USE OF GENOME DATA IN NEWBORNS AS A STARTING POINT FOR LIFE-LONG PRECISION MEDICINE

STEVEN E. BRENNER

Department of Plant and Microbial Biology, University of California
Berkeley, CA 94720, USA
Email: brenner@compbio.berkeley.edu

STEPHEN KINGSMORE

Rady Pediatric Genomic and Systems Medicine Institute
San Diego, CA 92123, USA
Email: SKingsmoreMDDSc@rchsd.org

SEAN D. MOONEY

Department of Biomedical Informatics and Medical Education, University of Washington
Seattle, WA 98105, USA
Email: sdmooney@uw.edu

ROBERT NUSSBAUM

Invitae Corporation
San Francisco, CA 94107, USA
Email: robert.nussbaum@invitae.com

JENNIFER PUCK

Professor, Departments of Immunology and Pediatrics, UCSF
San Francisco, CA 94143, USA
Email: Jennifer.Puck@ucsf.edu

Rare genetic disorders affect millions of individuals worldwide. Many of these disorders can take decades to correctly diagnose. Because of this, genome sequencing of newborns raises a substantial opportunity to identify genetic disorders before they present symptoms, and to identify patient risks at the start of life. Many of these disorders can take decades to correctly diagnose. Because of this, genome sequencing of newborns raises a substantial opportunity to identify genetic disorders before they present symptoms, and to identify patient risks at the start of life. This workshop will report on efforts to screen newborns using genetic sequencing technologies, and attendant biomedical informatics and computational biology approaches.

1. Introduction

> *Over the course of the next few decades, DNA sequencing will lead to each baby's genome being sequenced, and used to shape a lifetime of personalized strategies for disease prevention, detection and treatment.*
>
> FS Collins, Wall Street Journal, July 8, 2014.

Within the past year, the director of the NIH updated the 1990s projection by Walter Gilbert, who postulated that by 2030-2040 the parents of every newborn child in the developed world would leave the hospital with their child's genome on a CD-ROM.

The analytic cost of whole human genome (WGS) and whole exome sequencing (WES) continues to out-pace Moore's law and is currently approaching $1,000 for WGS. Given the high burden of treatable genetic diseases in children, there is considerable interest in using WGS to diagnose and treat genetic diseases of infancy, both as an extension of current newborn screening programs (NBS) and as a pioneering implementation of precision medicine for affected infants. Early diagnosis of genetic diseases can dramatically decrease morbidity and mortality, as evidenced by current, federally mandated newborn screening (NBS) programs for ~60 treatable genetic diseases, which identifies ~5000 affected babies per year at ~10 days of life. With the emergence of WGS and whole exome sequencing, the National Institutes of Health and state NBS programs are considering the value of genome sequence information in addition to currently performed tests with the goal of improving accuracy, reducing false positives and false negatives, or identifying disorders not currently screened for.

For NBS programs, some diseases screened for currently have positive predictive values well below 5%, so improving specificity could have a substantial positive societal impact. In order to evaluate the use of deep sequencing within context of newborns has many technical, computational, informatics and ethical questions that have arisen. This workshop will focus on challenges of population based genetic sequence diagnosis for newborn disorders, the future possibilities of WGS as an enabling technology for lifelong precision medicine, and current experience with use of diagnostic WGS in the delivery of precision medicine in acutely ill infants. The challenges addressed include:

- Determination whether accurate genome or exome sequences can be obtained using existing blood spot material and testing pipelines.
- Improvement in the annotation of pathogenic genetic variants in genes associated with metabolic and other disorders.
- Evaluation of whether genetic information can reduce false positive and false negative rates by identifying cases of nonpathogenic variants and heterozygous carriers of who are not anticipated to have symptomatic disease.
- Collection of patient-level rich phenotypic data through ontologies (and other means) and integration into national repositories and the electronic medical record.

