

What You Need to Know

■ Complex disorders, such as cancer, diabetes and hypertension, have multiple genetic and environmental risk factors. It is very challenging to identify these risks.

■ If a specific genetic variant is different at a statistically significant frequency among a sample of individuals with a disease than among a sample of unaffected individuals, there is an “association” between the disease and the specific DNA variant.

■ By testing a number of genetic variants for association, it is possible to uncover new risk factors for disease.

■ New technologies allow for the simultaneous testing of hundreds of thousands of variants across the genome. This type of approach is known as a Genome-Wide Association (GWA) study.

■ GWA studies have recently identified new genetic risks for over 60 complex traits and disorders. Surprisingly, most of the genetic factors confer only a modest increase in risk.

If you want to know more:

Genome-wide association study fact sheet - <http://www.genome.gov/20019523>

The following links explain the structure and application of GWA studies. They are posted on *Scitable*, an online cornucopia of genetic reviews and scientific discussions. Sponsored by Nature Publishing Group, *Scitable* is an outstanding way to “plug in” to current topics in genetics and genomics.

<http://www.nature.com/scitable/topicpage/Complex-Diseases-Research-and-Applications-748>

<http://www.nature.com/scitable/topicpage/Genome-Wide-Association-Studies-GWAS-and-Obesity-752>

BIOTECH 101

Genome-wide association studies searching for genetic needles in haystacks of data

Introduction

Over the last twenty five years, much progress has been made to identify the genetic causes of what are commonly known as “single gene disorders”. These diseases, which include cystic fibrosis, sickle cell anemia, phenylketonuria and Huntington disease, result from changes in the DNA sequence of a single gene. These findings have led to more accurate risk analysis, better testing approaches and, in some instances, more effective methods of treatment. However, single gene disorders account for only a minor portion of all genetically-influenced disease. Disorders present at birth like cleft lip or spina bifida as well as many



adult-onset diseases such as cancer, dementia, diabetes and hypertension are not caused by mutations in a single gene. Known as complex disorders, these diseases result from interactions between a few to possibly hundreds of genes and environmental triggers. The combination of specific genetic variations (alleles) in a person’s genome that increase risk, followed by exposure to predisposing environments, leads to the disease. When compared to single-gene diseases, complex disorders as a whole are much more common, affecting an estimated two-thirds of all Americans. These diseases place a heavy burden on the healthcare system.

Consequently, identifying the genes connected to complex disorders is a major goal for biomedical research. Traditional methods of determining the genes responsible for single-gene disorders do not work well for complex diseases, because no one gene’s contribution is strong enough to be shown with previously-existing methods. Fortunately, a technique known as genome-wide association (GWA) has made headway in identifying these complex genetic risks. GWA allows a scientist to examine hundreds of thousands of genetic variants that span the human genome – a previously unfathomable accomplishment. In this edition of Biotech101, we will explore GWA studies and discuss how their findings have caused researchers to rethink their ideas on the genetic contribution to complex disorders.

Understanding Association

The basic premise behind GWA studies is straightforward: if a specific genetic variation increases the risk of developing a disease, that variation will occur more frequently, and hold up under rigid tests for statistical significance, in individuals who have the disease compared to those not affected. In other words, there is an association between the specific allele and the incidence of disease. In reality, association can occur for a number of reasons - not simply because the variant directly increases risk. For the purpose of this discussion therefore, think of association as a flag that calls attention to a specific region as being possibly related to disease formation.

In association studies, scientists use a process called genotyping to examine alleles from a genetic region where the DNA sequence is known to vary. Genotyping is performed on two different groups - a collection of affected patients (known as cases) and a group of disease-free individuals (called controls). At each DNA site that is genotyped, the frequency of identified alleles is compared between the cases and controls. Any site where the allele frequency significantly differs between cases and controls is analyzed more closely. The DNA sequence surrounding the genotyped site is studied. Genes within this region may be screened to determine if there is a difference in activity level or protein structure between cases and controls. The goal is to identify a genetic change among the cases that is involved in disease development. If the results can be replicated in other, independently-collected sets of cases and controls, the gene and specific risk-modifying allele are accepted as a contributor to disease.

Genome-wide association studies

Historically, each association study genotyped a small number of genetic regions, selected because previous work identified the region as a candidate site. This was usually because the genes in the region were thought to have some biological reason for being related to the disease. Instead, if variable locations could be genotyped on a genome-wide basis, association studies could identify genetic contributions without the limitation of prior findings or biases. Such genome-wide association (GWA) studies had been hampered by three factors: small sample sizes, too few variable genetic sites for testing and a high genotyping cost.

