J Mol Genet Med 2017, 11:2 (Suppl)
DOI: 10.4172/1747-0862-C1-016

GeneHancer, a comprehensive human genomic enhancer annotation database: Key to interpreting intergenic regions in whole genome sequences

Doron Lancet, Simon Fishilevich, Yaron Golan, Dvir Dahari, Marilyn Safran, Ron Nudel and **Dana Cohen**
Weizmann Institute of Science, Israel

This year, the $1,000 whole genome target became a reality, leading to an avalanche of whole genome sequencing (WGS) projects worldwide. Yet, deciphering the ensuing data poses a major challenge, necessitating significant improvements in the annotation and interpretation of intergenic variants. Enhancers are remote regulators of gene expression, whose aberrations underlie certain diseases. They are by far the most abundant functional entities in non-coding DNA, hence are of great significance to WGS analyses. We present GeneHancer, a novel database of human enhancers and their inferred target genes, in the framework of GeneCards. GeneHancer computes combinatorial likelihood scores for enhancer-gene pairs, and displays charts of probable enhancers for every human gene. It thus makes it possible to map variants to enhancers, and forms a basis for their gene-phenotype interpretation. These capacities will soon be incorporated into GeneCards suite's next generation sequencing tools, TGex and VarElect. GeneHancer integrates information on ~258,000 enhancer entries from three genome-wide enhancer databases. The integration includes: 213,000 elements from the Ensembl regulatory build, predicted based on the ENCODE and Roadmap Epigenomics projects; 43,000 elements from FANTOM, identified via enhancer RNA (eRNA) transcriptomic and 1,700 elements from the VISTA Enhancer Browser validated by transgenic mouse assays. GeneHancer thus portrays 237,000 integrated enhancers, with ~17,500 derived from more than one source. Additional scheduled integration includes DENdb predicted enhancers in cell-lines. GeneHancer subsequently links enhancers to genes, using the following methods: Employing FANTOM expression correlation between eRNAs and candidate target genes; Making use of GTEx expression quantitative trait loci (eQTLs), exploring genetic association between variants within enhancers and the expression of candidate target genes; Seeking across-tissue expression correlation (PMID 27048349) between a transcription factor interacting with enhancers, and candidate target genes and using Hi-C chromosome conformation capture data as indications of genomic looping between a gene promoter and enhancer. The individual scores based on the above four methods, along with gene-enhancer genomic distances, form the basis for GeneHancer's combinatorial likelihood scores for enhancer-gene pairs.

dlancet@yahoo.com