Published in final edited form as:

Birth Defects Res A Clin Mol Teratol. 2009 January ; 85(1): 6–12. doi:10.1002/bdra.20546.

Genetic Tools and Algorithms for Gene Discovery in Major Congenital Anomalies

Patricia K. Donahoe^{1,*}, Kristin M. Noonan¹, and Kasper Lage^{1,2}

¹Pediatric Surgical Research Laboratories, Massachusetts General Hospital and Department of Surgery, Harvard Medical School, Boston, Massachusetts

²Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark

Major congenital malformations are a leading cause of infant mortality in this country, as improved obstetrical and neonatal care have successfully addressed many other previously lethal disorders (Centers for Disease Control and Prevention [CDC], 2008). Patients with congenital anomalies of the brain, heart, lung, and skeleton as well as gastrointestinal, craniofacial, and other organs fill as many as one third of all beds in children's hospitals in the developed world (McCandless et al., 2004), with little known about the underlying mechanisms. Of the 4 million births per year in the United States, 3% (120,000) have major congenital abnormalities (CDC, 2008). Survivors require extensive and repeated hospitalizations, often with complex reconstructive surgery, to achieve a good quality of life. For others, only prolonged supportive care can be offered. The economic, social, and emotional burden is enormous (CDC, 2007). It is clearly an imperative to discover new ways to diagnose and treat these patients. Even one discovery that ameliorates a single condition can have a substantial effect if it lessens or even averts the phenotype—for example, the preventive effect of folic acid on reducing the prevalence of neural tube defects (de Wals et al., 2007; Wald, 2004; Oakley et al., 2001; Czeizel et al., 1992).

Recent progress made in discovering genomic alterations requires expensive genetic and genomic tools, novel statistical and integrative analyses, a wide variety of molecular functional assays to analyze gene variations and to conceive potential therapeutics, and a well-organized infrastructure in which to conduct broad studies and to archive patient and sample collections. We have applied new and evolving genetic tools for gene discovery in the study of congenital diaphragmatic hernia (CDH) that we think are applicable to other congenital anomalies. Although a genetic contribution to congenital anomalies has long been recognized, traditional linkage cannot be applied to these early lethal disorders such as CDH. However, increased recurrence risk in siblings (Piacentini et al., 2005), the 20-fold occurrence in consanguineous unions of having multiple affected children (Stoltenberg et al., 1999), the association of malformations with chromosomal abnormalities (Poer et al., 2005), and a recent discovery of patients with de novo mutations in crucial developmental genes (Ackerman et al., 2005; Kantarci et al., 2006) offer correlative proof that a genetic component is responsible. Standard cytogenetic analyses have been augmented by array comparative genomic hybridization (aCGH) to discover de novo or inherited structural variants such as microdeletions and microduplications. High-density single nucleotide polymorphism (SNP) analyses (genotyping) are used both for homozygosity mapping and to assess the contribution of SNP differences or similarities in individual cohorts being analyzed concurrently with our colleagues studying

*Correspondence to: Patricia K. Donahoe, M.D., Director, Pediatric Surgical Research Laboratories, Massachusetts General Hospital, 185 Cambridge Street, Boston, MA 02114. donahoe.patricia@mgh.harvard.edu.
Presented as the Keynote Address at the Lewis B. Holmes Festschrift in Boston on May 10, 2008.

heart, brain, and neurosensory hearing disorders, each having collected significant cohorts of patients for each major anomaly (Fig. 1). We will integrate these findings using novel computational and functional analyses. It is clear that such approaches will be necessary to make significant changes in the care of these patients, but that the technologies discussed are rapidly evolving.

Genomic Technologies Applicable to Analyses of Congenital Malformations: Selecting the Platform in a Moving Stream

