# I. Title: Collection of Circulating Fetal DNA from Maternal Blood and from Cervical Fluid

## II. Study Center Collaboration

a. Lead Center: Wisconsin Study Center (Waukesha County Vanguard Center). Investigators: Aoy Mitchell, PhD, (Medical College of Wisconsin); Steven Leuthner, MD, (Medical College of Wisconsin); Maureen Durkin, PhD, DrPH, (University of Wisconsin – Madison)

b. Collaborators: David G. Peters, PhD, Hyagriv N. Simhan, MD, (University of Pittsburgh), Andrew DeWan, PhD, (Yale University), James Swanson, MD, Pathik Wadhwa, MD, PhD (University of California – Irvine Orange County Vanguard Center)

## III. Research Plan

### A. Goals

The goals of this proposal are to optimize fetal DNA yields from maternal circulation and from cervical fluid, to provide measurements of fetal DNA concentrations, and to assess the use of these samples for sequence and methylation analyses within the NCS. The intended result of this effort is to develop scientifically robust methods for the detection, evaluation, and storage of fetal DNA in maternal blood and from cervical fluid in NCS participants.

### B. Specific Aims and Methods

Specific Aim 1. Feasibility and evaluation of multisite maternal blood and cervical fluid collection and neonatal saliva collection.

*Feasibility study:* The target sample size will be 200 NCS participants. Initially these participants will include only those from the Waukesha County and Orange County original vanguard sites where collection of biospecimens are underway. In the event biospecimen collection is initiated by the University of Pittsburgh and Yale University sites during the project period, participants from those sites will also be included. Analysis of the specimens will take place at the Medical College of Wisconsin (blood-DNA quantification, fetal DNA concentration measurement), University of Pittsburgh (cervical fluid; maternal blood-first trimester; methylation analysis), and Yale University (neonatal saliva/buccal swabs, whole genome amplification genotyping).

*Blood Collection and Sample Tracking:* Samples will be obtained at up to 4 time points: pre-pregnancy, first trimester, third trimester, and post-delivery and shipped to Dr. Mitchell's laboratory at the Medical College of Wisconsin (MCW). Our abbreviated protocol is as follows:

1. Obtain up to 20 ml of blood via a peripheral or central blood draw into the purple top EDTA tubes provided in the kit.
2. Within 2hrs of blood collection, spin blood in a centrifuge at 1,600x g for 10 minutes.
3. Without disturbing the buffy coat, transfer plasma from the 10 ml EDTA tubes to the prelabeled (with ID code) flip-top tubes.
4. Transfer at least 500ul of buffy coat using pipet into one of the prelabeled (with the ID code and as "buffy coat") screw-top tubes also provided in the kit and store the tube in -20$^{o}$C freezer.
5. Spin plasma at 16,000 g for 10 minutes.
6. A pellet may or may not be visible after the spin. Last, transfer plasma (supernatant) without disturbing the pellet, to the prelabeled (with ID code) 2ml microcentrifuge screw-top tubes, also provided in the kit. All of the tubes will then be stored in the -20$^{o}$C freezer.
7. The frozen plasma sample and buffy coat will then be sent, on dry ice in a provided shipping container, to MCW to process.

*Cervical Fluid Collection and Tracking:* Cervical fluid will be collected from participants at the NCS sites in the 1st trimester in Orange County and 3rd trimester in Waukesha County, and shipped to Pittsburgh. Our abbreviated active protocol is as follows:

1. At the time of speculum exam, three Dacron swabs are placed in the cervix and left for 10 seconds to achieve saturation.

2. The swabs are each placed in a plastic container with 350 µl of PBS (final dilution 1:5).
3. The specimens will be stored in liquid nitrogen for transport. For processing, samples will be placed in a centrifuge filter unit (Spin-X; Costar, Cambridge, MA, USA), and centrifuged at 12,000 rpm for 20 minutes at 4°C. These samples will be shipped on dry ice to the University of Pittsburgh to process.

*Neonatal Saliva Collection and Tracking*:

1. At the time of the postnatal follow-up exam (2-4 weeks following birth) a research assistant will perform the saliva collection on the newborn using two Oragene CS1 collection sponges (DNA Genotek, Ottawa, Canada), one on either side to soak up as much saliva as possible.
2. Sponges will be deposited into the OG250 Oragene DNA Self-Collection Kit (DNA Genotek, Ottawa, Canada) and store at room temperature until shipped to Yale.

