The National Children's Study Genetics Workshop

March 23, 2007
Bethesda, Maryland

The National Children's Study Genetics Workshop

March 23, 2007
Hilton Washington DC/Rockville Hotel and Executive Meeting Center

List of Participants

Peter Scheidt	National Institutes of Health, Director, National Children's Study
Sarah Knox	National Institutes of Health, Senior Scientist, National Children's Study
Alan Guttmacher	National Institutes of Health, Deputy Director, National Human Genome Research Institute
Randy Jirtle	Duke University Medical Center, Professor, Department of Radiation Oncology
Pathik Wadhwa	University of California-Irvine, Director, Departments of Psychiatry, Obstetrics and Gynecology, and Pediatrics
Jeff Murray	Univeristy of Iowa College of Medicine, Professor of Pediatrics, Department of pediatric and Biological Sciences
Nancy Cox	University Chicago, Professor, Department of Human Genetics
Doug Wallace	University of California-Irvine, Donald Bren Professor of Molecular Medicine, Center for Molecular and Mitochondrial Medicine and Genetics
Bernie Devlin	Western Psychiatric Institute and Clinic, Associate Professor of Psychiatry and Human Genetics
Jim Swanson	University of California-Irvine, Principle Investigator, Child Development Center
Joel Hirschorn	Children's Hospital Boston, Assistant Professor of Genetics, Department of Genetics
Andrew Feinberg	Johns Hopkins University, Professor, Departments of Medicine, Molecular Biology, and Genetics and Oncology

David Valle Johns Hopkins University School of Medicine, Director, Departments of Pediatrics, Ophthalmology, and Molecular Biology and Genetics

Highlights of the NCS Study Design and Methods

Dr. Scheidt began the meeting with an overview of the National Children's Study (NCS) and discussed the goal for the conference. Last September, at a meeting in UC-Irvine, the opportunities for research in genetics and genomics in large longitudinal studies were discussed. The goal for this conference was focused specifically on the genetic research needs of the NCS.

An overview of the NCS was presented by Dr. Scheidt. The aim of the NCS is to study potentially harmful environmental exposures for the extent of harm, and association with disease, as well as to study healthy development in children. The NCS is also meant to serve as a resource for future studies. Most of the current NCS genetics hypotheses focus on gene x environment interactions, e.g., on individuals with particular genetic variants whose disease risk increases with environmental exposure. One example Dr. Scheidt discussed was variation in paroxinase genes and interaction with pesticides to alter developmental outcomes.

The thirty NCS hypotheses are organized according to priority exposures and outcomes. Priority exposures include physical, chemical, biological and psychosocial factors, as well as behavioral factors such as diet and smoking, all of which are capable of interacting with genotype to influence gene expression. Some priority outcomes for NCS include pregnancy outcomes, asthma, obesity and development, neurocognitive development and behavior, injury, and reproductive development.

The NCS is a national probability sample that will allow all exposure and outcome data to be representative of children of all subgroups of the United States. The sample is highly clustered in order to characterize neighborhoods and communities where children live. There is center-based implementation to allow for broad input and to ensure capability. The plan is to recruit 100,000 children and follow them from birth until age 20.

Dr. Scheidt also presented an overview of the planned visits, biological specimens to be collected, and schedule for the study (data collection to begin in 2008).

In his presentation, Dr. Scheidt emphasized the objectives of the meeting. These were:

1) To assure a framework for study of the role of genes and genomics in the development of health and disease.
2) To determine whether the study design, data collection, and specimens are appropriate for this purpose.
3) To evaluate whether the planned study design, data collection, and specimens assure optimal data for genetic analyses.

Dr. Scheidt also discussed the boundaries or limitations of the NCS. While the NCS is a large prospective cohort, of 100,000 children, even a study population of this size could experience limited statistical power while exploring complex gene x environment interactions. Second, given the large study size, cost is a major consideration. When selecting specimens, processing and analyses, cost, must of necessity, be included as one criterion. Therefore, wherever possible in the NCS, a nested case-control design will be performed. The final boundary for study is to minimize burden to participants—the risk to participants should not exceed minimal risk level and confidentiality is of the utmost importance.

