

Genetics, pipelines and personalised medicine

Genetic linkage studies have long been the standard for researching the genetics associated with heritable diseases. Recently, however, scientists have begun to realise the benefits of whole genome association studies, which enable them to view the entire human genome in greater detail. Enhancements in genotyping and information technology make whole genome association studies feasible: the greater resolution of information provided makes the approach a desirable alternative to the traditional method. This article provides an overview of whole genome association and genetic linkage studies, contrasting the methods and providing examples of scenarios where one approach may be preferable to the other.

By Dr John W. Hooper

The root cause of many common diseases is the interaction among several genes and between these genes and the patient's environment and lifestyle. For more than a decade, billions of dollars have been spent to identify genes that predispose to common diseases ('disease genes'). A comprehensive map of interacting disease genes (a 'GeneMap') would lead to an unprecedented understanding of the genesis of a common disease. This would lead to targets for new and better drugs to fill depleted pharma pipelines with drug candidates that directly address the root causes of the disease, or to diagnostics that are predictive of disease or drug response. The Holy Grail arising from such discoveries would be a Theranostic, the combination of a drug tailored to patients with a particular genetic profile, and a molecular diagnostic for selecting patients with that profile. This is often referred to as Personalised Medicine (see Figure 1). The drug would have a high probability of producing an optimal response and a low probability of causing an adverse reaction.

Personalised Medicine differs considerably from the current 'one drug fits all' approach to therapy by dividing the patient population into several genetic groups and providing a different drug to each group, specific to the genetic profile of each group.

Gene discovery, knowledge and technology

The 'standard' approach to gene discovery over the past decade has been studies of families in which the disease occurs frequently, using a principle known as linkage. Such studies were successful in identifying disease-susceptibility genes in rare, familial diseases such as cystic fibrosis and sickle-cell anaemia – diseases caused by mutations in a single gene – but they were mostly unsuccessful in more common heritable diseases where many genes interact to produce disease susceptibility and none has a strong effect on its own. Some success was achieved by decode Genetics, a company conducting family linkage studies using the unique genetic heritage of Iceland to enhance the detectability of disease genes. However, more often than not family linkage-based discoveries failed to corroborate in confirmatory studies.

Another method, the candidate gene approach, involves a search for causative mutations within a gene that is 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