Array comparative genomic hybridization is a powerful, but currently expensive, tool used to detect structural differences in the genome, or copy number variations (CNVs), in the form of microdeletions or microduplications. Our first analyses were done on the spectral array platform, which has one probe per megabase (Mb). Since two probes are needed to confirm a microdeletion or microduplication, it has a 2-Mb effective resolution. Analyses were repeated on patients with CDH using the Agilent 244 K oligonucleotide aCGH platform, which uses 1 probe per 9 Kb across the whole genome, and 1 probe per 7.4 Kb in coding regions. Three, or possibly two, probes are required to validate results, thereby giving a 15- to 30-Kb effective resolution for this platform. When used to analyze 52 patients with complex CDH, a total of 68 variants were found, with 47 in coding regions. Variants, validated by quantitative PCR assay, will be examined in parents. By comparison, the spectral array generated 3 variants in 39 patients, each confirmed on the Agilent 244K platform. The Affymetrix 6.0 array (Affymetrix, Santa Clara, CA), which detects both SNP variations and CNVs, has approximately 1 million copy number probes across the approximately 3 billion base pairs of the human genome, or 1 probe per 3 Kb. Three to five probes are required for each call, giving the Affymetrix Array a 15- to 30-Kb effective resolution. The Affymetrix 6.0 chip also contains 906,600 SNP probe sets with perfect match oligonucleotide probes for alleles A and B and a mismatch probe for each allele to detect nonspecific binding.

Algorithms to Interpret Genome-Wide Data

Much of the success in the use of genomic tools depends on the reliability of complex computational interpretation. Researchers at the Broad Institute of the Massachusetts Institute of Technology and Harvard University developed a suite of tools to detect SNP genotypes, common CNVs, and rare de novo CNVs from the Affymetrix 6.0 chip. This SNP genotype algorithm, Birdseed, is an evolution of the robust linear modeling genotype-calling algorithm (Rabbee et al., 2006). It performs a multiple-chip analysis to estimate the signal intensity for each allele (AA, AB, or BB) of each SNP, then makes genotype calls by giving confidence scores for every individual at every SNP. Because it is a clustering algorithm, it performs best when run on many samples simultaneously. Another algorithm detects rare de novo CNVs, and another serves to integrate CNV information with SNP genotypes (Weiss et al., 2008). In addition to these algorithms, the Partek Genomics Suite (Partek, St. Louis, MO) provides comprehensive Affymetrix-compatible software for copy number and SNP analysis; Affymetrix also recommends the Chromosomal Copy Number Analysis Tool for copy number calls.

In using genome-wide screens to detect SNPs significantly associated with diseased individuals (cases) versus controls for genome-wide association studies, the crucial issue is obtaining enough power to identify associated alleles that have an effect, given the number of cases and controls used in the study. Although large studies require considerable effort and expense, recent discoveries in obesity (Frayling et al., 2007) and type 2 diabetes (Zeggini et al., 2008) have shown the feasibility of this strategy. In contrast, for rare congenital disorders the most difficult hurdle of analysis is the lack of large numbers of patients for study, suggesting that a combination of genome-wide association studies with other strategies should be considered

for this subset of important diseases to improve the modest statistical power anticipated for each technique.

Platform Comparisons

In collaboration with the laboratory of Charles Lee (Iafrate et al., 2004), we have compared current aCGH platforms to select those best suited for use in congenital anomalies. For the pilot phase of the National Institutes of Health National Human Genome Research Institute 1000 Genome Project with the Sanger Center, three trios (affected proband plus both parents), will be whole genome sequenced at 20× coverage using a Solexa sequencing platform (Illumina, Inc., San Diego, CA, USA) at the Broad Institute and at the Sanger Center. As part of a platform comparison, these individuals will also be analyzed by different array-based technologies including Agilent 244 (Agilent Technologies Inc., Santa Clara, CA, USA), Illumina (1M) (Illumina Inc., San Diego, CA, USA), the 2.1-million Feature NimbleGen HD2 aCGH (Roche Nimblegen, Madison, WI, USA), plus the new Affymetrix 6.0 (Affymetrix Inc., Santa Clara, CA, USA) and Illumina Infinium Human 1M Beadchip genotyping arrays that provide both CNV and SNP data. This comparison will help us select the optimal CNV platform to use going forward, especially in resource-constrained settings.

Congenital Diaphragmatic Hernia

Congenital diaphragmatic hernia is a common human birth defect characterized by high mortality and morbidity owing to the lethal components of pulmonary hypoplasia and pulmonary hypertension that accompany the diaphragm malformations. We have used CDH as a model anomaly to identify microdeletions or microduplications (not detectable by conventional G-banded karyotyping) as major causes of CDH. We enrolled patients with CDH at Massachusetts General Hospital for Children and at Children's Hospital in Boston, where we have an established infrastructure of surgeons, geneticists, nurses, and study coordinators at both hospitals augmented by a parent support group, the Association of Congenital Diaphragmatic Hernia Research, Advocacy, and Support, and an international community of pediatric geneticists. More than 350 patients with CHD, their parents, and their healthy siblings have been recruited.

