Genomic contributions to Mendelian disease

Aravinda Chakravarti

Johns Hopkins University School of Medicine, McKusick–Nathans Institute of Genetic Medicine, Baltimore, Maryland 21205, USA

Genomic research has two quite distinct faces. On the one hand, it produces large, curated, reference data sets through numerous networks of investigators for community use—although this aspect has great and widespread utility, it does not inspire per se. On the other hand, it allows an unbiased genome-wide view that is exciting precisely because it habitually uncovers biology that we were hopelessly ignorant about. Consequently, I am sanguine that the search for Mendelian disease genes by exomic and genomic sequencing will produce more than a long and comprehensive list of genes and associated disease mutations. Importantly, we are likely to hear new and surprising biological stories.

Human geneticists have long devoted their energies to understanding, diagnosing, and treating disorders that display a clear and Mendelian (i.e., single-gene) pattern of inheritance. Nevertheless, as Victor McKusick showed through painstaking cataloging, this list of genetic disorders is neither small nor based on extensive genetic evidence (McKusick 1998). Mendelian inheritance of rare traits and diseases has defined patterns of segregation with well-defined quantitative risks of *recurrence*; but the vast majority of McKusick's entries are based on astute clinical observations of a handful of patients, not extensive quantitative analysis. In other words, in McKusick's catalog, the many rare disorders and syndromes are good hypotheses, not proven examples, of "Mendelian Inheritance in Man."

This is precisely the situation where a genomic approach is desirable.

Since 2009, technological advancements in sequencing and the ability to select desired segments of the genome have made rapid sequencing of the entire human exome feasible for individual laboratories (Ng et al. 2009). These advances have spurred the discovery of mutations and genes in more than 40 Mendelian disorders using exome and genome sequencing of a small number of cases. Today, any investigator or clinician who has a few well-characterized patients with a rare disorder (sometimes, even a single family) has a very real chance of identifying the genetic mutation underlying that disease. This is a particular boon for the numerous clinical entities where only a handful of patients are available worldwide, too scant a number for any formal mapping analysis. Knowing the mutation(s) in an implicated gene is very useful for the annotation of the human genome sequence, for a deeper exploration of the biology of that gene, to understand how its function is compromised in disease (pathophysiology), and for thinking how to mitigate the biochemical dysfunction in disease. There is no doubt that we will see a rapid rise in our understanding of the genetic basis of Mendelian disorders, and the human genome, over the next few years, and *Genome Research* is expecting to be a natural home for publishing these advancements. As a preview, in this issue, Erlich and colleagues use whole exome sequencing and disease-network analysis to associate a mutation in a novel gene, *KIF1A*, with hereditary spastic paraparesis cases from a single inbred family.

I suspect that the universe of genes and their "Mendelian" mutations revealed will be more exciting than a mere catalog of defects and their functional meaning. In the short term at least, from these studies, I foresee three types of challenges that we will meet in an unbiased manner: namely, (1) What is the total burden of Mendelian disease? (2) What are the inheritance patterns of rare diseases? (3) What is the spectrum of mutations that lead to Mendelian disease? Fundamentally, these answers will teach us much about the nature, frequency, and phenotypic effects of deleterious mutations in our genomes. In more ways than one, these studies will be one "functional" complement to the variation catalogs from the 1000 Genomes Project (The 1000 Genomes Project Consortium 2010).

It is commonly assumed that the total incidence of Mendelian disease is <5%, but this number is low if all of the human genome's ~20,000 genes carry at least one typical dominant or recessive deleterious allele with disease incidence between ~1:10,000–50,000 live births. If one includes dosage mutations, which tend to be more frequent, then the estimate is yet more discrepant. These estimates are guesses from our crude understanding of the human mutation rate and the distribution of deleterious alleles across human genes, and there are many uncertainties here. First, it is likely that many genes simply do not accommodate the survival of mutations to birth and are embryonic lethals. On the other hand, copy number variants across the genome are compatible with survival since only ~5% of our genome has not been found to be dosage variant in controls and individuals with intellectual disability or autism. This suggests that we can dispense or duplicate at least one copy of the lion's share of our genes and survive without a Mendelian disorder. Second, we simply do not know the human mutation rate accurately, nor how it varies across genes—especially for mutations that lead to a recognizable phenotype. It is quite likely that many genes mutate at very low rates and Mendelian disorders will map to only a subset of genes. Identifying which genes contribute to diseases in live births and beyond, and why, is an important piece of currently unknown biology. Such data, in turn, would provide a deeper understanding of the human mutation rate and how it is affected by genomic and chromatin features. Third, it is suspected that the human germ-line mutation rate increases with paternal age: Thus, there is insufficient chance for the majority of fathers to produce and transmit mutations. Given the steady increase in paternal age at conception in the past few centuries, has this affected the frequency of Mendelian disease?

