## Background

The prevalence of maternal cigarette smoking during pregnancy is about 14% in the United States despite warnings of its effects on the developing fetus. More generally, it is thought that many complex diseases and disorders have origins *in utero*, resulting from adaptations to the gestational environment and/or from exposure to toxicants, such as those that result from maternal smoking during pregnancy. Epigenetics is one mechanism by which *in utero* exposures may influence health outcomes of the offspring. For example, exposures may affect the epigenetic regulation of imprinted genes that are critical to normal growth and development and thereby increase risk of adverse health outcomes.

Imprinted genes are those whose regulation is either maternally or paternally controlled. Normally, one of the two alleles of an imprinted gene is silenced via methylation at differentially methylated regions (DMRs) located on that allele. If the gene is maternally expressed, the paternal allele is silenced; if it is paternally expressed, the maternal allele is silenced. Hence when such a DMR is assayed, a normal individual's methylation fraction (measured using the DNA of many cells from the subject) should be near 50%.

A well-characterized imprinted domain on human chromosome 11p15.5 contains the genes for paternally expressed Insulin-like Growth Factor II (IGF2) and maternally expressed H19. Deregulation of IGF2 expression has been linked to overgrowth disorders, obesity and cancer. Imprinted expression and transcription of IGF2 are regulated in large part through the patterns of differential methylation of at least two regulatory DMRs, one of which is located near the H19 promoter (H19 DMR) and the other upstream of the three IGF2 promoters that are subject to imprinting (IGF2 DMR). Both DMRs have been shown to exhibit altered methylation in cigarette smoking-related malignancies.

The goal of this analysis is to examine the influence of a developing child's *in utero* exposure to maternal cigarette smoking byproducts on percent methylation at the IGF2 DMR using umbilical cord blood samples. The data are from a multi-ethnic birth cohort established to facilitate study of the effects of early exposures on epigenetic profiles and phenotypic outcomes. The IGF2R DMR methylation is measured as the percent of the subject's (offspring's) alleles that are methylated in a sample of their DNA and, hence, takes values between 0 and 100. Methylation at this locus was assayed twice for 294 of the 314 subjects and only once for the remaining 20. The 608 DNA samples were arrayed on 22 96-well plates and each plate was processed separately. Each plate has eight rows, designated 'A' through 'H,' and twelve columns, designated 1 through 12; only a subset of wells were used on each plate.

## Data

There are 314 rows where each row corresponds to one subject (offspring and mother) with columns corresponding to:

**age** Maternal age coded 'lt30' (younger than 30 at delivery), '30to39' (30 to 39 years old at delivery), and 'ge40' (more than 40 years old at delivery).

**BMI** Maternal body mass index measured as weight before pregnancy (in kg) divided by height (in meters) squared and coded '0' (=less than 30) or '1' (greater than or equal to 30).

**smoke** Maternal cigarette smoking during pregnancy; coded '0' if did not smoke and '1' if smoked during early pregnancy and stopped *or* if smoked throughout pregnancy.

**gestage** Gestational age of the infant. Coded '1' if less than 37 weeks and '0' if greater than or equal to 37 weeks.

**gender** Infant's gender ('1'=male, '0'=female).

**edu** Mother's education level (coded 'ltHS' for less than high school, 'ltCollege' for high school/GED, and 'geCollege' for at least some college)

**race** Mother's race/ethnicity ('AA'=African American, 'EA'=Caucasian or 'Other').

**methyl1** the first replicate measurement of the subject's child's methylation level.

**methyl2** the second replicate measurement of the subject's child's methylation level. 20 subjects do not have a second measurement so this is missing.

**plate1** (and 'plate2') plates on which the subject's first and second replicate measurements were made, respectively. 'plate2' is missing for 20 subjects who have only a single measurement.

**row1** (and 'row2') row in which the subject's first and second replicate measurements were placed on the plates whose IDs appear in 'plate1' and in 'plate2,' respectively. 'row2' is missing for 20 subjects.

**column1** (and 'column2') column in which the subject's first and second replicate measurements were placed on the plates whose IDs appear in 'plate1' and in 'plate2,' respectively. 'column2' is missing for 20 subjects.

**well1** (and 'well2') well in which the subject's first and second replicate measurements were placed on the plates whose IDs appear in 'plate1' and in 'plate2,' respectively. 'well2' is missing for 20 subjects.

## Objectives

A child's methylation level may depend on gender and race and may be associated with maternal smoking or unmeasured exposures associated with maternal age, maternal BMI, maternal education level, and/or gestational age. Explore and analyze the data, characterizing the relationships between percent methylation in the offspring and maternal smoking status adjusting for any other covariates. It is also of interest to determine if the relationship between smoking and methylation is the same across race and gender.