Towards Incorporating Genetics in the ECHO-wide Cohort

Council of Councils
7 September 2018

Matthew W. Gillman, MD, SM
Director, Environmental influences on Child Health Outcomes (ECHO)

Lynn R. Goldman, MD, MS, MPH
Chair, ECHO External Scientific Board
Dean, Milken Institute School of Public Health, George Washington University

External Scientific Board

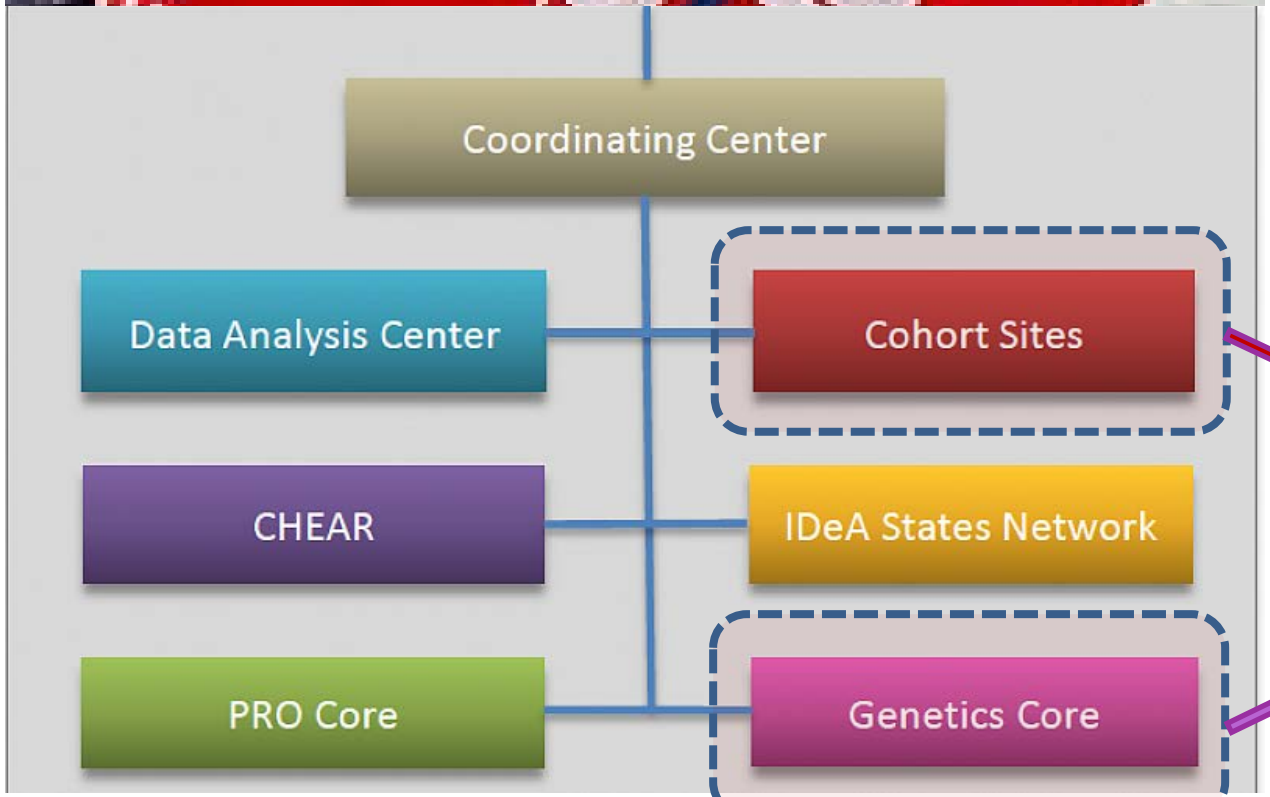
Initial Membership

- Working group of Council of Councils
  - 1 Council member
    - Children's Environmental Health Network
  - 3 Academic leaders
    - Genetics, toxic environment, neighborhood and social factors
    - NCS, IOM, FDA, CDC, NIH, Gates, etc.
  - 1 Parent, nominated by March of Dimes
  - 1 AI/AN representative, nominated by NIH Tribal Advisory Council
  - 1 Clinical trials expert

External Scientific Board

Requested Counsel

- Ensuring early and sustained successes
- Using funds wisely
- Attending to numerous strata of stakeholders
- Building a culture of collaboration and synergy
- Harmonizing data across disparate cohorts
- Capitalizing on expertise within as well as outside NIH
- Incorporating all ECHO components under one umbrella
  - Genetics Core