Dr. Scheidt ended his presentation with the following points to guide the discussion:

- Why should the NCS be investigating genetics; for what outcomes or conditions?
- Which biological specimens are needed?
- What is the most optimal timing for specimen collection?
- How frequently should specimens be collected?
- What specimen processing is required?
- What sample size is needed for adequate power?
- How much will it cost?

Data Access

After Dr. Scheidt's presentation, the issue of data access was discussed.

- Dr. Scheidt stated that the plan is for the NCS to allow public access as quickly as possible following NIH guidelines for access.
- Dr. Knox commented on the importance of protecting the confidentiality of participants and the difficulty of de-identifying data when a large amount of genetic information, as well as personal information of participants is available. This is especially problematic when this information can be connected to personal health and demographic information. One approach for addressing this concern is phased access to data where only smaller selected de-identified data sets are publicly available and more complete data sets are made available to scientists under proscribed circumstances.
- Dr. Guttmacher commented that much of the genetics community has an idea of broad access to data. Genotyping efforts such as the Genetic Association Information Network (GAIN); the Genes, Environment and Health Initiative (GEI); and the Framingham Cohort all moved rapidly to allow open access to researchers.

Summary of the Gene / Environment Meeting in Irvine, CA

Dr. Pathik Wadhwa presented a summary of the meeting at UC-Irvine in 2006. The aim of the meeting was to discuss scientific issues related to the investigation of gene x environment interactions and their implications for health, development and disease susceptibility. The questions that investigators were to address during the course of the meeting included the most important scientific issues relating to genetics, best and future

methods for genetic research, and practical ways to incorporate these methods into a birth cohort study.

The organizing framework of the questions that investigators explored during the meeting was based on the following questions:

- Which diseases are most relevant to study from a genetic perspective?
- Why are gene x environment interactions important?
- Which environmental influences are most relevant to gene expression?
- Why should the NCS study genetics in relation to development?
- Which genomes, aspects of genetics, or classes of genes?
- Which time points during development?
- Which statistical or mathematical modeling approaches should be implemented?
- Which techniques or measurement methods? Which applications of systems biology?
- Biological samples: which ones, from whom, when, how to collect, how to process, how to analyze

The diseases and outcomes of interest that were identified as priorities for NCS research include those disorders which are a major burden on children and young adults. For most of these common disorders, disease susceptibility is a complex interaction of multiple genetic and environmental factors and therefore, the focus of genetic questions should be on interactions. This is only possible when both environmental and genetic exposures are measured in the same cohort. Another characteristic of the diseases of interest to the NCS is a major developmental origin component with latent and varying periods of susceptibility to environmental influence.

The concept of predictive adaptive responses was also discussed. As an organism develops, it responds to the environment. In some contexts the response may be adaptive, but in other contexts the response may be maladaptive. The discrepancy between early and late effects may relate to when an exposure is harmless or harmful (Gluckman and Hanson, Science, 2004). This also applies to when environmental influences may be protective. For example, birth phenotypes are related to genotypes, prenatal environments, and mediated by epigenetics that may alter the relationship between phenotype and genotype. The post-natal environment and influences throughout childhood alter the association between genotype and adult phenotype / risk susceptibility. The predictive adaptive response deals with the congruence between prenatal and post-natal environments.

The question of which genetic components to examine was raised. Responses included nuclear DNA (nDNA) and mitochondrial DNA (mtDNA). mtDNA is an often overlooked biomarker, which may be particularly important to investigate due to the role of the mitochondria in energetics. Somatic mutations accumulate in the mtDNA of post-mitotic tissues. In addition, there is clear geographic variation in mtDNA sequences. It is important to recognize that mtDNA genetics requires different approaches than nuclear DNA.

Possible genomes suggested to study included the fetus/child, mom, gametes, father, grandparents (in particular the maternal grandmother), and microorganisms of the reproductive tract (meta-genome). There are several other aspects of genetics that would be of interest to investigate including:

- Structure -- DNA sequence variations (SNPs and haplotypes) and copy number variations
- Access to structure – alterations in DNA structure, such as epigenetic changes, that affect folding of DNA and access to genome – which may alter gene expression
- Function – gene expression (RNA/protein)
- Epigenetics and mutations in DNA – timing of collection for these markers and processing are issues to be addressed

In addition to genetics, different aspects of environmental exposure were discussed for investigation, including both the maternal and external environment. Importantly, in mammalian pregnancy, one of the key regulators of the maternal environment is the fetus. This leads to the concept of reciprocal determinism – each stage of fetal development is both a consequence of and influence on the environment.