- Improved capture and diagnosis of undiagnosed and affected infants for whom early intervention is warranted to minimize morbidity and mortality.
- Selection of disorders for screening, including and beyond the 31 conditions recommend by the HRSA advisory committee
- Ethical challenges surrounding widespread sequencing and storing genetic data of children for clinical diagnosis.
- Public education and engagement.

2. Workshop Presenters

Robert Nussbaum, MD
Invitae Corporation

Title: Newborn screening and genome sequencing

Newborn screening for inborn errors of metabolism was first instituted in the 1960's for phenylketonuria using blood obtained from newborns that is soaked into filter paper. Screening expanded over the next 30 years by the introduction of individual tests designed to screen for individual disorders until the 1990s when tandem mass spectroscopy (MS/MS) was first applied. MS/MS allows the identification of many disorders in parallel using a single assay that measures a host of abnormal metabolites in the newborn blood spot. The NSIGHT project, a four-site project funded by NICHD and NHGRI, is now looking at a potential next step in newborn screening: identifying many more actionable, treatable diagnoses in the newborn period through exome or genome sequencing of newborn blood spot DNA. Outstanding questions include:

- the positive predictive value of abnormalities found in newborn screening,
- sequencing's utility and cost as compared to standard mass spectroscopic methods to see whether it could replace standard screening,
- if not replacing current screening, could sequencing complement standard screening methods and make them more specific, and
- finally to explore the ethical and legal aspects of generating sequence information from newborns under a nearly mandatory public health regimen when some of these data may have significant medical relevance but are not germane to the mission of newborn screening.

Biography: Robert L. Nussbaum, M.D. is Chief Medical Officer of Invitae Corporation and also on the clinical faculty of the Dept. of Medicine, UCSF. As a board-certified internist and medical geneticist, Dr. Nussbaum has dedicated his career to improving the care of individuals with hereditary disorders. He has played leadership roles in the Cancer Genetics and Cardiovascular Genetics Programs at UCSF as well as serving as inaugural PI of the UCSF U19 project to investigate the role of deep sequencing as an adjunct to newborn screening. Dr. Nussbaum is a

member of the National Academy of Medicine and American Academy of Arts and Sciences and is past president of the American Society for Human Genetics. He was co-discoverer of the first inherited form of Parkinson's disease, and his basic research has focused on efforts to identify the pathogenesis and genetic contributions to the disease. Prior to joining UCSF, Dr. Nussbaum was chief of the Genetic Disease Research Branch of the National Human Genome Research Institute, National Institutes of Health.

Stephen Kingsmore, MB, ChB, BAO, DSc, FRCPath
President of the Rady Pediatric Genomic and Systems Medicine Institute.

Title: Integrating deep phenotyping and genome sequencing to enable precision medicine in neonatal intensive care units

Genetic diseases and congenital abnormalities are the leading cause of death both in neonatal intensive care units, and among infants in general (children aged <1 year). Until the advent of clinical WGS and whole exome sequencing (WES), timely molecular diagnosis of suspected genetic disorders had been largely precluded in acutely ill infants by virtue of profound clinical and genetic heterogeneity, and tardiness of results of standard genetic tests. However, it is now possible to decode an ill infant's genome in 24 hours (STATseq) and to use clinicopathologic correlation software to evaluate the likelihood that their symptoms are the result of any of the 4,500 known monogenic disorders. We and others have recently reported rates of molecular diagnosis of 25-73% in retrospective case series of infants and children with diseases of possible monogenic etiology by proband or trio WGS/WES. However, there are immense challenges in scaling STATseq for general use, and in timely implementation of STATseq as part of NICU workflows that have the potential to change outcomes in infants with genetic diseases.

