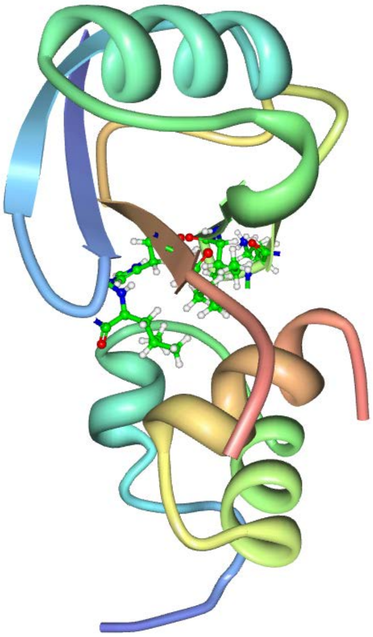
TECHNOLOGY SUMMARY



Reference
8810-7389

Inventor(s)
Andrew K. Wong
Ho Yin (Antonio) Sze-To
Dennis Zhuang
En-Shiun Annie Lee

Patent status
U.S. Provisional patent application

Stage of development
Working prototype and validating application data available.

Contact
Scott Inwood
Director of Commercialization
Waterloo Commercialization Office
519-888-4567, ext. 33728
sinwood@uwaterloo.ca
uwaterloo.ca/research

Biosequence Gap Pattern Discovery and Knowledge extraction and confirmation system

Background

With only approximately 1-2% of the genome accounting for coding regions, the other 98% consists of non-coding regions. While 80% of the non-coding region is suspected of having regulatory function, there is still no clear functionality to the non-coding regions. Being able to identify and determine the purpose and functionality of non-coding areas can aid in the discovery of regulatory functions for the transcription and translation of various genes or may provide insight to diseases that don't have evidence of functional coding variants.

Description of the invention

In bioinformatics analysis, the identification of gaps is important in sequence matching as they could represent a mutation that could offer insight to the discovery of new functionality. Waterloo researchers have developed a novel algorithm that discovers highly conserved, non-redundant, and statistically significant patterns, including complementary foldable patterns, which can then be used to find other pattern clusters.

The effectiveness and efficiency of Gap Pattern Discovery relies on the development of novel pattern redundancy concepts to generate a large number of redundant flexible gap patterns. Identified gap patterns are then converted into consensus patterns, which serve as inputs within an integrated software framework that automates the search of existing databases and literature and extracts validation information related to the the identity and functionality of these gap pattern elements identification.

Advantages

- Doesn't require any prior knowledge of the sequence to identify highly conserved non-redundant patterns and co-occurrence patterns in coding/non-coding DNA; unlike other methods such as Minimotif Miner and SLimDisc
- Can determine interacting and binding mechanisms of potential regulatory clusters, and Protein-DNA binding sites
- Faster than other methods such as MEME, Gibbs and qPMS7 because of its linear time and space algorithm

Potential applications

- Develop drugs and treatments for diseases that have unknown regulatory functions
- Discover previously unknown gene regulatory functions
- Analyze large scale genomic, transcriptomic, and proteomic data