Today's focus

ECHO Overall Scientific Goal

Answer solution-oriented questions about effects

of

broad range of early environmental exposures

on

child health and development

Health Outcomes

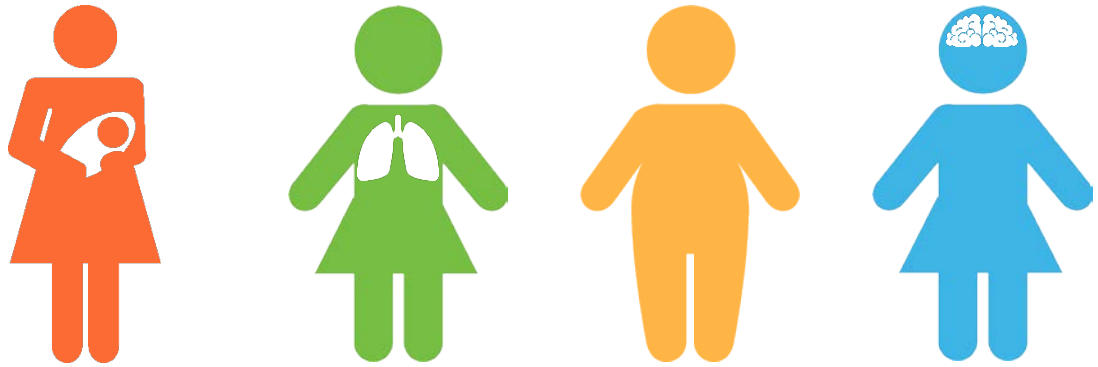
Focus on high-impact conditions throughout childhood and adolescence

PRE-, PERI- AND POSTNATAL

UPPER AND LOWER AIRWAY

OBESITY

NEURO- DEVELOPMENT



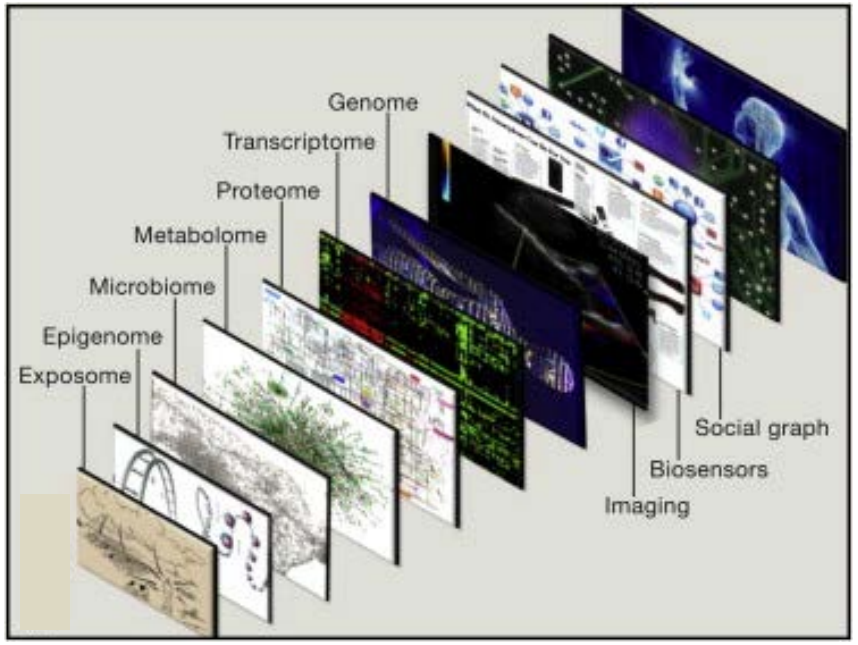
POSITIVE CHILD HEALTH

ECHO Cohorts

- 7 years starting FY2016
- Create ***ECHO-wide Cohort***
  - Start with existing cohorts of mothers and children
    - All continue follow-up of children
    - Some also still recruiting
  - Establish single data platform to conduct etiologic and prediction research
  - Harmonize existing measures & standardize new measures
  - >50,000 children and their families

