GENOMIC ANALYSES: A NEONATOLOGY PERSPECTIVE

C. Michael Cotten, MD, Geoffrey S. Ginsburg, MD, PhD, Ronald N. Goldberg, MD, and Marcy C. Speer, PhD

Advances in medicine have lowered the viability limits of preterm birth, but complex multifactorial diseases such as respiratory distress syndrome (RDS), nosocomial sepsis, necrotizing enterocolitis (NEC), intracranial pathologies (bleeding, ischemia, maldevelopment), retinopathy of prematurity (ROP), and bronchopulmonary dysplasia (BPD) threaten survival and optimal neurodevelopment. Genetic variations that are clinically insignificant among term infants may contribute to preterm infants' susceptibility to these complex diseases. The recent complete sequencing of the human genome [1,2] provides methods for assessing how genomic and environmental factors interact to contribute to risk of complex diseases.[3,4] Identification of associations between complex diseases, genetic variations, and the interactions between variations in multiple genes and environmental factors could lead to better understanding of disease pathophysiology as well as specific prevention and treatment strategies, based on genotype.[3]

Studies measuring frequency of variations in candidate genes for factors suspected of contributing to disease pathophysiology in cases compared with frequency in control subjects are one of many ways of addressing the link between genetic variation and disease. These candidate gene association studies are increasingly seen in the medical literature. Our goals in this review are to provide clinicians with an introduction to studies of genomic contributions to disease susceptibility, provide examples of candidate gene association studies of complex diseases of preterm infants, and provide guidelines for the structure and assessment of future candidate gene studies.

GENETIC CONTRIBUTIONS TO COMPLEX DISEASES

Clinicians treat common "complex diseases" that result from interactions between multiple genes, developmental, and environmental factors such as microbes, oxygen, and malnutrition.[3,5] Rather than being causative, genetic variations are more likely to influence susceptibility to complex diseases such as NEC and BPD. Identifying which diseases have plausible genetic contributions justifying exploration first requires assessment of the evidence that supports the hypothesis that genes are involved, and the likelihood that a genetic component to the trait/disease can be identified with current technologies and available samples. The samples and technology issues are touched on later in this review. Here we discuss the plausibility of the contribution of genetic variation to complex disease susceptibility.

Evidence for a genetic contribution to disease risk comes from approaches such as twin studies, in which disease concordance among monozygotic is greater than that seen in dizygotic twins. Infectious diseases,[6,7] RDS,[8-10] BPD, intraventricular hemorrhage, and NEC[11] are complex diseases in which genetic contribution to risk has been suggested by twin studies. Animal models in which genes are manipulated with a resulting clinical disease phenotype also provide evidence of genetic contributions to complex disease. Examples include manipulation of surfactant protein B gene with resultant severe RDS[12] and manipulation of the IGF-1 gene causing ROP.[13]

Another approach to documenting genetic contributions to complex disease is to establish familial aggregation. Studies of familial aggregation, or clustering, can be difficult in preterm infants because of difficulties obtaining samples from multiple generations and extended families. They have been limited to studies among siblings and parents of preterm infants. Familial recurrence of very specific severe RDS phenotypes has led to identification of particular rare genetic variants associated with disease.[14-16]

SELECTING CANDIDATE GENES

Given a possibility of genetic contribution to complex disease risk, investigators can begin searching for the particular genetic variants that add risk. Most often, candidate

BPD	Bronchopulmonary dysplasia	ROP	Retinopathy of prematurity
NEC	Necrotizing enterocolitis	SNP	Single nucleotide polymorphism
RDS	Respiratory distress syndrome	TDT	Transmission disequilibrium test

From the Department of Pediatrics, Division of Neonatology, Duke University School of Medicine, Durham, North Carolina; Center for Genomic Medicine, Duke Institute for Genome Sciences and Policy, Department of Medicine, Duke University School of Medicine, Durham North Carolina; and Department of Medicine, Center for Human Genetics, Duke University School of Medicine, Durham, North Carolina.

Supported by ES11961 and the George and Jean Brumley, Jr, Neonatal-Perinatal Research Institute, Department of Pediatrics, Duke University Medical Center, Durham, North Carolina.

Submitted for publication Aug 2, 2005; last revision received Dec 1, 2005; accepted Jan 4, 2006.