Temporal aspects are also very important and include life cycle, trans-generational and evolutionary aspects of time.

Statistical and mathematical modeling approaches to analyzing complex genetic data were discussed, including approaches to epistasis and gene x environment interaction with complex common disorders, maternal-fetal, gene x gene and gene x environment interactions, and consideration of evolutionary background. During pregnancy and early development, there are two genomes driving development, the mother's and the child's. There are many issues in modeling these complex interactions.

Systems biology was discussed as an approach for measurement and analysis, including the emerging technologies and advances in informatics. An important consideration for systems biology approaches is the issue of repeated measures over time.

At the UC-Irvine meeting, different samples were proposed to be collected including:

- Maternal DNA, RNA, proteins
- Maternal urine (includes epithelial cells)
- Maternal cervicovaginal fluid (epithelial cells, or meta genome of the reproductive track)
- Maternal cell lines from uterine tissue (c-section)
- Amniotic fluid
- Fetal trophoblasts
- Tissue from fetal loss
- Placental tissue
- Cord blood
- Umbilical cord

After Dr. Wadhwa's presentation, longitudinal sampling to examine flora in saliva and stool was discussed.

- Dr. Murray suggested that a longitudinal collection of saliva and stool to examine the gut and oral flora over time would be interesting to examine for interaction with obesity.
- The optimal timing (during life cycle) needs to be determined, but Dr. Murray suggested that these specimens should be obtained early, or in puberty.

Phenotypic Plasticity and Epigenetics

Dr. Andrew Feinberg presented on the phenotypic plasticity of development and epigenetics of human disease. With epigenetics, phenotype arrives from genotype through a programmed change.

Epigenetics involves heritable changes in gene expression and function through modification of chromatin but not DNA sequence. These epigenetic modifications are maintained during cell division. Dr. Feinberg mentioned a review article that was recently accepted to Nature summarizing different epigenetic modifications of genes (Feinberg, A. Phenotypic plasticity and the epigenetics of human disease. *Nature* 2007;447: 433-440).

There are several types of epigenetic modifications of genes:

- DNA methylation – generally high amounts because it is a general pattern
- Histone modifications – these modifications are for activated and silenced genes
- Modification of chromatin factors such as trithorax and polycomb
- Other modifications of chromatin structure (e.g., nucelosome compaction, higher loop structure)

Dr. Feinberg emphasized that epigenetics and developmental biology are the same thing. Unlike DNA sequence, the epigenome has a life cycle and tissue-specific marking. Epigenetic markings are distinct in stem cells, different tissue types, aging and cancer.

The biological impact of epigenetics is demonstrated by a single gene epigenetic disorder, Rett syndrome. In this syndrome, normal development is followed by loss of developmental milestones. In patients with this syndrome, there is a failure to maintain and continue developmental modifications, and genes critical to normal development fail to be silenced. Another example of a disease where epigenetic changes have been shown to have a role is cancer. In cancer, there are many epigenetic alterations that include both hypo- and hyper-methylation, which result in abnormal silencing and activation of genes. Cancer epigenetics may also be linked to aberrant stem cell differentiation.

Genomic imprinting occurs when expression of a particular gene only occurs from one allele due to epigenetic modifications of the other allele. Sometimes, epigenetic changes may occur that result in a loss of imprinting (LOI). One such change leads to a double dose of IGF-2 that may cause childhood disorders.

There is also evidence that epigenetics is related to aging. There are reports of time-dependent loss of environmental responsiveness (Dan Longo, NIA). If epigenetics alter plasticity of response, there may be a failure to respond appropriately to environmental influences.

In the NCS, it will be possible to study which environmental factors result in epigenetic changes and how epigenetic changes early in development influence subsequent developmental trajectories. To accomplish this, the NCS must take measurements at the right time points, study epigenetic changes over time, determine the important environment influences, and examine whether the epigenetic changes are related to phenotypes.

