Coupling genomics and human genetics to delineate basic mechanisms of development

Leslie G. Biesecker, MD

This article, based on a presentation given by the author at the third Asan-Harvard Medical International Symposium on "Genomics and Proteomics: Impact on Medicine and Health" in Seoul, Korea, July 3–4, 2001, discusses an iterative translational research approach to delineate the basic mechanism of human development. The study of humans to increase the understanding of mammalian development has critical advantages that make its limitations acceptable for certain types of studies. For instance, by looking at families affected by birth defects, researchers can gain insight into the basic mechanisms of development and how genes program organisms to assume their permanent, or adult, morphological shapes. A number of malformation syndromes have some overlapping manifestations, despite being phenotypically and, in some cases, genetically distinct. What can researchers learn from this? The author's research group clinically and genetically analyzed families affected with the Pallister-Hall syndrome. The researchers then went on to look at the McKusick-Kaufman syndrome, a disorder that is more common among the Old Order Amish of Lancaster County, Pennsylvania, in an attempt to understand more about genes, genetic pathways, and syndrome families. *Genet Med* 2002:4(6, Supplement):39S–42S.

Key Words: *human development, syndrome families, Pallister-Hall syndrome, McKusick-Kaufman syndrome, Old Order Amish of Lancaster County, Pennsylvania*

Among a number of potential approaches to delineate mechanisms of human development, the study of families affected by birth defect syndromes has numerous advantages. The goal is to understand basic mechanisms of development and how genes program organisms to assume their permanent, or adult, morphological shapes. This approach integrates human studies with model systems, such as fruit flies, mice, worms, and other more simple organisms. The approach is iterative, which is to say that it is necessary to cycle those data back to the clinic and again to the laboratory to understand both the problems that our patients suffer and the mechanisms of development.

The study of humans to increase the understanding of mammalian development has critical advantages that make its limitations acceptable for certain types of studies. First, extremely detailed study of the organism is possible. There is no animal on this planet that is studied in as exquisite detail as the human. This study has been going on for over a century: The beginning of the century was marked by early morphologic studies,[1] the middle with detailed embryonic and fetal pathologic work,[2] and in the third part of the century, the detailed clinical study of human malformations and birth defects.[3] This comprises a vast storehouse of information that no animal system comes close to matching. The range of phenotypes in humans is enormous. There are many more human phenotypes than there are model system phenotypes, primarily attributable to the high level of medical and lay interest in the understanding of these disorders, which has resulted in development of a number of sophisticated analytic tools. Not only is the number of phenotypes large, but there are numerous subtle malformations and functional abnormalities as well. These disorders are thoughtfully catalogued in a number of medical databases, although work needs to be done to integrate a number of disparate sources. The challenge for the clinical researcher is to exploit these advantages and apply as many of them as possible to the study of patients to address important biological questions and improve health.

Despite its power, using humans as the primary research subject has limitations. In the study of human development, it is (obviously) not appropriate to experimentally manipulate the system to test hypotheses. In addition, there is limited access to tissues and organisms during development, and it is possible to monitor development only at certain time points and using certain techniques, most of which are noninvasive. The final limitation or disadvantage is that study of humans is both expensive and slow. The American medical care system is very sophisticated but extraordinarily expensive. Clinical research costs are generally comparable to clinical care costs and thus are very expensive as well. In addition, as a genetic model system, the human has a long generation time, so families must be followed over long periods of time to study additional affected individuals and natural history and to determine modes of inheritance.

The typical approach to human translational research is to take a basic laboratory discovery and creatively apply that

knowledge to effect an improvement in clinical care. This is no small task, yet there is much more that can be derived from working at the clinical-bench research interface. The goal is to make clinical research more iterative and impart more feedback and bidirectionality within the system. The end result of this is to extract the maximum possible amount of basic knowledge from patients and supply the maximum amount of beneficial knowledge back to the patients. For example, once the molecular etiology of a syndrome is known, the researcher can go back to the clinic and look at the patients again to determine whether this molecular etiology can tell us something important about the patients. The researcher can then use these clinical and molecular data together to develop novel hypotheses about mechanisms of development, modifying factors, environmental variables, or other disorders with overlapping manifestations. The goal is to keep the cycle in motion, continually extracting new information and improving the care of our patients.