The ECHO-wide Cohort

Many people, many layers of data, many stages of life course

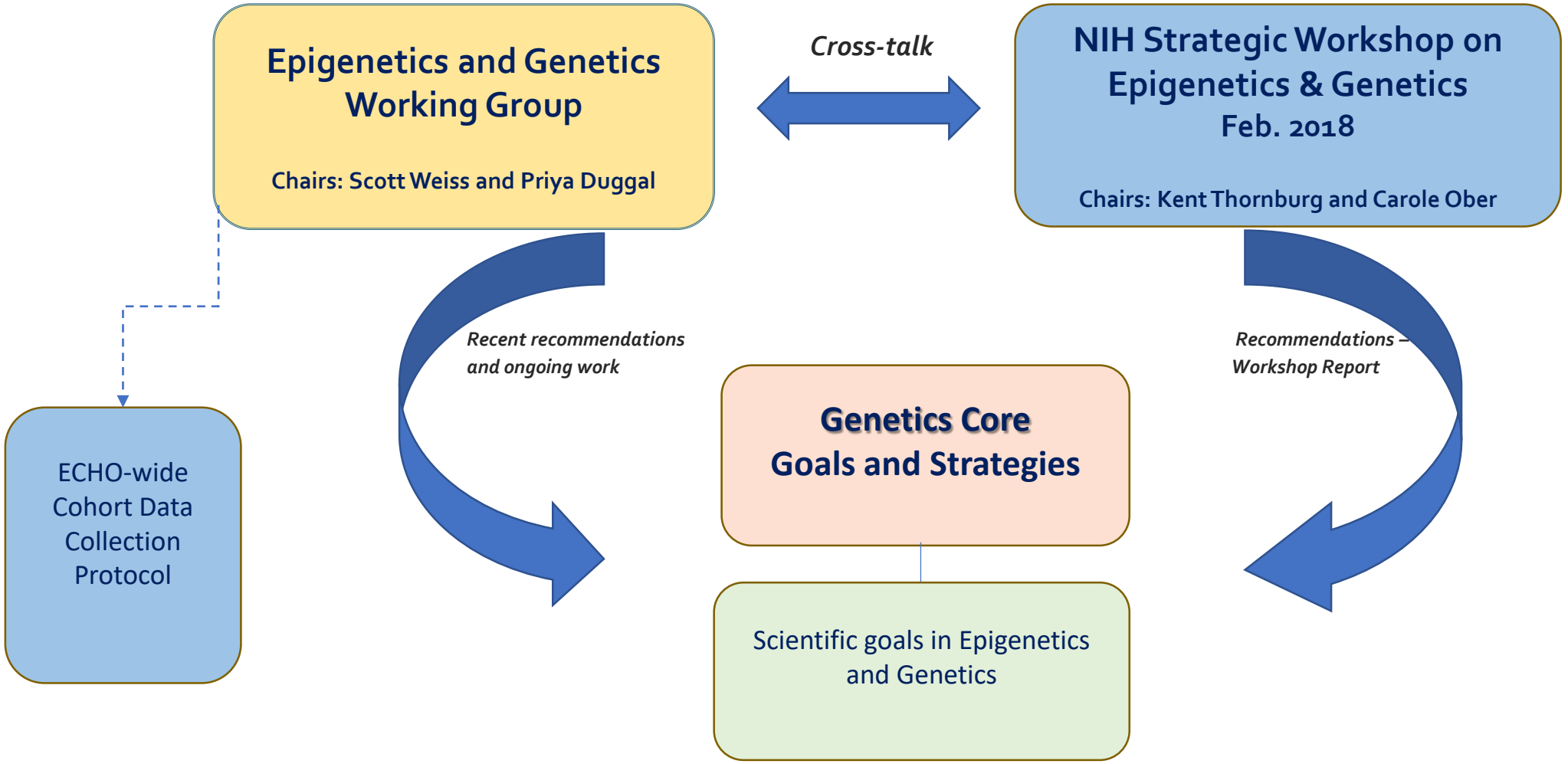


ECHO-wide Cohort

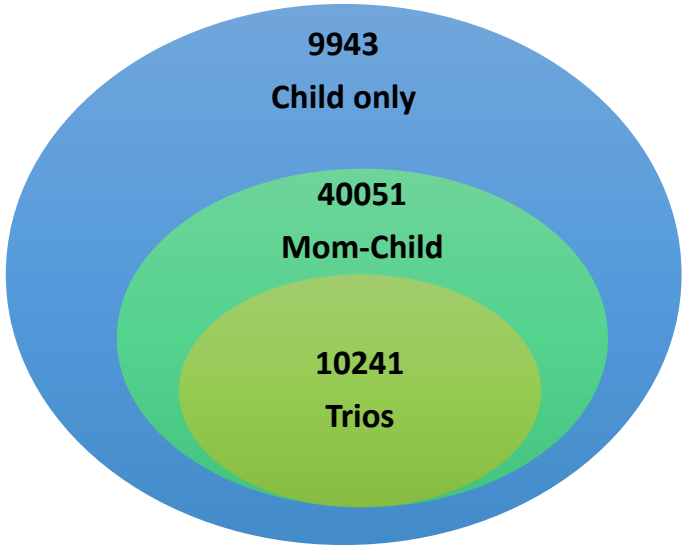
- Steering Committee Ratified Data Collection Protocol
  - In the field fall ’18
  - Genetics is an essential element
    - Core services for quality, harmonization, timeliness
    - 50,000 children
      - + moms (most)
      - + dads (few)
  - Epigenetics, other ‘omics are optional, recommended
    - [No core services]

