

File Size

File size limits

GREAT currently supports test files with up to 500,000 test regions and background files with up to 1,000,000 regions. Each must be less than 50 MB in size. Compressed data must decompress to plain text files at most 50 MB in size.

GREAT version 1.2 supported at most 200,000 test regions and did not support compressed data as input.

Handling large data sets

By default GREAT displays data in a ["Significant by Both" View](#) that only shows terms significant by both the binomial test over genomic regions and the hypergeometric test over genes. Large data sets can cause a large fraction of all genes to be selected via the regulatory domain association rules. This often results in saturation of the hypergeometric test over genes such that no hypergeometric test results are significant. The binomial test over genomic regions is robust to large data sets, however.

There are two ways to circumvent the saturation of the hypergeometric test:

1. Restricting the input set to a few thousand regions (i.e. by picking the most robust peaks generated by a peak-calling tool) eliminates saturation of the hypergeometric test.
2. Alternatively, results for large data sets can be viewed in the ["Significant by Region-based Binomial"](#) or ["Full" View](#) and terms enriched due to many regions clustered around one or few genes can be filtered by using the observed gene hits display filter.

This warning is not applicable when the foreground/background test is used since the [hypergeometric test is over genomic regions rather than genes](#).