NIH Strategic Workshop Report
Epigenetics and Genetics in the ECHO Program
February 20 - 21, 2018
Bethesda, Maryland

Planning Committee

Carole Ober, PhD
Chair, Department of Human Genetics
University of Chicago

Kent Thornburg, PhD
Chair, Director of Center for Developmental
Health
Knight Cardiovascular Institute
Oregon Health and Science University

Lawrence Brody, PhD
Director, Genomics and Society Division
National Human Genome Research Institute
(NHGRI)

Manjit Hanspal, PhD
Program Officer, ECHO
Office of the Director (OD)

Emily L. Harris, PhD, MPH
Deputy Associate Director
Epidemiology and Genomics Research Program
National Cancer Institute (NCI)

Tania Lombo, PhD
Program Officer, ECHO Office of the Director
(OD)

Kim McAllister, PhD
Genes, Environment, and Health Branch
National Institute of Environmental Health
Sciences (NIEHS)

Table of Contents

Background and Goals of ECHO Program

The National Institutes of Health launched the Environmental Influences on Child Health Outcomes (ECHO) program in 2016 with the overarching goal to enhance the health of children for generations to come. The ECHO program supports multiple synergistic, longitudinal studies by leveraging, harmonizing, combining, and performing innovative analyses from existing and new maternal/pediatric cohort data. ECHO comprises 62 grant awards; 110 principal investigators and over 1200 investigators in total; and academic and related institutions in 44 states, the District of Columbia and Puerto Rico. Together, more than 83 ECHO observational cohorts, with an anticipated combined enrollment exceeding 50,000 children from diverse populations across the United States, will leverage rich existing and new data and biospecimens.

One of the main deliverables of the ECHO program is the ECHO-wide Cohort data collection protocol, which specifies standardized collection of new data and sharing harmonizable existing data from 83 pediatric cohorts. Using the data, the goals of ECHO will be directly related to address the effects of a broad range of early life environmental exposures (e.g., physical/chemical, societal, psychosocial, behavioral, biological) on four key pediatric outcomes with high public health impact, namely pre-, peri-, and postnatal outcomes; upper and lower airway; obesity; and neurodevelopment; as well as an innovative fifth outcome, positive health, which reflects the positive attributes of healthy growth and development.

Within the core idea of the ECHO program lies the understanding of gene-environment interactions and the potential connection with the epigenome. Genes or environment are not usually sufficient to completely explain a disease process but instead they are one component of a larger complex molecular system that can contribute to the resulting phenotype. Generating and integrating -omics data from cells and tissues and exposure data are prerequisites for addressing the central questions of the ECHO program. ECHO program defines exposure period as conception through age 5 years while outcomes are throughout childhood and adolescence.

Executive Summary

The ECHO RFA set the expectation of genotyping the children to be the minimum programmatic requirement of genetic analysis for research participants in the ECHO-wide cohort protocol. The genotyping of parent-child trios and the addition of epigenetic profiling is optional. The analysis of the genetic constitution (genotyping) of ECHO children is an essential piece to understand how the environment and genes together influence children's health and development. Because of the important role of maternal influences on fetal well-being, understanding how the mother's genotype affects the fetus is also important. Therefore, the ECHO-wide cohort protocol requires the collection of samples for DNA isolation from mothers and children, as essential elements.

Additional approaches may be supported by ECHO to develop a greater understanding of the mechanisms through which genetics, the environment and their interactions impact health and disease outcomes (e.g., epigenomics, transcriptomics). A strategic workshop was held in February 2018 (see Appendix 1) to assist in planning the Genetics Core and define long term goals of the genetic and epigenetic studies in ECHO.

During the two-day meeting, subject matter experts identified key research opportunities, promising methodologies, and gaps that the ECHO program could address given the unique nature of its 83 cohorts. Participants included genomic and epigenomic, epidemiologic, social science, and population science researchers, NIH staff, government stakeholders and 15 ECHO investigators.

