Revolution Postponed

The Human Genome Project has failed so far to produce the medical miracles that scientists promised. Biologists are now divided over what, if anything, went wrong—and what needs to happen next

By Stephen S. Hall

DECADE AGO BIOLOGISTS AND NONBIOLOGISTS alike gushed with optimism about the medical promise of the $3-billion Human Genome Project. In announcing the first rough draft of the human “book of life” at a White House ceremony in the summer of 2000, President Bill Clinton predicted that the genome project would “revolutionize the diagnosis, prevention and treatment of most, if not all, human diseases.”

A year earlier Francis S. Collins, then head of the National Human Genome Research Institute and perhaps the project’s most tireless enthusiast, painted a grand vision of the “personalized medicine” likely to emerge from the project by the year 2010: genetic tests indicating a person’s risk for heart disease, cancer and other common maladies would be available, soon to be followed by preventives and therapies tailored to the individual.

Even before the first full sequence of DNA “letters” in human chromosomes was deciphered, a well-funded genomics juggernaut—armed with powerful sequencing and mapping technologies, burgeoning databases and a logical game to “mine miracles,” as Collins put it, from the genome—set out to identify key genes underlying the great medical scourges of humankind.

Fast-forward to 2010, and the scientific community finds itself sobered and divided. The problem is not with the genome project itself, which has revolutionized the pace and scope of basic research, uncovered heretofore hidden purpose in what used to be called “junk DNA” and even detected traces of Neandertal DNA in our genomes. Cancer researcher Bert Vogelstein, echoing a widespread sentiment, says, “The Human Genome Project has radically changed the way we do science.”

IN BRIEF

In the year 2000 leaders of the Human Genome Project announced completion of the first rough draft of the human genome. They predicted that follow-up research could pave the way to personalized medicine within as few as 10 years.

So far the work has yielded few medical applications, although the insights have revolutionized biology research.

Some leading geneticists argue that a key strategy for seeking medical insights into complex common diseases—known as the “common variant” hypothesis—is fundamentally flawed. Others say the strategy is valid, but more time is needed to achieve the expected payoffs.

Next-generation methods for studying the genome should soon help resolve the controversy and advance research into the genetic roots of major diseases.
The problem is that research springing from the genome project has failed as yet to deliver on the medical promises that Collins and others made a decade ago. Tumor biologist Robert A. Weinberg of the Whitehead Institute for Biomedical Research in Cambridge, Mass., says the returns on cancer genomics “have been relatively modest—very modest compared to the resources invested.” Harold E. Varmus, former director of the National Institutes of Health, wrote recently in the New England Journal of Medicine that “only a handful of major changes … have entered routine medical practice”—most of them, he added, the result of “discoveries that preceded the unveiling of the human genome.” Says David B. Goldstein, director of the Center for Human Genome Variation at Duke University: “It’s fair to say that we’re not going to be personalizing the treatment of common diseases next year.”

Perhaps it was unreasonable to expect miracles in just 10 years (the predictions of genome project promoters notwithstanding). Behind today’s disappointment, however, lies a more disturbing question: Does the surprisingly modest medical impact of the research so far indicate that scientists have been pursuing the wrong strategy for finding the genetic causes of common diseases? This strategy, at root, involves searching for slight variations in the DNA text of genes that could collectively increase an individual’s risk of acquiring a common disorder. For years many scientists have pursued the hypothesis that certain common variants would be especially prevalent in people with particular illnesses and that finding those variants would lead to an understanding of how susceptibility to major, biologically complex diseases, such as type 2 diabetes and atherosclerosis, gets passed down from one generation to the next. Could the failure to find genetic variations with much effect on disease mean the “common variant” hypothesis is wrong?

This question has opened a fissure in the medical research community. On one side, leading genome scientists insist the common variant strategy is working. Recent research identifying genetic clues to disease has been “mind-blowing” over the past three years, says Eric S. Lander, director of the Broad Institute (an affiliate of the Whitehead Institute), and “we haven’t even scratched the surface of common variants yet.” He says the medical revolution will come as technologies improve—in time for our children if not for us. The revolution, in other words, is just running late.

