

IBM Discovery Could Shed New Light on Workings of the Human Genome

25 April 2006

```
>3'UTR OF ENST00000245836|ENSG0000010196
-(162)-TTTCTTGATTTTTTCAG-(104)-ATACTGATTTA
GTATT-(146)-TGGATGAAAAATATTT-(591)-CTCAC
CTGCTCACTGCA-(6)-GCCTTCTGGGTTCAAG-(2)-AT
TCCTGACCTCAAG-(19)-TCCCAAAGTCTGGGATTA
ATCCAGCACTTTGGGA-(14)-GATCACGAGGTCAGG
ACT-(5)-GGCTGAGGCAGGAGAAT-GGTGTGAACCCG
AAAAAT-(69)-ACTCCCATCCTAATAC-(137)-CGGTGC
CTGGGCGTGGTG-(8)-CTGTAATCCCAGCTACT-(5)-
AAGAGCAAAACT-(58)-TTAGCCAGGCGTGGTG-(16)
CTCCAGTCTGGG-(2)-ACAGAGTGAGACCCCA-(50)-
TAACAACATTTTACAT-(37)-AGCAATTATTTTTAAA-(
G-(9)-TGTAGTCTGGCTACT-(15)-GGAGGATCGCT
TACT-(23)-CAGTCTCACTGTGTG-(4)-GGATGGAGT
GA-(15)-TTGGCCAGGCTAGTCT-(2)-AATTTCTGACC
GGTGAT-(8)-CTCGGCTCACAAAGTG-CTGGGATTA
T-(2)-AGACCAGCCTGGCCAA-(11)-CCTGTCTGTACA
GC-(1)-GAGGTTGCAGGGAGCC-(13)-CGCACTCCAG
GTGAAT-(10)-CAATAAAATACTATTC-(10)
```

This is an example of several tens of "pyknons" that are shared in various arrangements and combinations by the sequences of the 3' untranslated regions (3' UTRs) of 10 distinct human transcripts. These pyknons come from a set of 127,998 pyknons that are present in the intergenic/intronic regions of the human genome and have additional instances in the untranslated and protein-coding regions of 30,675 transcripts from 20,059 human genes. A 3' UTR is part of a gene's transcript that is not translated into an amino acid sequence. Source: The Journal Proceedings of the National Academy of Sciences. Credit: IBM

IBM today announced its researchers have discovered numerous DNA patterns shared by areas of the human genome that were thought to have little or no influence on its function and those areas that do.

As reported today in the *Proceedings of the National Academy of Sciences* (PNAS), regions of the human genome that were assumed to largely contain evolutionary leftovers (called "junk DNA") may actually hold significant clues that can add to scientists' understanding of cellular processes. IBM researchers have discovered that these regions contain numerous, short DNA "motifs," or repeating sequence fragments, which also are present in the parts of the genome that give rise to proteins.

If verified experimentally, the discovery suggests a potential connection between these coding and non-coding parts of the human genome that could have a profound impact on genomic research and provide important insights on the workings of cells.

"Our goal is to apply advanced computational techniques to analyze the workings of processes and systems, in this case the function of the human genome," said Ajay Royyuru, head of the Computational Biology Center at IBM Research. "Using these tools, we've been able to shed new light on parts of the DNA that were traditionally thought of as not having a specific purpose. We believe the innovative application of technology can provide further understanding in the life sciences at large."

The IBM team used a mathematical tool called pattern-discovery, often applied to mine useful information from very large repositories of data in both business and scientific applications, to sift through the approximately six billion letters in the non-coding regions of the human genome and look for repeating sequence fragments, or motifs.

Among the millions of discovered motifs, the team identified approximately 128,000 that also occur in the coding region of the genome and are significantly over-represented in genes involved in specific biological processes such as cell communication, regulation of transcription, transport and others. In fact, copies of one or more of these motifs can be found in over 90 percent of all known human gene sequences, as well as some genes of other animals where they associate with similar biological processes.

The report on this work "Short blocks from the non-coding parts of the human genome have instances within nearly all known genes and relate to biological processes" by Isidore Rigoutsos, Tien Huynh, Kevin Miranda, Aristotelis Tsirigos, Alice McHardy and Daniel Platt of IBM's T. J. Watson

Research Center, Yorktown Heights, NY appeared on April 24th in the early edition of the journal PNAS.

Source: IBM

APA citation: IBM Discovery Could Shed New Light on Workings of the Human Genome (2006, April 25) retrieved 18 June 2021 from <https://phys.org/news/2006-04-ibm-discovery-human-genome.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.