The twin goals of the ECHO workshop were to:

- Inform the ECHO Program and NIH leadership on scientific strategy, key questions and approaches for the future Genetics Core.

- Provide recommendations to define the long-term scientific opportunities on genetics and epigenetics within the ECHO-wide Cohort.

The present report summarizes the main scientific highlights discussed during the two days of the meeting, as well as recommendations provided by small group discussions during the second day. Key recommendations and potential impact of future research efforts at ECHO included the following:

- Array-based genotyping in all ECHO participants, with or without whole exome sequencing (WES), followed by centralized quality control and imputation, and making all genotypes should be made available to all investigators for downstream analyses.

- Whole genome sequencing (WGS) of a subsample of individuals from ethnicities or races that are included among the ECHO cohorts but not included in the 1000 Genome Project or TOPMed consortia, should be performed to establish panels for imputation of those participants' genotypes in ECHO.

- Epigenetics (DNA methylation arrays) should be performed in subsets with available age- and tissue-specific samples to create reference panels in cells relevant to ECHO, such as cord blood, placenta, etc., to facilitate imputation of epigenetic profiles in other samples (i.e., prediction by

genotypes). The same approach can be considered for transcriptomics, metabolomics, etc. in the future.

- Single cell sequencing, or single cell epigenetics, should be performed in subsets of samples to generate more accurate estimates of cell-specific expression or methylation for de-convoluting cell composition in complex cell mixtures (e.g., blood cells and placenta), parallel to studies described in #3 above.

- The development of methods for integrative analyses of 'omic data, including but not limited to transcriptome, epigenome, metabolome, microbiome data, is required to unveil the causality of the complex childhood outcomes that ECHO is seeking to understand.

- Store separated maternal and cord plasma samples to allow future levels of analysis as the technology to study Extracellular Vesicles unfolds.

Highlights from Scientific Sessions

The two-day workshop consisted of presentations and open discussions on each of the key scientific aspects related to genetics and epigenetics within the ECHO program. Presenters were asked to discuss their own science related to the session's theme and to give their perspective on how the ECHO program could respond to critical questions on how epigenetics and genetics play a role in the childhood outcomes that are a focus for this Program. It became evident that in the long-term future, the scientific community would benefit from the richness of ECHO data sets to better understand the impact of early life development on diseases in other life stages and through the lifespan.

The next section highlights discussions from all workshop sessions. All participants provided insights about the following expectations and key questions:

- EXPECTATIONS
  - Inform the ECHO Program and NIH leadership on scientific strategy, key questions and approaches for the future ECHO Genetics Core.

  - Provide recommendations to define the long-term scientific opportunities on genetics and epigenetics within the ECHO-wide cohort.

- KEY QUESTIONS
  - What would be the most informative approach (e.g., GWAS vs sequencing) for the genetics core?

  - What laboratory and statistical/bioinformatic infrastructure are needed to answer the key ECHO questions?

  - How should we leverage the diversity of the ECHO populations in prospective genetic analysis, for example to study the effects of ethnicity or race on disease trajectories?

Developmental Epigenetics & Environment Perspectives within ECHO

It is now clear that gene function is not based entirely on the DNA sequence inherited from parents. With new discoveries in gene regulation and interactions with microbial communities, we must understand the epigenetic influences that work in tandem with genetic background to determine how environmental conditions affect short- and long-term vulnerabilities to disease. ECHO is a unique program in the world, providing the opportunity to study a large number of cohorts that span the stage of pre-pregnancy, through pregnancy, infancy and childhood and beyond to find associations between a broad array of environmental exposures and positive and adverse health outcomes. The participants of the workshop agreed that there are several messages that should be considered by scientists and leaders in the ECHO program that reflect developmental aspects of the plastic epigenome. These include how the roles of the genome, epigenome and metagenome are modified by the effects of diet, social stress, and toxic chemicals. As an outcome of the workshop discussions on this topic, there are two essential research areas that ECHO should consider in the future are described as follows:

