

Poster B-61

Genome-wide Comparison of DNA Melting Bubbles and Ensembl's Genomic Annotations



Authors:

Geir Ivar Jerstad (*Dept. of Informatics, University of Oslo*)

Eivind Tøstesen (*Dept. of Tumor Biology, Norwegian Radium Hospital*)

Torbjørn Rognes (*Dept. of Informatics, University of Oslo*)

Eivind Hovig (*Dept. of Tumor Biology, Norwegian Radium Hospital*)

Short Abstract: The statistical occurrence of DNA melting bubbles can be represented using stitch profile diagrams. By computing genomic stitch profiles, a direct comparison between physical bubbles and genomic annotations is facilitated. Visualisation of the overlap has been achieved by the integration of a stitch profile view into the Ensembl genome browser.

Long Abstract:

During the last 5-10 years, an expansion of the concept of genomes has been emerging. While genomic DNA has mainly been considered from the base sequence point of view, the importance of considering also the 3D folded structures of chromatin at different length scales, as well as the epigenetic features, have become more appreciated recently when realizing that sequence does not, on its own, fully determine heredity and genomic function. Correspondingly, one can distinguish the analog genome from the digital genome in terms of the information content: The base sequence constitutes the 1D digital genome, while the 3D analog genome is represented by continuous physical quantities, such as spatial position inside the nucleus, polymer conformations, densities and gradients, etc. However, most genome browsers still present a 1D picture.

One aspect of the analog genome are the DNA bubbles that appear at a relatively small length scale. The ability of the two DNA strands to open up and form local bubbles that can later hybridize reversibly, is fundamental in transcription and replication as well as in most DNA-based technologies. Bubbles have mainly been studied in terms of thermally induced DNA melting in vitro and using calculated melting profiles.

We present a novel computational approach to DNA melting called stitch profiles [1-3], in which the statistical occurrence of bubbles is represented in terms of the corresponding alternative segmentations of the sequence. A stitch profile diagram shows a set of possibly overlapping segments ('stitches'), each indicating either an open or closed conformation of the double helix. Thus, a stitch profile represent the physical properties in an annotation style, which is closer in form to genomic annotations than the usual melting profile graphs. By computing genomic stitch profiles, a direct comparison between physical bubbles and biological annotations is facilitated, and it enables more detailed questions to be asked about the overlap and correlations.

Two bioinformatical strategies are employed here: Visualisation and large-scale statistical analysis. Visualisation has been achieved by the integration of stitch profile diagrams into a

local installation of the Ensembl genome browser in a modified version [4]. The idea is to add a window to the existing views that show a stitch profile for the sequence region being browsed. A visual exploration shows different types of correlations between stitch profiles and coding/non-coding regions in both the human genome and in *Saccharomyces cerevisiae*. Another window has also been added to show the melting map (T_m profile) of the human genome. Statistical analysis is performed to quantify the correlations and to test hypotheses regarding their possible origins. What roles do bubbles play in the biological processes of the genome?

[1] E. Tøstesen, G. I. Jerstad and E. Hovig (2005): Stitchprofiles.uio.no: Analysis of partly melted DNA conformations using stitch profiles. Nucl. Acids Res., 33, w573-w576.

[2] E. Tøstesen (2005): Partly melted DNA conformations obtained with a probability peak finding method. Phys. Rev. E, 71, 061922.

[3] [Http://stitchprofiles.uio.no](http://stitchprofiles.uio.no).

[4] G. I. Jerstad (2006): Merging the physical properties of DNA with genomic annotations in Ensembl. Master's thesis, University of Oslo.