Phenotypic classification of CDH patients from all sources is entered into a confidential database derived from a parental questionnaire, medical record review, and physical examination. Sixty percent of all CDH patients are classified as isolated cases, with only diaphragm malformations, while the remaining 40% are considered complex, with recognized syndromes that include CDH, nonsyndromic associations, or chromosome abnormalities (Pober, 2007). DNA samples are obtained from blood, skin, amniotic fluid, frozen tissue, and formalin-fixed pathology specimens. Most patients have either an Epstein-Barr (EB) virus-transformed lymphoblast or a primary fibroblast cell line established. Cell lines are also created for both parents and siblings, when indicated.

To identify genes responsible for CDH, we sequenced in the patient cohort multiple candidate genes that, when mutated in transgenic animal models, produced the CDH phenotype. We also sequenced candidate genes in relevant molecular pathways essential for stage-specific lung development and for genes in regions of recurrent chromosomal aberrations associated with CDH. In addition, consanguineous and multiplex families were sought worldwide. Children from consanguineous families were genotyped by homozygosity mapping in which genome-wide SNP analysis was used to determine regions of shared homozygosity, or “identity by descent.” To delineate regions of microdeletions or microduplications, aCGH was performed on DNA of patients with syndromic or complex CDH. Intervals were narrowed by fluorescence in situ hybridization or multiplex ligation-dependent probe amplification (Kantarci et al., 2007b). We identified a critical locus associated with CDH at 1q41-q42.12 for Fryns syndrome

(Kantarci et al., 2006) and are working with colleagues (Briggs et al., 2004) to narrow a previously detected locus at 15q26.1-q26.2 and another at 8p, both long known to be linked to CDH (Scott et al., 2007; Kantarci et al., 2007b; Klaassens et al., 2005; Shimokawa et al., 2005). As has already been shown for other birth defects, we anticipate that genes found to be involved in syndromic CDH will elucidate molecular pathways that are involved in nonsyndromic or isolated CDH. For example, mutations in *IRF6* cause the syndromic form of cleft lip with or without cleft palate known as *van der Woude syndrome*, whereas over-transmission of a common *IRF6* SNP increases the risk for isolated cleft lip with or without cleft palate (Zuccherro T M et al., 2004). Furthermore, we also predict that nodal points in molecular pathways will serve as targets for therapeutic intervention to alleviate the phenotype, or to prevent the disorder.

Consanguinity: Link to Congenital Diaphragmatic Hernia in Donnai-Barrow Syndrome

An important breakthrough in CDH gene discovery came from analysis of a consanguineous family with multiple affected siblings with Donnai-Barrow syndrome (DBS) (Kantarci et al., 2007a; Donnai and Barrow, 1993). SNP analysis showed a region (*2q23.3* to *2q31.1*) indicative of shared homozygosity, or identity by descent. Eight additional families were recruited in the Middle East and Europe (Kantarci et al., 2007a).

Microsatellite marker analysis in these multiplex kindreds refined the region to 18 Mb, containing 51 genes between *D2S2299* and *D2S2284*, which generated a 2-point (logarithm 10 the ratio odds of linkage between Loci) LOD score of 4.31 and multipoint LOD score of 6.243. Sequencing of the genes in that region revealed mutations in a large 79-exon gene (low density lipoprotein receptor related protein 2 [*LRP2*]), which encodes megalin, a multiple ligand receptor. All patients were found to excrete large amounts of low molecular weight urinary proteins, including retinol binding protein, which can now serve as a surrogate marker for DBS and supports the previous evidence that the retinoic acid pathway is a contributor to the CDH complex (Babiuk et al., 2004; Chen et al., 2003). Mutations were found in all kindreds with DBS at various positions across the 79 exons. Resequencing for 182 patients with CDH (Noonan et al., unpublished) and for 50 patients with absent corpus collosum (Walsh et al., unpublished) has only recently been completed and is undergoing analysis. Thus, genes detected in a consanguineous population can be retested in larger patient cohorts from nonconsanguineous unions with the hope of uncovering molecular pathways responsible for isolated malformations.