Key features of developmental epigenetics: The inherited genome is influenced by environmental factors that regulate DNA methylation patterns, changes in histone proteins, and expression of non-coding RNAs. In addition, during pregnancy, the mother, the placenta, and the fetus are exposed to living bacteria, bacterial wall fragments, and bacterial nucleic acids that interact with cells in each compartment. Recent investigations have suggested that the placenta and fetus is not sterile and that the maternal, placental, and fetal microbiomes are involved in normal and perhaps, abnormal development of offspring. The interaction of these modifiers will be different from one individual to the next, in part because of the genetic make-up of the host genome. Methylation patterns in the fetus are established in early stages of development when highly methylated genes in sperm and egg undergo erasure of marks early in embryogenesis; erased marks are then reinstated with further development. During these early developmental stages, the embryo and fetus are the most vulnerable to exposures that could potentially alter the epigenetic status of a broad range of cell types—each in their own way. Thus, it would be ideal to assay specific changes in the epigenome as early in development as possible to differentiate between beneficial changes that are markers of positive health trajectories and aberrant epigenetic changes that they may one day be reversed. Placenta, umbilical cord tissue, and cord blood provide the first practical opportunity to sample tissues that might reflect the infant's prenatal exposures.

Chemical toxins in the environment may alter epigenetic status of mother and offspring: Among the hazardous air pollutants that are encountered by parents and children are arsenic, benzene, cadmium, chromium, formaldehyde, lead, mercury, and polycyclic aromatic hydrocarbons (PAH). These compounds have specific effects including interference with organogenesis and elevation of disease risks in offspring. As the world becomes more industrialized, these chemicals are becoming more prevalent in the environment. Epigenetic studies to determine how environmental factors affect the regulation of genes that are important to the health of offspring could be informative. The research community should target a systematic protocol that cohorts can follow to standardize the investigation of environmental toxins as they relate to maternal and offspring health. The ECHO-wide cohort protocol and its design represents an excellent opportunity to fill this knowledge gap by keeping in mind chemical toxins and epigenetic factors that alter mother and offspring gene expression and health outcomes. Such analysis would propel useful results to the ECHO program and community in general.

Assessing Genetic Variation in ECHO

Underlying all major goals of ECHO is the availability of high quality, genome-wide characterization of genetic variation in all participants. While straightforward in concept, its implementation poses many challenges, including the choice of platform (whole genome sequencing, exome sequencing, genotyping arrays), type of variation surveyed (SNVs, indels, CNVs), and implementation of bioinformatic tools (data processing, variant calling and imputation pipelines, analysis tools). Each of these considerations was discussed at the workshop and summarized below.

Choice of platform: Whole genome sequencing (WGS), even at low coverage, will provide the most information on rare and common genetic variation and provide the greatest flexibility for downstream analyses, but at the highest cost. Whole exome sequencing (WES) would provide excellent resolution of all variation within the coding region and potentially identify rare variants with larger effect sizes. However, because >90% of the variants associated with common, complex phenotypes lie outside of the exons, WES would miss most of the variation relevant to the phenotypes and traits relevant to ECHO. In contrast, genotyping arrays survey common variation (>1%) across the entire genome and newer platforms provide good coverage in worldwide, ethnically diverse populations. For example, in the TOPMed experience, over 100 million missing genotypes (SNPs and indels) were imputed with high accuracy down to minor allele frequencies of 0.0001; this approach misses less than 1% of the variants detected by WGS using their robust reference panels. Imputation accuracy and all downstream analyses of genotype data will be vastly improved if all participants are genotyped on the same platform and imputation is performed on large sample sizes. Adding WES to genotyping with arrays would increase the ability to detect rare coding variants with potentially larger effects.

