Overview

When is GREAT (1) useful?

GREAT infers biological meaning for a set of known or assumed cis-acting non-coding genomic regions by analyzing the annotations of the nearby genes. Many experimental and computational screens produce sets of interest for GREAT.

One natural application of GREAT is to analyze data from a chromatin immunoprecipitation (ChIP) experiment with a transcription factor of interest. To hypothesize the processes that involve a given transcription factor:

1. Identify transcription factor binding sites via ChIP-seq
2. Use GREAT to find annotations enriched among the genes near the binding sites.
3. Hypothesize that the transcription factor helps regulate the processes whose annotations are highly enriched.

Why should I use GREAT instead of other annotation tools such as DAVID (3) or GO::TermFinder (4)?

Other annotation enrichment tools are gene based. The test set consists of a list of genes, and the tools provide annotations more common in the test set than in a background set of genes. This does not accurately model test sets of genomic regions because gene-based tests do not account for biases in the assignment of genomic regions to genes. Genes in gene deserts have larger regulatory domains than genes spaced closely to each other. In other words, a random genomic region is more likely to be assigned to a gene in a gene desert simply because deserts provide large regions where the gene is the nearest one. GREAT more accurately models this situation. Thus, it more accurately calculates enrichments for a set of genomic regions.

GREAT also includes numerous ontologies providing a range of annotations. Many other tools use only the Gene Ontology, but it is useful to consider other types of annotation, such as protein domains and pathways.

References


