies on the role of the epigenome in gene-environment interactions have yielded significant insights into mechanisms by which early life environmental exposures influence health and disease across the life course.

Walker received a Bachelor of Arts in molecular, cellular, and developmental biology from the University of Colorado-Boulder and a doctorate in cell biology from the University of Texas Southwestern Medical School, and received additional postdoctoral training as a staff fellow at NIEHS. She is a fellow of the Academy of Toxicological Sciences and the American Association for the Advancement of Science, and a member of the National Academy of Medicine. She currently serves on the National Cancer Institute Board of Scientific Advisors and the National Academy of Sciences committee on "Gulf War and Health: Generational Health Effects of Serving in the Gulf War." She is also the director of the new NIEHS T32 Training Program in Precision Environmental Health.

Fred Wright
North Carolina State University

Fred Wright, Ph.D., is Goodnight Innovation Distinguished Professor in the departments of statistics and biological sciences at North Carolina State University, and director of the Bioinformatics Research Center. His research interests include genomic bioinformatics, toxicogenomics, and the statistical principles underlying high-dimensional data analysis. Wright has been principal investigator on numerous grants, with activities ranging from development of new methods of gene mapping to expression-quantitative trait mapping for multiple tissues. He is director of an active NIEHS training grant in bioinformatics. He has authored or co-authored more than 150 scientific publications and is a fellow of the American Statistical Association and of Delta Omega, the public health honor society.

Participant List

Name	Organization	Email
Allen, Janice	NIEHS	allen9@niehs.nih.gov
Balshaw, David	NIEHS	balshaw@nih.gov
Birnbaum, Linda	NIEHS	birnbaumls@niehs.nih.gov
Boyles, Abee	NIEHS	abee.boyles@nih.gov
Bradfield, Christopher	University of Wisconsin-Madison	bradfield@oncology.wisc.edu
Bushel, Pierre	NIEHS	bushel@niehs.nih.gov
Carlin, Danielle	NIEHS	danielle.carlin@nih.gov
Castranio, Trisha	NIEHS	trisha.castranio@nih.gov
Chen, Jonathan	Stanford Department of Medicine	jonc101@stanford.edu
Collins, Jennifer	NIEHS	collins6@niehs.nih.gov
Collman, Gwen	NIEHS	collman@niehs.nih.gov
Craven, Mark	University of Wisconsin	craven@biostat.wisc.edu
Drew, Christie	NIEHS	DrewC@niehs.nih.gov
Duncan, Chris	NIEHS	duncancg@niehs.nih.gov
Federer, Lisa	NIH Library	lisa.federer@nih.gov
Finnell, Richard	Baylor College of Medicine	richard.finnell@bcm.edu
Florance, Valerie	National Library of Medicine	florancev@mail.nih.gov
France, Suzanne	NIEHS contractor: MDB, Inc.	sfrance@michaeldbaker.com
Garmire, Lana	University of Hawaii Cancer Center	lgarmire@gmail.com
Garton, Amanda	NIEHS	amanda.garton@nih.gov
Gatti, Daniel	The Jackson Laboratory	dan.gatti@jax.org
Gauderman, Jim	University of Southern California	jimg@usc.edu
Gray, Wesley	Southern University Baton Rouge	dr1wggray@aol.com
Heacock, Michelle	NIEHS	heacockm@niehs.nih.gov
Herring, Amy	Duke University	amy.herring@duke.edu
Hines, Ron	U.S. Environmental Protection Agency	hines.ronald@epa.gov
Holmgren, Stephanie	NIEHS	holmgren@niehs.nih.gov
Hubbard, Alan	University of California, Berkeley	hubbard@berkeley.edu
Humble, Michael	NIEHS	humble@niehs.nih.gov
James, Crystal	Tuskegee University	cjames@tuskegee.edu
Joubert, Bonnie	NIEHS	bonnie.joubert@nih.gov
Kulkarni, Resham	NIEHS	kulkarnird@niehs.nih.gov
Laborde, Danielle	HERMES LLC	dlaborde@hermesllc.com
Lan, Tuo	The University of Iowa	tuolan@uiowa.edu
Lawler, Cindy	NIEHS	Cindy_Lawler@hotmail.com
Luginbuhl, Sarah	NIEHS	sarah.luginbuhl@nih.gov
Miranda, Marie Lynn	Rice University	provost@rice.edu
Moakley, Maggie	NIEHS contractor: MDB, Inc.	moakleymg@nih.gov
Park, Yong-Moon (Mark)	NIEHS	mark.park@nih.gov
Patel, Chirag	Harvard University	Chirag_Patel@hms.harvard.edu
Pettibone, Kristi	NIEHS	pettibonekg@niehs.nih.gov
Popovich, Nicole	NIEHS	nicolepm@niehs.nih.gov

Quackenbush, John	Dana Farber Cancer institute, Harvard School of Public Health	johnq@jimmy.harvard.edu
Schmitt, Charles	NIEHS	charles.schmitt@niehs.gov
Shatz, Maria	NIEHS	shatzm@niehs.nih.gov
Shreffler, Carol	NIEHS	Shreffl1@niehs.nih.gov
Skalla, Lesley	NIEHS contractor: MDB, Inc.	lskalla@michaeldbaker.com
Stingone, Jeanette	Icahn School of Medicine at Mount Sinai	jeanette.stingone@mssm.edu
Teitelbaum, Susan	Icahn School of Medicine at Mount Sinai	susan.teitelbaum@mssm.edu
Thigpen Tart, Kimberly	NIEHS	kimberly.thigpentart@nih.gov
Thomas, Laura	NIEHS	laura.thomas@nih.gov
Trottier, Brittany	NIEHS	brittany.trottier@nih.gov
Tuyishime, Steven	NIEHS	steven.tuyishime@nih.gov
Van Horn, John	University of Southern California	jvanhorn@usc.edu
Voelker, Kerri	NIEHS contractor: MDB, Inc.	kvoelker@michaeldbaker.com
Walker, Cheryl	Baylor College of Medicine	Cheryl.walker@bcm.edu
Wright, Fred	North Carolina State University	fred_wright@ncsu.edu
Xu, Karen	University of North Carolina at Chapel Hill	kxu.karen@gmail.com

Workshop on Developing a Data Science Competent EHS Workforce