After Dr. Feinberg's presentation, the types of specimens to collect for study of epigenetics were discussed.

- Dr. Murray suggested that stem cells (cord blood) be collected as baseline material to track epigenetic changes longitudinally.
- Other specimens suggested included: cord blood, umbilical cord, Wharton's jelly, and amniocytes (from mothers who had amniocentesis).
- Dr. Hirschorn suggested that for epigenetic studies it is optimal to obtain a large amount and a variety of tissues.
- Dr. Feinberg mentioned the possibility of storing blood (cryopreservation) and isolating viable components later as a source for study of epigenetic changes. Lymphocytes are a good specimen source for studying modifications.
- Other sources include buccal swabs and hair follicles (which should be easy to obtain because children lose hair).

Drs. Swanson and Scheidt discussed costs for collection and storage of all of these specimens. Dr. Wallace pointed out that having the specimens may generate additional funding in the form of grants.

The optimal preservation of specimens for epigenetic studies was discussed.

- Dr. Wallace suggested that specimens should be stored in liquid nitrogen. Glycerol should be added to cord blood prior to freezing in order to maintain live cells.

Dr. Swanson asked Dr. Feinberg about studies of regression in autism. Dr. Feinberg responded that he is performing twin studies in monozygotic twins to study autism. This led to a discussion by the group about including twin studies in the NCS.

- Dr. Hirschorn asked if NCS had plans to enrich for twins. He pointed out that for general quantitative traits measured in everyone, the sample size of NCS is likely to be reasonable for twin studies. However, for looking at twins discordant for particular diseases, without enrichment, the NCS sample size is small. One approach for addressing this is to find shared environments in which heritability was different in order to identify modifiers of heritable effects.
- Dr. Hirschorn also made a technical point about the placenta. By sampling the placenta, it will be possible to determine if study participants are monozygotic or dizygotic twins at birth. Therefore, instead of actively enriching for twins, the

NCS could more extensively follow those children at birth identified as monozygotic twins.

The timing of specimen collection for tracing epigenetic changes and association with exposure was discussed.

- Dr. Jirtle noted that specimens should be collected at different developmental stages. In addition, multiple samples should be taken at multiple ages.
- In current plans for the NCS, blood will be collected at multiple time points in pregnancy and early childhood.

Mitochondrial DNA

Dr. Douglas Wallace presented on mitochondrial DNA. Dr. Wallace emphasized that there is a need to change paradigms related to physiology. Diseases are not simply the result of tissue-specific changes. These old paradigms ignore both environmental and whole body interactions. There is a need to examine what is special about being alive, that which is separate from simply structure – which is energetics. Human cells are the symbiosis of nuclear cytosol which encodes structure (Mendelian genes) and mitochondrial DNA which encodes energy. All of human wiring is due to energy, and energy is created by the mitochondria.

Mitochondrial DNA (mtDNA) is essential due to the role of the mitochondria in energy generation and balance. The human mitochondrial genome consists of approximately 1500 genes, 37 of which are maternally inherited mitochondrial DNA and the remainder of which are encoded in the nuclear DNA (nDNA). There are hundreds of mitochondria and thousands of copies of mtDNA in each cell. On the order of 5000 copies of mtDNA are inherited from mother to child.

Mitochondrial DNA could be 1%, 25% to 80% or 100% mutant. Each percent of mutated mtDNA may be associated with different phenotypes. In addition, it is a matter of chance which number of copies of mutated mtDNA will be transmitted to daughter cells. Therefore, mtDNA genetics is stochastic rather than deterministic.

Every function in humans is related to the mitochondria. For instance, mitochondrial function influences eating behavior, because of the role of mitochondria in producing energy. If calories are limited, the mitochondria inform cells to mobilize fat and up regulate oxidative phosphorylation. Thus, mitochondria can be fundamentally different in different populations.