Strategic Planning for Genetics Core

for the ECHO-wide Cohort



DNA Availability from all 84 Cohorts	Sample Size	Number of cohorts Contributing
Children		
Number of **existing** children	33955	57
Expected number of **existing** children	10191	39
Expected number of **new** children	16059	47
Total	**60,205**	**80 unique cohorts**
Biological Mothers		
Number of **existing** Moms	28452	48
Expected number of **existing** Moms	9451	37
Expected number of **new** Moms	15198	43
Total	**53,101**	**78 unique cohorts**
Biological Fathers		
Number of **existing** Dads	6683	22
Expected number of **existing** Dads	1702	21
Expected number of **new** Dads	1856	27
Total	**10,241**	**41 unique cohorts**
Child- Mother- Father Biological Triad		
Number of **existing** Trios	6683	21
Expected number of **existing** Trios	1702	24
Expected number of **new** Trios	1856	28
Total	**10,241**	**41 unique cohorts**
Child –Mother Biological Dyad (child-mother dyad should only be reported if there is not a parental trio)		
Number of **existing** mother-child Dyads	16757	44
Expected number of **existing** mother-child Dyads	8144	33
Expected number of **new** mother-child Dyads	15150	41
Total	**40,051**	**77 unique cohorts**

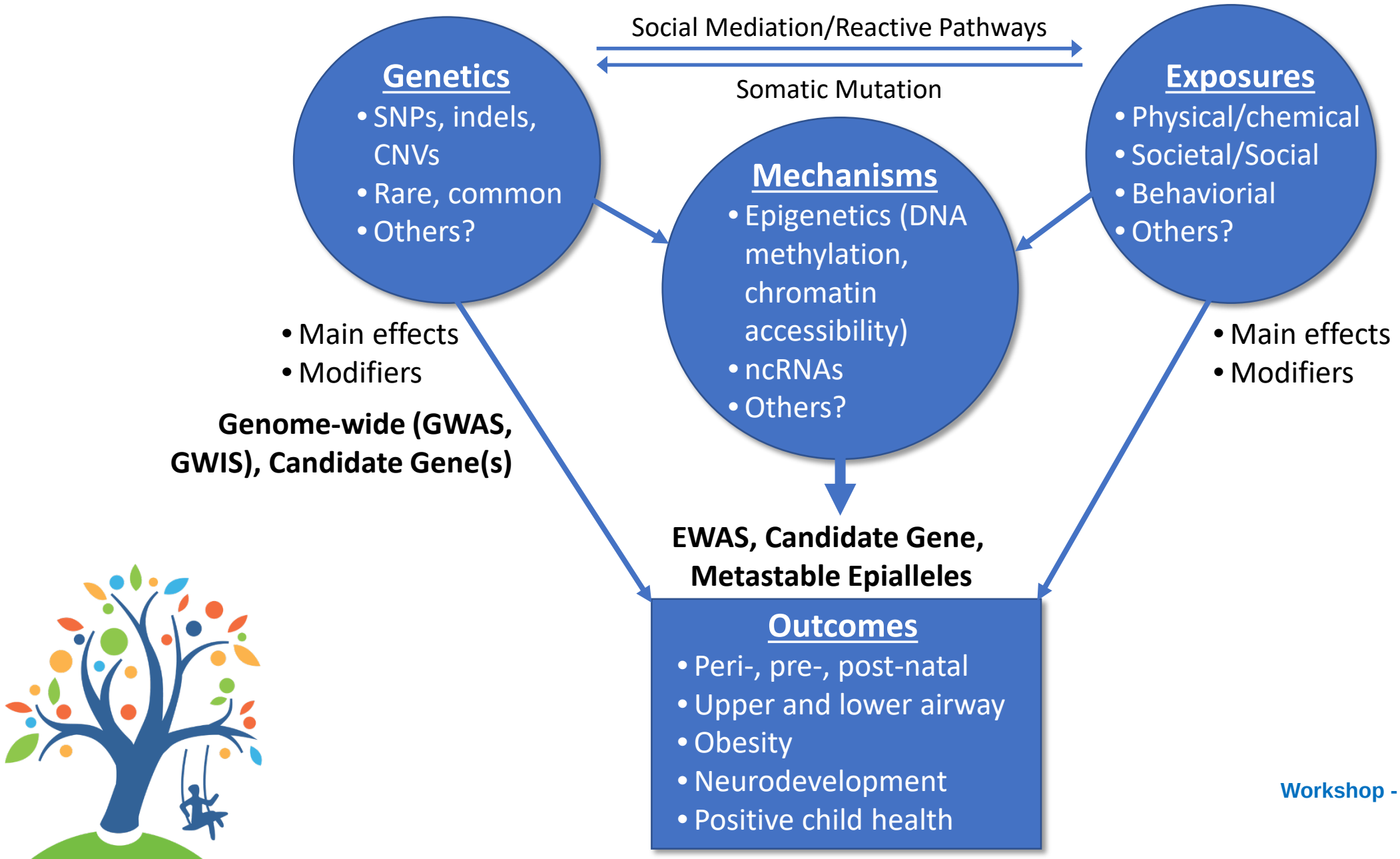


Race/Ethnicity among ECHO cohort participants with existing genetic data

	Non Hispanic	Hispanic
White	46%	14%
Black	14%	1.1%
Asian	2.5%	0.2%
American Indian/Alaskan Native	2%	0.5%
Multiple Races	5%	5%
Native Hawaiian/Pacific Islander	0.1%	0.1%
Total	73%	27%

- **The majority of samples are from self reported White ancestry, both Hispanic and non-Hispanic.**
- **Self reported Black individuals are the next largest group, with a total of 15% of samples.**
- **10% of individuals self-report as multiple race.**
- **Asians represent 2.7%**

Concept Map for ECHO Workshop



Workshop Considerations

1. Genotyping of the ECHO children and moms is the minimum programmatic requirement for genetic analysis.
2. Because of the important role of maternal influences on fetal well-being, genotyping the mother is also critical.
3. Additional approaches (e.g., epigenomics, transriptomics) may lead to greater understanding of the mechanisms through which genetics, environment, and their interactions impact health and disease outcomes
   - Optional in ECHO Cohorts

Strategic Workshop Goals

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

