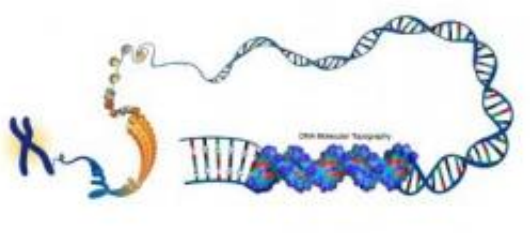
Researchers develop a structural approach to exploring DNA

March 12 2009



The many topographical levels of a chromosome. Genomic DNA is shown streaming out from a chromosome (left), progressively unfolding as chromatin, the 30-nm filament, nucleosomes, the DNA double helix, and finally the letters representing the nucleotide sequence. Although it is the molecular topography of the DNA helix that is recognized by proteins, current methods of genome analysis mostly focus on the order of nucleotides. Credit: NHGRI

A team led by researchers from Boston University and the National Institutes of Health has developed a new method for uncovering functional areas of the human genome by studying DNA's three-dimensional structure -- a topographical approach that extends the more familiar analysis of the sequence of the four-letter alphabet of the DNA bases.

Unlike the well-understood genomic sequences that code for proteins and comprise about two percent of the human genome, the remaining 98 percent is the non-coding portion, which encodes many functions.

However, little is known about how this functional non-coding information is specified.

In a study which appears today in the online edition of *Science,* the researchers focused on examining the non-coding regions of the genome for areas that are likely to play a key role in human biological function.

To do this, the researchers developed a method which incorporates information about the structure of DNA to compare sequences of genomes from humans and 36 mammalian species that included the mouse, chimpanzee, elephant and rabbit.

By examining the shapes, grooves, turns and bumps of the DNA that comprises the human genome, the team discovered that 12 percent of the human genome appears to be constrained by evolution. That's double the six percent detected by simply comparing the linear order of DNA nucleotides (A, T, G, and C, the familiar letters that make up the genome). The huge increase stems from finding some DNA sequences that differ in the order of nucleotides, but have very similar topographical shapes, and so may perform similar functions.

They went on to show that the topographically-informed constrained regions correlate with functional non-coding elements better than constrained regions identified by nucleotide sequence alone.

"By considering the three-dimensional structure of DNA, you can better explain the biology of the genome," said Thomas D. Tullius, Boston University professor of chemistry who has spent more than 20 years developing ways to map the structure of the human genome. "For this achievement Stephen Parker, a Boston University graduate student, deserves much of the credit for his development of the algorithm that incorporated DNA structure into evolutionary analysis."

Bringing a molecular biologist's point of view and expertise in comparing the genomes of different species was Elliott Margulies, an investigator at NHGRI's Genomic Technology Branch. "Proteins that influence biological function by binding to DNA recognize more than just the sequence of bases," he said. "These binding proteins also see the surface of the DNA molecule and are looking for a shape that allows a lock-and-key fit."

In their Science paper the researchers also explored how small genetic changes, or variations, known as SNPs (Single Nucleotide Polymorphisms) could prompt structural changes that might lead to disease. In studying these mutations from a database of 734 non-coding SNPs associated with diseases, such as cystic fibrosis, Alzheimer's disease, and heart disease, they found that disease-associated SNPs produced larger changes in the shape of DNA than SNPs not associated with a disease.

The new research findings on evolutionary conservation of DNA structure stem from recent progress in analyzing the functional elements in a representative fraction of the human genome. That study, known as ENCODE (ENCyclopedia of DNA Elements), organized by the National Human Genome Research Institute (NHGRI), challenged the traditional view of the human genetic blueprint as a collection of independent genes. Instead, researchers found a complex network of genes, regulatory elements, and other DNA sequences that do not code for proteins.

The study determined, for the first time, where many types of functional elements are located, how they are organized, and how the genome is pervasively made into RNA. The current research on genome structure and function is based on some of the ENCODE findings, noted Tullius, whose work in developing the new technology was funded through the ENCODE project.

Source: Boston University