Integrating Multiscale Geospatial Environmental Data into Large Population Health Studies – Presenter Biosketches and Presentation Abstracts

Presenters

Rick Woychik

RICK WOYCHIK, Ph.D., became the Director of the National Institute of Environmental Health Sciences (NIEHS), one of the National Institutes of Health (NIH), and the National Toxicology Program (NTP) on June 7, 2020. In these roles, Woychik oversees federal funding for biomedical research to discover how the environment influences human health and disease. As a mammalian geneticist, Woychik has had a number of noteworthy accomplishments. His laboratory was the first to clone and characterize the gene called *agouti*, which provided molecular insights into obesity and the satiety response in the brain. Additionally, his laboratory was the first to identify a gene mutation associated with polycystic kidney disease, which provided insights into this molecular biology of this important human disease. Also, his laboratory was the first to determine that a member of the protocadherin family was associated with the hearing loss in a mouse model that was ultimately paved the way to better understanding the molecular basis of Usher syndrome type 1F in humans. More recently his research program has been focused on investigating the molecular mechanisms associated with how environmental agents influence the epigenetic control of gene expression.

Carolyn Hutter

CAROLYN HUTTER, PhD., is Division Director for the Division of Genome Sciences at the National Human Genome Research Institute (NHGRI) of the National Institutes of Health (NIH). Dr. Hutter joined the NIH in 2012, serving as a program director in the National Cancer Institute (NCI) Epidemiology and Genomics Research Program before transferring to NHGRI in 2013. Prior to NIH, she was a senior staff scientist at the Fred Hutchinson Cancer Research Center and Lecturer at the University of Washington, where her research focused on large-scale consortia work for genome-wide association studies and on gene-environment interactions for cancer and other complex diseases.

Duncan Thomas

DUNCAN THOMAS, Ph.D., is Professor of Biostatistics in the Department of Preventive Medicine, and Verna R. Richter Chair in Cancer Research at the University of Southern California Keck School of Medicine. He received his Ph.D. from McGill University in 1976, where he continued as a faculty member until his recruitment to USC in 1984. There he served as the Head of the Biostatistics Division until 2013 and co-directed the Southern California Environmental Health Sciences Center and the Cancer Epidemiology Program in the USC/Norris Comprehensive Cancer Center. His primary research interest has been in the development of statistical methods for environmental and genetic epidemiology, with numerous collaborations in both areas. On the environmental side, he has been particularly active in radiation carcinogenesis and air pollution health effects research, notably as one of the senior investigators on the Southern California Children's Health Study and the Women's Environmental Cancer and Radiation Exposure (WECARE) study and as a member of President Clinton's Advisory Committee on Human Radiation Experiments. On the genetic side, he is a coinvestigator in the NCI's Colon Cancer Family Registry, the Genetic Analysis Workshop, the ENDGAME consortium to develop methods for genome-wide association studies, and past President of the International Genetic Epidemiology Society. Dr. Thomas has numerous publications, including the textbooks *Statistical Methods in Genetic Epidemiology* (Oxford University Press, 2004) and *Statistical Methods in Environmental Epidemiology* (Oxford University Press, 2009). He currently directs a program project grant on "Statistical methods for integrative genomics in cancer." These three broad areas of interest make him uniquely qualified to address methodological challenges in studying gene-environment interactions.

Abstract:

Starting with some examples of simple gene-environment interactions with candidate genes and established environmental risk factors, I will expand to the challenges of gene-environment-wide interaction studies (GEWIS), use of genetic risk scores, and the exposome. High-dimensional omics data, such as the metabolome and microbiome offer the prospect of studying both mediation and more complex interactions. A recurring theme will be the formal incorporation of external biological

pathway ontologies and experimental models of gene-environment interactions directly into the analysis of epidemiologic studies.

Joshua Apte

JOSHUA APTE, Ph.D., is an Assistant Professor at UC Berkeley, jointly appointed in the Department of Civil and Environmental Engineering and in the School of Public Health. Dr. Apte's research focuses on the intersection of air quality, sustainability, and environmental justice, with an emphasis on the development of new methods for quantifying air pollution exposures. His group uses field measurements, air quality models, and satellite remote sensing to quantify air pollutant emissions and concentrations, and their resulting spatial patterns, human exposures, and public health consequences in US communities and around the world. Before coming to UC Berkeley, he was previously on the faculty at the University of Texas at Austin, the ITRI-Rosenfeld Postdoctoral Fellow at Lawrence Berkeley National Laboratory, and a Fulbright-Nehru Fellow at the Indian Institute of Technology, Delhi. He holds MS and PhD degrees from the Energy and Resource Group at UC Berkeley and a ScB in Environmental Science from Brown University.