Types of variation: Three categories of variation were discussed, single nucleotide variants (SNVs), insertion/deletion variants (indels), and copy number variation (CNVs). The first two types of variation are well interrogated by all three of the platforms described above and represent the majority of variation in human genomes. Structural variations, including CNVs, are more difficult to call and require longer sequence reads and different sequencing technologies than those required for SNV and indel calling. At the present time, the most optimized technologies for studying structural variants still miss a significant portion of the variants. Because of the nature of this type of variation, such as higher mutation rates compared to SNVs, imputation of CNVs will be less reliable and require extensive methodological development. Thus, although structural variation may be important for specific outcomes, such as autism or developmental delay, surveying this class of variation in all the ECHO cohorts is hard to justify at this point.

Implementation of bioinformatic tools: It was recommended that all genotype data be processed at one location, where genotype QC, imputation, and variant annotation are performed before data are shared back with the ECHO investigators. The utility of these data would be further enhanced by visualization platforms that facilitate the 'look-up' of any variant in any cohort or across cohorts that includes basic annotations with respect to location and putative function.

Approaches to Integrated -Omics in ECHO

Generating and integrating -omics data from cells and tissues and exposure data beginning in utero and through adolescence (at least) are prerequisites for addressing the central questions of the ECHO Consortium, but present significant challenges. Several concepts were discussed about how best to leverage the unique characteristics of ECHO and address challenges that would benefit researchers more broadly.

Because it is not feasible to generate primary multi-omic data on relevant cells types in all ECHO participants, strategies for imputing -omic data based on well-defined reference panels were discussed. This approach requires creating reference panels in a subset of individuals that utilize genetic variation to impute -omic readouts (e.g., methylome, transcriptome) in relevant cells or tissues (e.g., placenta, cord blood, neonatal blood spots), as has been done in the GTEx program for imputing gene expression in adult tissues. Once reference panels are created, then genetic data alone can be used to impute predicted methylation and gene expression in these relevant cell types. One can then ask, which traits or exposures correlate with methylation or gene expression patterns at birth or what is the expected methylation or gene expression pattern predicted by genetic variants associated with a trait of interest. A major advantage of this approach is that once the panels are created only genotypes are needed to predict cell- and age-specific -omic readouts in individual studies. The same approach could potentially be used for other -omic readouts, such as the metabolome and proteome.

The ages of ECHO participants represent a critical slice of the total lifespan. Therefore, using data from large biobanks can provide invaluable data on the full range of effects of the genetic variants or -omic readouts measured early in life on health and disease throughout life. In particular, genetic predictors of -omics, exposures, or early life phenotypes can be used in a phenome-wide association study (PheWAS) in large biobank data to assess whether these early life predictors also predict health outcomes throughout life, such as heart disease and cancer. In particular, assessing the lifelong pleiotropic effects of genetic variations associated with methylation levels, transcript levels, or exposures in early life would inform our understanding of the early life origins of health outcomes through the life course. ECHO has the opportunity to contribute to the understanding of genetic features during ages 0-5, which then can be applied to information obtained from biobanks to determine the degree to which those same alleles influence later life outcomes. The latter will be a key contribution from ECHO to the scientific basis of the Developmental Originis of Human Disease (DoHaD) research.

Most tissue samples collected in the ECHO cohort are composed of many cell types. Better (larger) panels for de-convoluting the heterogeneous cell composition of placenta tissue and various blood samples collected at different ages (e.g., cord blood, cord blood mononuclear cells [CBMCs], whole blood, peripheral blood mononuclear cells [PBMCs]) would provide better estimates for predicting cell proportions, and therefore improve ability to differentiate true differences DNA methylation levels, transcript abundance, etc. per se, from those that are confounded by variation in cell composition. Alternatively, single cell sequencing in fewer samples may provide the best resolution for predictive models.

