AT-A-GLANCE

Methods Used in Medical and Population Genetics

Over a decade ago, scientists completed the sequencing of the human genome and laid out a "parts list" for life. This Herculean task was a huge advance toward transforming medicine through genomic insight, but the effort didn't reveal the role of each of those parts in human health and disease. By studying natural genetic variation among people, whether in a small family or a large, diverse population, researchers today can gain insight into the function of genes and genetic variation in human biology, illuminate the genetic roots of disease, and potentially discover new therapeutic avenues.

Scientists in the Broad Institute of MIT and Harvard's Program in Medical and Population Genetics primarily study common, complex diseases for which many genes contribute to risk of an individual getting the disease, such as type 2 diabetes, heart disease, and inflammatory bowel disease. Additionally, researchers study rare, "Mendelian" diseases where a single genetic defect can cause illness.

Researchers at the Broad use a variety of techniques to look for relationships between DNA variants and traits or disease risk. A core aim of the program is to connect genotype, or the particular genetic makeup of an organism, to phenotype, an organism's observable traits. In other words, researchers seek to uncover the genetic underpinnings of traits, and then understand the functional effects of those genetic variations, especially with regard to human illness.

Historically, it has been difficult to pinpoint the genes that underlie common diseases because the impact of each DNA variant is often quite small. To bring these subtle disease risk factors to light, scientists conduct "association studies" on a great number of people, to identify variants that are found more often in people with a trait or disease than those without. This approach requires powerful analytical and statistical methods, many developed at the Broad Institute and shared openly with researchers around the world.

However, correlation — in the form of association — does not equal causation. After identifying the DNA changes associated with a trait, scientists can then develop and apply phenotypic assays, or experimental measurements, often in large-scale screening studies, to test the impact of those variants on cells or animal models and home in on the true causal variants that influence a trait. These so-called "functional studies" can potentially reveal new therapeutic avenues; the DNA variants may become therapeutic targets themselves, or they may highlight an important molecular pathway with therapeutic potential.

Broad scientists use some of the following methods in the study of medical and population genetics.

METHODS TO IDENTIFY GENETIC VARIATION ASSOCIATED WITH TRAITS OR DISEASE

- **GENOME-WIDE ASSOCIATION STUDIES (GWAS):** These studies examine the frequency of common variations within the human genome to determine which locations in the genome may be linked to a specific phenotype. To study these variations, researchers scan strategically selected sites of the

BROAD INSTITUTE

Office of Communications
www.broadinstitute.org
617-714-7151 • communications@broadinstitute.org
415 Main Street • Cambridge, MA 02142

genome that are known to vary considerably across the population, taking note of single nucleotide polymorphisms (SNPs) – single-letter variations in the genetic code. SNPs found to be significantly more common in people with a trait than in those without are considered to be "associated" with that phenotype. Where the associated SNP resides in the genome can provide valuable clues about the genes and mechanisms that may be contributing to the phenotype being studied. GWAS can be conducted with standard genotyping arrays known as SNP chips, which contain more than a million genetic markers spread across the genome, or can be customized to target genes or pathways of interest. For example, the MetaboChip targets genetic regions involved in metabolic, cardiovascular, and anthropometric traits such as body mass index. SNP arrays can also be used to identify copy number variants – extra or missing segments of DNA – and their association to disease through the GWAS approach.

- **WHOLE-EXOME ASSOCIATION STUDIES:** In whole-exome sequencing, only the protein-coding part of the genome, the "exome," is sequenced. Researchers can identify variation in protein-coding sequences, and then compare cases and controls to identify variants that appear more often in patients with a disease or trait versus those without. Whole-exome sequencing is a cost-effective approach when scientists want to focus on protein-coding mutations and get very deep coverage of the genome, improving their statistical power to call, or identify, variants. Compared to SNP arrays, sequencing offers a more complete picture of genetic variation, including single-letter substitutions, copy number variants, rare variants, insertions, or deletions in the genome. However, a drawback of exome sequencing is that is misses the "non-coding" regions of the genome, which have been shown to harbor disease-associated variation in other studies.

- **LOW-PASS WHOLE-GENOME ASSOCIATION STUDIES:** To capture both coding and non-coding variation in the genome, researchers can sequence the entire human genome, known as the "whole genome." This method is much more costly than exome sequencing, so researchers often do whole-genome sequencing at "low-pass," meaning each segment of DNA is not sequenced very deeply. Advantages of whole-genome sequencing are that it offers complete coverage of the genome and can potentially capture rare mutations that could be driving disease. It can also be used to do genetic "fine-mapping" to more precisely locate a variant of interest, and to test whether GWAS signals are driven by something rare that correlates with common variants.

METHODS TO IDENTIFY FUNCTIONAL EFFECTS OF GENETIC VARIANTS

- **EXPRESSION QUANTITATIVE TRAIT LOCUS (eQTL) STUDIES**: An eQTL is a regulatory element of the genome that controls the activity of a gene. eQTL studies aim to correlate genetics with gene expression and can be used to learn more about genomic sites identified in association studies. In eQTL studies, scientists ask whether a genetic variation causes the expression of other genes to increase or decrease, suggesting that the variant plays a regulatory role. Initially these studies used expression microarrays to measure genome-wide gene expression, but more recent studies use RNA sequencing to quantify the activity of genes. Because gene expression is tissue-specific, these studies are typically done in cells or tissues thought to be most relevant to the disease of interest. An example of a large-scale eQTL effort is the Broad's Genotype-Tissue Expression (GTEx) project, which provides a comprehensive atlas of gene expression and regulation across multiple human tissues.

- **FUNCTIONAL PROFILING USING GENETIC PERTURBATIONS:** To directly test the functional effects of hits from association studies, researchers can either knock genes down using RNA interference (RNAi) or knock them out with CRISPR-Cas9 genome editing technology. The effects of these perturbations can then be observed using phenotypic assays in cellular or animal models. Scientists can also conduct unbiased whole-genome screens using RNAi or genome editing to observe the effects of perturbing thousands of sites across the genome, using disease-relevant functional assays. Because of its ability to target single mutations in the genome and highlight important molecular pathways, CRISPR-Cas9 technology is a powerful experimental tool to help bridge correlation to causation and shed light on the roots of disease phenotypes.

For more on medical and population genetics work at the Broad Institute, see our website.

Primer videos on a variety of topics from the Broad's MPG program are available on YouTube or downloadable on iTunesU.

Last updated April 2015