There are numerous examples of this type of iterative translational research that can be used to illustrate this concept. The example cited here constitutes a family of syndromes, that is, a number of malformation syndromes that have some overlapping manifestations but are phenotypically and, in some cases, genetically distinct. The first disorder is the Pallister-Hall syndrome (PHS), which comprises hypothalamic hamartoma, imperforate anus, laryngeal anomalies, and central polydactyly, with shortening of the terminal digits.[4,5] The disorder is inherited in an autosomal dominant pattern and has significant interfamilial variability. Remarkably, there are only five disorders with central polydactyly, and each of those is rare.[6] For that reason, there was little knowledge, either clinical or basic, on this malformation or on PHS. My research group performed a clinical and genetic analysis of families affected with this disorder and determined that it is caused by mutations in the *GLI3* zinc finger transcription factor gene.[7] This result was startling to us because approximately 7 years previously, alterations in that gene had been shown to occur in patients who have a different malformation syndrome, the Greig cephalopolysyndactyly syndrome (GCPS).[8] The gene was cloned in 1990 because those patients had balanced translocations that interrupt the gene. GCPS is a distinct disorder of human development caused by haploinsufficiency of *GLI3*.

A number of questions arose that sent us immediately back to the clinic. The immediate challenge was to confirm that these syndromes were distinct (which they clearly are) and to carefully delineate how they were distinct. The next challenge was to search the scientific literature to develop a model that could explain how two distinct phenotypes are caused by mutations in a single gene. It turned out that the *Drosophila* homologue of *GLI3* is *Cubitus interruptus*, or *ci*, and mutations in *ci* cause a wide range of phenotypes in fruit flies.[9] The mildest end of that spectrum of abnormalities is abnormal wing vein patterning, whereas more severe mutations cause early larval death (http://flybase.harvard.edu). *ci* functions in a genetic pathway downstream of the hedgehog signaling molecule. *ci* resides in the cytoplasm, and upon hedgehog signaling, it mi-

grates to the nucleus and activates the transcription of downstream genes. Alternatively, ci protein can be cleaved into a shorter form, which is a negative controller of downstream genes, and this short form translocates to the nucleus to turn off the expression of downstream genes.[10]

The combination of what the patients and their mutations were telling us, together with what was known about fruit flies, led to the proposal that the *GLI3* protein was cleaved by the same mechanism as was the ci protein.[11] Furthermore, the truncation of the *GLI3* protein, if constitutive, would be expected to have a different clinical consequence than would haploinsufficiency of the gene. This distinct biological mechanism was hypothesized to result in a different syndrome, PHS instead of GCPS. The model proposes that the PHS mutation changes the quality of the signal, but not the quantity. In contrast, the GCPS mutation changes the quantity but not the quality. The model has been shown to be substantially correct. It has been shown that *GLI3*, like *ci*, is proteolytically processed.[12] Thus the study of a disorder affecting only a few hundred or perhaps a thousand people can not only benefit those patients directly, but can illuminate basic mechanisms of mammalian development.

The story could stop there but continues into the next iterative step of this translational research project. We hypothesized that in syndrome families that share certain malformations, these malformations should be caused by mutations in genes in connected developmental pathways. That is, mutations in genes that comprise a pathway should produce human malformation syndromes that have overlapping manifestations. This is an example of where the enormous amounts of catalogued human data are critically important. The accomplishment of this next step required going into the literature to find the disorder to match the hypothesis. This match turned out to be the McKusick-Kaufman syndrome (MKS).[13] This disorder is inherited in an autosomal recessive pattern and comprises central polydactyly (though postaxial polydactyly may be more common), congenital heart disease, and hydrometrocolpos (congenital uterine outflow obstruction that leads to uterine dilatation and secondary compromise of other organs). This disorder is again rare, with fewer than 100 case reports in the medical literature. However, it was clear that this disorder was much more common among the Old Order Amish of Lancaster County, Pennsylvania.[14] This is a closed population of religiously conservative persons who immigrated to the United States in the mid-18th century. They now number between 50,000 and 100,000 persons and are all descended from only several hundred founders. Because of this population structure, the group has an increased frequency of a number of disorders that are inherited in an autosomal recessive pattern. The population structure is thus well suited to homozygosity mapping approaches.

