Poster H-70

Computational identification of structural variation in genomes

Authors:
David Fredman (*Bergen Center for Computational Science*)
Boris Lenhard (*Bergen Center for Computational Science*)

Short Abstract: We present a new approach to explore structural variation (duplication, deletion, inversion) in annotated genomes. Our algorithm classifies sequence as prone to structural variation based on sequence attributes. We provide a detailed view of genomic features associated with structural variation and novel sites for targeted identification in the laboratory.

Long Abstract:

Copy number variation of sequence fragments has recently been discovered as a new and major form of polymorphism in the human genome. Initial studies (1-3) show that a substantial portion of the genome is affected, that these regions often harbor genes involved in immune response and biotic stimuli(4), and that they are subjected to positive selection (5). Additional sequence copies that are not reflected in the genome assembly also complicates the use of SNPs and microsatellites in ongoing association studies.

The goal of our study is to discriminate between genome regions that are prone to amplification, deletion and inversion in any given genome, and those that are not. To this end, we develop a bioinformatics scoring and prediction method that utilizes a large set of attributes based on single nucleotide polymorphism, LD, repeats, conservation, genes and sequence and other features. In the process, we identify those features that are associated with such regional plasticity.

We have applied this method to sets of known structural variants and control regions across the entire human genome. Preliminary results show that sequence heterozygosity, conservation and SNP profiles are all correlated with structural variation. Identification of novel sites of frequent copy number variation for further testing in the laboratory is under way.

Our findings will aid in targeting genomic regions for high resolution typing of structural variation, identifying possible subfamilies of structural variance, surveying new genomes for their structural variation propensity and generate information on regions where genotyping is hampered by excessive variation. A detailed view of genomic features associated with structural variation will also help in generating hypotheses concerning possible underlying mechanisms.

1. Fredman, D. et al. Complex SNP-related sequence variation in segmental genome duplications. Nature Genet. 36, 861–866 (2004).
2. Sebat, J. et al. Large-scale copy number polymorphism in the human genome. Science 305, 525–528 (2004).

3.Iafrate, A. J. et al. Detection of large-scale variation in the human genome. Nature Genet. 36, 949–951 (2004).
4.Feuk, L. et al. Structural variation in the human genome. Nat Rev Genet. 7(2), 85-97 (2006).
5.Nguyen DQ. et al. Bias of Selection on Human Copy-Number Variants. PLoS Genet 2(2): e20 (2006)