Resequencing

Combining our discovery of megalin mutations in patients with DBS with other loci known to be involved in CDH, we are resequencing a number of genes in our entire cohort of patients with CDH. Megalin presents a remarkable clue pointing to the retinoic acid pathway (Chen et al., 2003), and *DISP1* in chromosome 1q42 implicates the Sonic hedgehog pathway (McCarthy et al., 2003; Bellusci et al., 1997), both long suspect in the etiology of CDH (Loscertales et al., 2008; Roberts et al., 1995). The former raises the possibility of retinoic acid modulations as preventative or as therapeutic for CDH. New connectivity maps of genes from regions with microdeletions and knockouts, which produce CDH phenotypes in animal models, suggest that resequencing genes such as *RAP* (4q16), *DISP1* (1q42), *HLX1* (1q42), *MEF2A* (15q26), and *STRA6* (15q26) (Web links, <http://www.genecards.org/>) in the entire cohort may be revealing. These genes and their interaction matrices will be expanded by discovery based on CNV and SNP analyses.

We hope to select the optimal genomic tools to evaluate children with a number of congenital anomalies, with the expectation of incorporating these technologies as routine standard of care in the diagnostic armamentarium of physicians responsible for these patients. Novel computational algorithms used to study the integration of genes mapped to susceptibility loci, and their expressed proteins, will allow us to predict otherwise unexpected molecular pathways that contribute to congenital anomalies.

Computational Biology

Unbiased attempts to identify molecular pathways contributing to the causative genes in CDH entails strategies in which genomic data is integrated with a human protein-protein interaction (PPI) network created by pooling human interaction data from a many protein databases and from model organisms (Fig. 2) (Lage, et al., 2007). To reduce inherent noise, we devised and tested a network-wide confidence score (von Mering et al., 2002) to produce a network containing more than 400,000 unique interactions between more than 10,200 human proteins. Of these, more than 70,000 are high-confidence interactions.

We then generated a human PPI network spanning a set of 54 proteins known to be involved in the CDH related phenotypes in humans or mice and queried for first and second order interaction partners (Barabási and Oltvai, 2004). Next, we assigned a score to the network based on the fraction of CDH-related proteins in the total amount of proteins in the second order network (number input/total) (Fig. 2) (Bergholdt et al., 2007; D'Hertog et al., 2007; Lage, et al., 2007). Compared to random networks, the CDH network score was highly significant ($p < 0.00001$), suggesting that CDH could indeed be caused by defects affecting different proteins in one or several shared pathways and that network approaches could be used in combination with genetic data to identify proteins and systems involved in the etiology of the disease (Goh et al., 2007). These results encourage us to include computational network approaches that integrate CNV and SNP data in future investigations of CDH and related pathways and genes.

Figure 3 illustrates how loci related to a particular phenotype can be used in concert with PPI data to elucidate putative genes and molecular pathways associated with that phenotype. In the study by Bergholdt et al. (2007), genetic loci from a large sibling study of type 1 diabetes were interrogated for direct physical interactions between proteins from the different loci. The resulting networks were subjected to statistical testing to identify a small subset of networks significantly connecting proteins from multiple associated loci. These networks indicate significant functional association between specific genes in associated loci, thereby incriminating specific genes and pathways in the molecular etiology of type 1 diabetes. A similar strategy can be used in congenital anomalies to deduce networks composed of genes from loci that have been found, with either CNV or SNP technologies, to be associated with the anomalies. We are currently pursuing this approach in CDH.

Four Congenital Anomalies

We are exploring the hypothesis that common embryonic and genetic insults can affect multiple organ systems that develop simultaneously or at different times during embryonic development. Mutations at different points in a finite number of key developmental pathways can affect multiple organ systems either synchronously or asynchronously. In addition to detecting anomaly-specific variations in important developmental pathways (such as Sonic hedgehog, WNT, retinoic acid, Notch, TGF β , or FGF), we may also find shared pathway defects across anomalies. These latter variations could impart differences in severity and cause organ-specific phenotypes, because of the relative nodal position of the gene defect in a specific molecular pathway or the availability of compensatory pathways that may vary with the stage of gestation. To assess this hypothesis, we have employed the Affymetrix 6.0 chip for a

genome-wide association study of ~200 patients from each of four collaborative groups studying major congenital malformations including (1) the Seidman laboratory studying heart (Morita et al., 2008; Alcalai, et al., 2007), (2) the Walsh laboratory studying the brain (Shen, et al., 2007; Amadio et al., 2006), (3) the Morton group examining neurosensory hearing loss (Morton et al., 2006; Lee et al., 2008), and (4) our group examining lung/diaphragm malformations (Pober, 2007; Kantarci et al., 2007a, 2007b). Most of these patients are the only affected individuals in each family, and most often they are born to unrelated parents. This “meta-analysis” across malformations will search for patterns of variation held in common among the various disorders (Passutto et al., 2007; Golzio, et al., 2007), as well as those specific to each disorder. SNP array data will be integrated with CNV data from the same 6.0 chip for each group (Morrow et al., 2008), and then compared across all of the major malformations. CNV results from aCGH platforms, where available, will be compared to CNV data derived from the Affymetrix 6.0 chips, and the data will be integrated.

