Deciphering the human genome

Martin Frith
Research Scientist
Sequence Analysis Team

The ultimate aim of my research is to understand the genetic codes for making humans and other organisms. Therefore, I am especially interested in deciphering the DNA codes that control gene expression in embryonic development. Since these codes often reside in regions of the genome that are highly conserved between species (e.g. humans and fish), I am working with Naoki Osato to find over-represented motifs in these regions. Encouragingly, some of the motifs that we are finding are known targets of developmental transcription factors, including Hox factors.

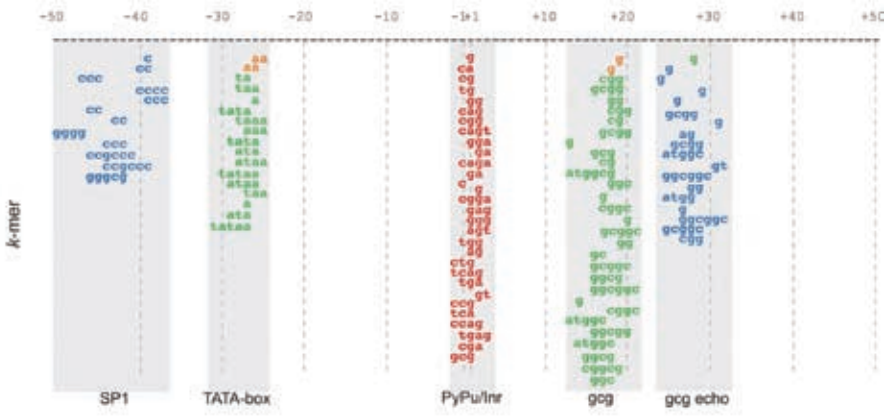
I am also studying the organisation of transcription initiation, which is the first step in gene expression. Using large-scale transcript start sequence data, I was able to show that initiation events are organised in layers of clusters within clusters on the chromosomes. Moreover, I discovered a DNA code that explains the precise selection of transcription start sites within small clusters. Now, I am trying to understand which cluster layers are regulated when gene expression changes.

Finally, I am developing new methods for sequence comparison and alignment. Such methods are fundamental for much of computational biology (including the projects mentioned above), but existing software, such as BLAST, has difficulty with the large-scale datasets produced by new DNA sequencing technologies. Working with Paul Horton, I have developed new software, called LAST, which can compare two human-genome-sized datasets on a standard desktop computer in about one day. Key to LAST's efficiency is an interesting data structure called a suffix array. I am also working on an important but overlooked aspect of sequence alignment: how to detect and avoid inaccurate parts of alignments.

In summary, my research consists of developing and applying methods to decipher the biological information in genome sequences.

Reference

1. http://www.cbrc.jp/~martin/



A code for transcription initiation