As in most large birth cohort studies, standardized tissue collections, especially of placenta tissue, is essential for minimizing the type 1 error rate in all downstream analyses. Thus, great attention to sample collection protocols should be made for all prospective collections.

Statistical Approaches with relevance for ECHO

A primary goal of the ECHO consortium is to disentangle the effects of genes, prenatal and early life environments (defined broadly to include physical and chemical, social, behavioral, and biological), and their interactions on lifelong health outcomes. Several approaches for addressing these goals, along with their advantages and disadvantages were discussed. Three approaches were considered: transgenerational inheritance, Mendelian randomization, and gene-environment interactions. These approaches provide complementary information on the effects of environmental exposures or other modifiers and of biological or genetic factors on disease or health outcomes.

Studies of transgenerational inheritance aim to disentangle environment-mediated from genetic-mediated inheritance of transmitted traits. However, because of the significant methodological and practical challenges of 'proving' transgenerational inheritance in humans, especially in the absence of germ cells (sperm or oocytes) from parents and multigenerational families, this approach is less amenable to addressing questions in the ECHO cohorts.

Mendelian randomization can be used to infer causality and in combination with mediation analysis can increase our understanding of mechanism. These approaches consider the pathways through which exposures, as predicted by instrumental variables (genetic variation such as SNVs, CNVs in this case) are associated with a trait of interest (e.g., obesity, type 2 diabetes). Mendelian randomization, with or without mediation analysis, will be a powerful tool for addressing a question central to the ECHO mission: Do genetic variants exert their effects directly on traits of interest or through their primary effect on intermediate phenotypes, 'exposures', and/or epigenetic marks of exposure. The power of these approaches will be enhanced if all ECHO cohort participants are genotyped on a common platform and genotypes imputed using the same reference panels, imputation software, quality control checks, and analytical pipelines.

Another approach that will reveal potential sites of gene-environment interactions is methylation (or other epigenetic mark) quantitative trait locus (meQTL) mapping. Epigenetic marks (e.g., methylation) can serve as molecular sensors of exposures and provide clues about mechanism; studies in cord blood or placenta especially can provide information on the in utero "exposome." In particular, genetic variants that are associated with epigenetic marks (e.g., meQTLs) may be the sites of gene-environment interactions. The power of these studies in the ECHO cohorts will be greatly enhanced if all individuals are both genotyped and assessed for the epigenetic mark using the same methods, including choice of cell types, collection of specimens, genotyping platform, and analysis tools.

Addressing Social and Behavioral Aspects within ECHO

Among the goals of the ECHO program is to better understand genetic influences on childhood health and development and how gene expression patterns are influenced by environmental factors. This conference addressed issues regarding the importance of social determinants during developmental stages of life, the mechanisms by which stresses can affect children, the barriers that make such research difficult, and potential strategies for future investigations within the field. Desired outcomes of ECHO studies include finding associations of growth patterns, social stressors, consequent epigenetic changes, and neurological patterns that predict adverse outcomes; and finding factors that can be modified to ameliorate or prevent negative consequences in children as they become adolescents and adults.

The Significance of Social Stress and Adverse Outcomes in Children: Epidemiological data have demonstrated that associations between social circumstances and children's mental health and behavior al stresses associated with low social economic status during fetal life, in infancy and in early childhood lead to lifelong vulnerabilities for chronic diseases such as obesity, diabetes, and heart disease. The physiological underpinnings of such environmental stresses have been demonstrated in animal models, but it has been difficult to design studies in infants and children to discover associations of specific adverse environmental factors with specific outcomes. However, there are well accepted examples of associations between social adversities and detrimental health outcomes. Growing up in low socio-economic status is associated with a pro-inflammatory immune profile, which is consistent with studies showing that stressors in the prenatal environment lead to pro-inflammatory gene expression profiles in offspring. Thus, inflammation is a key biological outcome of exposure to a stressful environment

Genetic and Epigenetic Mechanisms as Determinants of Adverse Outcomes: Distinguishing the roles of genetic background and epigenetics in humans is increasingly possible. There are many studies showing differences in DNA methylation patterns associated with poor social circumstances as well as with specific chemical exposures. These data suggest that such interactions are important targets for study in ECHO cohorts.