In addition to the basic population structure, the Amish have a strong interest in genealogy. Our need to draw accurate extended pedigrees for linkage calculations required development of methods to accurately extract genealogy information, check it for accuracy, and construct valid pedigrees. To address

this problem, the Amish genealogy database was developed, which is a unification of the three largest genealogies, which together comprise nearly 300,000 people.[15] The next step was to build a software tool to be able to query that database.[16] This software requires input of a list of families who share the attribute of having children affected with the same genetic disease and outputs the simplest pedigrees that connect these families. These pedigrees are then processed further to determine which is the simplest pedigree with the most recent founder couple.

The gene for MKS was mapped with a whole genome scan with 385 markers.[17] This study defined a region of homozygosity on chromosome 20p, converted that to a physical map, sequenced BAC clones from gaps in the map, defined the gene content, and searched for mutations in the genes in that region. However, this is where the strength of the Old Order Amish genetic structure becomes a liability. The mutation screen identified two substitution mutations in a single mutant chromosome in a candidate gene with no known function. Because the Amish are a closed isolate, it is impossible to prove the sequence variants are pathologic. For this reason, it was necessary to identify a sporadic (non-Amish) case of MKS. When the authors of published case reports were contacted, there were three typical responses: the patient(s) was (were) deceased or they were lost to follow up, but in the rest, the diagnosis had been changed. The reporting physicians noted that as the patients got older, it was determined that they no longer held a diagnosis of MKS but instead had a different disorder. Eventually, a newborn girl with MKS was identified and found to be a compound heterozygote for a substitution on one allele and a 2-bp deletion on the other allele in the same gene that was identified in the Amish.[18] This confirmed that mutations in this novel gene cause the MKS.

Again, at this point, the standard approach would be either to move on to another disorder or to study the pathophysiology and biology of this protein. However, the intriguing fact, as noted above, was that many of the patients who were previously diagnosed with MKS had their diagnoses changed. Surprisingly, all of the diagnoses were changed to the Bardet-Biedl syndrome (BBS),[19] and this has been noted by others as well.[20] As these children got older, they developed mental retardation, pigmentary retinopathy or blindness, diabetes, and obesity. Before this work, BBS was thought to be a multiple malformation syndrome that was completely unrelated to MKS. The critical observation is that many of the manifestations that are specific to BBS are age-dependent, so although MKS was diagnosed in the nursery, the development of other features with time necessitated a change of the diagnosis of BBS. At that time, there were five loci in the genome known to harbor (unidentified) genes that cause BBS, but 10% of the BBS families did not link to any of those five loci. On this basis, we hypothesized that BBS and MKS were a phenotypic continuum and that a subset of BBS patients have mutations in the *MKKS* gene. To test this hypothesis, samples of BBS deoxyribonucleic acid not known to map to the five loci were sequenced and mutations were identified in patients.[21,22] These mutations are different from the Amish mutant allele. Because the BBS patients with mutations in *MKKS* all have at least one null mutation, we hypothesize that the Amish MKS allele is a hypomorphic allele to the allele that causes BBS in the non-Amish. At this point, it is necessary to return to the clinic and carefully review patients with BBS and MKS to determine if there is significant overlap in these phenotypes. Again, this is another iteration of the approach of taking basic science data back to the clinic, reformulating a hypothesis, and asking additional questions to advance our knowledge of our patients.

We began this project by studying a syndrome of extreme rarity, PHS. The causative gene led us to an allelic syndrome (GCPS) and a mechanism of disease (*GLI3* processing). By following these phenotypes, we entered into the study of another rare disorder, MKS. The causative gene for MKS showed that it is allelic to a form of BBS. Thus this iterative approach to human developmental biology is beginning to spin a web of understanding of genes, genetic pathways, and syndrome families. Although such an approach to biology cannot be considered high-throughput, it is productive and valuable. Key insights into normal and abnormal development have been made, and our clinical knowledge of these syndromes has been significantly expanded. This iterative approach to human development is not inexpensive, fast, nor glamorous work, yet it is productive, rewarding, and has yielded significant benefits for the subjects.