Biography: Stephen Kingsmore grew up in Northern Ireland during the 'troubles'. He moved to the US after medical school and trained in internal medicine at Duke. After positionally cloning a few disease genes in academia in the '90s, he switched to the genomics industry for a decade. He was CEO of the National Center for Genome Resources in Santa Fe while it became a leader in the development of bioinformatic tools for next-gen sequencing for AgBiotech applications. From 2011-2015 he was the Director of Genomic Medicine and Executive Director of Medical Panomics at Children's Mercy Hospital in Kansas City. Recently he became President of the Rady Pediatric Genomic and Systems Medicine Institute. This is a new research institute which is affiliated with the Rady Children's Hospital and UCSD.

Jennifer Puck, MD
Professor, Departments of Immunology and Pediatrics, University of California San Francisco

Title: Using newborn dried blood spots to obtain deep sequence data for enhanced, early diagnosis

Our center has developed methodology for extracting DNA from dried blood spots obtained in newborn nurseries that is of sufficient integrity and quantity for whole exome or whole genome sequencing. Modifications to standard DNA processing have been made to optimize the generation of deep sequence data. Deep sequencing and analysis is possible using de-identified samples from infants with positive metabolic newborn screens. Moreover, additional archived newborn dried blood spot samples have been obtained with informed consent from individuals affected with clinically significant primary immune system disorders. The latter sample set addresses the question of whether newborn screening by deep sequencing could improve early detection of diverse immunodeficiency diseases, making possible optimal treatment and avoidance of infectious complications.

Biography: Dr. Puck is a Professor of Immunology in the Dept. of Pediatrics at UCSF and is also a member of the UCSF Inst. for Human Genetics and an associate of the Berkeley Innovative Genomics Initiative. Her basic and translational research program focuses on inherited human immune disorders. Noting the advantages in survival and outcome for infants with severe combined immunodeficiency (SCID) diagnosed early in life, Dr. Puck conceived and developed a newborn screening test using the universally collected dried blood spots to detect SCID. DNA extracted from the blood spots is assayed by PCR to quantitate T cell receptor excision circles (TRECs), a biomarker for the generation of a normal diverse repertoire of T cells. Absent or low TRECs suggest SCID. SCID screening, now adopted in over half of the states in the US, allows infants affected with SCID and other conditions with insufficient T cells to be detected early and treated. Dr. Puck directs the UCSF Jeffrey Modell Diagnostic Center for Primary Immunodeficiencies. She serves on the Medical Advisory Committee of the Immune Deficiency Foundation, the Committee on Primary Immunodeficiency Disease of the International Union of Immunological Societies, the Board of Scientific Councilors of NIAID, and the Steering Committees of the Primary Immune Deficiency Treatment Consortium (PIDTC) and the US Immunodeficiency Network (USIDNET). A member of the American Society of Clinical Investigation (ASCI), Association of American Physicians (AAP), American Pediatric Society (APS) and Institute of Medicine, she received the Abbot Award in Clinical and Diagnostic Immunology from the American Society of Microbiology in 2013 and the Colonel Harlan Saunders Award for Lifetime Achievement in Genetics from the March of Dimes in 2014.

Steven E. Brenner, PhD

Professor, Department of Plant and Microbial Biology, University of California, Berkeley and Adjunct Professor, Department of Bioengineering and Therapeutic Sciences, UCSF

Title: Analysis challenges of newborn genome sequences

A hallmark of current newborn screening programs is their outstanding sensitivity with false negatives on the order of one per million babies screened and impressive specificity. Both the sensitivity and the specificity are essential in order, respectively, to meet the public health goals of the program at a reasonable financial and human cost. This is achieved in large part through careful optimization of highly-sensitive mass spectrometry instruments which are detecting a molecular intermediate phenotype resulting from the genetic abnormalities. Identifying such diseases from genetic information is intrinsically more challenging. The analytic instrumentation for genome variant calling is less mature than the mass spec technology. More importantly, identifying disease phenotypes from genotypes can be vastly more challenging than identifying disease phenotypes from molecular phenotypes. Preparatory studies have revealed widespread limitations of even the most authoritative databases of mutations causing those diseases incorporated in newborn screening; they have numerous variants reported as pathogenic that are present in unaffected individuals. It seems likely that they also overlook variants that may cause disease. Computational predictive methods fare far worse. Health disparities may arise due to differential ability to reliably identify pathogenic mutations in individuals from different ethic backgrounds. This talk will include progress on methods to overcome these challenges, and approaches that may help present genome sequencing as a useful technology associated with newborn screening.

