

Making Connections

GOOGLE-LIKE SEARCH TOOL LINKS DISEASE-RELATED GENES TO POSSIBLE DRUG TREATMENTS.

Scientists have begun connecting the dots between genes and diseases, but translating what they are learning into effective treatments has been difficult. Now, HHMI investigator Todd R. Golub and his colleagues have created what they dub a “connectivity map” to link diseases, genes, and small drug molecules.

“The problem is that the language typically used to describe diseases is based on signs and symptoms from the patient and on what disease cells look like under the microscope,” says Golub, who is based at the Broad Institute of the Massachusetts Institute of Technology and Harvard University. “Small molecules and drugs are described in a totally different language—the language of chemistry.”

To dissolve that barrier, Golub’s team has effectively created a computerized search engine—not unlike Google, says Golub. The “user,” or investigator, first compiles a list of genes that are known to be important in the disease being studied. That list is used to query a database of drug-responsive genes, or genes that were turned on or off in cell lines that were individually treated with various drugs. The goal is to find a “hit,” or gene match, between the disease genes and the drug-responsive genes.

In their work, described in the September 29, 2006, issue of *Science*, Golub’s team confirmed known connections and found



new ones, including possible drugs for prostate cancer and a particular form of childhood leukemia.

The map is far from complete, however. Only 164 drugs and biologicals were used in this study to represent a broad range of available drugs and compounds. They hope to soon add every drug that is approved for marketing by the U.S. Food and Drug Administration. In the long run, they hope to expand the database to include information from all the inhibitory RNAs that target each of the approximately 20,000 to 25,000 genes in the human genome. ■ - JACQUELINE RUTTIMANN

IN BRIEF

UNRAVELING DNA REPLICATION

Viruses are helping scientists understand the various components of DNA replication, a process important in maintaining and repairing our genomes. Viruses are masters of inserting their genomes into their host’s genome, which they accomplish by altering components of the host’s DNA replication protein machinery. Until now, even though various molecules associated with the process had been identified, how the components worked together remained unclear.

In the November 29, 2006, issue of the *EMBO Journal*, HHMI professor Ellen Fanning of Vanderbilt University reports using the viral protein SV40 T antigen (Tag), which helps initiate replication by unwinding a host cell’s double-stranded DNA, to look at how replication protein A (RPA) is involved in DNA replication. When researchers combined both molecules with pieces of single-stranded DNA 8 nucleotides and longer, they found that both molecules bound to the DNA, with Tag binding first and then helping to load RPA onto the strand. This coupling of DNA unwinding and RPA loading helps elucidate how other elabo-

rate protein assemblies can be placed on the genome.

WE’RE MORE DIFFERENT THAN WE THOUGHT

New research by Stephen W. Scherer, an HHMI international research scholar at The Hospital for Sick Children and the University of Toronto, and colleagues shows that at least 10 percent of genes in the human population can vary in the number of copies of DNA sequences they contain—a finding that disagrees with current thinking that the DNA of any two humans is 99.9 percent similar in content and identity. This discovery of the extent of human genetic variation is expected to change the way researchers think about genetic diseases and human evolution.

Genes usually occur in two copies, one inherited from each parent. The researchers found approximately 2,900 genes with variations in the number of copies of specific DNA segments. These differences in copy number can influence gene activity and ultimately an organism’s function. Scherer’s team compared DNA from 270 people of Asian, African, or European ancestry that had been

compiled in the so-called HapMap, a DNA database previously used to map the single nucleotide changes in the human genome. The team looked for duplicated or deleted DNA sequences, which they call copy number variations (CNVs). Not only were the changes common, they also were large—some a million or so nucleotides long. They reported their findings in the November 23, 2006, issue of *Nature*.

Most research on genetic disease and evolution has focused on small alterations, called single nucleotide polymorphisms. It may be, says Scherer, that some diseases and aspects of evolution are instead caused by CNVs.

CRYSTALLIZING A NEW APPROACH TO PROTEIN STRUCTURE

To better understand a protein’s function, researchers often use x-ray crystallography to determine its three-dimensional architecture. With this technique, crystallized proteins are first bombarded with x-ray beams. The x-rays pass through and bounce off atoms in the crystal, creating a diffraction pattern, which can then be analyzed to determine the three-dimensional shape of the protein. One of the