Barriers and future strategies to study of social determinants of adverse outcomes: Distinguishing the roles of family dynamics from the effects of poverty, crime, and poor educational systems is difficult. In addition, poorer communities may have different exposures to environmental chemical toxicants that complicate the analysis of social circumstances. Finding special opportunities within ECHO where it is possible to disaggregate these factors may be important in sorting the individual components that underlie the social determinants of adverse outcomes.

In order to quantify the role of individual stressors that alter childhood mental health and later adverse outcomes, several strategies that can be considered: twin/adoption studies, studies where economic status of a community has changed, studies where school systems have changed, peer behavior and indices reflecting peer contexts. It might be possible to acquire data regarding specific aspects of social determinants that can be associated with single tissue (blood cells or buccal swabs) epigenetic profiles, as well as neurobehavioral studies (ranging from simple questionnaires to MRI). Because diet is a powerful regulator of epigenetic processes, collecting dietary data may be highly important. ECHO should consider different aspects when aiming to include social determinants within the scope of future research: effect size, timing of sampling, availability of surrogate tissues, tissue heterogeneity, choosing

epigenetic technology to yield the information being sought, sex, ethnicity, age, statistical power, type of analysis, and statistical approaches to be used.

Ethnic Diversity at ECHO: Impact on Design

As the ECHO program is designed to learn as much as possible about early life environmental stressors and adverse outcomes among children, it is desirable to interpret data with full consideration of ethnic and racial backgrounds. To accomplish this goal, it will be important to consider both the genetic variation associated within identifiable social/racial groups and to gain as much information as possible regarding epigenomic associations with specific stressors within those groups. There are two features of this topic that offer new insight into potential protocol design to capture this crucial information: the use of genetic diversity associated with evolving ethnic and social groupings and using ethnicity as a substrate for assessing epigenetic responses to environmental exposures.

Genetic ancestry can empower genomic studies: Modern genomic analysis has led to a greater understanding of the relationships between genes and geography. This has allowed commercial companies to estimate a person's ancestry based on their genotypes. Rare and common variants have been used to reveal population structures within Europe and in other regions of the globe. It is now possible to combine genetic information, which can inform ancestry estimates, with disease outcomes. Such modeling might become an important tool for the ECHO-wide cohort protocol. Workshop participants highlighted the need to understand ancestry as a tool for discovering disease-associated variants, but ancestry estimates by themselves may provide valuable information on disease risk trajectories that are influenced by both genetic and environmental factors correlated with race, ethnicity, or ancestry. This would also inform our understanding of biological and social determinants of health in children during development.

Epigenetic factors associated with exposure to stressors: Different ethnic groups have very different rates of disease. Some of these differences are due to genetic risk variants that may occur at different frequencies between groups. However, there may also be epigenetic changes that add risk for disease. For example, there are a number of changes in methylation marks among children exposed to maternal smoking, as well as cadmium, lead, and arsenic. ECHO cohorts offer the opportunity to study a number of epigenetic marks taking into account race and ethnicity, as described above. These studies could include assessments of DNA methylation, non-coding RNAs, and histone modification. Future technologies will offer better definitions of accessible chromatin and states of differentiation using single cell analyses of relevant tissues.

Emerging Technologies of relevance for ECHO

It is important for the ECHO program to ensure that cohort studies are gleaning as much information as possible related to the genetic and epigenetic determinants of vulnerability for adverse outcomes in children. To this end, cohort leaders must be aware of the technology landscape and be ready to include new technologies that will benefit data acquisition and processing and provide new insights into disease vulnerability. Two aspects of technology that are not yet in routine use by most laboratories include those that target: 1) enhancers as tools for understanding the effects of environmental stressors and providing new avenues for therapies; and 2) extracellular vesicles that are released from the placenta into maternal blood to evaluate placental/fetal health, evaluate maternal targets, diagnose maternal disease severity and provide opportunity of amelioration of adverse maternal or placental conditions.