Reprint requests: C. Michael Cotten, MD, Box 3179, DUMC Department of Pediatrics, Division of Neonatology, Durham, NC 27710. E-mail: cotte010@mc.duke.edu

J Pediatr 2006;148:720-6

0022-3476/$ - see front matter

10.1016/j.jpeds.2006.01.006

genes are selected for investigation on the basis of biologic plausibility, for example, examining variations in cytokine genes for associations with disease phenotypes in which immunity and inflammation are thought to be major contributors to disease[17-20] or examining polymorphisms in the OATP 2 and UGT1A1 genes that play roles in transport and conjugation, respectively, of unconjugated bilirubin, for association with extremes of unconjugated hyperbilirubinemia.[21]

Strategies that broaden the search for association genes using genome-wide studies of thousands of genes, rather than small selections of "usual suspects" with suspected biologic plausibility, to identify chromosomal regions associated with disease risk are on the near horizon and await compilation of larger sample populations.[5,22] These studies take advantage of the property of linkage disequilibrium. Linkage disequilibrium occurs when two genetic variants are inherited together through multiple generations. Linkage disequilibrium allows mapping of chromosomal regions of interest by tracing known variants or "markers." If a marker occurs more frequently among cases than among control subjects, genetic loci in that region of the chromosome are potential sources for the true disease-associated variant. Such broad net studies could be used to test thousands of markers on hundreds of genes involved with particular physiologic pathways that may be involved with disease.

POTENTIAL PITFALLS OF CANDIDATE GENE ASSOCIATION STUDY DESIGN

Candidate gene association study results must be interpreted with caution. Any significant observation in such a study could be valid or false, and causes for spurious associations are numerous. Possibilities include a single nucleotide polymorphism (SNP) statistically associated with disease that is simply in linkage disequilibrium with the true functional but unidentified variant. Further problems in study design such as poor choice of control subjects (eg, genotype frequency among older gestation infants without disease versus frequency among lower gestation infants with disease) or problems with population stratification within the study population (high occurrence of a SNP in one population of particular continental ancestry versus low occurrence in another)[23] threaten validity of positive association studies. Additional challenges such as accounting for genotyping errors in study design and acknowledging need to assess proportions of genotypes given allele frequencies among cases and control subjects mandate genetic epidemiology input for study design, data analysis, and peer review.[22,24–29]

As resources and study populations increase, family-based case-control designs such as the transmission disequilibrium test (TDT) of case-parent triads will provide additional evidence for inherited disease. The TDT detects transmission of suspected disease susceptibility alleles or genotypes to affected offspring. Such studies avoid problems with population stratification among an unrelated study population by using limited family gene pools and also allow assessment of maternal and imprinting effects.[30] The Haataja et al[31] TDT study of surfactant protein A polymorphisms is an example from neonatology.

EXAMPLES OF CANDIDATE GENE STUDIES AMONG PRETERM INFANTS

Respiratory Distress Syndrome

Respiratory distress syndrome is a complex disease whose incidence and outcome are believed to involve interactions of multiple factors including gestational age, sex, race, and treatment strategies.[9,12] Epidemiologic investigations suggest a genetic contribution to RDS susceptibility.[12,32] Hallman et al[10] and Cole et al[14] have summarized the genetic influences in RDS susceptibility, with discussion of particular surfactant protein genotypes that cause severe RDS,[19] along with more recent work indicating genetic variations in surfactant pathways that contribute to the risk of severe RDS.[20]

Severe surfactant protein B (SP-B) deficiency was the first reported genetic cause of severe RDS phenotype[33,34]; however, the problem is rare relative to RDS associated with prematurity.[14,35] Investigators have found evidence for contribution to RDS risk from far more prevalent polymorphisms in the SP-A gene, alleles $6A^2$ and $1A^0$, and the SP-B Ile131Thr genotype. The most convincing results have come from Finnish populations.[10] The largest study (184 premature infants with RDS and 500 premature control infants) showed that RDS risk was associated with the two SP-A alleles, both with frequencies of greater than 50%. RDS disease phenotype was well defined by clinical, radiographic, and/or pathologic criteria. The association of the SP-A1 $6A^2$ risk allele with RDS was enhanced with the specific SP-B genotype, SP-B Ile131Thr.[36] Haataja et al[31] also performed transmission disequilibrium tests (a family-based association test) in a homogenous Finnish population, demonstrating that the $6A^2$-$1A^0$ haplotype was seen more often in affected infants, whereas the SP-A1 $6A^2$ allele was seen less often in families with infants born before 32 weeks without RDS.

For RDS, pathophysiologic mechanism can be plausibly associated with the surfactant protein genotypes. The risk alleles $6A^2$ and $1A^0$ have been associated with low levels of SP-A mRNA,[37] and low levels of SP-A are associated with more severe outcomes.[38,39] Infants with the risk allele may have problems with the pro–SP-B carbohydrate recognition domain of SP-A and the amount of SP-A.[36,40,41] Infants with higher amounts of SP-A may be protected against surfactant inactivating factors present in injured lungs.[39,42,43] In the future, SP-A content of surfactant replacement based on genotype could improve outcome for preterm infants with an increased-risk, low–SP-A genotype.

Rova et al[44] linked SP-B polymorphisms with BPD but not the SP-A alleles. This study included cases with BPD, defined as requirement for supplemental oxygen caused by deficient lung function at the postmenstrual age of 36 weeks (for subjects of gestation <32 weeks at birth) or at 56 days for those born at 32 weeks. Control infants were divided into those with and those without RDS. Even after accounting for