Abstract:

Over the past decade, a wide array of newer approaches for measuring and estimating spatially resolved air pollution exposures have moved into the mainstream. These techniques include mobile monitoring, dense deployments of stationary low-cost sensors, satellite remote sensing, and predictive models. This talk will highlight recent advances in these geospatial exposure assessment technologies, emphasizing connections and complementarities among multiple methods.

Susan Anenberg

SUSAN ANENBERG, Ph.D., M.S., is an Associate Professor of Environmental and Occupational Health and of Global Health at the George Washington University Milken Institute School of Public Health. Dr. Anenberg studies the health implications of air pollution and climate change, from local to global scales. Dr. Anenberg has been a Co-Founder and Partner at Environmental Health Analytics, LLC, the Deputy Managing Director for Recommendations at the U.S. Chemical Safety Board, an environmental scientist at the U.S. Environmental Protection Agency, and a senior advisor for clean cookstove initiatives at the U.S. State Department. Her research has been published in top academic journals such as *Science, Nature,* and *Lancet Planetary Health.* She has also led or contributed to many science-policy reports on air quality and climate change published by U.S. EPA, World Bank, World Health Organization, United Nations Environment Programme, and others.

Abstract:

Dr. Anenberg will provide a summary of the workshop sessions on using geospatial data for advancing environmental justice and health equity, personalizing exposure science, and preparing for and responding to environmental disasters. She will also summarize challenges and opportunities identified by breakout groups on data availability, data integration, training and capacity building, and privacy and ethics**.**

Rena Jones

RENA JONES, Ph.D., M.S., is an environmental cancer epidemiologist in the Division of Cancer Epidemiology and Genetics at NCI, where she co-chairs the Geographic Analysis Working Group. She is an Associate Editor of the International Journal of Health Geographics and Adjunct Assistant Professor of Environmental Health Sciences at the Yale School of Public Health. Her intramural NCI research program investigates the role of environmental contaminants, including air and water pollutants, light at night, and other industrial and agricultural exposures, in cancer etiology. She seeks to improve long-term environmental exposure estimates in these studies by optimizing the spatial accuracy of residential addresses and exposure data sources, characterizing, and incorporating information on participant mobility and time spent in microenvironments, and

leveraging information from surveys, regulatory monitoring, and other geo-referenced datasets. Jones' research also focuses on validation studies on to evaluate the quality of surrogate exposure metrics in epidemiologic studies of cancer.

Abstract:

Integrating multiple exposure data sources to assess lifetime exposure is a key challenge for studies investigating the role of environmental factors on cancer etiology. Dr. Jones will discuss her work in epidemiologic studies to characterize long-term exposure to common environmental pollutants, including light at night and drinking water contaminants, with various geospatial data sources and methods. She will also describe her research to estimate and incorporate residence histories and population mobility information in longitudinal exposure assessments to support studies of cancer.

Dustin Duncan

DUSTIN DUNCAN, Sc.D., (he/they) is a social and spatial epidemiologist, studying how neighborhood characteristics influence population health and health disparities. Dr. Duncan's intersectional research focuses on Black gay, bisexual, and other sexual minority men (SMM) and transgender women of color. His research has a strong domestic focus--including in New York City, Chicago and the Deep South (e.g., New Orleans) --and his recent work spans the globe such as in West Africa, especially with Columbia's International Center for AIDS Care and Treatment Programs (ICAP). In addition to HIV epidemiology and sleep epidemiology, his current interests include characterizing the COVID-19 epidemic in the United States and globally, especially among marginalized populations. Notably, his group completed the N2 COVID Study, where they surveyed 226 Black SMM and Black transgender women in Chicago on various aspects of COVID-19 from April to July 2020. Methodologically, his research utilizes a geospatial lens to apply advanced geographic information systems, web-based and real-time geospatial technologies, and geospatial modeling techniques. Working in collaborations with scholars across the world, he has nearly 200 high-impact scientific articles, book chapters, and books and his research has appeared in major media outlets including U.S. News & World Report, The Washington Post, The New York Times and CNN. Dr. Duncan's work has been funded by the National Institutes of Health, the Centers for Disease Control and Prevention, the HIV Prevention Trials Network, the Robert Wood Johnson Foundation, the Verizon Foundation, and the Aetna Foundation. He currently leads two NIH-funded R01 studies, as well as studies funded by other sources, and mentors K and other awards of junior scientists. He has received several early career and distinguished scientific contribution awards including from the Harvard University T.H. Chan School of Public Health and the Interdisciplinary Association for Population Health Science (IAPHS). In 2020, he received the Mentor of the Year Award from Columbia University Irving Medical Center's Irving Institute for Clinical and Translational Research.