On the other side, a growing chorus of biologists has begun to insist that the common variant strategy is flawed. In a hotly debated essay this past April in Cell, geneticists Mary-Claire King and Jon M. McClellan of the University of Washington argued that “the vast majority of [common] variants have no established biological relevance to disease or clinical utility for prognosis or treatment.” Geneticist Walter Bodmer, an elder statesman of British science, has flatly called the strategy of looking at common variants “scientifically wrong.”

As some genome scientists celebrate the progress made so far, others who look at the same results see mostly failure and are now asking, Where do we go from here? The pursuit of an answer may take medical research down completely new avenues for understanding human disease and how it is passed down through the generations.

**On one side, leading genome scientists insist the “common variant” strategy for finding clues to the causes of common disorders is working.**

**DISAPPOINTMENT**

The common variant hypothesis seemed like a reasonable bet when it was first advanced in the 1990s, proposing that many familiar human maladies might be explained by the inheritance of a relatively small number of common gene variants. Genes have traditionally been defined as stretches of DNA that encode proteins. The variants might be thought of as slightly different, mutated texts of the same gene, altering either the protein-coding part of the DNA or the nearby DNA that regulates the rate of gene “expression” (protein synthesis). Proteins carry out many tasks in cells, and deficiencies in their function or concentration can disrupt molecular pathways, or chains of interactions, important to health.

The belief that common variants would be helpful in understanding disease had a certain evolutionary logic. The rapid and recent population explosion of ancestral humans tens of thousands of years ago “locked” many variants in the human gene pool, Lander says. The bet was that these common variants (common usually meaning appearing in at least 5 percent of a given population) would be fairly easy to find and that a relatively small number of them (from several to perhaps dozens) would shape our susceptibility to hypertension, dementias and many other widespread disorders. The disease-related genetic variants and the proteins they encode, as well as the pathways in which they played crucial roles, could then become potential targets for drugs.

From the very beginning, however, the scheme was met with some dissent. In 1993 Kenneth M. Weiss, an evolutionary biologist at Pennsylvania State University, paraphrased Leo Tolstoy’s famous line about families, from his novel *Anna Karenina*, to make a point about the genetics of complex diseases: “All healthy families resemble each other; each unhealthy family is unhealthy in its own way.” The point, which Weiss and Columbia University statistical geneticist Joseph D. Terwilliger made repeatedly, was that common variants would probably have very small biological effects; if they did powerful harm, natural selection would have prevented them from becoming common in the population. Rather they argued that susceptibility to biologically complex diseases probably derives from inheritance of many rare disease-promoting variants that could number in the hundreds—perhaps thousands—in any given individual. In Tolstoy’s idiom, ill people are genetically unhappy in their own way. Coming from a self-described “lunatic fringe,” the argument didn’t win many converts.

The obvious way to see who was right would have been to sequence the full genomes of diseased and healthy individuals and, using powerful computers, identify DNA variations that turned up in patients with the given disease but not in control subjects. In contrast to standard genetic research of the past, which relied on having a biology-based suspicion that a particular gene played
The Logic behind Genome Studies

Much research into the genetic contributions to common diseases has started with the seemingly logical assumption that DNA variants occurring frequently in the human population would be at fault. Some argue, though, that this reasoning is faulty.

The Starting Point

The Human Genome Project identified the sequence of nucleotide pairs, or DNA building blocks, in the human genome, based on DNA from several volunteers. A single pair consists of a nucleotide (A, C, T or G) on one strand of the DNA double helix and its complement on the opposite strand (C always pairs with G; A with T). Related work revealed many single-nucleotide polymorphisms, or SNPs—chromosomal locations where a nucleotide pair in one person can differ from that in another person (below)—and it identified “common” SNPs, ones that vary in many people.