References

1. Gilbert SF. The embryological origins of the gene theory. *J Hist Biol* 1978;11:307–351.
2. Warkany J. Congenital malformations: notes and comments. Chicago: Year Book Medical Publishers, 1971.
3. Smith DM. Recognizable patterns of human malformation. Philadelphia: Saunders, 1970.
4. Biesecker LG, Graham JM Jr. Syndrome of the month: Pallister-Hall syndrome. *J Med Genet* 1996;33:585–589.
5. Hall JG, Pallister PD, Clarren SK, Beckwith JB, Wigglesworth FW, Fraser FC, Benke PJ, Reed SD. Congenital hypothalamic hamartoblastoma, hypopituitarism, imperforate anus, and postaxial polydactyly: a new syndrome? Part I: clinical, causal, and pathogenetic considerations. *Am J Med Genet* 1980;7:47–74.
6. Winter RM, Schroer RJ, Meyer LC. Hands and feet. In: Stevenson RE, Hall JG, Goodman RM, editors. Human malformations and related anomalies. New York: Oxford University Press, 1993:805–843.
7. Kang S, Graham JM Jr, Olney AH, Biesecker LG. *GLI3* frameshift mutations cause autosomal dominant Pallister-Hall syndrome. *Nat Genet* 1997;15:266–268.
8. Vortkamp A, Gessler M, Grzeschik K-H. *GLI3* zinc finger gene interrupted by translocations in Greig syndrome families. *Nature* 1991;352:539–540.
9. Kinzler KW, Ruppert JM, Bigner SH, Vogelstein B. The *GLI* gene is a member of the Kruppel family of zinc finger proteins. *Nature* 1988;332:371–374.
10. Aza-Blanc P, Ramírez-Weber F-A, Laget M-P, Schwartz C, Kornberg TB. Proteolysis that is inhibited by hedgehog targets Cubitus interruptus protein to the nucleus and converts it to a repressor. *Cell* 1997;89:1043–1053.
11. Biesecker LG. Strike three for *GLI3*. *Nat Genet* 1997;17:259–260.
12. Wang B, Fallon J, Beachy P. Hedgehog-regulated processing of *Gli3* produces an anterior/posterior repressor gradient in the developing vertebrate limb. *Cell* 2000;100:423–434.
13. Lurie IW, Wulfsberg EA. The McKusick-Kaufman syndrome: phenotypic variation observed in familial cases as a clue for the evaluation of sporadic cases. *Genet Couns* 1994;5:275–281.
14. Kaufman RL, Hartmann AF, McAlister WH. Family studies in congenital heart disease, II: a syndrome of hydrometrocolpos, postaxial polydactyly and congenital heart disease. *Birth Defects Orig Artic Ser* 1972;8:85–87.
15. Agarwala R, Schäffer AA, Tomlin JF. Towards a complete North American Anabaptist Genealogy II: analysis of inbreeding. *Hum Biol* 2001;73:533–545.
16. Agarwala R, Biesecker LG, Hopkins KA, Francomano CA, Schäffer AA. Software for constructing and verifying pedigrees within large genealogies and an application to the Old Order Amish of Lancaster County. *Genome Res* 1998;8:211–221.

17. Stone D, Agarwala R, Schäffer AA, Weber JL, Vaske D, Oda T, Chandrasekharappa SC, Francomano CA, Biesecker LG. Genetic and physical mapping of the McKusick-Kaufman syndrome. *Hum Mol Genet* 1998;7:475–481.

18. Stone D, Slavotinek A, Bouffard G, Banerjee-Basu S, Baxevanis A, Barr M, Biesecker L. Mutations of a gene encoding a putative chaperonin causes McKusick-Kaufman syndrome. *Nat Genet* 2000;25:79–82.

19. Beales P, Elioglu N, Woolf A, Parker D, Flinter F. New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey. *J Med Genet* 1999; 36:437–446.

20. David A, Bitoun P, Lacombe D, Lambert JC, Nivelon A, Vigneron J, Verloes A. Hydrometrocolpos and polydactyly: a common neonatal presentation of Bardet-Biedl and McKusick-Kaufman syndromes. *J Med Genet* 1999;36:599–603.

21. Katsanis N, Beales PL, Woods MO, Lewis RA, Green JS, Parfrey PS, Ansley SJ, Davidson WS, Lupski JR. Mutations in *MKKS* cause obesity, retinal dystrophy and renal malformations associated with Bardet-Biedl syndrome. *Nat Genet* 2000;26:67–70.

22. Slavotinek A, Stone E, Mykytyn K, Heckenlively J, Green J, Heon E, Musarella M, Parfrey P, Sheffield V, Biesecker L. Mutations in *MKKS* cause Bardet-Biedl syndrome. *Nat Genet* 2000;26:15–16.