Dr. Wallace described the mechanism of function of mitochondria, including the reactions of oxidative phosphorylation that generate ATP for cells. A toxic bi-product of these reactions is the reactive oxygen species (ROS). Because the mitochondria store a large amount of oxidative stress, lots of mutations occur in mtDNA. These mtDNA mutations are associated with many diseases. Over 30 different pathogenic mtDNA mutations have been related to disease.

In addition to mutations in mtDNA, lots of polymorphisms in mtDNA have been identified. Some of these polymorphisms are very ancient, non-neutral, adaptive variants. Mutations in mtDNA also accumulate with aging because mitochondria are constantly replicating and accumulating mutations.

Only a limited number of mtDNA lineages are present in the Americas. There appear to be 2 different groupings – lineages due to temperate and non-temperate climates. These different lineages may be due to the need for cells to either make more heat or more ATP depending on the climate. Every time people transition to new climates, cell energetics must adapt to meet new challenges. These may be related to haplogroups of mtDNA mutations.

Dr. Wallace commented that epigenetic biology is also linked to mtDNA and mechanisms of energetics.

The clinical relevance of mtDNA, in particular to NCS hypotheses, is related to energetics. If energetic pathways are uncoupled, such that an individual consumes a huge excess of calories, but is not doing work, the extra fuel produces oxygen radicals. However, if work is being performed, or the calories are being used to create heat, mtDNA is protected from damage. Several diseases appear to be mtDNA diseases, including Alzheimer's, sepsis, and asthma.

After Dr. Wallace's presentation, Dr. Knox asked about which specimens should be collected to investigate mtDNA. The group suggested the following:

- Stem cells
- Whole blood
- Amniocites
- Specimens from C-sections
- Urinary epithelial cells

Dr. Cox and Dr. Wallace discussed whether to study polymorphisms or somatic mutations in mtDNA.

- Dr. Wallace commented that 4 polymorphisms were related to energetic phenotypes and may be of interest. However, severe mtDNA mutations may be lost in the blood.
- New mtDNA mutations may be detected in urinary epithelial cells.
- Dr. Wallace warned that somatic mutations in mtDNA may be difficult to study due to the fact that these only increase exponentially after age 35.

Power and Gene x Environment Interactions

Dr. Swanson presented on statistical power to detect main effects of gene x environment interactions in the NCS. In the review of NCS hypotheses, and evaluation of the asthma outcome, Dr. Swanson explored issues related to statistical power. He presented a figure from a recent review by Manolio and Collins (Hum Hered. 2007; 63(2): 63-6). In this review, power calculations were provided in terms of the minimal odds ratio detectable

for the disorder. Dr. Swanson asked the author of the Quanto program to perform power calculations for both main effects and interactions using this program. He concluded that nested case-control studies from within NCS, with a sample size of 100,000, will be well powered for some disorders, but less well powered for others. However, Dr. Swanson emphasized, that many previous studies had much smaller population sizes than NCS. Currently, NCS is reviewing hypotheses and performing more detailed power analyses.

Gene x Gene and Gene x Environment Interactions

Dr. Knox raised the point that most common diseases and conditions are complex disorders, involving risk that is modulated by multiple genes and exposures. She mentioned that one of the goals for the meeting was to receive feedback on the types of analyses that can actually be done with the specimens available in a minimal risk study, e.g., blood, saliva, urine, etc. Much of the animal work on environmental exposure and gene expression has been performed on tissues (e.g., brain) that won't be available in this non-invasive study.

Dr. Devlin indicated that the statistical methods for analyzing complex gene x gene or gene x environment interactions are advancing. The limitation in these studies is not the tools available for analyzing the data, but the statistical power of many studies. Dr. Devlin also raised the issue of whether nested case-control designs were appropriate for these studies.

Dr. Jirtle asked whether statistical methods can differentiate between multigenetic disease and a single genetic change which gives rise to multiple phenotypes (such as Gude mice).

- Dr. Cox suggested that with the right data, it would be possible to sort out this question. Data would include breeding data and good information about the linkage of genes.

The issue of the lack of reproducibility of linkage studies was discussed.

