Scientists find a new class of small RNAs and define its function

January 26 2009

Researchers at Cold Spring Harbor Laboratory (CSHL) announced today the discovery of a new class of small RNAs. At the same time, they reported that their discovery suggests the presence of a strikingly novel biochemical pathway for RNA processing in which these and possibly other small RNAs are produced. The research, which is part of a multinational project called ENCODE, also provided information concerning the biological function of the new short RNA class.

The team's findings, which appeared online January 25th, ahead of print, in the journal *Nature*, significantly improve our understanding of how functional information is stored in the genome. The work at CSHL was spearheaded by Professors Thomas Gingeras, Ph.D., a leader of ENCODE, and Gregory Hannon, Ph.D., a world-renowned expert in small RNAs.

"These results are a good illustration of why the ENCODE project was established," says Dr. Gingeras. "They show how collaborative projects can reveal functional elements and mechanisms embodied in the genome that have never before been described."

Exploring vast, non-coding regions of the genome

At the conclusion of the Human Genome Project in 2003, scientists published a final draft of the DNA sequence found within healthy human cells - an assemblage of roughly 3 billion "As" "Ts" "Cs" and "Gs." While justifiably proud of the feat, genome scientists knew that

the most interesting part of their task was just beginning.

Using the published 2003 sequence, they were able to specify across the entire genome which stretches of DNA comprised genes - regions that act as blueprints for the manufacture of proteins. To the surprise of many, those regions accounted for only about 2% of the genome. Following that realization, most of the remaining 98% began to look more like terra incognita than conquered territory.

To define the full set of genomic elements that perform functions in living cells and to hunt down their location amidst the thicket of genes and non-coding DNA, a multinational project known as ENCODE (an acronym for Encyclopedia of DNA Elements) was initiated in 2003. Recent research by Professor Gingeras, who has played a major role in the project, has revealed that nearly all of the genome is converted into various types of RNA molecules, a process once thought to be restricted to protein-coding genes. What roles, if any, each of these new types of RNA play within the cell is now an important topic of research.

The world of small RNAs gets bigger

As one of the hubs of ENCODE, Gingeras's laboratory at CSHL is part of the effort to catalogue the entire long and short RNA output of cells. Focusing on two ENCODE- targeted human cell lines in the newly announced results, Gingeras's group, in collaboration with Hannon's laboratory, used powerful genome-sequencing techniques to zoom in on small RNA molecules and select potentially new types of RNA for further analysis.

Small RNAs are one RNA subtype among several that have been discovered during the last decade. As a group, they are distinguished by the fact that they do not "code" for proteins, and are physically smaller than coding RNAs. In the small RNA group selected by the CSHL

scientists for further analysis, an abundant type is one that Gingeras's group recently identified as arising specifically from transcription start sites - gene regions also known as promoters, where the synthesis of protein-coding RNA molecules begins.

These promoter-associated small RNAs, or PASRs, can be contrasted with a new species just discovered by Gingeras, Hannon and colleagues, a type they call non-PASRs. The latter originate at sites distant from those where PASRs are generated. Both types of small RNAs were observed to have undergone "capping," a chemical modification that makes them stable and impervious to degradation. "This quality," Hannon observes, "lengthens their lifespan in the cell, a clue that suggests these small RNA classes may have significant biological duties."

Curiously, PASRs and non-PASRs may not be initially synthesized in their "short" form. The CSHL team proposes a model in which mature long RNAs are cleaved followed by a capping of the newly generated long RNA fragment. This is followed by the clipping of the end of the capped long RNA to produce a short RNA product.

Small RNAs can act as "off" switches at "on" sites

Now that these new capped small RNA types have been discovered, the question naturally arises: what do they do? Using a human gene called MYC as a model, the team studied how the presence of PASRs at the start site of a gene impacted its expression, i.e., the way it manifested itself in a living cell. The researchers found that if the level of expression of PASRs was increased, the expression of the MYC gene was reduced. PASRs thus seem to modulate the production of mature RNA transcripts.

The function of non-PASRs is unclear at the moment. This class of

RNAs "could possibly participate more globally in a bookkeeping or quality-control mechanism by which the cell keeps track of the genes it is expressing -- its transcriptional output," according to Gingeras.

This work and the future contributions of the ENCODE project have a larger significance in understanding the genetic roots of human disease. "Unless we obtain much more information about non-protein coding sequences of the genome and learn how various functional elements in the genome impact the production of proteins, we won't fully be able to understand the biological and clinical effects of disease-causing mutations," Gingeras emphasizes.

Reference: "Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs" will appear in the January 25th issue of Nature. The full citation is: Katalin Fejes-Toth, Vihra Sotirova, Ravi Sachidanandam, Gordon Assaf, Gregory J. Hannon, Philipp Kapranov, Sylvain Foissac, Aarron T. Willingham, Radha Duttagupta, Erica Dumais and Thomas R. Gingeras.
This article is available online at www.nature.com/nature/journal/ … ull/nature07759.html (doi is doi:10.1038/nature07759)

Source: Cold Spring Harbor Laboratory