The Studies and Results

Investigators hoped that they could identify gene variants responsible for major diseases by comparing nucleotides at common SNPs throughout the genomes of people with and without a disease. SNP variants, or “alleles,” and nearby protein-coding genes tend to be inherited together, and so researchers expected that SNP alleles occurring much more frequently in people with a disease would point to common gene variants important to the illness. These genome-wide association (GWA) studies uncovered many SNP alleles related to specific diseases. So far, though, the variations found have typically accounted for only a small fraction of disease risk.
a role in a disorder, such “agnostic” comparisons would presumably shine light on any and all DNA culprits, including those not previously suspected of being important. But 10 years ago it was technologically impossible to undertake such an approach, and the common variant hypothesis—if correct—offered a shortcut to discovering genes that contributed to common diseases.

Genome scientists guided by the common variant hypothesis began planning large-scale studies, known as genome-wide association studies (often called GWAS, or “gee-waz”), that relied on landmarks in DNA known as single-nucleotide polymorphisms, or SNPs (pronounced “snips”), to uncover common gene variants important in disease. SNPs, which occur throughout chromosomes, are sites in DNA (not necessarily within genes) where a single code letter in one person’s DNA can differ from the letter at that same spot in another person’s DNA. The plan was to examine large numbers of SNPs that often vary between people to see which versions occurred frequently in people with particular disorders. The SNPs statistically linked to disease would then lead researchers to nearby gene variants (inherited along with the landmarks) that could account for the association.

The plan, however, required the assembly of an atlas, as it were, of common human SNPs. Over the past decade or so biologists have gathered increasingly large numbers of SNPs to guide their search for the genetic roots of diseases, beginning in 1998 with the SNP Consortium (which assembled maps of these landmarks on each human chromosome) and progressing to the HapMap (which catalogued a larger swath of SNPs called a haplotype). In the past five years genome-wide association studies have looked at hundreds of thousands of common SNPs in the genomes of tens of thousands of individual patients and controls in the search for SNPs linked to common diseases.

This is where the rift in the biology community occurs. Lander and others hail the recent discovery of common, disease-associated SNPs as a portal to medically important pathways. To be sure, a flood of recent papers from huge genome consortia have uncovered hundreds of common SNPs related to such diseases as schizophrenia, type 2 diabetes, Alzheimer’s and hypertension. Francis Collins, in a recent appearance on PBS’s The Charlie Rose Show, claimed scientists have “figured out” how almost 1,000 of those common gene variants “play a role in the risk of disease, and we have used that information already to change our entire view of how to develop new therapeutics for diabetes, for cancer, for heart disease.” Others point out, however, that the data have not been very useful so far in predicting disease risk. In type 2 diabetes, for example, association studies analyzing 2.2 million SNPs in more than 10,000 people have identified 18 SNPs associated with the disease, yet these sites in total explain only 6 percent of the heritability of the disease—and almost none of the causal biology, according to Duke’s Goldstein.

In 2008 Goldstein told the New York Times: “It’s an astounding thing that we have cracked open the human genome and can look at the entire complement of common genetic variants,
and what do we find? Almost nothing. That is absolutely beyond belief.” This past summer Goldstein spoke of the common variant/common disease hypothesis as a thing of the past: “We have entered and left that field, which explained less than a lot of people thought it would.”

David Botstein of Princeton University offers much the same verdict on the strategy of creating a haplotype map: “It had to have been done. If it had not been tried, no one would have known that it didn’t work.” The $138-million HapMap, he says, was a “magnificent failure.”

Walter Bodmer, who was among the first to propose the genome project in the 1980s and is a pioneer of the association studies that have dominated recent genomics, asserts that searching for common gene variants is a biological dead end. “It is almost impossible to find what the biological effects of these variant genes are, and that’s absolutely key,” he says. “The vast majority of [common] variants have shed no light on the biology of diseases.”