- Dr. Cox suggested that issues around linkage studies are different than those faced by the NCS. The main problem with most linkage studies is that the ability to detect the association is not within the range of the studies, or that many linkage studies are underpowered. Therefore, in many linkage studies the effects observed are not reproducible. Dr. Cox emphasized that effect sizes expected are small.
- Dr. Hirschorn indicated that while it is known that effect sizes are small, for both association and linkage studies, it is also estimated that there are a number of genetic factors associated with disease risk. There is solid evidence of multiple genetic loci playing a role in complex traits.
- Dr. Knox pointed out that another problem with linkage studies is that they test for linkage to a specific genotype without consideration of the environmental context. Inconsistencies in results can arise when environmental factors influence the effects of polymorphisms of a single genotype in different ways, such that one and the same allelic variant can express differently depending on its environmental context. A study which does not demonstrate linkage may underestimate the risk associated with a specific genotype because of multiple

environmental factors working in opposing ways in the same population to mask effects. Leaving environment out of the equation can lead to contradictory results.

Dr. Swanson asked about the situation in gene x environment interactions, where an environmental exposure causes one variant to be protective and another variant to be associated with increased risk of disease, therefore resulting in a null main effect.

- Dr. Hirschorn indicated that canceling of main effects is a very particular case, and that while the exposure could modulate the main effect, the frequency of both the genetic variant and the exposure will likely differ across populations making it unlikely that an association might be completely masked by an interaction.
- Dr. Hirschorn emphasized that while gene x environment interactions are important to study, the NCS should not abandon main effects in their analyses. He indicated that he had seen few good examples of gene x environment interaction in the absence of main effects.

Studying differential imprinting was discussed.

- Dr. Cox indicated that methods are being developed for studying imprinting used in human genetic studies for association. She commented that the limitation is not in statistical modeling but in having the appropriate data.
- Dr. Wadhwa mentioned that an example of specimens required for imprinting would be data from grandparents.
- Dr. Wallace emphasized the importance of obtaining adequate family history data for clinical geneticist to interpret results. He asked whether clinical records would be obtained for NCS. Clinical records would be useful in studying the trajectory of disease.

Due to the importance of obtaining adequate family history data in genetic studies, different tools for collecting family history data were discussed.

- Dr. Scheidt mentioned that the plan for the NCS is to use a questionnaire adapted from NHANES.
- The group commented that there were no efficient, validated, family history instruments for genetic studies.
- Dr. Guttmacher mentioned that the US Surgeon General with the support of HHS, had developed a web-based tool for family members to input their own data. There is evidence that family history data obtained in the home may be better quality than interview data. Unfortunately, in terms of validity, family history instruments are not validated because there is no gold standard. Dr. Guttmacher suggested that this electronic tool might be used to supplement data already being obtained in the NCS.
    - Dr. Guttmacher demonstrated the tool at Familyhistory.hhs.gov
    - Some clinicians are beginning to use the tool
    - Focus groups found the new version of the tool user-friendly
    - The tool is available in Spanish or English
- With regard to family history, Dr. Cox suggested obtaining access to medical records. For some diseases, questionnaire data is sufficient, however for others,

access to medical records is crucial. If, at the time of consenting study participants, access is granted, medical records are more affordable.

The importance of timing of measurements was discussed.

- Dr. Devlin emphasized the importance of measuring continuous quantitative traits as opposed to dichotomous traits. If phenotypes are measured in real time, they will be easier to detect, and it is better to study these things prospectively.
- Dr. Knox indicated that this is exactly what is being planned in the NCS. There is a large battery of assessments which measure a broad range of exposures and outcomes quantitatively and prospectively on all participants. However, case-control studies will be utilized where appropriate, not only to contain costs but to provide the most efficient use of the limited quantity of biospecimens.

Whole Genome Scans

Whole Genome Scans vs. Candidate Gene Approaches

- Dr. Cox noted that in the context of preserving scarce genetic resources and cost savings, it would be much more cost efficient to do large scale genotyping. Candidate gene approaches, while scientifically interesting, are not cost-effective.
- Dr. Guttmacher noted that costs for whole genome genotyping have declined dramatically.
- Dr. Guttmacher mentioned that in the next 2-3 years analytical methods for whole genome association studies will get dramatically better.
- Dr. Hirschorn commented that statistical tools exist to assess common variation in the genome. With regard to gene x environment interactions, most studies are underpowered, although in the NCS, this may be less of a problem. Therefore, Dr. Hirschorn believes a goal of the NCS should be to do high throughput genotyping to obtain as much information as possible.
- High throughput genotyping may be done retrospectively on stored samples, in order to take advantage of declining prices and improved technologies.