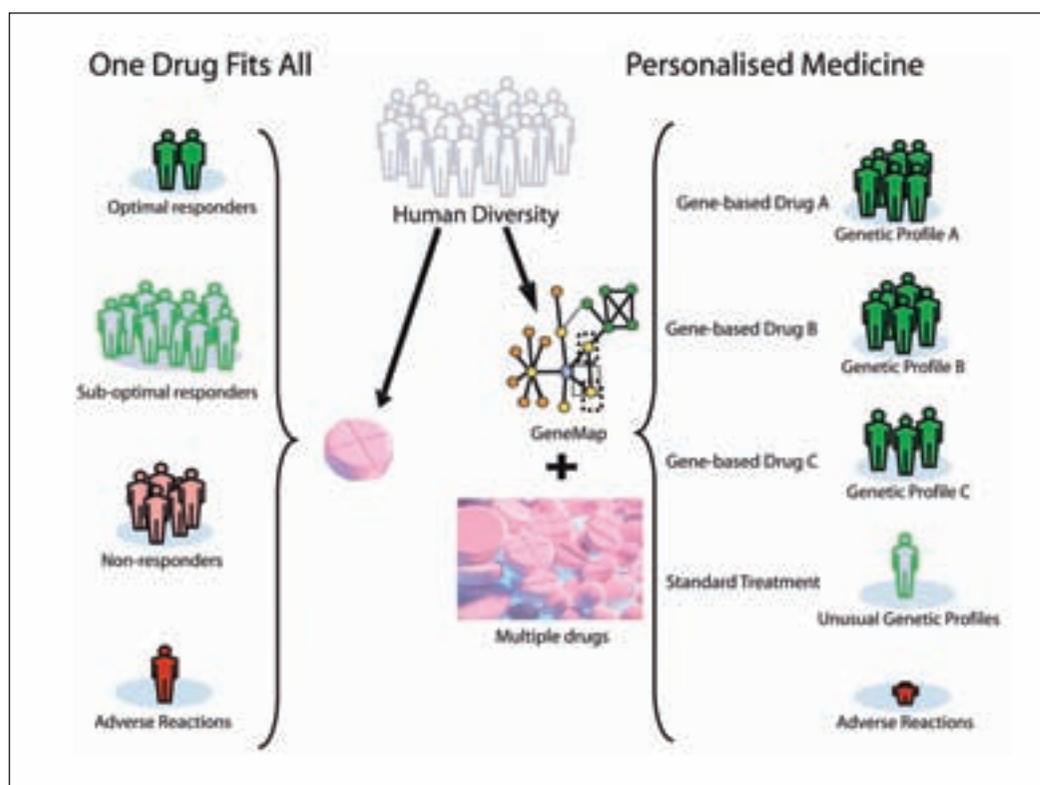


Figure 1
Current paradigm vs
Personalised Medicine

of these have proved to be a necessary or sufficient factor in a common disease. The key objective, 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.

First and second generation population genetics

In the late 1990s, with the race to map the human genome making headlines in the mainstream media, small well-financed genetics-based companies were forging large deals with Pharma companies. In almost all cases, these deals were based on the discovery of a single putative disease gene, or even sometimes less well characterised DNA variations that appeared to be disease-associated. With the completion of the first draft of the human genome sequence in 2001, it had become clear that much more was needed in terms of tools and technology to realise the promise of genetics in drug discovery. The bubble burst, and pharma companies became disillusioned with most of the first generation of genetics companies. Generally, minimal value was generated from these alliances. The bad taste remains today in some pharma companies. Since the sequencing of the human genome, knowledge and technology have evolved rapidly and the tools required for more powerful

approaches to disease gene discovery are becoming available.

While the human genome still retains many secrets, it is now sufficiently well documented to allow precise identification of the locations of disease genes. Several million mutations, most of which we all carry but which vary from person to person, have now been documented on a large scale. One type of these mostly harmless variations, known as SNPs, is being used to track down disease genes, enabled by technology platforms that generate millions of SNP analyses (genotypes) daily. Currently there are several groups offering high throughput SNP genotyping, with Affymetrix and Illumina both providing products containing more than 500,000 SNPs. Illumina markets a system based on microscopic beads to which fragments of DNA bind and fluoresce, and the Affymetrix platform is based on chips that rely on similar principles. The platforms vary in the genome coverage of their maps, the extent of missing data and in their accuracy, but all have costs that are more than 100 times lower than what was available only a few years ago, with very high throughput. For example, with eight BeadArray Readers from Illumina, Genizon BioSciences now has a genotyping platform with a throughput of more than 220 million SNP genotypes per day.

These SNP genotyping platforms, together with

Genomics

maps of SNPs generated by the HapMap project and other sources, make it possible to conduct studies comparing DNA from hundreds to thousands of patients with that from controls, typically generating hundreds of millions of genotypes. The computing power required to analyse the terabytes of genotype data involved in these studies has also recently become sufficiently fast and affordable, although extensive customised software is required to achieve viable output. Thus, genome-wide association studies (GWAS), the gold standard for disease gene discovery, are now achievable leading to the promise of novel targets for therapeutic intervention that act on the root cause of disease.