During the mid 2000s, these obstacles were largely overcome. Researchers formed large collaborative studies to pool

patient samples, the International HapMap (profiled in the Spring 2008 Biotech 101) identified millions of new variable DNA sites known as SNPs, and new genotyping technologies caused costs to plunge. Where previously researchers were limited to simultaneously testing only a handful of variable sites at a time, new so-called “gene chips” could genotype 600,000 or more SNPs in a single experiment. Suddenly, scientists were performing association studies on thousands of individuals using genotyping data from nearly one million DNA sites! New genetic contributors were quickly identified for both type 1 and type 2 diabetes, Crohn disease, rheumatoid arthritis, macular degeneration and heart disease. This rapid pace of discovery has continued into 2009 and to date, more than 150 variants or genetic regions have been associated and replicated for greater than 60 complex diseases or traits. Most of the newly associated genes have not previously been linked to the disease of interest. Intriguingly, some genetic regions have been associated with multiple disorders, suggesting common chemical pathways that influence a number of different processes.

Limitations of GWA

As the number of GWA studies increase, some important limitations have come to light. For example, the contribution by a single genetic variant to the overall clinical picture is often small. Increases in risk identified from these GWA studies are generally quite modest – in the 1.2 to 1.5-fold range or less. Consider recent findings related to human height, known to be a strongly genetic trait. A number of large-scale GWA studies identified over 50 regions of the genome that impact height, but their sum total explains less than 5% of the overall variation in height, suggesting that each genetic variant adds or subtracts only fractions of an inch to our total height. In the same way, studies of genes that influence body mass index identified eight genes linked to BMI. Individuals with the “heaviest” genetic variants at these genes only weighed on average about 4.5 pounds more than in-

dividuals with “average” variants for these genes. These individual contributions are much smaller than researchers had hoped, making both simple diagnostic tests and the resulting clinical treatments more challenging. Rather than testing for five or six genetic risk factors, tens or even hundreds of incremental genetic partners may need to be examined.

Although strong predictive genetic factors have not been uncovered, combining the effects of moderate influence can still identify individuals with substantially altered disease risk. In addition, GWA studies have provided important insight into the biological pathways

that underlie genetic disease. Seventeen genetic contributors to type 2 diabetes have been identified by this approach - fourteen of these were previously unknown. These variants may be useful in tailoring therapeutic options, especially as they relate to drug choice and the design of new pharmaceutical targets. For example, variation in a gene known as *KCNJ11* has been associated with diabetes. This gene produces the sulfonylurea receptor, a protein that sits on the outer membrane of cells and is a key target for diabetes drug therapy.

Recent studies also suggest that many disease-influencing genetic alleles may be quite rare. As such, these rare variants are not included on current gene chip technologies, which traditionally genotype more commonly varying alleles. Consequently, a number of genetic contributors may yet be unrecognized. This is likely to be a short-term limitation however; as emerging technologies in DNA sequencing continue to drive down costs, many believe GWA studies will shift from genotyping to direct genome sequence-based data, allowing us to see any variant at any position instead of interrogating the genome only for specific variants at specific positions. At that point, identifying even the most rare of variation becomes feasible.

Like so many aspects of this field, GWA studies have undergone a head-spinning

Are all genetic variants bad?

Although the primary focus on this edition of Biotech 101 is on identifying genetic alleles that increase risk, there are clearly other alleles that provide a protective, or risk-decreasing effect. Appropriately designed GWA studies should detect these variants as well. In fact, identifying changes to a genetic recipe that reduce disease risk offer excellent opportunities to better understand disease pathways and design medications that mimic the effect of the protective change.

: Continued on page six

Biotech 101 continued:

increase in scale. The ability to search for genetic association across the genome in a single experiment is an amazing technological feat. The results of GWA have generated substantial excitement among the genomics community, yet have also introduced concerns about the relatively modest influence of most genetic contributors. Even so, many associations have provided novel understanding into the pathway of disease development and are opening the door to advances in prevention and treatment of common disorders.

GWA Studies at HudsonAlpha

HudsonAlpha currently runs three genotyping machines in the lab of Dr. Devin Absher, from a company called Illumina. These machines use glass slides that contain up to one million short DNA pieces on them, so we are essentially conducting one million experiments on a person's genome every time we expose one human sample to the microarray. Dr. Absher, his lab staff, and collaborators have so far conducted a GWA study in coronary artery disease as part of the Atherosclerotic Disease, Vascular Function, and Genetic Epidemiology (ADVANCE) study. The lab is now moving on to assess GWA in bipolar disorder, a disease which clearly has some genetic component but has symptoms that are very difficult to define precisely. HudsonAlpha is also discussing collaborations with Vanderbilt University and others around the country to genotype samples for GWA studies in other diseases.

By Dr. Neil Lamb
director of educational outreach
HudsonAlpha Institute for
Biotechnology