The issue of data dissemination was discussed again during this session.

- Dr. Swanson commented that a plan for analysis and sharing of data from whole genome scans will need to be developed.
- Dr. Cox raised the issue that an infinite number of issues can be addressed with data from whole genome scans.

A strategy for whole genome scans for studying genetic variation was discussed.

- Dr. Cox detailed how the NCS will be obtaining data on quantitative traits and the usefulness of looking at whole genome association with these types of traits.
- Dr. Devlin suggested that whole genome association analyses should also include candidate genes which are known to have roles in illness to increase the statistical power of analyses. He cautioned that in a situation where association studies are completely exploratory, unless a gene x environment interaction is strong, studies will be underpowered.

- Dr. Cox responded that the field has not yet exhausted studies of main effects for childhood illnesses. The NCS is the first study to be well enough powered to address main effects.
- Dr. Cox also suggested that it may be important to focus on regions of the genome more strongly associated with a phenotype. Some of these regions may show selection and variation that are important to the research questions.

Maternal Grandparents

DNA obtained from maternal grandparents would be an important specimen to collect for genetic studies.

- According to Dr. Hirschorn, grandparents are important to include if genetics questions include maternal-fetal interactions. DNA obtained from grandparents will allow evaluation of parent origin effects. The grandparent DNA allows a disentangling of maternal genome effects from interaction effects.
- In addition, according to Dr. Jirtle, studying maternal grandparents allows an examination of trans-generational inheritance of epigenetic diseases.
- Dr. Murray emphasized the utility of obtaining DNA from grandparents because it is useful for learning about maternal effects on imprinting.

The expense of including grandparents and approaches for obtaining this DNA was also discussed.

- It was stated that due to cost, collecting DNA from grandparents in the NCS is not currently planned, except for ancillary studies.
- Dr. Wadhwa mentioned that the National Geographic project collected buccal cell DNA. One approach would be to ask grandparents to collect buccal cells. However, Dr. Wadhwa expressed a concern that grandparents' time may be more limited than parents.
- Dr. Murray stated that he believed the expense of obtaining DNA from grandparents was exaggerated because blood draws would not be needed. All they would have to do is mail saliva kits.
- Dr. Knox noted that public/private partnerships may be another source of funding for these studies.

Thrifty Genotypes and Developmental Origin Hypotheses

- Dr. Wallace explained that thrifty genotypes are related to the hypothesis that certain ethnic groups have a higher predisposition to obesity or cardiovascular disease due to genotypes that provide a selective advantage in calorie limited environments. In modern environments, there is an excess of calories with which this genome is not equipped to deal, leading to obesity.
- Dr. Swanson explained that thrifty genotypes are an example of a gene-environment mismatch hypothesis.

- Dr. Wadhwa mentioned that in order to explore these genotypes, it would be of interest to look at environmental exposures both early and late in the life cycle.
- Dr. Wallace mentioned that there is also interest in whether lifestyle of the mother during gestation is capable of imprinting phenotypic outcome in the child.

RNA studies

Stability of RNA for gene expression studies was discussed.

- Dr. Knox mentioned that RNA, extracted from both whole blood and leukocytes is stable for only 4-6 months. For longer storage, amplification is necessary but is also costly.
- Dr. Jirtle indicated that if cell viability is maintained, cells will be a source of RNA and may be stored long term.
- Dr. Wallace mentioned that the best approach might be storing viable cells. He mentioned problems with storing whole blood for RNA experiments. Dr. Wallace also pointed out that there are solutions that may be used to inactivate RNAases, to stabilize RNA in whole blood specimens. However, these solutions would degrade protein in the specimen.
- Dr. Hirschorn suggested converting RNA into more stable substrate prior to storage.
- Dr. Hirschorn also suggested that given the instability of RNA, it might be worth an initial investment.

Dr. Jirtle mentioned that the best use of RNA is microarray expression studies.