For these carefully phenotyped patients, the era of SNP and CNV discovery will be expanded with computational analyses that will integrate these two powerful detection engines across the genome. These integrations will be augmented by protein interaction data from public databases, such as gene expression libraries and multiple annotated text libraries (Lage et al., in press, PNAS), to reveal logical pathways for gene discovery. We anticipate that this march toward understanding the genetic basis of major congenital anomalies will lead to theoretical, and ultimately, experimentally validated treatment strategies.

Because karyotyping is one of the few diagnostics covered by third party payors and available for the benefit of our patients with congenital anomalies, one of our missions after establishing the optimal platforms is to work toward reducing the costs and to lobby for the use of appropriate genomic tools now remunerated only by research or philanthropic funds. Furthermore, hospital-wide infrastructure must be established for orderly phenotyping, genotyping, and integrated bioinformatic analyses of all patients with congenital anomalies.

Acknowledgments

This project is supported by NICHD (HD 055150) and The Broad Institute (Harvard and Massachusetts Institute of Technology), Cambridge, Massachusetts.

We gratefully acknowledge the generous contributions of Drs. Charles Lee, Christopher Walsh, Ahmed Teebi (Qatar), and Lihadh Al-Gazali (United Arab Emirates), as well as Sibel Kantarci, M.D., and R. Sean Hill, Ph.D, who are now Harvard Medical School faculty members. We are also grateful to Barbara Pober, M.D., for her wisdom and guidance, and Caroline Coletti for her editorial expertise in preparing this manuscript. We look forward to future comparative analyses with Cynthia Morton, M.D., and her colleagues, Christine Seidman, M.D., and Jonathan Seidman, Ph.D.

References

- Ackerman KG, Herron BJ, Vargas SO, et al. Fog2 is required for normal diaphragm and lung development in mice and humans. *PLoS Genet* 2005;1:58–65. [PubMed: 16103912]
- Alcalai R, Seidman JG, Seidman CE. Genetic basis of hypertrophic cardiomyopathy: from bench to the clinics. *J Cardiovasc Electrophysiol* 2007;19:104–110. [PubMed: 17916152]
- Amadio JP, Walsh CA. Brain evolution and uniqueness in the human genome. *Cell* 2006;126:1033–1035. [PubMed: 16990130]
- Babiuk RP, Thébaud B, Greer JJ. Reductions in the incidence of nitrofen-induced diaphragmatic hernia by vitamin A and retinoic acid. *Am J Physiol Lung Cell Mol Physiol* 2004;286:L970–973. [PubMed: 14729516]
- Barabási AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet* 2004;5:101–113. [PubMed: 14735121]
- Bellusci S, Furuta Y, Rush MG, et al. Involvement of Sonic hedgehog (Shh) in mouse embryonic lung growth and morphogenesis. *Development* 1997;124:53–63. [PubMed: 9006067]

