GREAT.STANFORD.EDU: IMPROVED FUNCTIONAL INTERPRETATION OF CIS-REGULATORY REGIONS

D. M. Bristor* and C. Y. McLean

Department of Computer Science, Stanford University, 321 Beckman Center
Stanford, CA 94305, USA
**Email: bristor@stanford.edu*

Many tools have been developed over recent years to measure gene enrichments. Given a subset of genes, for example from the result of a microarray analysis, these tools assess the surprise (p-value) associated with any given gene annotation from a fixed vocabulary such as the Gene Ontology (GO). We present GREAT (Genomic Regions Enrichment of Annotations Tool), the first computational tool which correctly analyzes whole genome cis-regulatory data. GREAT takes as input not a set of genes, but a set of arbitrary genomic regions. These regions are assumed to be associated with the cis-regulation of one or more nearby genes. They can be the result of a functional assay such as ChIP-seq, or an evolutionary survey such as ultraconservation. Our tool associates that input set of genomic regions with nearby genes. It performs several annotation enrichment tests. When assessing statistical significance, these tests properly account for biases in the sizes of the regulatory domains associated with each gene, and for clustering of genomic regions near individual genes. GREAT analyzes enrichment of input regions against 20 ontologies including phenotype data, biological pathways, disease, transcription factor motifs, and gene expression data. Correspondingly, GREAT provides a method to detect likely functions of a set of unannotated genomic regions, and is a unique tool for the study of cis-associated genomic regions. GREAT is available as a web tool at http://great.stanford.edu. Data can be also be submitted directly from the UCSC Genome Browser.

---

* Corresponding author.