Abstract:

The field of neighborhoods and health (sometimes referred to as spatial epidemiology) has grown exponentially in the last twenty years, especially since the publication of the first edition of *Neighborhoods and Health* edited by Ichiro Kawachi and Lisa Berkman in 2003. There is a large literature on the influence of neighborhoods on health to date. Many reviews and technical reports of accumulated neighborhoods and health research have been conducted. For example, in 2017, Bauermeister and colleagues conducted a review on multi-level studies of neighborhoods and HIV prevention and care outcomes among sexual minority men who have sex with men. However, relatively few neighborhood studies have focused on sexual minority populations. Neighborhood studies that have taken an intersectional approach, e.g., by focusing on sexual identity as well as race/ethnicity simultaneously are even more scarce. This talk aims to: 1) provide an historical overview of neighborhoods and health research, 2) examine recent directions in neighborhoods and health research, and 3) touch on methodological areas including the issue of spatial misclassification. The talk will overview applications of geospatial methods to study neighborhoods in population health and health disparities research with examples from one of Dr. Duncan’s prospective cohort studies, which use novel geospatial methods. The N2 (Neighborhoods and Networks) Cohort Study is a cohort study including 600 HIV-negative and HIV-positive Black gay, bisexual and other sexual minority men in Chicago IL, Jackson MS, and New Orleans and Baton Rouge LA (Grant Numbers: R01MH112406 and U01PS005122). The cohort focuses on

advanced global positioning systems (GPS) methods to understand how activity space neighborhoods (defined as the local areas within which people move during their daily activities) influence HIV prevention and care outcomes.

Rima Habre

RIMA HABRE, Sc.D., is an Associate Professor of Environmental Health and Spatial Sciences at the University of Southern California. Her research aims to understand the effects of co-occurring environmental exposures, air pollution mixtures and social stressors on the health of vulnerable populations across the life course. She develops methods to advance personal exposure assessment using personal monitoring (e.g., wearables, sensors), geolocation, and machine-learning based spatiotemporal models. Dr. Habre co-chairs the Geospatial Working Group in the national NIH Environmental Influences on Child Health Outcomes (ECHO) program. She co-leads the Exposure Sciences Research Program in the Southern California Environmental Health Sciences Center (NIEHS Core Center). Dr. Habre is also MPI/Project Lead and Director of Exposure Assessment in two large research centers at USC investigating the effects of air pollution exposure during pregnancy and the postpartum period on maternal and child health. She received her Doctor of Science in environmental health with a concentration in exposure science from the Harvard T.H. Chan School of Public Health.

Abstract (Day 1, Session 1):

Personal exposure to air pollution and other environmental risk factors is highly variable in space and time and depends on human time-activity and mobility patterns, behaviors, indoor and outdoor sources. Especially in large population studies, personal exposure to air pollution is often approximated by static, outdoor residential concentrations predicted with varying degrees of spatiotemporal resolution and accuracy. This talk will demonstrate how advances in personal monitoring and geospatial technologies are allowing researchers to assess highly personalized and spatiotemporally resolved exposures within actual activity spaces to inform precision environmental health. These advances ultimately lead to a better understanding of contextual effects, temporality, and important microenvironments driving adverse exposures and health effects. They also significantly reduce exposure measurement error and amplify statistical power to detect relevant exposure-health associations.

Abstract (Day 2, Session 4):

The NIH ECHO Program aims to understand the effects of a range of early life environmental influences on child health and development over the life course. Its mission is to enhance the health of children for generations to come. It brings together over 50,000 participants from existing pediatric cohorts and clinical trials and focuses on five major pediatric outcomes. As a result, ECHO provides a unique opportunity to leverage a wide range of geospatial data resources, methods, and modeling approaches to understand the effects of environmental and social risk factors in a nationwide, geographically diverse cohort. This talk will provide an overview of major geospatial needs, challenges, and opportunities within the ECHO consortium.

Yang Liu

YANG LIU, Ph.D., M.S., received his PhD in environmental science and engineering from Harvard Graduate School of Arts and Sciences in 2004, and completed his postdoctoral training at Harvard T.H. Chan School of Public Health. He is the Gangarosa Distinguished Professor and Chair of the Gangarosa Department of Environmental Health at the Rollins School of Public Health of Emory University. His research interests include satellite aerosol retrieval and product design, applications of satellite remote sensing in public health research, potential impacts of global climate change on public health, machine learning and spatial statistics.

Abstract:

Integrating Multiscale Geospatial Environmental Data Into Large Population Health Studies

Satellite remote sensing data has become a powerful tool to assess large-scale population exposure to various environmental risk factors to human health such as air pollution. Various exposure models driven by satellite data and other gridded information dramatically expand the spatial coverage of exposure estimates to regions of the world with sparse or only recent ground monitoring network. This talk will provide an overview of emerging satellite instruments, their data products, and potential applications in environmental health research.