Stem Cells

- Dr. Wallace and Dr. Jirtle mentioned that collecting stem cells would be a valuable resource for the NCS. These cells may be obtained from umbilical cord or aborted fetuses.
- Dr. Wallace suggested that if cells are preserved right away and stem cells are kept alive, these cells in culture would be a resource for studying mtDNA. Cells may be exposed in culture and studied for modifications of genes.
- Stem cells may be stored frozen in liquid nitrogen.

Cell Lines

Different sources of material were suggested for the generation of cell lines:

- Whole blood
- Umbilical cord
- Placenta

Dr. Wallace commented that these materials may be stored at -80°C or in liquid nitrogen.

Gametes

- Dr. Jirtle asked if the NCS plans to collect gametes. Dr. Knox said a lot of concern had been expressed that asking for sperm would harm recruitment. One of the VCs is developing plans for a pilot study to test the feasibility of collecting sperm.
- Dr. Jirtle emphasized that gametes are very important for imprinting studies to determine if the epigenetic modification is environmental or inherited from mother and father. In colon cancer, it is estimated that 20% of patients have a loss of imprinting. It is unclear whether this association was not in place properly from the previous generation or is a result of exposure. Such a question cannot be addressed without the gametes.
- Dr. Scheidt commented that the only gamete specimens available to the NCS would be sperm. Studies suggest that approximately 20% of the fathers may not be the biological fathers. Therefore, the sperm collection may be inefficient and a large burden. In addition, sperm was not directly related to any specific NCS hypothesis.
- Dr. Wadhwa mentioned that there is a lot of data about damage to sperm in response to environmental contaminants.

Other Genetic Material

Dr. Wadhwa indicated that current techniques are able to capture approximately 40% of the fetal DNA in maternal blood.

Phenotype Measurements

- Dr. Cox mentioned a limitation in studying genotype-phenotype relationships, which is that the definition of relevant phenotypes is not always clear. A strength of the NCS is the plan for rich phenotype assessment, providing an opportunity for studying those relationships.
- A concern raised by Dr. Swanson was the role of thrifty genotypes contaminating measurements of phenotypes.
- Dr. Hirschorn suggested that twin studies will be useful for assessing phenotypes in different environments.

Conclusions and Summary

Dr. Scheidt led the discussion summarizing the meeting. He highlighted the following points:

- Candidate gene association studies will be performed on basic DNA. The Study should have adequate power for many main effect associations as well as interactions but for many other non-main effect associations, the power will be inadequate.
- Given the many genetic association studies to be considered and conducted, it is far more efficient to perform genome-wide association scans. The cost and efficiency of these scans are improving rapidly and, like buying technology, they should not be performed until needed.

- Viable cell lines for genetic studies will be the most dependable and versatile means of performing genetic studies
- Twins represent a very special opportunity requiring extra data and effort. Care should be taken to examine placentas with twins and assure maximum follow-up and data collection.
- Grandparent family history and DNA specimens are important and strongly recommended not only for traditional family studies but especially for epigenetic marker studies to determine how and when imprinting occurs.
    - Grandparent DNA can be efficiently obtained by mail in saliva specimens
    - Gamete specimens (semen, eggs not obtainable) are urged for the same reason. A possible alternative strategy is to seek semen specimens from fathers of twins.
- RNA for study of gene expression is important and urged. However, RNA is not stable. Preservation of RNA degrades protein. Obtaining the capability for doing RNA studies must be prospective and viable cells are needed.

- Mitochondrial DNA is an important consideration for an increasingly large number of conditions. Studies of mtDNA can be performed with multiple specimens (see above).

- The Surgeon General's Family History tool (http//.familyhistory.hhs.gov) is a computer based format for systematically obtaining family history that may be useful for the NCS.

- Based on experience, all participants strongly recommended not offering an opt-out for use of genetic data. Separating out use of genetic data reduces use of the data far more than its inclusion reduces overall participation.

- Finally there was a discussion of linking with other cohorts. NCS is currently linked with a childhood cohort consortium of childhood cancers. There are several other cohorts including the Danish, Chinese and ALSPAC cohorts. There was also discussion of collaborating with the South Hampton Women study.