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

Scientific Opportunities and Recommendations

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 (children, mothers and potentially fathers).

The workshop recommended:

Array-based genotyping in all ECHO participants (± exome sequencing), with centralized QC as well as imputation, and making all genotypes available to all investigators for downstream analysis

Additional Recommendations

1. **Whole genome sequencing** of subsamples of individuals from ethnicities or races that are not represented in the 1000Genomes or TOPMed consortia to establish panels for genotype imputation in those participants and ECHO cohorts.
2. **Epigenetic studies** (DNA methylation) in subsamples with available age- and tissue-specific samples to create reference panels in cells relevant to ECHO (e.g., cord blood, placenta) to facilitate imputation in all participants (i.e., predictions of epigenetic marks from genotypes). This can be extended in the future to other 'omics (transcriptomics, metabolomics, etc.)

Additional Recommendations

3. **Single cell sequencing (or epigenetics)** to generate more accurate estimates of cell-specific expression (or methylation) for deconvoluting cell composition in complex cell mixtures (e.g. cord blood, placenta)
4. **Methods development** for integrated analyses of 'omic data to unveil the causality of childhood outcomes that ECHO is seeking to understand
5. **Storage of maternal and cord blood plasma** for future studies of extracellular vesicles in relevant tissues as placenta.

Recommendations from ECHO PI-led Epigenetics & Genetics Working Group

Whole genome genotyping (WGG) on all ECHO participants (children, parents)

- This will generate high throughput, accurate genotypes that can be compared across all individuals and cohorts.
- The WGG genotype common variants. The sample size for ECHO (n~100,000) should be powered for common variants with smaller effect sizes for many different complex disease outcomes.
- Focus on 'gene x environment' will be on common variants.

MEGA array as GWAS platform of choice

- Includes more variants overall
- Contains better coverage for non-European populations which are present in ECHO.

Recommendations from ECHO PI-led Epigenetics & Genetics Working Group

Whole Genome Sequencing (WGS) on a subset of individuals (n= 100-500 per ancestry).