John Haynes

JOHN HAYNES, M.S., serves as Program Manager for Health and Air Quality Applications in the Applied Sciences Program of the NASA Earth Science Division at Headquarters in Washington, DC. John entered NASA Headquarters in 2002 through the Presidential Management Fellowship (PMF) program. As required by the PMF program, John completed two detail assignments during his fellowship (NOAA and the US House of Representatives). John converted to a civil service management position at NASA Headquarters in August 2004 upon graduation from the PMF program. John graduated from the University of South Alabama in 1999 with a B.S. in meteorology. In 2002, he graduated with an M.S. in meteorology from the University of Oklahoma. The first portion of his thesis work ("Analysis of Warm Season Morning Convection across the Southern Great Plains") was published in the December 2003 edition of *Weather and Forecasting*. The second portion of his thesis work ("The Evolution of Morning Convective Systems over the U. S. Great Plains during the Warm Season. Part II: A Climatology and the Influence of Environmental Factors") was published in the March 2008 edition of *Monthly Weather Review*. John has received several awards during his tenure at NASA including a NASA Aviation Safety and Security Program Award, two NASA Group Achievement Awards, a One NASA Award, a Team Excellence Award, and the 2020 NASA Exceptional Service Medal. John has also been honored by his alma mater (the University of South Alabama) as an Exceptional Alumnus of the School of Meteorology.

Abstract:

Emerging issues in environmental health produce a significant health burden, especially across low- and middle-income countries. According to the World Health Organization, 24% of global deaths are attributed to modifiable environmental factors. To better understand the potential environmental exposures on human health, novel data and technology sources should propel scientific inquiry to examine and mitigate health risks. Earth-observing satellites offer real-time data that monitor temporal and spatial changes, due to natural and anthropogenic causes, within global ecosystems. Multidisciplinary collaborations that integrate cross-cutting research designs, methods, and tools can expand the boundaries of field applications and strengthen evidence-based findings. In this presentation, I will provide an overview of NASA Earth-observing satellites, including the upcoming TEMPO and MAIA missions, which highlight scientific innovation to better understand the dynamic processes of global ecosystems. This information can complement collaborative community efforts to enhance risk assessment and public health surveillance programs. I will also offer approaches that can foster community partnerships, further strengthen public engagement, and ultimately protect population health.

Adam Szpiro

ADAM A. SZPIRO, Ph.D., is an Associate Professor of Biostatistics at the University of Washington. His primary research focus is statistical methods for environmental epidemiology, especially spatial and spatiotemporal statistics as they apply to observational air pollution studies. He develops methods for spatiotemporal air pollution prediction, characterizing and correcting for measurement error from spatially misaligned data, integrating large-*p* machine learning methods with spatial and spatiotemporal statistics, accounting for unmeasured spatial and temporal confounding, and analyzing mixtures of air pollution and other environmental toxicants. Dr. Szpiro has been a core investigator of the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air) since 2006 and leads the Data Center for the Environmental Influences on Child Health Outcomes (ECHO) Pathways project. He received a Doctor of Philosophy in Applied Mathematics from Brown

University focused on stochastic processes and numerical methods for solving partial differential equations and he completed a post-doctoral fellowship in Biostatistics at the University of Washington.

Abstract:

Measurement error with spatially misaligned data occurs in air pollution cohort studies when exposure data are not available at study subject locations, so exposures need to be predicted by spatial statistics using monitoring data from different locations. We decompose the error into classical-like and Berkson-like components, which share some features in common with standard classical and Berkson error but differ in fundamental ways. Importantly, we do not assume that the measurement error has classical-like and Berkson-like components. Rather, these properties follow directly from how the exposures are predicted and how the uncertainty in these predictions propagates through the health effect analysis. Both types of error inflate health effect SEs and can lead to bias, including effect transfer in multi-pollutant settings. Bias can be away from or toward the null and the magnitude does not necessarily correspond to how accurately each pollutant in a multi-pollutant study is predicted. We describe a comprehensive correction approach, including requirements and conditions for successful implementation. We illustrate our results in simulations and in an analysis of the association between PM2.5 and NO2 and elevated blood pressure in the NIEHS Sisters Study.

Perry Hystad

PERRY HYSTAD, Ph.D., is an Associate Professor in the College of Public Health and Human Sciences at Oregon State University, USA where he leads the Spatial Health Lab (http://health.oregonstate.edu/labs/spatial-health). His research focuses on environmental exposure assessment and epidemiology. He has developed new methods to assess environmental exposures for large population-based health studies using new technologies and data science approaches. He leads several grants that are focused on air pollution, green space, healthy built environments, and climate change, with the goal of informing effective policy and prevention activities.