*Case Report Form (CRF):* The Waukesha County Vanguard Center will develop a database for tracking and managing the specimens collected for this project. A case report form will be generated for each participant. This form will include the NCS participant ID, date of sample collection, and results of DNA analysis of infant and other information as needed. The following data will be extracted from the NCS Study Management System:

a. Mother's age
b. Mother's ethnicity and race
c. Mother's weight
d. Gestational age and expected due date
e. Mother's physical activity level
f. Mother's medical history and conditions (limited)
g. Prior pregnancies and date
h. Environmental exposures (e.g., alcohol, smoking)

<u>Specific Aim 2. DNA isolation. Optimization of maternal blood and cervical fluid collection and processing, focusing on sample stability, decreasing cost, and increasing fetal DNA yield and concentration.</u>

*DNA isolation.* DNA from maternal white blood cells (MCW) and buccal swabs (Yale) will be extracted using standard protocols (i.e. PureGene, Qiagen).

*Detection of Fetal DNA in Maternal Blood Samples.* While we already have a protocol for blood collection, sample processing, and fetal DNA isolation, we will continue to investigate parameters such as different blood collection tubes and different methods of DNA extraction/purification from blood. The current protocol utilizes the QIAamp Circulating Nucleic Acid kit. We will test these protocols for total plasma DNA yield and quality and fetal DNA concentrations.

*Size fractionation to increase fetal DNA concentrations (Yale).* It has also been shown that it is possible to increase the concentration of fetal DNA by separating the smaller DNA fragments from larger fragments by both gel electrophoresis (8, 10, 11) and isotachophoresis (12). This has been taken one step further by applying whole-genome amplification to the size fractionated DNA samples to increase the total amount of fetal-enriched DNA fraction followed by successful single locus genotyping (11). Here we propose to compare methods and assess the feasibility of whole-genome genotyping to obtain fetal genotypes from maternal serum.

*Total plasma DNA yield (MCW):* Each plasma sample will undergo a DNA extraction protocol and the total amount of DNA present in the sample will be calculated using standard quantification methods such as quantitative real time PCR using TaqMan, picogreen, or Nanodrop spectrophotometer analysis.

*Detection of Fetal DNA in Plasma (MCW) and Cervical Fluid Samples (Pittsburgh)*

DNA will be analyzed for SRY target gene and B-globin as the endogenous control using the following primers and probes. *SRY:* Forward primer – 5'TGGCGATTAAGTC AAATTCGC-3' Reverse primer – 5' – CCCCCTAGTACCCTGACAATGTATT-3' Probe – 5' – FAM- AGCAGTAGAGCAGTCAGGGAGGCAGA- TAMRA-3'

*B-globin*: Forward primer – 5' – GTGCACCTGACTCCTGAGGAGA-3' Reverse primer – 5' – CCTTGATACCAACCTGCCCAG-3' Probe – 5' – FAM- AAGGTGAACGTGGATGAAGTTGGTGG-TAMRA-3' (Lo, et al, 1998). Each reaction will be analyzed in triplicate and run against the endogenous control. A calibration curve of known male input DNA will be run in duplicate and in parallel with the unknown samples. The conversion factor of 6.6pg DNA per genome equivalent will be used to express the data as copy numbers. The qPCR will be performed using the 7900HT Sequence Detection System (Applied Biosystems). Reference: Lo, et al. (1998). *Am J. Hum. Genet.* 62: 768-775.

Another approach that will be used to quantify fetal DNA is by detecting heterozygous SNPs when the mother is known to be homozygous. The detection platforms that used at MCW to quantify fetal DNA are cycling temperature capillary electrophoresis and 2 next generation sequencing platforms (Roche 454, Illumina GA2).

Specific Aim 3. Fetal DNA analysis and method comparison.