NEW WAYS FORWARD

THE CURRENT ARGUMENT over the common variant hypothesis is not just an arcane scientific debate. It suggests at least one alternative way forward for solving what many are calling the “missing heritability” problem, at least for the short term. Bodmer, for instance, has been urging researchers to train their sights on rare genetic variants. The boundary where common ends and rare begins is not exact—“rare,” by Bodmer’s definition, refers to a particular genetic mutation that occurs in a range from 0.1 to 1 or 2 percent of the population (a frequency well below the resolution of most current genome-wide association studies). But the main idea of the hypothesis is that gene variants that have large disease-related effects tend to be rare, whereas those that are common almost always exert negligible or neutral effects.

This same argument surfaced in the controversial Cell essay, by King and McClellan, that this past spring stirred up so much animosity in the genome community—an essay Lander dismisses as an “opinion piece.” King (who has found hundreds of rare variations in the BRCA1 and BRCA2 genes that cause familial breast cancer) and McClellan (who has similarly found many rare variants contributing to the genetics of schizophrenia) are suggesting a “new paradigm” for understanding complex diseases. They suggest that most of these diseases are “heterogeneous” (meaning that many different mutations in many different genes can produce the same disease), that most high-impact mutations are rare, and that many rare genetic variants are relatively recent additions to the gene pool. Rare variants identified in patients could thus lead researchers to specific molecular pathways related to a particular disease, and the biological understanding of those pathways could suggest new therapeutic interventions.

Bodmer, the Cell authors and others point to the work of Helen H. Hobbs and Jonathan C. Cohen as a model for using biology as a guide to uncovering medically significant information buried in the genome. The Hobbs-Cohen approach focuses on extreme cases of disease, assuming that rare gene variants that strongly perturb biology account for the extremity and will stand out starkly. They also pick and choose which genes to examine in those people, based on a knowledge of biology. And, they sequence specific candidate genes, looking for subtle but functionally dramatic variations between people, rather than using SNP associations, which can indicate the genetic neighborhood of a disease-related gene but often not the gene itself.

In 2000, when the big news in the genome field was the race between J. Craig Venter, founder of the biotech company Celera Genomics, and NIH scientists to produce the first rough draft of the human genome sequence, Hobbs and Cohen quietly embarked on a project known as the Dallas Heart Study to help uncover the causes of heart disease. Cohen, a South African physiologist, had studied cholesterol metabolism (its synthesis and breakdown) for many years. Hobbs, trained as an M.D. and now a Howard Hughes Medical Institute investigator at the University of Texas Southwestern Medical Center at Dallas, had done research in the laboratory of Michael S. Brown and Joseph L. Goldstein, who shared a Nobel Prize in 1985 for their work on cholesterol metabolism, which in turn laid the groundwork for the development of the popular class of cholesterol-lowering drugs known as statins.

Hobbs and Cohen set their scientific compass according to a biological “intuition” that represented a strategy completely different from almost everyone else working in genomics. They recruited some 3,500 residents of Dallas County (half of them African-Americans) and then gave them intensive medical workups. They did not just focus on the genome (although they dutifully collected everyone’s DNA) but gathered very precise measures for many factors that can contribute to coronary artery disease: blood chemistry (including cholesterol numbers), metabolism, body fat, cardiac function, arterial thickening (assessed through high-tech imaging) and environmental influences. Over the course of two years they compiled a massive, highly detailed database of individual physical traits—what geneticists call “phenotypes.”

After that, they concentrated their genomic attention on people with particularly dramatic phenotypes—specifically with extremely high or low numbers for high-density lipoproteins (HDL, often called the “good” cholesterol) or for low-density lipoproteins (LDL, the “bad” cholesterol). And there was nothingagnostic about their search through the genome. As Cohen puts it, “We came at this from a more functional standpoint.”

As they reported in Science in 2004, they first looked at patients with very low HDL concentrations in the blood, a condition that increases risk for heart disease. They knew of three genes involved in rare disorders of cholesterol metabolism, and so they compared DNA sequences from those genes in the very low HDL patients and people with high HDL levels, finding several rare variants linked to the extremely depressed HDL levels. They also reported that mutations in the affected genes “contribute significantly” to low HDL values in the general population.