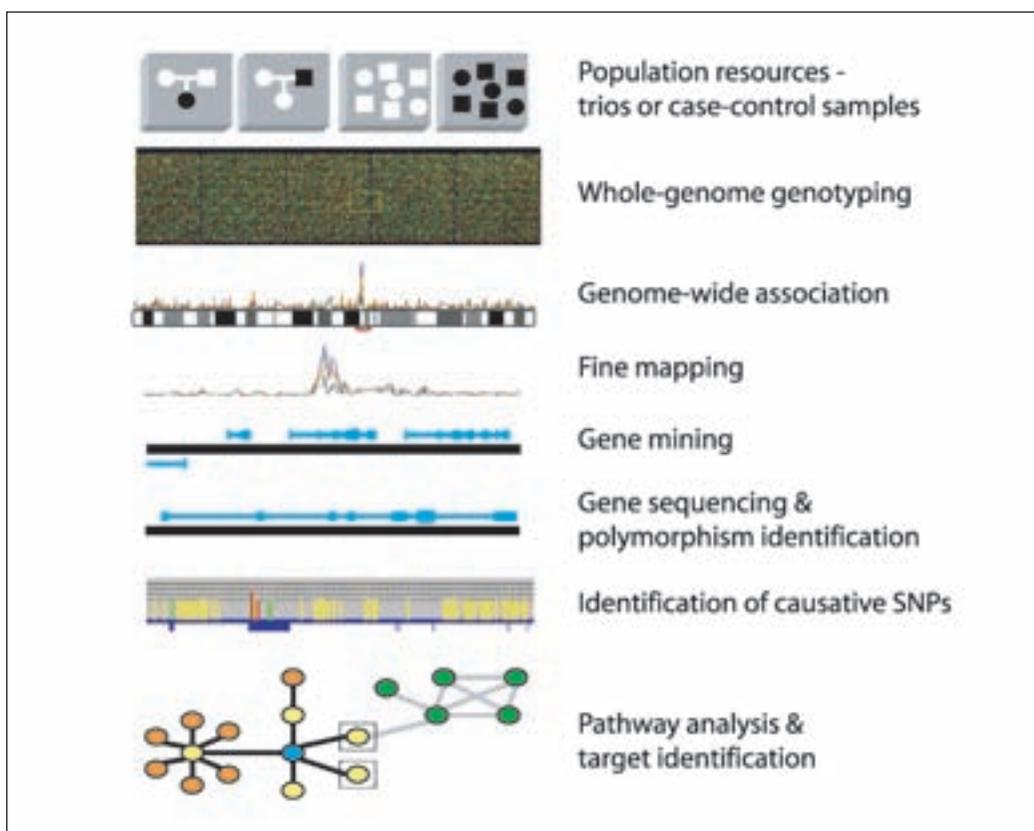
GWAS involves scanning the whole human genome in unprecedented depth using unrelated patients, either as case-control cohorts or in family trios, utilising hundreds of thousands of SNPs markers located throughout the human genome. Statistical algorithms are applied that compare the frequency of either single SNP alleles or multi-marker haplotypes, formed by combinations of SNPs, between disease and control cohorts. This analysis identifies regions (loci) with statistically significant differences in allele or haplotype frequencies between cases and controls, pointing to their role in disease (Figure 2).

Advantages of GWAS

Genome-wide association studies have several advantages over alternative disease gene discovery methods. In contrast to candidate gene studies, which select genes for study based on known or suspected mechanisms of disease, a GWAS involves a comprehensive scan of the genome in an unbiased fashion and, hence has the potential to identify totally novel susceptibility factors. In comparison to family linkage based approaches, association studies have two key advantages. First, they are able to capitalise on all meiotic recombination events in a population, rather than only those in the families studied, and hence association signals are localised to small regions of the chromosome containing only a single to a few genes, enabling rapid detection of the actual disease susceptibility gene. Second, a GWAS permits the identification of disease genes with only modest increases in risk (a severe limitation in linkage studies), the very type of genes one expects for common disorders. Due to these advantages, genome-wide association studies can identify multiple interacting disease genes and their respective pathways, thus providing a comprehensive understanding of the etiology of disease.