Abstract:

This presentation will review additional data and methods needed to complement geospatial models to enhance causal inference. This includes the potential of collecting existing retrospective time-activity data from smartphones; novel geospatial exposure assessment approaches based on imagery, social-media data, and connected vehicles; and casual inference methodology that is enabled by large geospatial data. The trade-offs between different approaches and implications for measurement error, bias, and confounding will be discussed.

Maria Argos

MARIA ARGOS, Ph.D., is an associate professor of epidemiology at the University of Illinois at Chicago School of Public Health. Maria's research focuses on the environmental and molecular epidemiology of chronic diseases, with particular emphasis on the health effects of metal exposures. Her research has a strong focus on identifying molecular pathways altered by environmental exposures that influence disease risk by leveraging high-dimensional molecular data and biomarker data obtained from biological specimens collected in the epidemiological setting and data linkages to geospatial data. She is the principal investigator of the Bangladesh Environmental Research in Children's Health (BIRCH) cohort that aims to evaluate early life health effects of metal mixtures and a principal investigator of the Illinois Precision Medicine Consortium enrollment site of the *All of Us* Research Program.

Abstract:

This talk will describe data gaps for characterizing human exposure to drinking water contaminants in public and private sources in the United States, present collaborative work conducted to assess organic and inorganic drinking water exposures and discuss implications for evaluating the link of geospatial exposure data with human health.

Roel Vermeulen

ROEL VERMEULEN, Ph.D., is Professor of Environmental Epidemiology and Exposome Science, with a joint appointment at Utrecht University and the Julius Center, University Medical Center Utrecht (UMC Utrecht). His scientific research focuses on environmental risk factors for non-communicable diseases with a strong emphasis on integrating epidemiology, high quality exposure assessment, and molecular biology into multidisciplinary investigations. Vermeulen is the project coordinator of the Dutch exposome research programme Exposome-NL, the European research project EXPANSE (that focuses on the urban exposome),the Utrecht Exposome Hub and the Open Research Infrastructure Exposome-Scan. He is also the principal investigator of several large case-control and prospective occupational and general population studies.

Abstract:

Understanding how our environment might affect our health requires a system understanding on both the external and internal exposome and their relation to health. Systematic mapping of the measurable components of the exposome, combined with network science might provide such an understanding and facilitate the identification of new (upstream) drivers of health and the development of more effective intervention strategies.

Scott Weichenthal

SCOTT WEICHENTHAL, Ph.D., is an Associate Professor in the Department of Epidemiology and Biostatistics at McGill University in Montreal, Canada. His research program is dedicated to identifying and evaluating environmental risk factors for chronic illnesses such as cancer and cardiovascular disease. To support this objective, Dr. Weichenthal develops new approaches to population-based exposure assessment and examines how the urban built environment influences environmental exposures at both the individual and population-level. His past studies have examined the health effects of air pollution from biomass burning, traffic, as well as the oxidative potential of fine particulate air pollution (PM2.5). His current research is examining the use of deep learning models in estimating environmental exposures on both a local and global scale.

Abstract:

This presentation will provide an overview of our recent work using images and audio data paired with machine learning models to predict environmental exposures on both a local and global scale. We will provide examples using aerial images to predict global variations in annual average PM2.5 concentrations, state-level variations in NO2 in Belgium, and near-real time variations in local outdoor ultrafine particle concentrations and noise levels. Our ongoing work aimed at operationalizing this approach using a network of sensors will also be discussed.

Cole Brokamp

COLE BROKAMP, Ph.D., is an Assistant Professor of Pediatrics at Cincinnati Children's Hospital Medical Center and the University of Cincinnati College of Medicine interested in geoinformatics, environmental and population health, and fairness in precision medicine. Dr. Brokamp's research is focused on understanding the heterogeneous effects of environmental exposures and community characteristics on childhood psychiatric and neurobehavioral health outcomes. Dr. Brokamp is also interested in developing new methods and technologies to support environmental and population health research, including statistical computing tools for geocoding and geomarker assessment, high resolution spatiotemporal air pollution exposure assessment models, and causal inference machine learning methods.

Abstract:

High resolution spatiotemporal interpolation and prediction models are increasingly common, but their practical usage for exposure assessment is hindered by (1) large data files sizes or long computation times, (2) the need for technical expertise to

implement in new study populations, and (3) privacy restrictions around sharing relatively precise spatiotemporal locations of study participants. We extended the free and open-source DeGAUSS (Decentralized Geomarker Assessment for Multi-Site Studies) framework to deal with gridded spatiotemporal datasets, allowing for decentralized, private, and reproducible exposure assessment. This approach only downloads the coarse spatiotemporal slices of data needed for exposure assessment, significantly reducing the size of exposure data that needs to be transferred. Specifically, spatiotemporal gridded data is broken into smaller files based on their geohash and calendar year, allowing us to satisfy HIPAA's Safe Harbor definition of deidentified data. This approach to sharing spatiotemporal exposure assessment models makes them portable and reproducible, satisfying the recent acceptance and implementations of FAIR (findable, accessible, interoperable, reusable) data principles and efforts to mobilize computable biomedical knowledge. As a free and open-source software package, we encourage others to utilize this framework for any spatiotemporal exposure assessment.