A. Evaluate factors that may affect fetal DNA levels (examples include gestational age, maternal age, maternal weight, maternal race and ethnicity, maternal and fetal medical conditions, prior pregnancies, and environmental exposures (e.g., maternal smoking, alcohol use, diet and exercise). Currently, there is a lack of robust data on fetal DNA percentages in maternal plasma in early pregnancy. We will analyze the distribution of fetal DNA percentages at four different time points during pregnancy across the targeted sample of 200 participants at the 4 specified time points- prepregnancy, 1st trimester, 3rd trimester, and at the 2-4 week home visit, therefore yielding up to 800 samples, and collect data to determine factors that may affect fetal DNA percentage. Examples of such factors include but are not limited to gestational age, maternal age, weight, and ethnicity. *The March request for Letters of Intent issued by the NCS Program Office stated: "Data which address questions regarding how early the target species are detectable, at what concentration, and how long they persist, are of particular interest." This aim specifically addresses this question*.

B. Examine methylation profiles of maternal lymphocytes and maternal plasma DNA, maternal cervical fluid. DNA methylation patterns have distinct tissue specificity. Such specificity provides an opportunity for the discovery of nucleic acid sequences whose expression via differential DNA methylation patterns are fundamentally distinct between the early and late gestational placental villi and paired/gestational age-matched maternal blood cells (MBC). This comparison is significant because it is the nucleated portion of the maternal hematopoietic system that is thought to be the primary source of the maternal component of cell free plasma nucleic acids. Similarly, apoptotic bodies derived from the placental villus are thought to be the primary source of fetal nucleic acid in maternal plasma. Therefore, functional genomic differences between these tissues can potentially be exploited to identify biomarkers for the selective enrichment of fetal nucleic acids. The team at the University of Pittsburgh has previously undertaken a comprehensive analysis of tissue type specific DNA methylation patterns between CVS and MBC and therefore has generated considerable insight into placenta-derived nucleic acids in maternal plasma at the level of DNA methylation in normal pregnancies. We propose conducting a study to discern feasibility of detecting fetal methylation patterns from maternal plasma and cervical fluid. Specifically we will use our preliminary data to select gene specific epigenetic markers that distinguish the chorionic villi from maternal lymphocytes. We will use the blood samples collected to validate the unique presence of these placental markers in plasma. Conversely we will select other gene specific epigenetic markers that are uniquely maternal. In both instances we will determine whether the levels of these unique fetal/maternal biomarkers vary between the first and third trimesters. We will also determine their concentrations in lymphocyte and plasma samples obtained from pre-pregnant and post-delivery samples.

In addition to these gene-targeted methods we will undertake a limited number of genome-wide assays in which we will compare DNA methylation patterns in DNA samples from matched pairs of lymphocytes and plasma. These assays will be performed using the Humanmethylation27 microarray platform, commercially available from Illumina. This is a 27,000 probe microarray that can detect methylation differences in thousands of human gene promoters and/or CpG islands. This platform technology will be used in a proof of concept study to identify DNA methylation differences in 12 lymphocyte and 12 matched plasma samples, thereby assessing concordance and discordance in methylation profiles.

We will also test the hypothesis that fetal DNA is present in cervical fluid samples. To achieve this goal we will focus on pregnancies with a male fetus. Dacron swabs will be used to collect cervical fluid samples. Cervical fluid and its associated cellular content will be recovered by centrifugation and DNA extracted using the Qiagen. We will utilize quantitative real time PCR to detect and quantify Y chromosome-derived SRY gene sequences.

Specific methods used in this phase of the study are as follows:

*(i) Gene Specific DNA Methylation Analysis*. Reactions will be designed using the Epidesigner software package (www.epidesigner.com) (Sequenom, San Diego, CA). PCR reactions will be carried out in a 384 well format as follows. To each reaction will be added 1.42μL $ddH_2O$, 0.5μL 10x Hot Star Buffer (Qiagen) (15mM $MgCl_2$, Tris-Cl, KCl, $(NH_4)_2SO4$, pH 8.7), 0.04μL dNTP mix (25mM each), 5 U/μL Hot Star Taq (Qiagen). Primers will then be added to a final concentration (each) of 1μM and 1μL bisulphite converted DNA (1ng/μL per reaction). Reactions will be incubated as follows: 94°C for 15 minutes then 45 cycles of 94°C for 20 seconds, 56ºC for 30 seconds (temperature adjusted according to primer Tm), 72°C for 1 minute followed by 72°C for 3 minutes. Reactions will be treated with Shrimp

alkaline phosphatase, *in vitro* transcribed and analyzed according to the manufacturer's instructions (Sequenom). Fully methylated DNA controls will be obtained from Millipore- (CpGenome Universal Methylated DNA, part S7821).