The power to detect association between genetic variation and disease is a function of several factors,

Figure 2
GWAS process



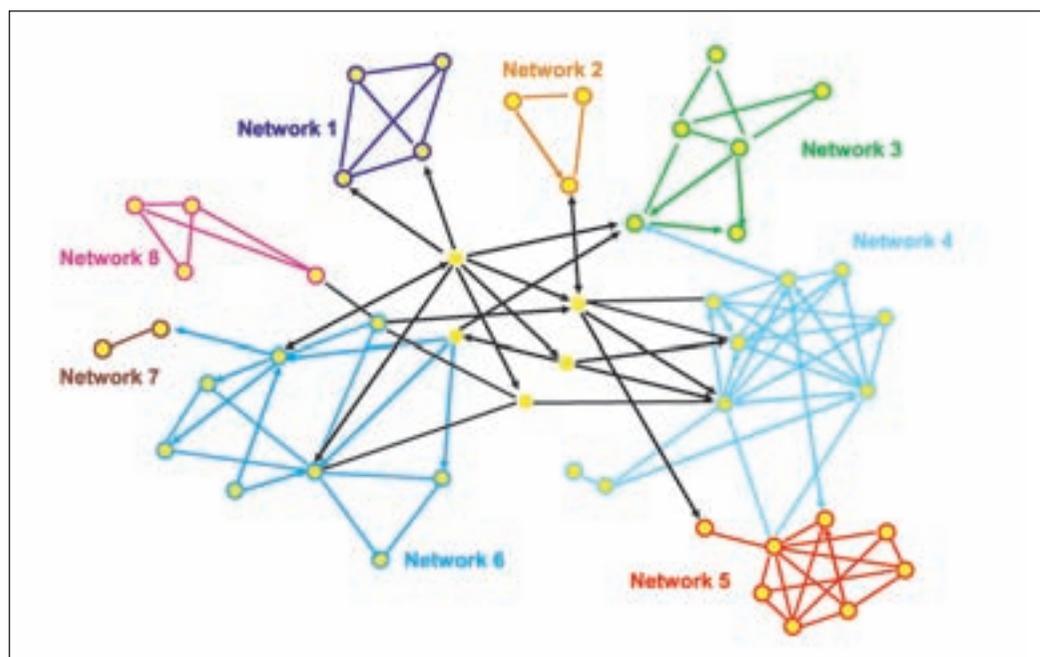


Figure 3
A GeneMap for psoriasis

including: the frequency of the risk allele or genotype, the relative risk conferred by the disease associated allele or genotype, the correlation (LD) between the genotyped marker and the risk allele, sample size, disease prevalence and genetic heterogeneity of the sample population. Key success factors include sufficiently large sample sizes, rigorous phenotypes, comprehensive SNP maps, accurate high throughput genotyping technologies, sophisticated IT infrastructure, and rapid algorithms for data analysis.

The second generation of population genetics is now with us. There is a swelling chorus of key opinion leaders moving away from their first love, family linkage studies, to their trophy brides, GWAS. Several GWAS are under way using the results of the HapMap project, large DNA collections and high throughput genotyping technologies.

Founder populations and the first whole genome association study

For two or more years, there has been a consensus that discovery of disease genes in common diseases using GWAS would require a thousand to several thousand patient DNA samples, a comparable number of controls, and the analysis of between 300,000 and 1,000,000 SNPs for each DNA sample. Further, the SNPs would have to be distributed across the human genome in a manner that reflects the extent to which blocks of DNA are shared among members of a population. These blocks vary from several million 'letters' of the genetic code to a thousand or less

such letters. The lower the extent of genetic sharing in the population, the more SNPs required.

In 2003, the first genotyping platform with a throughput sufficient to process a GWAS became operational at Perlegen. Scientists at Genizon reasoned that although not all of the components required for a successful GWAS were available at that time, use of a founder population would compensate for the deficiencies. The members of such populations tend to share larger blocks of DNA inherited from common ancestors than is found in general populations. They are also thought to carry fewer mutations per gene, since only a limited number of such mutations were brought into the population by the founders.

Quebec is such a founder population. Its members have descended from a small group of approximately 2,600 French immigrants who arrived in Quebec in the period 1608 to 1763. That population expanded, in relative isolation with minimal intermarriage, to six million people today. To detect disease genes, this population should require fewer SNPs than would a general population and the limited number of mutations should make such genes more readily detectable.