Alison Motsinger-Reif

ALISON MOTSINGER-REIF, Ph.D., is the Branch Chief of the Biostatistics and Computational Biology Branch at the National Institute for Environmental Health Sciences. She received her PhD in Human Genetics and MS in Applied Statistics from Vanderbilt University, and was faculty at North Carolina State University before coming to NIEHS in 2018. Overall, her group focuses on the development and application of modern statistical approaches for understanding the etiology of common, complex diseases and other clinically relevant traits. Her work includes approaches for gene-environment interactions, machine learning methods, and genetic association mapping. She has published over 180 peer-reviewed publications in a range of interdisciplinary journals.

Abstract:

The Personalized Environment and Genes Study, formerly the Environmental Polymorphisms registry is a diverse, North Carolina based cohort with extension exposure data, including both geospatial estimates and survey questionnaires. Recently, whole genome sequencing data has been collected, and ongoing efforts will merge this data with electronic health records. Initial work in leveraging these diverse data have revealed associations with several common, complex diseases.

Andrea Baccarelli

ANDREA BACCARELLI, M.D., Ph.D., is the Leon Hess Professor and Chair of the Department of Environmental Health Sciences and serves as the Director of the NIH/NIEHS P30 Center for Environmental Health in Northern Manhattan, one of such 21 centers across the country. Dr. Baccarelli's work has supported international best practices for air pollution control developed by multiple agencies worldwide, and his findings have served as the basis for the Environmental Protection Agency's decision to enforce stricter guidelines for human exposure. Dr. Baccarelli's research investigates molecular mechanisms as pathways linking environmental exposures to human disease. Current projects investigate a range of mechanisms, including epigenomics, epitranscriptomics, extracellular vesicles and small non-coding RNAs, mitochondrial DNA, and the microbiome. Dr. Baccarelli was elected to the National Academy of Medicine for his pioneering work showing that environmental exposures adversely affect the human epigenome and has been included in the Web of Science list of highly cited, world's most influential scientists of the past decade.

Abstract:

The use of big data to study environmental exposure has grown alongside our capacity to generate massive genomic and multi-omic datasets. Over the past decade, large population studies, which often have a sample size ranging from 500 to 100,000 participants have progressively added omic data. Those omic analyises are often costly and the use of geospatial environmental data allows for a cost-effective approach to integrate a variety of exposure-related information. However, the use of omic data in conjunction with geospatial environmental data also present challenges, including those common to any

large population studies as well as some that are specific to molecular population studies. General challenges include availability of participants addresses and/or complete residential history, limited capacity to assess the actual location of the study participants during the time of concern for exposure assessment, and capacity to reconstruct past exposures for cohorts with long follow up. However, molecular studies have their own set of challenges. While the parent cohorts are often large, budget consideration have often constrained the sample size with omic data. This is an important issue because omic data are multidimensional and require higher sample size than other type of data analysis. Many cohorts today have multiple types of omic data but limited research has been conducted on how to analyze geospatial data together with highly dimensional omic data. Also, discovery-replication and meta-analysis of multiple study populations are popular on omic data, but lack of or uneven geospatial exposure data may cause additional constraints. Finally, use of geospatial exposure data is dependent on appropriate confounding control, but identification of potential confounders may be challenging with highly dimensional omic data, where tailoring confounding strategies to individual loci is not feasible.

Rachel Liao

RACHEL LIAO, Ph.D., is Executive Director of the International Common Disease Alliance (www.icda.bio), an international scientific community dedicated to accelerating progress translating from genetic maps to molecular mechanisms, to physiology and medicine. Additionally, Rachel has worked at the Broad Institute for the past 12 years and recently took the role of Head of Science Strategy in the Broad's Data Sciences Platform; immediately prior to this, she was Scientific Advisor to the Director of the institute.

Abstract:

Human genetics stands at a pivotal moment. The past several decades has seen enormous progress. Various efforts—the Human Genome Project, deep catalogs of genetic variation, to the GWAS revolution, and massive sequencing— have resulted in a wealth of knowledge about the genetics of common disease. This includes the discovery of more than 100,000 robust genetic associations to common diseases and traits, and important insights into the underlying basis of some diseases. Yet, there is much more that needs to be done. There is again a growing sense across the human genetics community that now is the time to articulate a vision for the next phase of common complex disease genetics —to accelerate progress in moving from Maps to Mechanisms to Medicine.