In 2005 Hobbs and Cohen turned their attention to people in the Dallas Heart Study who were found to have unusually low LDL levels. The researchers hit a genomic jackpot when they analyzed the DNA sequences of a gene called PCSK9, known to

© 2010 Scientific American
A number of scientists searching for the heritable influences on common diseases are pushing for research strategies that do not rely on massive statistical analyses of common SNPs, which, some argue, are not likely to be extremely informative about disease risk.

**Search for Rare Gene Variants**
Some researchers say that rare mutations in genes probably contribute more to disease than do the gene variants linked to common SNPs. Even if each person is ill because of a unique rare mutation, it is possible that many of those mutations affect genes whose encoded proteins work together (operate in the same “pathway”) to accomplish some important job in the body. Identifying the affected genes should suggest ways to compensate for a pathway’s disruption. One way to find important rare mutations is to fully sequence and compare all the protein-coding bits (exons) in the genes of diseased and healthy people (left). This approach, known as exome sequencing, is already under way in a number of labs.

**Look beyond Genes**
SNP studies and the search for rare variants focus on finding variations in the DNA sequences of protein-coding genes and thus in proteins involved in disease (top right). But other processes not involving changes in those sequences can also impair the manufacture of needed proteins and thereby predispose people to illnesses. Two of those are depicted here schematically (bottom right). Investigators are devising ways to examine those processes as well.

Protein synthesis (blue sequence) occurs when an enzyme called a polymerase transcribes a gene into messenger RNA, which is read out, or translated, into a protein by ribosomes. If a mutant version of the gene is transcribed into RNA, the resulting protein will also be abnormal.
be involved in cholesterol metabolism. Two mutations that silenced the gene correlated with the low LDL levels. In a follow-up study that analyzed data collected from populations in Mississippi, North Carolina, Minnesota and Maryland over a 15-year period, Hobbs and Cohen determined that African-Americans with one or another silencing mutation in PCSK9 have a 28 percent reduction in LDL levels and an astounding 88 percent reduction in the risk of coronary heart disease. In whites, a mutation in the same gene reduced LDL by 15 percent and reduced the risk of heart disease by 47 percent. Hardly any of the hundreds of genome-wide association studies have identified genes with such a large effect on disease risk.

Drug companies are already testing molecules that shut off the PCSK9 gene, or perturb the molecular pathway the gene affects, as a way to lower LDL and reduce the risk of heart disease in the general population. PCSK9, Hobbs says, is a “top-10 target” of virtually every pharmaceutical company now.

Acknowledging the small effect of genes identified by the common variant approach and heartened by the success of the Hobbs-Cohen work, David Goldstein and Elizabeth T. Cirulli, also at Duke, recently proposed expanding the search for medically important rare variants. One idea, for example, is to sequence and compare whole “exomes” in carefully selected people. The exome is a collection of actual protein-coding parts of genes (exons) in chromosomes, along with nearby regions that regulate gene activity; it does not include the stretches of DNA that lie between exons or genes. Cirulli and Goldstein also suggest looking for these rare variants within families affected by a common disease or in people who share an extreme trait, where significant DNA differences can more easily be identified. This work is already under way in many labs. “We are sequencing exomes in the lab every day,” University of Washington’s King says. Exome sequencing is a stop-gap strategy, though, until inexpensive, reliable whole-genome sequencing becomes available, probably in three to five years.

BEWARE THE RABBIT HOLE

A FEW BRAVE VOICES ARE SUGGESTING that the rabbit’s hole of human biology may go still deeper than a focus on DNA sequences and proteins can reveal. Traditional genetics, they say, may not capture the molecular complexity of genes and their role in disease. The vast areas of DNA that do not code for proteins, once dismissed as “junk,” are now known to conceal important regulatory regions. Some DNA stretches produce small bits of RNA that can interfere with gene expression, for instance. And chemical “tags” on DNA that do not change its sequence—that are thus “epigenetic”—can also influence gene expression and can be modified by environmental factors over the course of a lifetime. This environmentally modified DNA may even be passed on to offspring.