*(ii) Genome Wide DNA Methylation Analysis*. The HumanMethylation27 DNA Analysis BeadChip (Illumina) allows interrogation of 27,578 CpG sites based on the NCBI CCDS database (Genome Build 36) and also targets the promoter regions of 110 miRNA genes. Bisulphite conversion of DNA is carried out using the EZ DNA Methylation™ Kit (Zymo Research Corp., CA) to convert unmethylated cytosine nucleotides to uracil. Following denaturation with 0.1N NaOH, converted DNA samples are amplified by incubation at 37ºC for 20 hours in a proprietary amplification reaction mix. Amplified DNA is fragmented using vendor supplied reagents by incubation for 1 hour at 37ºC. Fragmented DNA sample is precipitated and resuspended in hybridization buffer. Infinium BeadChips are cleaned and activated by washing with ethanol, formamide and vendor supplied pre-hybridization buffers. DNA samples are denatured, applied to the Infinium arrays and hybridized 16-24 hours with rocking at 48ºC. Unhybridized and non-specifically hybridized DNA is washed away, the Beadchip surface is treated to facilitate primer extension reaction and single base extension performed on bound primers with labeled nucleotides. Bound DNA sample is washed away. Staining steps are performed to attach florescent dyes to the labeled nucleotides and the array surface is sealed to protect the dyes from atmospheric degradation. The final array is scanned using an Illumina BeadArray Reader and the data analyzed using Bead Studio 2.0.

*(iii) Biostatistical Analysis of Microarray Data*. On an Infinium array, each targeted CpG site is interrogated by 2 probes: probe A for unmethylation status, and probe B for methylation status. If a CpG site targeted by the probe pair is unmethylated, probe A should have strong signal, while probe B should have weak or no signal. On the other hand, if the CpG is methylated, probe A should have weak or no signal, while probe B should have strong signal. Thus, we shall first normalize the A probe signals and B probe signals separately, using the cyclic loess algorithm, then compute the log ratio of probe B to probe A: log(B/A).

To identify the candidate CpG sites, we first identify CpG sites that are either hypomethylated or hypermethylated in all MBC samples, regardless of the disease status. A CpG site is hypomethylated if log(B/A) of that site is significantly lower than log(1/2). It is hypermethylated if log(B/A) of that site is significantly higher than log(2). A CpG site is more methylated if log(B/A) of that site is significantly lower than 0. It is less methylated if log(B/A) of that site is significantly higher than 0. All these analyses will be done using the empirical Baeyesian method, with FDR controlled at 5%.

C. Investigate feasibility of whole genome genotyping. We will perform microarray genotyping on fractionated and whole-genome amplified DNA isolated from maternal serum. Each sample will also have a matched maternal lymphocyte-derived DNA sample and neonatal saliva-derived DNA sample.

For fractionation and whole-genome amplification, DNA extracted from maternal serum will be run on a 1% agarose gel and electrophoresis carried out at 80V for 1h. Using the molecular weight marker as a guide, the section of each lane containing the DNA of 90-500bp will be cut from the gel. The DNA will then be extracted from the agarose sections with the QIAEXII Gel Extraction Kit. To overcome the low yield expected following fractionation, the resulting DNA will then be subjected to whole-genome amplification using the Repli-g Whole Genome Amplification Kit from QIAGEN.

In order to compare the genotypes from maternal serum DNA and assess accuracy of the genotyping, we will use DNA isolated from lymphocytes to assess the accuracy of maternal genotypes. To assess the accuracy of fetal genotypes, comparison to genotypes from saliva-derived DNA will be performed. Genotyping of all samples will be done using the Affymetrix Genome-Wide Human SNP Array 6.0 (Affy 6.0) and will be performed at the Keck Microarray Resource Center at Yale University.

