ENCODE project: Yale team finds order amidst the chaos within the human genome

September 5 2012

The massive Encyclopedia of DNA Elements (ENCODE) unveiled Sept. 5 reveals a human genome vastly more rich and complex than envisioned even a decade ago. In a key supporting paper published in the journal *Nature*, the lab of Yale's Mark Gerstein, the Albert L. Williams Professor of Biomedical Informatics, has found order amidst the seeming chaos of trillions of potential molecular interactions.

The scientists show it is not just the gene, but the network that makes the human genome dynamic.

"We now have a parts list of what makes us human," Gerstein said. "What we are doing is figuring out the wiring diagram of how it all works."

What Gerstein has found is a regulatory network that has properties similar to, say, the connections in a social network or the organizational chart of a Fortune 500 company. Using sophisticated mathematical modeling, his team traced the cascade of a half million molecular interactions triggered by 119 transcription factors—special genes that can simultaneously activate or silence thousands of genes. The model shows that these transcription factors are wired together in a hierarchical fashion, with some factors operating like top-level executives, and some as middle managers or shop foremen. Together they regulate the 20,000 or so genes in the human genome.

By necessity, this hierarchical structure creates information-flow

bottlenecks at the level of "middle managers," which Gerstein's team shows work together to more efficiently regulate target genes and ease the bottlenecks. This means that the human genome is organized much more democratically than say, the top-down command system of the military, Gerstein says.

However, the "executive-level" transcription factors do tend to have the most influence in key functions such as driving gene expression, and also have better connections with other genes in different molecular networks. Attesting to their importance to survival, these "executives" tend to be more conserved across populations.

Gerstein notes that both the size and flexibility of the human genome makes it different than many other organisms studied so far. Model organisms such as worms or flies have a simpler diagram—a switch-like promoter close to a gene is responsible for all its regulation. But the ENCODE project shows dramatically that there are hundreds of thousands of more distant elements, known as enhancers, that can influence human gene action from afar. Gerstein's team found that networks regulated by enhancers tend to be wired differently than those regulated by nearby promoters.

"This wiring diagram gives us framework to interprets the many variants of personal genomes that don't directly affect genes," Gerstein said.

Key Yale contributors to this research include Koon-Kiu Yan, Chao Cheng, Xinmeng Jasmine Mu, Ekta Khurana, Joel Rozowsky, Roger Alexander, and Sherman Weissman.

The work was funded by the National Human Genome Research Institute.

Within the genome, sex does matter

Yale researchers studying the human genome say they can now tell how much "mom" and how much "dad" is genetically active in each of us.

These gender-specific markers may not determine which parent can take credit—or the blame—for the successes or shortcomings of their offspring; however, they could help explain differences in human populations.

"We can now track the relative genetic contribution of mom and dad," said Gerstein.

All human beings are born with two copies of the genome—one from the mother and one from the father. However, sometimes only one of the copies, or alleles, ends up being biologically active for a particular gene. Based on an analysis of the massive amounts of data generated by the ENCODE project, Yale researchers observed this occurs 10 to 20 percent of the time. Researchers did not analyze the functions of these maternal and paternal specific genes and regulatory networks. However, they did note that these "gender-specific" networks tend to be evolving more rapidly than other networks.

"Perhaps, they account for the differences we see among individuals," Gerstein said.

Fossil DNA resurrected in contemporary human genome

Among the oddities turned up during the exploration of the human genome are pseudogenes—stretches of fossil DNA, evolutionary remnants of an active biological past. Yale researchers using sophisticated data mining and statistical models have discovered that many of these genes may not be quite dead after all, as they report in the

journal *Genome Biology*.

These ancient genes no longer code for proteins that carry out life's functions. However, the Yale team shows many of them are resurrected to produce non-coding RNAs, which scientists now know are crucial to the activation and silencing of protein-coding genes throughout the genome.

"This is another example of nature not wasting resources, a story we see repeated time and time again throughout the 3 billion letters of our genome," said Gerstein, senior author of the paper.

The existence of pseudogenes illustrates how human evolution may have worked. The pseudogenes have been inherited from functional ancestors but rendered obsolete via a variety of genetic mechanisms. This is an ongoing procedure, and some pseudogenes could have "died" relatively recently in human history, Gerstein's team found. However, at the same time, some pseudogenes may have been resurrected and harbored an ability to produce tiny RNAs, some of which may have regulatory activity in an advantageous way. As a result, they remain preserved in the genome, note the scientists.

Provided by Yale University