Enhancers are 'promoters' of gene promoters that can act from distal sites, and are now emerging tools for understanding disease processes. They play critical roles in regulating gene expression depending on the tissue and developmental stage, particularly in response to environmental conditions. Experts at the ECHO workshop shared recent technological advances where enhancers can be functionally characterized; enhancers and ripe for study as therapeutic agents. Although the use of enhancer analysis is not at the stage where it can be applied to ECHO cohorts at present, it is likely that assays for enhancer activity will be developed that will greatly increase our understanding of environmental cause of disease vulnerability in children.

The human placenta releases proteins and other molecules, microvesicles, exosomes and apoptotic bodies into the maternal circulation beginning as soon as the maternal-placental circulation is established. Extracellular vesicles (EVs) may be as small as the low nanometer size range. There is evidence from studies in cancer metastasis that EVs may influence which distal sites metastatic cells will invade. New sorting methods are rapidly being developed and one can expect that rapid detection of placental signaling molecules in maternal plasma will be a technology that will become routine within the decade. Given the collection of blood and cord blood plasma from pregnant women at ECHO, investigators could propose new hypothesis to understand molecular communication between the mother and fetus mediated by extracellular vesicles. Workshop participants agreed that storing separated maternal and cord plasma samples at ultracold temperatures will allow new levels of analysis as the technology unfolds.

General Recommendations

At the end of the two-day workshop, all participants deliberated on the immediate and long-term recommendations for the NIH ECHO Program. All agreed on the tremendous contributions of the program and how important a role genetics and epigenetics will play to achieve its overarching goals. While the rich discussion brought interesting and diverse viewpoints on many topics, those listed below were accepted as high priority by the general audience:

Recommendations for ECHO's immediate future:

- Genotyping in all ECHO participants should be performed with an array-based technology, with or without exome sequencing, followed by centralized QC and imputation, and making all genotypes available to all investigators for downstream analyses.

- For ethnicities or races that are included among the ECHO cohorts but not included in the 1000 Genome Project or ToPMed consortia, whole genome sequencing of a subsample of these individuals will be required to establish panels for imputation of those participants in ECHO.

Recommendations for ECHO's long-term future:

- Epigenetics (methylation patterns) and possibly transcriptomics should be performed in subsets of samples to create reference panels in cord blood, placenta, etc. to facilitate imputation in cells and time points relevant to ECHO

- In-depth integrative analyses of transcriptomic, metabolomic, and microbiome data will help to unveil the causality of the complex childhood outcomes that ECHO is planning to understand.

- Single cell sequencing in subsets of individuals to generate more accurate estimates of cell-specific methylation or expression for de-convoluting cell composition in complex mixtures (blood cells, placenta) will improve interpretation of analyses of these tissues in ECHO.

- Store separated maternal and cord plasma samples to allow future analyses as the technology to study Extracellular Vesicles unfolds.

