Scientists reveal how biological activity is regulated in fruit fly and roundworm genomes

December 22 2010

Scientists today published catalogs of the fruit fly and roundworm's functional genomic elements: DNA sequences in the genome that carry the instructions and determine which genes are turned on and off at various times in different cells.

Initially sequenced as part of the Human Genome Project, the genomes of the fruit fly, *Drosophila melanogaster*, and the roundworm, Caenorhabditis elegans, are powerful models for understanding human biology and disease, as many functional genomic elements have been conserved across the vast evolutionary distances separating each organism. Scientists can now study functional genomic elements in the fruit fly and roundworm that are also present in humans to better understand how the human genome works in health and disease.

"These findings will enable scientists everywhere to carry out experiments in fruit flies and roundworms to better understand the relationship between molecular and biological activities in these animals," said NHGRI Director Eric D. Green, M.D., Ph.D. "What we learn from these model organisms will contribute greatly to our understanding about the genomic basis of health and disease in humans."

The papers reporting these new findings will appear in the Dec. 24 issue of *Science* and are authored by members of the model organism ENCyclopedia Of DNA Elements (modENCODE) Consortium, which is

funded by the National Human Genome Research Institute (NHGRI), part of the National Institutes of Health (NIH). In addition, more than a dozen companion modENCODE papers will be published online in the journals *Nature*, *Genome Research* and *Genome Biology*.

The fruit fly and roundworm modENCODE projects were launched in 2007 to complement the work being done by the ENCyclopedia Of DNA Elements Consortium (ENCODE) Project, which is building a comprehensive catalog of functional genomic elements in the human genome. In 2007, ENCODE completed a pilot project that developed innovative methods and technologies to find functional elements in about 1 percent of the human genome.

The modENCODE project takes advantage of many of the same tools and has developed some new ones to apply to the smaller – and therefore more tractable – genomes of the fruit fly and the roundworm. Unlike the researchers in the human effort, modENCODE researchers can conduct genetic experiments on flies or worms to validate the biological relevance of the functional elements they have identified.

To analyze the fruit fly and roundworm genomes, researchers studied many different cell types and developmental stages to produce the catalogs of functional genomic elements. In addition to genes that code for proteins, these functional elements include non-protein-coding genes; regulatory elements that control gene transcription; and DNA sequences that mediate the structure and dynamics of chromosomes.

In the newly published papers, the fruit fly and roundworm researchers report the discovery of hundreds of new protein-coding genes. For instance, in the roundworm genome there is now evidence for thousands of new and refined gene transcripts – instructions from genes that produce proteins – along with thousands of new non-protein coding RNAs (ncRNAs), which regulate gene expression.

"We now know when these genes are used in the life cycle and increasingly what cells the genes are used in," said Robert H. Waterston, M.D., Ph.D., senior author of the roundworm paper and chair of the Department of Genome Sciences, University of Washington in Seattle. "Putting the pieces together has begun to reveal how genes may work in concert to produce the marvelous biology of the roundworm and fruit fly."

"Identification of thousands of new gene transcripts has significantly increased our knowledge of the protein repertoire used in fruit flies," said Susan Celniker, Ph.D., co-author of the fruit fly paper and head of the Department of Genome Dynamics, Lawrence Berkeley National Laboratory, Berkeley, Calif., and also the lead on the project to identify fruit fly RNAs. "Our work provides new resources for studying development, sex determination and aging."

The researchers also examined the organization and structure of chromatin in the cells throughout the life stages of each organism. Chromatin is the protein superstructure that packages DNA and modulates which sections of the genome are accessible to regulatory molecules that convert the genetic code into cellular action. Both groups discovered specific chromatin signatures associated with the regulation of protein-coding genes. Unique chromatin signatures were associated with distinct regions of the genome that either turn genes on or off.

"Chromatin signatures are emerging as a powerful lens into the structure and function of the regulatory portion of the genome that controls cell activity," said Manolis Kellis, Ph.D., senior author of the fruit fly paper and associate professor of computer science, Massachusetts Institute of Technology, Cambridge.

Next, to identify sites responsible for controlling when genes are turned on during the development of an organism, and in which tissues genes

are used, the researchers searched across the genomes of worm and fly during key developmental stages. Primarily, they looked for transcription factors – regulatory proteins often found in specific tissues that control the expression of different genes. In both organisms, they found that many different regulatory proteins bind to the same, overlapping regions of the genome in both organisms, which they call highly occupied targets (HOT).

"Networks give one a different view of the genome than a linear parts list and potentially provide a way of connecting together many chromosome elements to give us insights into how the genome functions," said Mark Gerstein, Ph.D., first author of the roundworm paper and professor of biomedical informatics, Yale University, New Haven, Conn.

The ability to combine the functional data about the fruit fly and roundworm genomes allowed the researchers to construct predictive models that connect regulatory elements with gene-expression changes across specific life stages. The researchers were able to use these combined data to predict the function and expression of thousands of genes.

"The integration of data, from the transcriptome to chromatin to non-coding RNAs and DNA replication, combined into networks vastly increases the information about the genome available to researchers and provides a foundation for in-depth functional studies," said co-author of the fruit fly paper, Gary Karpen, Ph.D., director, Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, Calif.

"The results of modENCODE allow scientists to begin reading the genome sequences, moving from a list of letters to delineating words and punctuation marks," said Elise Feingold, Ph.D., an NHGRI program director overseeing the ENCODE and modENCODE projects.

"Making this effort a success required a high level of coordination and teamwork amongst the groups that enabled the synthesis and high-level view of the data presented in these publications," added Peter Good, Ph.D., who is also a program director for the ENCODE and modENCODE projects.

Over the next year, modENCODE researchers will work to identify additional functional genomic elements to expand the respective catalogs. Moreover, by combining and comparing data from worms, flies and humans, scientists hope to learn far more about the functional elements and regulatory principles in each genome, and gain new insights into human health and disease.

Provided by National Institutes of Health