- Array-based data had improvement on imputation calls when they used a small number of their own WGS samples.
- Important for some ethnicities or population isolates in ECHO and will also serve as a general reference.

Sequencing child-parent Trios

- Particularly those identified to have rare phenotypes or that are on opposite sides of phenotypic spectra.
- This will provide information on the effect of de novo mutations for some phenotypes that could then be explored more in depth in future studies.

- Other omics, specifically epigenetics be considered in the future given the emphasis on environmental determinants of health in ECHO and the longitudinal study design.

ESB recommendations

- **We largely are in agreement with several of the major recommendations of the two groups:**
  - Perform whole genome genotyping in all ECHO children and parents
  - MEGA array as platform of choice
  - Perform whole genome sequencing on subsets of individuals based on ancestry (100-500/ancestry group), to increase the validity of imputation calls for genotyping analyses
  - Perform Whole Genome Sequencing (WGS) for infant-mother-father triads perhaps as suggested for those with rare phenotypes or perhaps all 10,000 such triads.
  - Plan to add epigenomics, transcriptomics and metabolomics analyses in the future.
  - Support efforts to integrate 'omics into epidemiological analyses of exposure-behavior-outcome hypotheses.
- **Single Cell Sequencing**
  - A priority for NIH but not for ECHO –could be done in less time with smaller studies outside of ECHO.

ESB recommendations

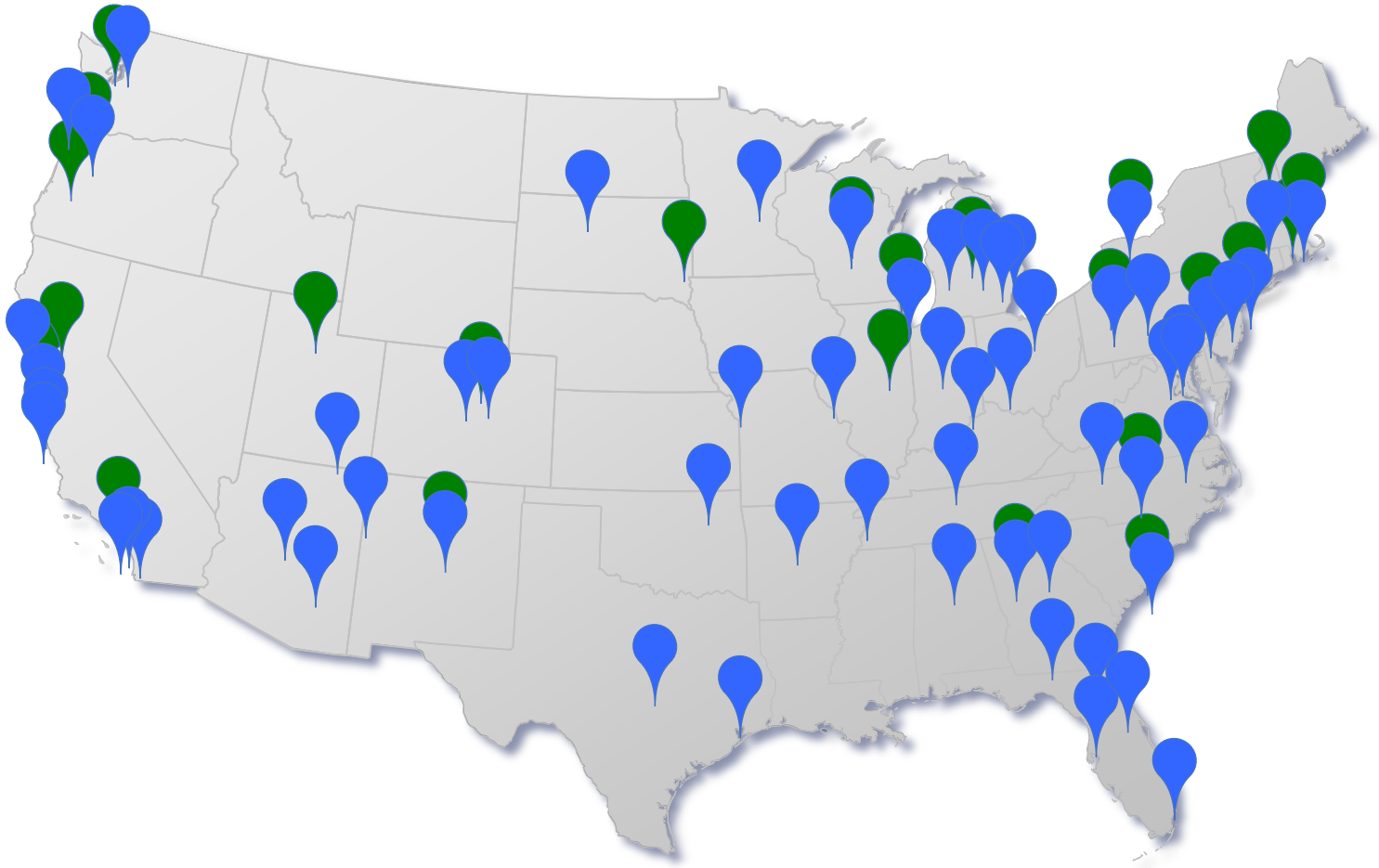
- **General advice**
  - To the extent possible, use the same platform and laboratory for analyses
  - Carry out centralized quality control and imputation
  - Assure data availability for all investigators
  - Articulate policies on return of individual results to participants
  - Be open to technological transformations and the possible need to validate new platforms in the future. Plan for this.
  - Policies for inclusion (or not) of sub-cohort sequencing data completed prior to ECHO
- **Sample repository**
  - Include samples for future ’omics analyses, i.e., epigenetics, RNA, metabolome and perhaps extracellular vesicles
- **Sharing**
  - Make both samples and “omics” results available to the broad scientific community. ARIES (Accessible Resource for Integrated Epigenomics Studies) from the ALSPAC (Avon Longitudinal Study of Parents and Children) is a possible model for sharing.

