For more than a decade, the race has been on to identify genes that predispose patients to common diseases. The Holy Grail arising from such discoveries would be a “theranostic,” a term for the use of diagnostic testing to diagnose the disease, to choose the correct treatment regime, and to monitor the patient response to therapy. This is often referred to as personalized medicine (see Figure 1). The “standard” approach to gene discovery has been through studies of families in which a particular disease occurs frequently using a principle known as linkage.

Such studies were successful in identifying disease-susceptibility genes in rare, familial diseases (i.e., cystic fibrosis and sickle-cell anemia) — diseases caused by mutations in a single gene. They were mostly unsuccessful in more common heritable diseases where many genes interact to produce disease susceptibility. Some success was achieved through family-linkage studies using the unique genetic heritage of Iceland to enhance the detectability of disease genes. More often than not, family-linkage-based discoveries failed to corroborate in confirmatory studies.

Another method, the candidate-gene approach, involves a search in patient DNA for mutations within a gene that are considered to be functionally involved in a putative disease-associated biochemical pathway. While candidate-gene studies gain from their targeted approach, they are inherently limited to the genes being investigated. Several disease-susceptibility genes have been identified this way, but none of these have proved to be a necessary or sufficient factor in a common disease. A comprehensive picture of the genes involved in a disease and how they interact appears to be beyond the scope of linkage and candidate-gene studies.

With the completion of the first draft of the human genome sequence in 2001, it was clear that more tools and technology were needed to realize the promise of genetics in drug discovery. Since then, knowledge and technology have evolved rapidly, and the tools required for more powerful approaches to disease-gene discovery are now becoming available. While the human genome still retains many secrets, it is sufficiently well documented to allow precise identification of the locations of disease genes. Several million mutations have now been documented on a large scale. The human genome comprises some 3 billion pairs of DNA, each made up of chemicals identified with the letters A, T, C, or G — virtually all of which are in the same sequence in every human. But scientists discovered that there are an estimated 10 million “spelling” differences, involving a single change where one DNA letter is replaced by another. These “spellings” are the single nucleotide polymorphisms or SNPs, some of which are thought to be related to disease.

**Value of the HapMap Project**

In 2002, researchers began a $100-million International HapMap Project to identify blocks of DNA that contain common variations in the human genetic structure; its goal was to find a quicker way to identify genes that cause disease. Nine research groups in five countries analyzed genetic patterns
in blood samples taken from people in Nigeria, Japan and China, and from people of northern and western European ancestry in the United States. HapMap determined that SNPs organize into DNA neighborhoods called haplotype blocks. Once the haplotypes were mapped, researchers had found a “shortcut” to identifying inherited gene sequences linked to disorders such as diabetes, heart disease, and cancer. Researchers can go directly to those blocks comprising about 10,000 or more base pairs and search for disease genes instead of having to search through all 3 billion DNA base pairs.

Today, SNP genotyping platforms with the ability to analyze large numbers of SNPs for each patient and each control have arrived. All have very high throughput at costs for consumables that are 100 times lower than a few years ago. These platforms — together with SNP maps generated by HapMap and other sources — make it possible to conduct studies comparing DNA from patients with that from controls, typically generating hundreds of millions of genotypes. Coincidentally, the computing power required to analyze the terabytes of genotype data involved in order to identify the locations of disease genes on the human genome has also become sufficiently fast and affordable. Thus, the gold standard for disease-gene discovery studies — whole genome association studies (WGAS) — is now becoming achievable.

With this second generation of population genetics, there is a swelling chorus of key opinion leaders moving away from their first love — family-linkage studies — to their trophy bride, WGAS. WGAS have much greater disease-gene detection power than family-linkage studies; they locate the disease gene approximately hundredfold more precisely; and they have greater power to detect the common but low-risk genes that are typical of common diseases.

The first WGAS
Towards the end of 2003, the first genotyping platform with sufficient throughput to process a WGAS became operational. Some scientists reasoned that although not all of the components required for a successful WGAS were available at that time, use of a founder population, similar to that used to enhance disease-gene detection using family-linkage studies, would compensate for the deficiencies. Both Iceland’s and Quebec’s “founder populations” descended from a small group of “founders” 10 to 20 generations ago, with minimal gene dilution from other populations. Founder-population members tend to share larger blocks of DNA inherited from common ancestors than do general populations, and they are thought to carry

Continues on page 25
fewer mutations per gene. To detect disease genes, this population should require fewer SNPs than a general population, and the limited number of mutations should make such genes more readily detectable.

In early 2004, a whole WGAS was initiated on Crohn’s disease in the Quebec founder population, which is more than 80% heritable. This was one of the few common diseases where disease genes had been unequivocally identified, providing positive controls. Approximately 140,000 SNPs were eventually used for data analysis in the Crohns’ WGAS, with dramatic results: Sixteen disease genes were detected, including the two known genes. About half of these have been corroborated in a German population — a proportion higher than expected, considering the mutations within a disease gene and that the incidence of any one mutation can differ considerably between populations.

**Current WGAS Initiatives**

Technology and genetic knowledge continue to evolve. In two to four years’ time, many disease genes will have been discovered using founder populations or in a second wave of non-founder population studies. In common diseases, comprehensive gene maps will have been created, leading to revolutions in drug discovery and development, disease diagnosis, patient therapy, and drug markets. Opportunities for drug intervention for subgroups of patients who carry particular risk factors could lead to selection of the best possible drug targets for each. New drugs that address these targets would be “nichebusters” — superior drugs that treat the root cause of the disease and not just the symptoms — as opposed to the one-drug-treats-all blockbusters of today.

With current technology, within a decade we may well see theranostic products that show great therapeutic benefit over existing drugs. As the knowledge of disease processes becomes more comprehensive, treatment regimes can be targeted to individual genetics. Diagnostics can be developed to determine lifetime disease-susceptibility risk, enabling proactive therapy and/or lifestyle changes that reduce or eliminate the impact of the disease, often before symptoms manifest. Gene-based drugs and diagnostics can be expected to transform the way medicine is practiced. Susceptibility to disease will be ascertained early in life, disease onset will be detected early and treated before significant damage is done, or therapeutics may be given to prevent disease in susceptible individuals. □

John W. Hooper, PhD, is president and CEO of Genizon Biosciences in St. Laurent, Quebec, Canada. This article was adapted from Dr. Hooper’s article “Genetics, pipelines, and personalized medicine.”