Simplifying complexity -- new insights into how genomes work

April 1 2010

(PhysOrg.com) -- A genome is a complex system of genes and factors that regulate them. A European research team has clarified how such dynamic systems work, leading to a new way to predict genetic regulators.

As an organism develops and interacts with its environment, suites of genes are constantly being turned on and off, orchestrating every aspect of life. Researchers worldwide are trying to understand transcriptional networks, the intricate webs of genes and regulatory agents that control the growth and functioning of every organism from Escherichia coli to Homo sapiens.

The EU-funded research project GENNETEC (for GENetic NETworks Emergence and Complexity) set itself the ambitious goal of developing a deeper understanding of all complex systems and then applying those insights to living organisms, including humans.

Among GENNETEC's accomplishments is a new way to predict which transcription factors - molecules that turn genes on or off - regulate particular genes. Their findings promise to boost research into the functioning of genetic networks in general, and into the dynamics of the human genetic system in health and disease.

"We're now in a better position to understand genetic regulation in human cells, for a lower cost and in a shorter time," says François Képčs, coordinator of the GENNETEC project.

Much like a pianist fingering particular keys or chords to play a melody, transcription factors bind to particular sites along a chromosome to turn nearby genes on or off. Decades of research have shown that the resulting patterns of gene expression direct a cell or organism's development, normal functioning, and responses to environmental challenges.

In addition, malfunctions in the genetic system can cause various diseases including cancer. "A disease might sometimes be considered an improper change in the dynamics of a network of interactions," says Képcs. "So understanding their properties and how to correct or control their dynamics is essential."

Trolling for transcription factors

Until now, the most effective way researchers had to try to match genes and provisional transcription factors was to look for short DNA sequences that were known to bind to specific regulatory molecules. This approach remains useful, says Képcs, but produces many false positives - potential regulatory relationships that prove false.

Following up on those false leads is wasteful. "Doing it with a pipette takes a long time and costs a lot of money," says Képcs.

The GENNETEC team decided to address that problem by studying a new and independent way to predict whether a gene is controlled by a particular factor.

In earlier research, Képcs and his colleagues discovered that genes that respond to the same transcription factor are often spaced regularly along a chromosome. They suspected that this periodic spacing is related to the way that DNA coils up inside the nucleus of a cell, and serves to optimise the functioning of related genes and transcription factors by

grouping them geographically.

Scientists are always more comfortable if they understand the mechanism that produces an observed regularity. The GENNETEC researchers used sophisticated numerical simulations of DNA folding to prove that the presence of those periodically spaced genes helps determine the structure of the folded or condensed strand of DNA.

They also found that the final shape, which brings related genes close together physically, is important for gene expression.

“What we discovered is that there is a clear link between chromosome structure and gene expression,” says Képčs, “a link that we can now predict in a very precise and workable way.”

Faster, more focused search

When the GENNETEC team combined their new positional predictor with the standard sequence predictor, they found that they could identify new gene-regulator relationships far more efficiently.

“Combining the two predictors allows us to predict the regulators of a particular gene much better, by cutting down on the false hits,” says Képčs. “We typically double the specificity of the prediction.”

One of the consortium partners, NorayBio, based in northern Spain, is developing a commercial software package that will allow researchers worldwide to apply this more powerful approach to deciphering genetic networks.

The consortium is also making a functional, but less sophisticated, version of the software available for free.

While Képčs is pleased with this new research tool, he emphasises that the consortium's fundamental research on complex systems is equally important. Their findings can be applied in fields as diverse as designing software that does only what it's supposed to do and engineering systems that, like cells, can respond optimally to a wide variety of situations.

"Cells have just one genome, but with that one genome they can cope with multiple challenges," says Képčs. "We can use this biological solution as inspiration to make a new generation of algorithms to address complex problems better than before."

More information: GENNETEC project - gennetec.csregistry.org/

Provided by ICT Results