

Embargoed News from the University of Washington in Seattle
UW Health Sciences/UW Medicine

Media contacts:

Leila Gray
206.543.3620 or 206.941.4506
leilag@uw.edu

Stephanie Seiler
206.713.0124
stephanie@geminibioprojectsllc.com

EMBARGOED UNTIL 1 P.M. EDT (17:00 GMT) WEDNESDAY, SEPT. 5:

Researchers unlock disease information hidden in genome's control circuitry

Data published in Science show that DNA changes associated with a wide range of diseases impact how the genome switches genes on and off rather than the genes themselves

Researchers at the University of Washington have determined that the majority of genetic changes associated with more than 400 common diseases and clinical traits affect the genome's regulatory circuitry. These are the regions of DNA that contain instructions dictating when and where genes are switched on or off. Most of these changes affect circuits that are active during early human development, when body tissues are most vulnerable.

By creating extensive blueprints of the control circuitry, the research also exposed previously hidden connections between different diseases. These connections may explain common clinical features, as well as offer a new approach for pinpointing the specific types of cells and tissues that either cause or are most affected by a particular disease. The findings provide a major paradigm shift for understanding the genetic causes of disease, and open new avenues for development of diagnostics and treatments. The findings appear in the Sept. 5 online issue of *Science*.

"Genes occupy only a tiny fraction of the genome, and most efforts to map the genetic causes of disease were frustrated by signals that pointed away from genes. Now we know that these efforts were not in

vain, and that the signals were in fact pointing to the genome's 'operating system' -- the instructions for which are hidden in millions of locations around the genome," said Dr. John A. Stamatoyannopoulos, associate professor of genome sciences and medicine at the UW. "The findings provide a new lens through which to view the role of genetics and genome function in disease."

The human genome's control circuitry is encoded in millions of regulatory regions -- short DNA sequences that are scattered throughout the 98 percent of the genome that does not specify the protein product of a gene. Specialized proteins, called regulatory factors, recognize specific DNA sequences in these regulatory regions, thereby creating switches that turn genes on and off. In many cases, these switches are located far away from the genes that they control. These distances have made it difficult to determine the relationship between specific switches and genes.

The researchers used a special molecular probe called a nuclease to detect all of the regulatory regions active in each cell type they studied. The specific nuclease they used -- called DNase I -- snips the genome where regulatory factors are bound to DNA. By treating cells with DNase I and analyzing the pattern of snipped DNA sequences using massively parallel sequencing technology and high-performance computers, the researchers were able to create comprehensive maps of all the regulatory DNA in many different types of cells. These maps were then analyzed with advanced software algorithms to sort through the data and expose previously hidden connections between disease-associated genetic variation and specific regulatory regions.

The regulatory mapping and analysis was conducted on 349 cell and tissue samples. These included samples from all major organs as well as 233 tissue samples from different stages of early human development. In total, nearly 4 million distinct regulatory regions were discovered, though only about 200,000 of these were 'on' in any particular cell type.

To make a connection with common diseases and clinical traits, the researchers analyzed genetic variants that had been strongly associated with diseases and traits through so-called genome-wide association studies, which compare genetic information between groups of people with or without a particular disease or trait. During the past decade, hundreds of genome-wide association studies involving hundreds of thousands of patients worldwide have been performed for over 400 diseases and traits. Nearly 95 percent of the time, these studies flagged genetic variants that were located outside

of gene protein-coding regions. Comparison of these data with the regulatory DNA blueprints yielded several key findings:

- 76 percent of disease-associated variants in non-gene regions are actually located within or are tightly linked to regulatory DNA. This suggests that many diseases result from changes in when, where, and how genes are turned on rather than changes to the gene itself.
- 88 percent of the regulatory regions that contained disease-associated DNA variants were active in early human development fetal development. Because many of these variants are associated with common diseases that occur in adults, the finding indicates that factors influencing the genome's regulatory circuitry early in development may impact the risk of developing particular diseases later in life.
- DNA changes associated with specific diseases tend to occur in the specific short DNA codes recognized by regulatory proteins involved in physiological processes related to the disease or the organs or cells affected by the disease. For example, DNA variants associated with diabetes tend to occur in the codes recognized by regulatory proteins that control various aspects of sugar metabolism and insulin secretion. Similarly, variants associated with immune system disorders, such as multiple sclerosis, asthma, or lupus, are found in specific recognition codes for proteins that regulate immune system function.
- Many seemingly unrelated diseases share common regulatory circuitry, including diseases that affect the immune system, different types of cancers, and a range of neuropsychiatric disorders.

The study also revealed a wealth of additional connections between genetic variants and disease that had been lurking within existing genome-wide association studies data. Viewing these data through the lens of regulatory DNA exposed thousands of variants that were highly selectively localized within regulatory DNA of disease-specific cell types. These variants had previously been ignored because the stringent selection criteria used in earlier studies did not take regulatory regions into account.

Another surprising finding was that the regulatory circuitry blueprints could be used to pinpoint cell types that play a role in specific diseases -- without requiring any prior knowledge about how the disease worked. For example, genetic variants associated with Crohn's disease

(a common type of inflammatory bowel disease) were found to be concentrated in the regulatory regions mapped in two specific subsets of immune cells -- the same cell types that took decades of prior research to be linked with development of tCrohn's disease. Applying this approach systematically will enable researchers to identify cell types not previously known to play a role in a particular disease, expanding our understanding of the disease process and potentially leading to new therapies.

###

The study was supported in part by the National Institute's of Health (NIH) Common Fund Roadmap Epigenomics Program (grant number U01ES01156), the National Human Genome Research Institute's ENCODE Project (U54HG004592), the National Institute of Child Health and Human Development (R24HD000836-47), the National Institute of Diabetes Digestive and Kidney Diseases (P30 DK056465), and the National Heart, Lung, and Blood Institute (R01HL088456). The NIH Common Fund supports a series of exceptionally high impact research programs that are broadly relevant to health and disease. Common Fund programs are designed to overcome major research barriers and pursue emerging opportunities for the benefit of the biomedical research community at large. The research products of Common Fund programs are expected to catalyze disease-specific research supported by the NIH Institutes and Centers.

This paper is one of several related to Project ENCODE coming from the University of Washington the week of Sept. 3. Joshua Akey will report on regulatory gene differences among individuals and populations. This differences are likely to affect physical characteristics and disease susceptibility. The regulatory genes also show the effects of evolutionary change from natural selection. Genome scientist and computer engineer William Noble contributed new developments in machine learning -- computer programs that draw from experience to improve their performance of tasks -- and artificial intelligence to the overall ENCODE project.