- Bergholdt R, Storling ZM, Lage K, et al. Integrative analysis for finding genes and networks involved in diabetes and other complex diseases. *Genome biology* 2007;8:R253. [PubMed: 18045462]
- Biggio JR Jr, Descartes MD, Carroll AJ, et al. Congenital diaphragmatic hernia: is 15q26.1–26.2 a candidate locus? *Am J Med Genet A* 2004;126A:183–185. [PubMed: 15057983]
- Centers for Disease Control and Prevention. Hospital stays, hospital charges, and in-hospital deaths among infants with selected birth defects—United States, 2003. *MMWR Morb Mortal Wkly Rep* 2007;56:25–29. [PubMed: 17230142]
- Centers for Disease Control and Prevention. Update on overall prevalence of major birth defects—Atlanta, Georgia, 1978–2005. *MMWR Morb Mortal Wkly Rep* 2008;57:1–5. [PubMed: 18185492]
- Czeizel AE, Dudas I. Prevention of the first occurrence of neural tube defects by periconceptional vitamin supplementation. *N Engl J Med* 1992;327:1832–1835. [PubMed: 1307234]
- Chen MH, MacGowan A, Ward S, et al. The activation of the retinoic acid response element is inhibited in an animal model of congenital diaphragmatic hernia. *Biol Neonate* 2003;83:157–61. [PubMed: 12660430]
- de Wals P, Tairou F, Van Allen MI, et al. Reduction in neural-tube defects after folic acid fortification in Canada. *N Engl J Med* 2007;357:135–142. [PubMed: 17625125]
- D'Hertog W, Overbergh L, Lage K, et al. Proteomics analysis of cytokine-induced dysfunction and death in insulin-producing INS-1E cells: new insights into the pathways involved. *Mol Cell Proteomics* 2007;6:2180–2199. [PubMed: 17921177]
- Donnai D, Barrow M. Diaphragmatic hernia, exomphalos, absent corpus callosum, hypertelorism, myopia, and sensorineural deafness: a newly recognized autosomal recessive disorder? *Am J Med Genet* 1993;47:679–82. [PubMed: 8266995]
- Frayling TM, Timpson NJ, Weedon MN, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 2007;316:889–894. [PubMed: 17434869]
- Goh KI, Cusick ME, Valle D, et al. The human disease network. *Proc Nat Acad Sci U S A* 2007;104:8685–90.
- Golzio C, Martinovic-Bouriel J, Thomas S, et al. Matthew-Wood syndrome is caused by truncating mutations in the retinol-binding protein receptor gene STRA6. *Am J Hum Genet* 2007;80:1179–87. [PubMed: 17503335]
- Iafrate AJ, Feuk L, Lee C, et al. Detection of large-scale variation in the human genome. *Nat Genet* 2004;36:949–951. [PubMed: 15286789]
- Kantarci S, Al-Gazali L, Hill RS, et al. Mutations in LRP2, which encodes the multiligand receptor megalin, cause Donnai-Barrow and facio-oculo-acoustico-renal syndromes. *Nat Genet* 2007a;39:957–959. [PubMed: 17632512]
- Kantarci S, Donahoe P. Congenital diaphragmatic hernia (CHD) etiology as revealed by pathway genetics. *Am J Med Genet* 2007b;145C:217–226. [PubMed: 17436295]
- Kantarci S, Casavant D, Prada C, et al. Findings from aCGH in patients with congenital diaphragmatic hernia (CDH): a possible locus for Fryns syndrome. *Am J Med Genet* 2006;140A:17–23. [PubMed: 16333846]
- Klaassens M, van Dooren M, Eussen HJ, et al. Congenital diaphragmatic hernia and chromosome 15q26: Determination of a candidate region by use of fluorescent in situ hybridization and array-based comparative genomic hybridization. *Am J Hum Genet* 2005;76a:877–882. [PubMed: 15750894]
- Lage K, Karlberg EO, Storling ZM, et al. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol* 2007;25:309–316. [PubMed: 17344885]
- Lee C, Morton CC. Structural genomic variation and personalized medicine. *N Engl J Med* 2008;358:740–741. [PubMed: 18272898]
- Loscertales M, Mikels AJ, Hu JK, et al. Chick pulmonary Wnt5a directs airway and vascular tubulogenesis. *Development* 2008;135:1365–1376. [PubMed: 18305003]
- McCandless SE, Brunger JW, Cassidy SB. The burden of genetic disease on inpatient care in a children's hospital. *Am J Hum Genet* 2004;74:121–127. [PubMed: 14681831]
- McCarthy RA, Argraves WS. Megalin and the neurodevelopmental biology of sonic hedgehog and retinol. *J Cell Sci* 2003;116(Pt 6):955–60. Review. [PubMed: 12584240]