Put simply, the very definition of a gene—not to mention a medically significant gene—is now vexed with multiple layers of complexity. What was once assumed to be a straightforward, one-way, point-to-point relation between genes and traits has now become the “genotype-phenotype problem,” where knowing the protein-coding sequence of DNA tells only part of how a trait comes to be.

In animal experiments, Joseph H. Nadeau, director of scientific development at the Institute for Systems Biology in Seattle, has tracked more than 100 biochemical, physiological and behavioral traits that are affected by epigenetic changes and has seen some of these changes passed down through four generations. “It’s totally Lamarckian!” he laughs, referring to the 18th-century biologist Jean-Baptiste Lamarck’s idea that acquired traits could be inherited.

As if that level of complexity were not enough, Nadeau has experimental evidence that the function of one particular gene sometimes depends on the specific constellation of genetic variants surrounding it—an ensemble effect that introduces a contextual, postmodern wrinkle to genetic explanations of disease. It suggests, Nadeau says, that some common illnesses may ultimately be traceable to a very large number of genes in a network or pathway whose effects may each vary depending on the gene variants a person has; the presence of one gene variant, say, can exacerbate or counteract the effect of another disease-related gene in the group. “My guess is that this unconventional kind of inheritance is going to be more common than we would have expected,” Nadeau says.

Exactly how powerful the aspects Nadeau addresses will turn out to be in disease remains unclear. In the meantime, a new generation of fast, cheap sequencing technologies will soon allow biologists to compare entire genomes, by which time the common versus rare variant debate may subside into ancient history. Far from casting a pall over the field, the current puzzle allows experts to compare entire genomes, by which time the common versus rare variant debate may subside into ancient history. Far from casting a pall over the field, the current puzzle allows biologists to compare entire genomes, by which time the common versus rare variant debate may subside into ancient history. Far from casting a pall over the field, the current puzzle allows biologists to compare entire genomes, by which time the common versus rare variant debate may subside into ancient history. Far from casting a pall over the field, the current puzzle allows biologists to compare entire genomes, by which time the common versus rare variant debate may subside into ancient history. Far from casting a pall over the field, the current puzzle allows biologists to compare entire genomes, by which time the common versus rare variant debate may subside into ancient history. Far from casting a pall over the field, the current puzzle allows biologists to compare entire genomes, by which time the common versus rare variant debate may subside into ancient history. Far from casting a pall over the field, the current puzzle allows biologists to compare entire genomes, by which time the common versus rare variant debate may subside into ancient history. Far from casting a pall over the field, the current puzzle allows biologists to compare entire genomes, by which time the common versus rare variant debate may subside into ancient history. Far from casting a pall over the field, the current puzzle allows biologists to compare entire genomes, by which time the common versus rare variant debate may subside into ancient history. Far from casting a pall over the field, the current puzzle allows biologists to compare entire genomes, by which time the common versus rare variant debate may subside into ancient history. Far from casting a pall over the field, the current puzzle allows biologists to compare entire genomes, by which time the common versus rare variant debate may subside into ancient history. Far from casting a pall over the field, the current puzzle allows biologists to compare entire genomes, by which time the common versus rare variant debate may subside into ancient history. Far from casting a pall over the field, the current puzzle allows biologists to compare entire genomes, by which time the common versus rare variant debate may subside into ancient history. Far from casting a pall over the field, the current puzzle allows biologists to compare entire genomes, by which time the common versus rare variant debate may subside into ancient history.


Put simply, the very definition of a gene—not to mention a medically significant gene—is now vexed with multiple layers of complexity.

Joseph H. Nadeau has documented inheritance of acquired characteristics in mice.