Genizon formed an alliance with Perlegen to conduct a GWAS in the Quebec founder population on Crohn's disease, a common disease that is more than 80% heritable. This was one of the few common diseases where disease genes had been previously unequivocally identified, providing positive controls. While the selection of SNPs was not

Genomics

ideal, and there was minimal knowledge of how these should be distributed across the human genome, the project was initiated in early 2004. Approximately 140,000 SNPs were eventually used for data analysis in the Crohn's GWAS, substantially fewer than is likely to be required for a general population; nevertheless this led to dramatic results. Sixteen disease genes were initially detected, including the two known genes. About half of these have been corroborated in a German population, a proportion higher than expected, considering mutations within a disease gene and the incidence of any one mutation can differ substantially between populations. Application of knowledge management systems and additional statistical analyses has revealed additional disease genes, either directly involved or through connecting biochemical pathways.

A better way

The Crohn's data from the genomes of 1,000 Caucasians was used to measure the variation in genetic sharing across the human genome in the Quebec founder population. This may be contrasted with the international HapMap project that has conducted a more in-depth evaluation of genetic sharing, but for which only 60 Caucasian individuals were used. The information from both sources was used to select a smaller set of SNPs, the 'Quebec LD Map', reflecting genetic sharing in Quebec, for application to a GWAS of psoriasis. This project was successful despite the use of only 60,000 SNPs versus the 140,000 used in the Crohn's study and the 500,000 or more likely to be required for a non-founder population. The disease gene discoveries in the psoriasis study, together with other genes revealed by applying knowledge management systems and new types of statistical analyses to the data from this study, resulted in the generation of a GeneMap for psoriasis (see **Figure 3**).

GeneMaps provide a comprehensive picture of the major and many minor genes involved in a common disease. They consist of disease genes (and their protein products) identified through genetic studies as well as additional proteins in their respective biochemical pathways. Unlike many potential disease genes from functional and expression maps, the genes in a GeneMap are backed by strong evidence from patients and are believed to be unequivocally involved in human disease.

Current GWAS initiatives

While Genizon was conducting GWAS on Crohn's disease, psoriasis and five other diseases, technology and genetic knowledge continued to evolve.

Genizon has revised its Quebec LD Map by customising a HumanHap BeadChip from Illumina containing ~317,000 SNPs with an additional 57,000 SNPs distributed according to the genetic sharing profile of the Quebec founder population, for a total of 374,000 SNPs on a single array. Genizon has conducted three genome-wide association studies – in attention deficit-hyperactivity disorder, schizophrenia and endometriosis – using this array. This has produced a GeneMap for each disease. These are still evolving in content and sophistication, but it is clear that they provide a unique and comprehensive picture of the genetic origins of these diseases.

A leader in the field, Leonid Kruglyak of Princeton University, recently remarked that: "Assessments of the HapMap resource suggest that the tools (to conduct GWAS) are now at hand..." in reference to application of GWAS to general populations. Several such studies were started in early 2006 in academic centres using the Affymetrix or Illumina platforms and up to 550,000 SNPs per DNA sample and, while early reports of results are somewhat disappointing, the factors discussed below will presumably soon be addressed, leading to a second wave of such studies.

Potential roadblocks

Most of the disease genes discovered by Genizon's genome wide association studies of five common diseases have small effects (increases in disease risk of 40% to 80%, compared to 300% or more in rare, single gene heritable diseases) and they occur frequently in the population. Both of these factors make detection challenging. Disease risk is likely to be even lower for these genes in non-founder populations due to a larger number of mutations in each gene. Taking the lower end of the Quebec risk range, detection of such genes in a GWAS of a non-founder population would require 1,000-2,000 patients and a comparable number of controls, all other factors being ideal. However, these numbers are based on the disease gene signals from SNPs being 100% co-inherited with the disease-inducing mutation. In fact the SNPs chosen for GWAS of non-founder populations are calculated to be typically only 50-80% linked with other mutations in the block of DNA they represent, and this evidence is based on only 60 genomes from a Utah Caucasian population from the HapMap project. The uncertainty of this link between SNPs and the disease-causing mutation could more than double the number of patients required for GWAS in general populations.