Biography: Steven E. Brenner is a Professor at the University of California, Berkeley, and also holds appointments at Lawrence Berkeley National Laboratory and at the University of California, San Francisco. As an undergraduate he studied in Walter Gilbert's laboratory at Harvard College. He received his M.Phil from the Department of Biochemistry at Cambridge University, and obtained a Ph.D. from the MRC Laboratory of Molecular Biology and Cambridge where he studied with Cyrus Chothia. After graduation Brenner had a brief fellowship at the Japan National Institute of Bioscience, followed by postdoctoral research supervised by Michael Levitt at Stanford University School of Medicine. Brenner's research is primarily in the area of computational genomics, covering topics in protein structure, RNA regulation, function prediction, metagenomics, and individual genome interpretation. He is founding chair of the Computational Biology graduate program at Berkeley. He is currently a director of the Human Genome Variation Society, and is a founding editor of PLoS Computational Biology. He has served two terms as a director of the ISCB and was a founding director of the Open Bioinformatics Foundation. His recognitions including being a Miller Professor, a Sloan Research Fellow, a Searle Scholar, an AAAS Fellow, and named the recipient of ISCB's Overton Prize.

Sean D. Mooney, PhD
Professor, Department of Biomedical Informatics and Medical Education, University of Washington

Title: Understanding the complex genetics of simple Mendelian traits

Most of the diseases included in newborn screening are considered classical Mendelian diseases. However, more careful study reveals that many if not all involve variable penetrance and expressivity, and several likely involve degrees of epistasis or other engagement of multiple genes. Our group has spent much effort development methods to identify and characterize pathogenic variants from sequencing projects. This talk will outline some of these challenges and successes in characterizing these more complex genetic relationships relevant to accurate newborn screening by genome sequencing. I will discuss our efforts to connect phenotype to disrupted molecular mechanisms to better predict clinically relevant mutations.

Biography: Prof. Sean Mooney is the Chief Research Information Officer (CRIO) of UW Medicine, a Professor in the Department of Biomedical Informatics and Medical Education, and an NIH funded researcher. Previous to his CRIO role, he was an Associate Professor and Director of Informatics at the Buck Institute for Research on Aging. He has a long history in managing the development of collaborative electronic systems supporting biomedical research. His interests focus on leading the next generation informatics tools for biomedical research and in understanding the underlying molecular causes of inherited genetic diseases and cancer. As an Assistant Professor, he was appointed in Medical and Molecular Genetics at Indiana University School of Medicine and was founder of the Indiana University School of Medicine Bioinformatics Core. In 1997, he received his B.S. with Distinction in Biochemistry and Molecular Biology from the University of Wisconsin at Madison. He received his Ph.D. in 2001 at the University of California in San Francisco under the mentorship of Dr. Teri Klein, and then was an American Cancer Society John Peter Hoffman Fellowship at Stanford University. He is funded by the National Library of Medicine and other NIH Institutes, mostly in the area of data science and translational medicine. He was part of the team that won the $150k 2000 Garage.com Student Business Plan Competition, where the proposed plan focused on web-based tools for drug discovery research.

3. Acknowledgments

We would like to thank the PSB 2016 chairs and Tiffany Murray of Stanford University for their support in organizing the meeting. Research was supported by NICHD and NHGRI under Award U19HD077627, NLM LM009722 (Mooney), the National Center For Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR000423 (Mooney) and

by Tata Consultancy Services. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.