Recent publications reporting the successful identification of disease-associated, presumably causal, mutations in Mendelian disorders should not lull us into thinking that this will always be the case. Although medical genetics has had many examples to suggest that several thousand Mendelizing clinical entities exist, careful quantitative analysis of the inheritance patterns does not exist for the vast majority of these rare disorders. Indeed, the typical rare disease patient occurs in a family with no other family history, and we preferentially ascertain multiplex families. In other words, we routinely undersample the simplex families unless we conduct a census. Consequently, there is no reason to believe that all rare disorders in McKusick's catalog will be the product of

a single gene mutation. Nature is seldom discontinuous. Although human geneticists usually speak about Mendelian or multifactorial entities, I suspect that exome or genome sequencing will reveal not only single-gene mutations, but also numerous cases of digenic, trigenic, and more complex inheritance. The intellectual challenges to the data interpretation, beyond bioinformatic analysis, will not be trivial since the results will question which mutations are causal, which are primary, and which are modifiers, although they will go a long way to explain phenotypic associations, comorbidities, variability in expressivity, and reduced penetrance. These disease sequencing projects might be the first unbiased survey of the magnitude of "Mendelian Inheritance in Man." These studies are very likely to also reveal new types of mutations, distribution in the genome by functional site, their genetic effects, and inheritance patterns in rare disorders. Such insights will be practically useful since they will provide an objective and concrete basis for accurate genetic counseling and for understanding why a phenotype maps to multiple loci. It will also make DNA-based diagnostics more challenging.

A major impediment in all of these studies is accurately recognizing the causal "mutation." Although this step is tacitly assumed to be simple, reality indicates otherwise since the vast majority of disease-associated mutations are missense (Stenson et al. 2009) and not readily recognized as contributing to the disease in question. As mentioned above, family information may be too limited in many cases to do any significant genetic analysis on these discoveries. A major challenge going forward is recognizing deleterious mutations based on the sequence itself. These predictions will be necessary in two distinct types of studies. In the first type of study, we wish to identify the disease gene and so need to recognize only a few rare severe mutations in a collection of patients. In the second type of study, we wish to understand the genotype–phenotype correlation and so require the identification of *all* mutations in a collection of patients with the same phenotype and who are suspected to have mutations in the same gene. Both of these will require new and innovative genomic predictive tools, an area that is likely to be catalyzed by the increased availability of exome sequence data from patients with Mendelian disease. We have long assumed that the majority of Mendelian

mutations are coding and only a small minority noncoding (or, reside in the proximal promoter region). These assumptions may be correct, but are biased since only mutations in coding sequences have been aggressively sought (in most cases). A broader genomic screen may yield surprising findings, revealing a more complete description of the sites in which mutations lead to a rare disease. Finally, almost all of human genetics assumes diploidy in both normalcy and disease. The importance of copy number and dosage abnormalities, however, has recently come to the fore. The true contribution of dosage alterations in Mendelian disease is unknown, but likely to yield new surprises and insights and a description of the full spectrum of genetic variation in the context of disease and transmission.

The overriding features of genomics research are comprehensiveness and unbiasedness: Some might term this a "holistic" approach. We are rapidly moving to a time when we will have the exome and genome sequences of tens of thousands of individuals both with Mendelian phenotypes and those without (controls). Beyond contributing to disease gene discovery and impacting some families immediately, the possibilities of recognizing the fundamental genomic truths revealed are truly tantalizing.

References

The 1000 Genomes Project Consortium, Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. 2010. A map of human genome variation from population-scale sequencing. *Nature* **467:** 1061–1073.

Erlich Y, Edvardson S, Hodges E, Zenvirt S, Thekkat P, Shaag A, Dor T. 2011. Exome sequencing and disease-network analysis of a single family implicate a mutation in KIF1A in hereditary spastic paraparesis. *Genome Res* (this issue). doi: 10.1101/gr.117143.110.

McKusick VA. 1998. *Mendelian inheritance in man. A catalog of human genes and genetic disorders*, 12th ed. Johns Hopkins University Press, Baltimore. http://www.ncbi.nlm.nih.gov/omim.

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, et al. 2009. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461:** 272–276.

Stenson PD, Mort M, Ball EV, Howells K, Phillips AD, Thomas NS, Cooper DN. 2009. The Human Gene Mutation Database: 2008 update. *Genome Med* **1:** 13.