GO enrichment of microarray data from maize root under drough

Roberto Wiilians Noda, Embrapa Milho e Sorgo, Brazil
Roberto Noda, Brazil; Daiane Mariele de Laat, Instituto Agronômico de Campinas, Brazil; Claudia Teixeira Guimarães, Embrapa Milho e Sorgo, Brazil; Thiago Corrêa de Souza, Universidade Federal de Alfenas, Brazil; Leonardo Lucas Carnevalli Dias, Universidade Federal de São João Del-Rei, Brazil ; Francisco Pereira Lobo, Embrapa Informática Agropecuária, Brazil; Jurandir Vieira de Magalhães, Embrapa Milho e Sorgo, Brazil; Roney Santos Coimbra, Fundação Oswaldo Cruz, Brazil; Paulo César Magalhães, Embrapa Milho e Sorgo, Brazil; Adhemar Zerlotini, Embrapa Informatica Agropecuaria, Brazil; Glória Regina Franco, Universidade Federal de Minas Gerais, Brazil; Newton Portilho Carneiro, Embrapa Milho e Sorgo, Brazil

Short Abstract:
Water-deficit is one of the most common environment stresses, which affect plants productivity and its geographic distribution. Plant mechanisms for water-deficit stress adaptation and response are often characterized by specifics genes activation. In this study microarray data from roots of a drought tolerant inbred maize line (Embrapa's breeding program) under two water regimes was analyzed. The analysis revealed that 746 genes were differentially expressed, of which 455 were up- and 291 were down-regulated. Twenty six genes differentially expressed were successfully amplified and evaluated using quantitative PCR (qPCR). These genes showed a moderate positive correlation (r = 0.39) between microarray (logFC) and qPCR results (2-??CT). No genes were found to display a divergent expression pattern in microarray and qPCR experiments, whereas some genes showed expression values very similar in both techniques. Using Gene Ontology, 353 differentially expressed genes were categorized within biological processes and the most representative categories were sorted to stress and carbohydrate metabolic process. GO enrichment analysis was performed comparing the GO of the differentially expressed genes found in this study with the GOs across all genes represented in the array. In Biological Process domain, three terminal categories significantly overrepresented: "response to stress", "response to abiotic stimulus" and "transport". For Cellular Component domain, the category "vacuole" was detected as GO differentially represented. This network information created valuable input for selection of candidates gene related to drought tolerance in maize under water stress which will be quite important to implement into the Embrapa's breeding program.

Top

Poster F54

wavClusteR: an R package for PAR-CLIP data analysis and substitution-based inference problems in genomics

Federico Comoglio, ETH Zurich, Switzerland
Cem Sievers, ETH Zurich, Switzerland; Tommy Schlumpf, ETH Zurich, Switzerland; Ritwick Sawarkar, ETH Zurich, Switzerland; Renato Paro, ETH Zurich, Switzerland

Short Abstract:
The Photo-Activatable Ribonucleoside-enhanced CrossLinking and ImmunoPrecipitation (PAR-CLIP) is a recently developed method for global identification of RNAs interacting with proteins. A strength of PAR-CLIP is the induction of specific T to C transitions at sites of protein-RNA interaction. However, current analytical tools do not distinguish between non-experimentally and experimentally induced transitions. In addition, geometric properties at potential binding sites are not taken into account. To fill this gap, we developed a two-step algorithm consisting of a non-parametric two-component mixture model and a wavelet-based peak calling procedure. Our algorithm can reduce the number of false positives up to 24% thereby identifying high confidence interaction sites. We provide an implementation of our algorithm in the R package wavClusteR. wavClusteR was successfully employed in conjunction with a modified PAR-CLIP protocol to study the functional role of nuclear Moloney leukemia virus 10 (MOV10), a putative RNA helicase interacting with Argonaute2 and Polycomb. In this poster, we present our method and discuss the general applicability of wavClusteR to other substitution-based inference problems in genomics.

Top

Poster F55

ARepA: automated repository acquisition for standardized high-throughput data retrieval, normalization, and analysis

Daniela Boernigen, Harvard School of Public Health, Harvard University, United States
Yo Sup Moon, Harvard School of Public Health, United States; Levi Waldron, Harvard School of Public Health, United States; Eric Franzosa, Harvard School of Public Health, United States; Curtis Huttenhower, Harvard School of Public Health, United States

Short Abstract:
Biological databases of high-throughput experimental results provide vast and growing resources for medical, and bioinformatic research. Open questions remain in how best to maintain such resources, access them computationally, meta-analyze their contents from hundreds of experiments, and do so reproducibly while maintaining computational best practices. We present ARepA, an extensible, modular Automated Repository Acquisition system for reproducible biological data acquisition and processing. ARepA allows configurable data access for any organism(s) from the GEO, IntAct, BioGRID, RegulonDB, STRING, Bacteriome, and MPIDB databases. A user can retrieve raw data and metadata from these repositories, normalize data files, and automatically process them in standardized ways (e.g. for network analysis). When retrieving data from six model organisms, ARepA currently produces more than 2M interactions (600K physical interactions, 4K regulatory interactions, 1.5M functional associations) and 2.7K gene expression data sets covering approx. 800K samples, accompanied by corresponding metadata and derived network data. We include biological examples demonstrating the utility of ARepA for integrative analyses. When focusing on human data, ARepA's metadata database allowed us to identify and standardize 12 human prostate cancer gene expression datasets from GEO, which were subsequently meta-analyzed across six different platforms. A subsequent co-expression network analysis correctly recovered the Nf?B signaling pathway along with new candidate genes with roles in prostate cancer. A similar example in mouse integrates 11 gene expression datasets selected by querying ARepA for metadata indicating germ-free and intestinal tissue conditions. Finally, multiple data types from three model microbes were integrated to assess differences in peptide secretion systems.

Top

Poster F56

Coupling read-based ChIP-seq analysis with expression data enriches for TF direct targets

Jonathan Cairns, University of Cambridge, United Kingdom
Shamith Samarajiwa, University of Cambridge, United Kingdom; Andy Lynch, University of Cambridge, United Kingdom; Simon Tavaré, University of Cambridge, United Kingdom

Short Abstract:
Understanding complex diseases, such as cancer, requires an explanation of the intricate regulatory networks working within a cell. A key part of this puzzle is to find genes that are direct targets of Transcription Factors (TFs). The assay ChIP-seq (Chromatin ImmunoPrecipitation, followed by high-throughput sequencing) detects TF binding events. The ENCODE consortium define all such events as ``functional", a definition that has caused much controversy (Graur et al, 2013) since,