Extra slides

ECHO Mission

Enhance the health of children for generations to come

ECHO Cohorts

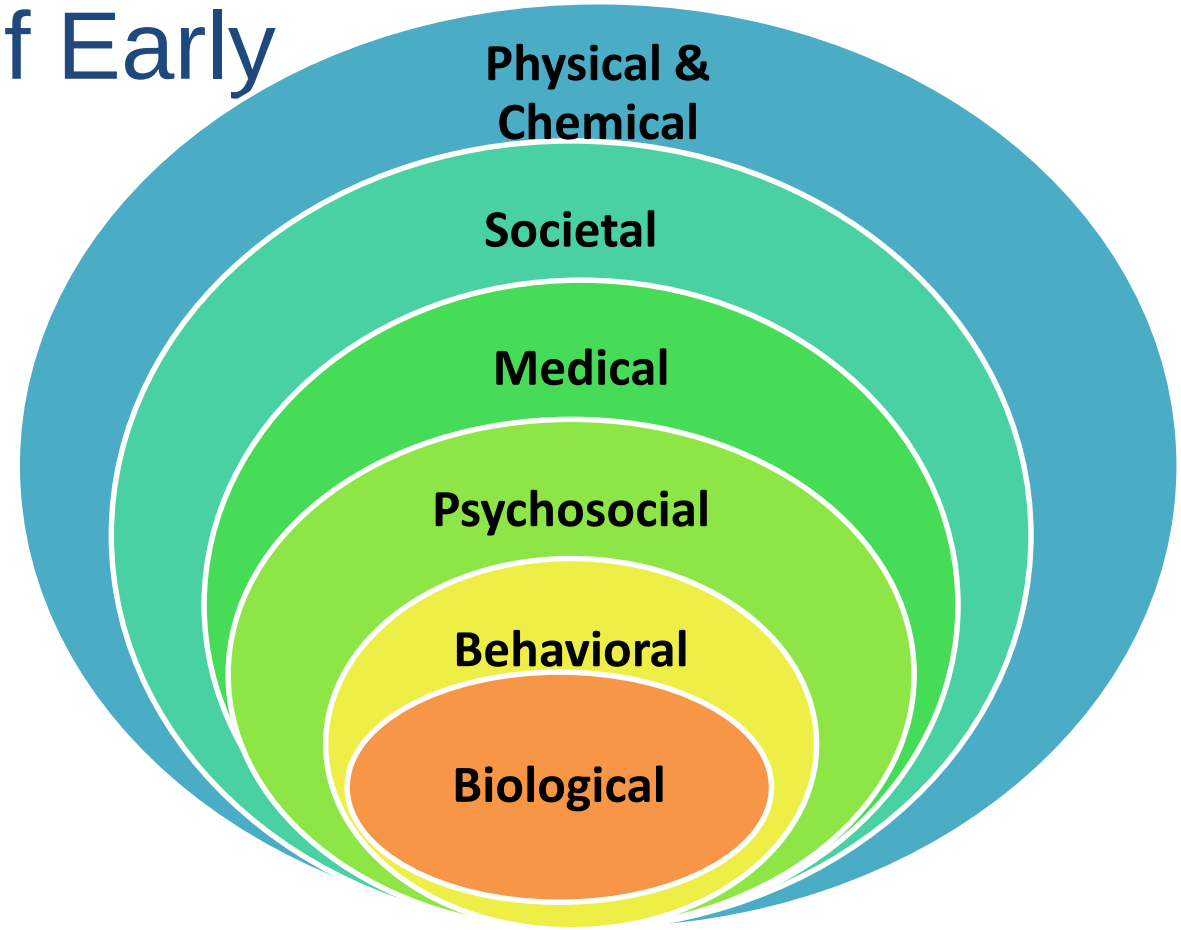


35→31 awards
60+ cohorts

-Majority started prenatally-

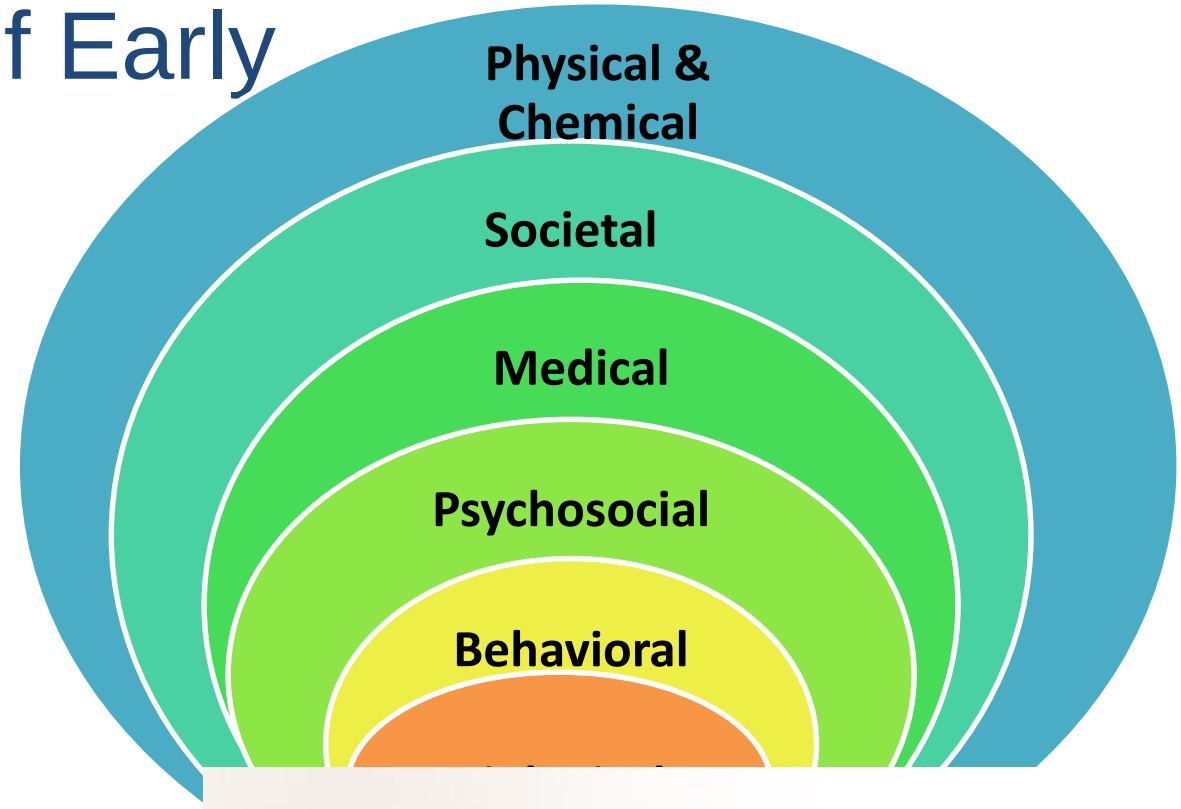
Broad Range of Early Environmental Exposures

From society to biology



Broad Range of Early Environmental Exposures

From society to biology



- *Exposures from conception to age 5 years*

The ECHO-wide Cohort

Weaving together many individual cohorts

Promise of ECHO-wide Cohort

- 50,000+ children and their families
  - Address research questions that no single cohort, or even a few cohorts, can answer alone.
- Solution-oriented research
  - Impact on policies, programs, practices
- Nationwide research collaboration
  - Resource for entire scientific community