- Morita H, Rehm HL, Menesses AD, et al. sarcomere protein gene mutations in pediatric cardiac hypertrophy. *N Engl J Med* 2008;358:1899–1908. [PubMed: 18403758]
- Morrow EM, Yoo SY, Flavell SW, et al. Identifying autism loci and genes by tracing recent shared ancestry. *Science* 2008;321:218–223. [PubMed: 18621663]
- Morton CC, Nance WE. Newborn hearing screening—a silent revolution. *N Engl J Med* 2006;354:2151–2164. [PubMed: 16707752]
- Oakley GP. Minimum fully effective dose of folic acid for the prevention of neural tube defects. *Epidemiology* 2001;12:475. [PubMed: 11416789]
- Piacentini G, Digilio MC, Capolino R, et al. Familial recurrence of heart defects in subjects with congenitally corrected transposition of the great arteries. *Am J Med Genet A* 2005;137:176–180. [PubMed: 16059940]
- Pober B. Overview of epidemiology, genetics, birth defects, and chromosome abnormalities associated with CDH. *Am J Med Genetics* 2007;145C:158–171. [PubMed: 17436298]
- Pober B, Lin A, Russell M, et al. Infants with Bochdalek diaphragmatic hernia: sibling precurrence and monozygotic twin discordance in a hospital-based malformation surveillance program. *Am J Med Genet* 2005;138A(2):81–88. [PubMed: 16094667]
- Rabbee N, Speed TP. A genotype calling algorithm for affymetrix SNP arrays. *Bioinformatics* 2006;22:7–12. [PubMed: 16267090]
- Roberts DJ, Johnson RL, Burke AC, et al. Sonic hedgehog is an endodermal signal inducing Bmp-4 and Hox genes during induction and regionalization of the chick hindgut. *Development* 1995;121:3163–3174. [PubMed: 7588051]
- Scott DA, Klaassens M, Holder AM, et al. Genome-wide oligonucleotide-based array comparative genome hybridization analysis of non-isolated congenital diaphragmatic hernia. *Hum Mol Genet* 2007;6:424–430. [PubMed: 17210672]
- Shen Y, Irons M, Miller DT, et al. Development of a focused oligonucleotide-array comparative genomic hybridization chip for clinical diagnosis of genomic imbalance. *Clin Chem* 2007;53:2051–2059. [PubMed: 17901113]
- Shimokawa O, Miyake N, Yoshimura T, et al. Molecular characterization of del(8)(p23.1p23.1) in a case of congenital diaphragmatic hernia. *Am J Med Genet* 2005;136A:49–51. [PubMed: 15937941]
- Stoltenberg C, Magnus P, Skrandal A, et al. Consanguinity and recurrence risk of birth defects: a population-based study. *Am J Med Genet* 1999;82:423–428. [PubMed: 10069715]
- von Mering C, Krause R, Snel B, et al. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 2002;417:399–403. [PubMed: 12000970]
- Wald N. Folic acid and the prevention of neural-tube defects. *N Engl J Med* 2004;350:2209–2211.
- Web Links/URLs: <http://www.genecards.org/>
- Weiss LA, Shen Y, Korn JM, et al. Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* 2008;358:667–675. [PubMed: 18184952]
- Zeggini E, Scott LJ, Saxena R, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature Genetics* 2008;40:638–645. [PubMed: 18372903]
- Zucchero TM, et al. Interferon regulatory factor 6 (*IRF6*) gene variants and the risk of isolated cleft lip or palate. *N Engl J Med* 2004;351:769–780. [PubMed: 15317890]

EDITORIAL

***RNAi, or RNA interference, is a technique which can be used to silence genes.

Figure 1.
Systems approach to gene discovery.

Figure 2.

The congenital diaphragmatic hernia (CDH) network. Proteins involved in CDH-related phenotypes in mice or humans are yellow. Interaction partners are gray. Of 54 CDH-related proteins, 41 can be placed in this network. Compared to randomly generated networks, this network has a highly significant network score ($p < 0.00001$, see text). Although this figure, due to its resolution, cannot be used directly to deduce functional relationships on the gene level, it illustrates the proximity of the CDH proteins to each other in a human protein-protein interaction (PPI) network. Together these results indicate that the CDH proteins might participate in one or several common pathways involved in human development and that

integrating genetic and PPI data could be a powerful way of identifying novel genes involved in CDH.

Figure 3.

Example of an integrative approach that combines genomic and protein protein-interaction (PPI) data. Regions that show statistical evidence for epistasis in type 1 diabetes (T1D) are identified indicating that these regions harbor genes playing a role in the etiology of this disease. Proteins coded by genes in these regions are extracted, followed by systematically searching for subnetworks significantly enriched for genes from the epistatic regions, which would suggest the presence of a functional cellular subnetwork that is associated with the disease in question. Upon identifying such subnetworks, relevant genes can be tested experimentally in patient material. (Bergholdt et al., 2007). Permission obtained from Genome Biology.