Web-Based GeneChip Analysis System

Manhong Dai, Pinglang Wang, Elvis Jakupovic, Stanley J. Watson and Fan Meng
Molecular and Behavioral Neuroscience Institute and Psychiatry Department
University of Michigan, Ann Arbor, MI 48109

Affymetrix GeneChip for expression analysis is probably the most widely adopted microarray platform for large-scale gene expression profiling projects due to their high manufacturing quality and relatively simple assay procedure. However, there is still no off-the-shelf solution that provides necessary capabilities for dealing with data security, data sharing and analysis procedure sharing for GeneChip data generated at different geographical locations in collaborative projects. Most commercial products only allow data sharing within intranet. Adding third-party analysis algorithms into existing software often requires major programming efforts. Furthermore, large expression profiling projects as well as the meta analysis of GeneChip data from different sources often require the processing of hundreds or thousands GeneChip CEL file together, which is not possible for typical research laboratories without heavy investment in high performance computing hardware and their maintenance.

We develop a Web-based GeneChip Analysis System (WGAS) that provides centralized data analysis and management for large-scale GeneChip data analysis. Our system is based on the BioConductor platform thus it enables easy integration of third-party analysis procedures. WGAS uses Tomcat as web server. Data analysis is performed at a centralized server cluster thus each research group can conduct large-scale analysis efficiently without the need to setup and manage server cluster locally. We also created a distributed network that allows encrypted data transfer across different geographic locations and within intranets in a highly secured manner

All sample descriptions and analysis parameters are tracked in the backend Oracle database. GeneChip CEL files and analysis results are stored in a Linux file server. WGAS uses a LDAP server for flexible user and group authentication. It accepts job submissions from multiple users simultaneously and if all CPUs in our clusters are busy, a new analysis job will be queued and the submitter will receive an e-mail when the corresponding analysis is finished. WGAS provides a Python client package to allow user upload GeneChip CEL files and sample description information under their own private accounts. Analysis results and parameters can be retrieved through the WGAS web interface.

The Web-based GeneChip Analysis System (WGAS) incorporates popular probe-, gene- and function category-level analysis methods through straightforward web interface. Some of the most useful functions are:

- Natively supports custom GeneChip probe set definition files for more accurate GeneChip data interpretation (Dai et al, Nucleic Acids Res. 33, e175).
- Has a unique ProbeFilter function for removing allele-specific probes or non-performing probes meeting users' criteria.
- Performing simultaneous alternative splicing analysis using exon-specific and gene-specific probe sets derived from the ENSEMBL database.
- Include optimized C implementations of the Gene Set Enrichment Analysis (Subramanian et al, PNAS, 102, 15545-50) and SigPathway (Tian et al, PNAS,102, 13544-9) methods to facilitate the detection of subtle coordinated expression change of functionally related genes.
- Provide extensive gene function annotation gene set-based analyses. Currently our function annotation covers more than 10,000 concepts from Gene Ontology, Cytoband, KEGG, Biocarta, and GenMapp. Users also have the flexibility to upload their own function annotations.
- Users can utilize any CEL files deposited in NCBI GEO through the WGAS web interface for re-analysis or meta analysis. WGAS retrieves NCBI GEO's CEL files every night and check if any existing celfiles are re-uploaded every week.