Genotype calls for DNA isolated from lymphocytes and saliva samples will be made using the standard Birdseed algorithm that is part of the Linux-based Affymetix Power Tools (APT) suite of software. This is a 2 dimensional Gaussian Mixture Model (GMM) that clusters diploid samples into the canonical SNP genotype classes AA, AB and BB. To make the genotype calls for the serum-derived DNA samples it will be necessary to extend the Birdsees algorithm to cluster samples into five possible genotypes and potentially incorporate information on relative proportions of maternal and fetal DNA to assign the most likely fetal genotype given the sample's distance from the cluster mean. The algorithm development will be greatly assisted by having the "true" genotype calls for the mother and offspring. To determine the accuracy of genotyping calls, we will compute the concordance of maternal and fetal genotype calls determined from the serum-derived DNA samples with those from saliva-derived DNA samples.

## IV. **Anticipated Deliverables**

a. Description of sample and data collection and sample tracking in the Vanguard sites.
b. Description of optimized DNA extraction protocols for WBCs, plasma, cervical fluid, buccal swabs.

c. Fetal DNA yield and fetal DNA concentration measurements from different time points.
d. Feasibility of using cervical fluid samples for fetal DNA analysis.
e. Feasibility of various methods of fetal DNA detection: methylation and whole genome amplification genotyping.
f. Pilot data evaluating factors affecting fetal DNA concentration (such as gestational age).
g. Analysis of costs and other resources needed.

## V. **Timeline of Activities**

| **Specific Aims** | **Apr - Aug 2010** | **Sept 2010-Jan 2011** | **Feb - July 2011** |
|---|---|---|---|
| Feasibility study: Waukesha and Orange County<br>A. Maternal blood<br>B. Maternal cervical fluid<br>C. Neonatal buccal swabs | | | |
| DNA isolation<br>A. WBC, plasma<br>B. Cervical fluid<br>C. Neonatal buccal swabs<br>D. Optimization of fetal DNA isolation for yield and cost | | | |
| DNA analysis and method comparison<br>A. Evaluate factors affecting fetal DNA concentration<br>B. Methylation profiles of maternal lymphocytes, plasma DNA, cervical fluid<br>C. Investigate feasibility of whole genome genotyping | | | |

## VI. **Projected staffing requirements and other required resources**

Waukesha County NCS Vanguard Center: Aoy Mitchell, PhD, MCW, will lead this effort at 20% FTE (project design and management). Other investigators and staff include: Michael Mitchell, MD, MCW. 2.5% FTE (co-investigator); Mats Hidestrand, PhD, MCW, 50% FTE (DNA quantification and concentration); Mary Goetsch, MS, MCW, 25% FTE (DNA extraction); Craig Struble, PhD, MCW, 2.5% FTE (bioinformatics); and Pippa Simpson, PhD, MCW, 2% FTE (statistical support); Steven Leuthner, MD, MCW subcontract principal investigator, 1%, and Maureen Durkin, PhD, DrPH, University of Wisconsin-Madison principal investigator; 10% FTE for database development and management. Other budget items include DNA extraction from plasma and lymphocytes and DNA quantification and fetal DNA measurement, shipping costs, and participant incentives for added speculum exam and specimen collections.

University of Pittsburgh NCS Vanguard Center: Pittsburgh's team will be led by David G. Peters, PhD, (project design and management) at 20% FTE; Hyagriv N. Simhan, MD, MSc at 2% FTE; Tianjiao Chu, PhD, (bioinformatics) at 5% FTE, and Annie Shaw, BS, (technical support) at 100% FTE. Additional costs include: Epityper (Mass Array) for Quantification of Methylation at Single Genes where each 384 well plate will permit analysis of 48 samples at 8 loci (including controls); Infinium Microarray Analysis of DNA Methylation (humanmethylation27); Amplification of SRY Sequences from Cervical Fluid to cover probes, Taq and other consumables (including plasticware) for these experiments; and General Reagents for DNA Extraction and Sample Management

Yale University NCS Vanguard Center: Andrew DeWan, PhD, is the Principal Investigator for the Yale Research Team (25% FTE). Additional personnel support is requested for William Murk, MPH,. (50% FTE), laboratory technician. Other costs include materials, microarray chips and service charges for chip processing.

University of California – Irvine Orange County NCS Vanguard Center: Shipping costs for specimens and participant incentives for added speculum exam and specimen collections.