Appendix

Workshop AGENDA: Day One (Tuesday, February 20, 2018)		
8:00 am- 8:30 am: Registration and Logistics		
8:30 am – 8:45 am	**Opening Remarks**	**Francis Collins, NIH**
8:45 am – 8:55 am	ECHO Program Overview	Matthew Gillman, ECHO Program Director
8:55 am – 9:05 am	Charge to Workshop Members	Carole Ober (University of Chicago) Kent Thornburg (OHSU)
9:05 am - 10:40 am: Developmental Epigenetics and Mendelian Randomization		
9:05 am – 9:30 am	Developmental Epigenomics & Environmental Exposures: Opportunities to Explore Mechanisms Driving Inheritance & Adaptation in the ECHO Cohort	Kjersti Aagaard (Baylor College of Medicine)
9:30 am – 9:55 am	Challenges and Opportunities for Investigating the Influences of Paternal and Maternal Exposures on the Child's Epigenome in Humans	Andrea Baccarelli (Columbia University)
9:55 am – 10:20 am	Using Mendelian Randomization Approaches in Mother-Child Cohorts	Marie France Hivert (Harvard Medical
10:20 am – 10:40 am: Discussion		
10:55 am – 12:35 pm: Interaction of Genetics and Epigenetics		
10:55 am – 11:20 am	Experiences from Imputing Transcriptomes for Association Studies in Biobanks: How Might Biobanks Help ECHO and How Might Imputation Extend Discovery from Epigenome	Nancy Cox (Vanderbilt University)
11:20 am – 11:45 am	Utility of Genetic and Epigenetic Data in Understanding and Improving Child Health	Daniele Fallin (Johns Hopkins University)
11:45 am – 12:10 pm	Using Epigenomics and Prospective Birth Cohorts to Understand Causes of Childhood Cancer	Zdenko Herceg (International Agency
12:10 pm -12:35 pm: Discussion		
12:35 pm -1:35 pm: Lunch on your own		
1:35 pm – 2:45 pm: Ethnic Diversity - Impact on Design		
1:35 pm – 2:00 pm	Ethnic Variation in Epigenetic Response to Environmental Exposures	Catherine Hoyo (North Carolina State
2:00 pm – 2:25 pm	Leveraging Diversity to Empower Genomic Studies	Eimear Kenny (Icahn School of Medicine at
2:25 pm – 2:45 pm: Discussion		
3:00 pm – 4:15 pm: Emerging Technologies in Genetics and Epigenetics		
3:00 pm – 3:25 pm	Functional Characterization and Therapeutic Targeting of Gene Regulatory Elements	Nadav Ahituv (UCSF)
3:25 pm – 3:50 pm	New Insights into Maternal-Placental-Fetal Communication	Yoel Sadovsky (Magee-Women's Research

	3:50 pm – 4:15 pm: Discussion	
4:15 pm – 5:15 pm	Day One Open Discussion and Wrap-Up	Kent Thornburg (OHSU), Carol Ober

Workshop AGENDA: Day Two (Wednesday, February 21, 2018)		
8:30 am – 8:40 am	**Opening Remarks**	**Matthew Gillman**
8:40 am – 8:55 am	Recap and Charge for the Day	Kent Thornburg (OHSU) Carole Ober (University of Chicago)
8:55 am - 10:05 am: Social and Behavioral Effects on Health Outcomes		
8:55 am – 9:20 am	What do we Mean by a Child's Environment?	Philip Shaw (National Human Genome
9:20 am – 9:45 am	The Social Epigenome in Human Health and Disease	Michael S. Kobor (University of British Columbia)
9:45 am – 10:05 am: Discussion		
10:20 am – 11:15 am: Genomic Data at ECHO		
10:20 am – 10:45 am	Challenges and Opportunities for Large-Scale Genetic Analysis: Lessons from Sequencing 100,000 Genomes	Goncalo Abecasis (University of
10:45 am – 11:05 am	Different Aspects of Genetic Diversity in Man and Mouse	Charles Lee (The Jackson Laboratory
am – 11:15 am: Discussion		
11:15 am -12:15 pm: Working Lunch		
12:15 pm- 2:30 pm: Group Activity: Discussion on Recommendations and Opportunities for ECHO Program		
2:45 pm – 3:00 pm	Reflections on the Workshop Goals	Kent Thornburg (OHSU)
3:00 pm – 3:10 pm: Directors' Next Steps and Closing Remarks		
3:10 pm: Adjourn		
3:00 pm – 4:15 pm: Emerging Technologies in Genetics and Epigenetics		