Over the past three years, discussions among scientists across the human genetics community have led to the decision to form an International Common Disease Alliance (ICDA) as a way to engage the community. ICDA has developed a white paper and recommendations (available at www.icda.bio) mapping out the next steps that could be taken to improve our understanding of the genetic causes of common complex diseases in order to benefit people around the world. This talk will describe the work ICDA has undertaken with the scientific community to begin convening the international community to implement these recommendations.

Lew Berman

LEW BERMAN, M.S., Ph.D., serves as a Computer Scientist in the NIH All of Us Research Program Division of Technology and Platform Development. In this role, he is responsible for new methods and technologies for aggregating clinical data on research participants. This includes data from electronic health records and extant linkable data sources. Prior to working at NIH, Lew worked for 9 years as an industry consultant focused on health data collection, survey research, and methods development to improve the efficiency of health studies. Previously, Lew served as the CDC National Health and Nutrition Examination Survey (NHANES) Deputy Director, Special Assistant for Community HANES, and Branch Chief for Informatics. He was responsible for the design, development, and deployment of field-based home and mobile exam center data collection systems, data dissemination, and research studies on dried blood spots, digital imaging, improving respondent participation, and data linkages / disclosure. Earlier in his career, Lew held positions at the National Library of Medicine and the US Naval

Research Laboratory where he focused on imaging archives and automated image feature detection. Lew has a bachelor's degree in Computer Science from the University of Maryland, a master's degree and a doctorate in Computer Science and Public Health from George Washington University.

Abstract:

In 2018 the NIH All of Us Research Program launched an effort to collect and study data from one million or more people living in the United States. This precision medicine initiative in part seeks to answer questions on how to deliver the appropriate treatment for each person. This effort will collect electronic health records, self-reported survey data, physical measurements, bio-samples, genetic and digital health data on participants ages 18 and above. This longitudinal program will run for 10 or more years. This talk will discuss the program status, design considerations, and current approaches, opportunities, and challenges to linking with extant data sources.

Geoffrey Ginsburg

GEOFFREY GINSBURG, M.D., Ph.D., is the founding director for the Center for Applied Genomics & Precision Medicine at the Duke University Medical Center and for MEDx, a partnership between the Schools of Medicine and Engineering to spark and translate innovation. His research addresses the challenges for translating genomic and digital health information into medical practice and the integration of precision medicine into healthcare. In 2017 he received Duke's Translational Research Mentorship Award and is a finalist in the NIH/BARDA Antimicrobial Resistance Prize. He was a member of the Advisory Council to the Director of NIH and is co-chair of the National Academies Roundtable on Genomic and Precision Health and is founder and president of the Global Genomic Medicine Collaborative, a not-for-profit organization aimed creating international partnerships to advance the implementation of precision medicine. He has recently served as a member of the Board of External Experts for the NHLBI, the advisory council for the National Center for Accelerating Translational Science, chair of the review for Genome Canada's Large Scale Applied Research Competition in Genomics and Precision Medicine, and the World Economic Forum's Global Agenda Council on the Future of the Health Sector. He is a founder of Predigen, Inc and MeTree&You, Inc. He was previously Vice President of Molecular Medicine at Millennium Pharmaceuticals, Inc and a faculty member at Harvard Medical School. He is the founder and co-chair of the Global Genomic Medicine Collaborative, a not-for-profit (501c(3)) dedicated to implementing genomic medicine with a focus on low resource settings around the globe. The G2MC operates the International 100K+ Cohorts Consortium of which he is a founding cochair along with Peter Goodhand (GA4GH).

Abstract:

Multiple large cohort studies involving hundreds of thousands of people have recently been launched in several regions worldwide. They are of great value for studying diverse populations and key demographic subgroups, rare genotypes and exposures, and gene-environment interactions. Each cohort is constrained, however, by its size, ancestral origins, and geographic boundaries that limit the subgroups, exposures, outcomes, and interactions it can examine. Combining data across large cohorts to address questions none of them can answer alone enhances the value of each and leverages the enormous investments already made in them to address pressing questions in global health.

Leaders of large-scale cohorts, with support from the National Institutes of Health and the Welcome Trust and in collaboration with the Global Alliance for Genomics and Health (GA4GH, https://www.ga4gh.org/), and the Global Genomic Medicine Collaborative (G2MC, https://g2mc.org/) have come together to form the International Hundred Thousand Plus Cohort Consortium (IHCC, https://ihccglobal.org/). As of May 2020, IHCC comprises 103 cohorts in 43 countries involving nearly 50 million participants. Collaborative efforts to date have focused on developing a query-able Global Cohorts Atlas and data sharing platform, identifying and piloting high-priority scientific projects, and developing a charter and governance structure to foster collaborations (https://ihccglobal.org/wp-content/uploads/IHCC-Membership-Agreement.pdf).