Other factors, including a patient cohort that

may include several different forms of the disease, controls who are contaminated with people who have or could eventually contract the disease under study, and genetic heterogeneity due to inclusion of patients from varied ethnic backgrounds, could further reduce the power, again requiring a doubling of the patient numbers. A further challenge is the development of algorithms that objectively distinguish between disease gene signals and the inevitable noise associated with all biological systems, and computer software to reliably apply these algorithms to the terabytes of data arising from a GWAS. The computational biology tasks associated with GWAS are not trivial issues, as recently noted by leaders in the field. In recent research Genizon has discovered that matching of patients and controls with respect to their genetic origins, as judged by their grandparents' origins, is crucial in order to distinguish signals from disease genes from those arising from population differences between patients and controls.

Definitive results from the first dense GWAS in non-founder populations are expected in the first quarter of 2007. If the factors defined above reduce gene detection power substantially, these studies may detect only the low hanging fruit. However, the data will allow better definition of genetic sharing in the populations concerned and better study design, perhaps leading to a second wave of studies with better distributed and denser SNP maps, larger cohorts and better matching of patients and controls, and thus greater statistical power.

Meanwhile, GeneMap construction continues at Genizon using GWAS and the power of the Quebec founder population. Even if the first wave of GWAS using non-founder populations is less successful than hoped, there is little doubt that in 2-4 years' time many disease genes will have been discovered using founder populations or in a second wave of non-founder population studies. In some or even many common diseases, comprehensive GeneMaps will have been created. These will lead to revolutions in drug discovery and development, disease diagnosis, patient therapy and drug markets.

Benefits of GWAS

The doors are opening to a new paradigm in drug development and therapy. The author believes that with current technology, within a decade we will see theranostic products that show great therapeutic benefit over existing drugs. As the knowledge of disease processes becomes more comprehensive through generation of GeneMaps, treatment regimes can be targeted to individual genetics. Diagnostics can be developed from causative muta-

tions in disease genes identified in genome wide association studies, 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. If gene-based theranostics lead to faster registration, as has been asserted by some regulators, and if the evolution of genetic knowledge and platform technology continues at its current breakneck pace, we may see such products, at least for life-saving situations, in only a few years. In a recent public presentation, Quintiles, a Contract Research Organisation, forecasted that pharmacogenomics-based clinical trials could result in a reduction in drug development time of more than five years and associated reduction in cost, facilitated through the identification and targeting of good responders for the drug.

Genome-wide association studies will lead to a plethora of new drug targets. While there may be no current shortage of targets in the pharma industry, the GWAS-derived targets will differ in that they are directly and unequivocally related to disease causation. Furthermore, the percentage of patients affected and the likely response of these patients to a drug tailored to the patient's genetic profile will be much more predictable. The prospect of pipelines filling again and the luxury of having choices among several validated and relevant targets may be expected to transform drug development. Gene-based drugs and diagnostics can also be expected to transform the way medicine is practised. Susceptibility to disease will be ascertained early in life, disease onset will be detected early and treated more effectively and safely and before significant damage is done, or therapeutics may be given to prevent disease in susceptible individuals.

DDW

Dr John Hooper is currently President and Chief Executive Officer of Genizon Biosciences, Inc. Prior to Genizon he was CEO of Phoenix International, which grew under his leadership from 20 to 2,600 employees in 16 countries and annual revenues of \$300 million. Phoenix was sold to MDS for \$500 million in April 2000. Dr Hooper has 33 years of relevant experience, including 28 at executive level, and a successful track record managing time-sensitive R&D for profit. He was twice named Entrepreneur of the Year for Quebec in the life sciences and was recently made a member of the 'Cercle Excelcia' for "An exceptional career and a remarkable contribution to the Quebec biotechnology and life sciences industries". Dr Hooper received his PhD in organic chemistry from the University of London.