Most of the cohort leaders had participants' consent to share data beyond the initial study investigators and were willing to share data more broadly, albeit with some limitations. The IHCC's vision is to accelerate the generation and application of population-based scientific knowledge on a global scale. Other cohorts, particularly from under-represented areas such as Africa, South America, and South Asia, are actively being sought to join the IHCC to assist us all in realizing the vast scientific potential of a worldwide collaboration for prospective cohort research.

Ewan Birney

EWAN BIRNEY, Ph.D., CBE, FRS, FMedSci, is Deputy Director General of EMBL. He is also Director of EMBL-EBI with Dr Rolf Apweiler and runs a small research group. Ewan completed his PhD at the Wellcome Sanger Institute with Richard Durbin. In 2000, he became Head of Nucleotide data at EMBL-EBI and in 2012 he took on the role of Associate Director at the institute. He became Director of EMBL-EBI in 2015. In 2020, Ewan became the Deputy Director General of EMBL. In this role, he assists the EMBL Director General in relation to engagement with EMBL Member States and external representation. Ewan led the analysis of the Human Genome gene set, mouse, and chicken genomes and the ENCODE project, focusing on non-coding elements of the human genome. Ewan's main areas of research include functional genomics, DNA algorithms, statistical methods to analyse genomic information (in particular information associated with individual differences in humans and Medaka fish) and use of images for chromatin structure. Ewan is a non-executive Director of Genomics England, and a consultant and advisor to a number of companies, including Oxford Nanopore Technologies, Dovetail Genomics and GSK. Ewan was elected an EMBO member in 2012, a Fellow of the Royal Society in 2014 and a Fellow of the Academy of Medical Sciences in 2015. In 2019, Ewan became a Board Member of the Biotechnology and Biological Sciences Research Council (BBSRC). He has received a number of awards including the 2003 Francis Crick Award from the Royal Society, the 2005 Overton Prize from the International Society for Computational Biology and the 2005 Benjamin Franklin Award for contributions in Open Source Bioinformatics.

Abstract:

Accurate and efficient data integration is an essential component for maximizing the utility of large-scale data sets for research, yet it remains a significant challenge to reliably harmonize complex data measurements across diverse cohorts on different computational infrastructure. Here I will describe some relevant aspects from some active research projects within our group and muse around the approaches that have worked well as well as providing example of current and future work into the improvement of method transferability across multiple environments.

Peter James

PETER JAMES, Sc.D., was trained in environmental health and epidemiology. Peter has focused his research on estimating the influence of spatial factors, including exposure to nature, the built environment, the food environment, air pollution, light pollution, noise, and socioeconomic factors, on health behaviors, mental health, and chronic disease. He has over a decade of experience working with large prospective cohort studies, including the Nurses' Health Studies, the Framingham Heart Study, and the Southern Community Cohort Study, where he has aided in the creation of many spatial exposure metrics and linked them to health data. He is developing methodologies to assess real-time, high spatio-temporal resolution objective measures of location and behavior by linking smartphone-based global positioning systems (GPS) and wearable device accelerometry data to understand how spatial factors influence health behaviors. Most recently, he is creating novel metrics of spatial factors by applying Deep Learning algorithms to Google Street View imagery.

Abstract:

This talk will describe the history of conducting epidemiologic research on environmental factors within the Nurses' Health Studies. The Nurses' Health Studies are nationwide prospective cohorts of over 230,000 female participants who have been

contributing data since the 1970s. Participant addresses have been geocoded at each questionnaire return, and researchers have linked multiple geospatial measures to these addresses, including air pollution, UV, light at night, noise, temperature, the built environment, green spaces, and socioeconomic factors. These geospatial measures have been linked to myriad health outcomes, including but not limited to cancer, cardiovascular disease, mental health, reproductive outcomes, cognitive function, physical activity, obesity, and mortality. Moving beyond these approaches, we are incorporating mobile health measures from smartphone and consumer wearable data to better understand the time-activity patterns of participants to develop personalized measures of exposure and behavior. We are also applying deep learning algorithms to nationwide and time-varying Google Street View imagery to create specific exposure metrics of the built and natural environment from a street-level perspective.

Gary Miller

GARY MILLER, Ph.D., is a leader in the exposome field, which strives to provide a systematic and comprehensive analysis of the non-genetic contributors to health and disease. He was the founding director of the HERCULES Exposome Research Center at Emory University, the first exposome-based research center in the U.S. He authored the first book on the topic, The Exposome: A Primer published by Elsevier. His research focuses on environmental drivers of neurodegeneration. His laboratory uses a variety of methods including transgenic mouse production, immunohistochemistry, neurotransmitter transport assays, high-resolution metabolomics, electrochemistry, and behavioral assays. His work is conducted in several experimental models from cultured neurons and C. elegans to mice and human studies.