Information Systems in the Life Sciences

PERP GeneChip data warehouse;

Implementation of a dynamic time series
query tool with graphical interface

by Srećko Sužnjević

Overview

- General
- PERP GeneChip data warehouse
- Affymetrix related information
- Focus on one of the tools:
  Single Gene Query Tool (SQLT)

General

DNA databases have relatively low dimensionality Genome is a linear code	RNA expression and protein databases need to be able to handle very high dimensional data Time, tissue, cell type and genes are interrelated variables

---

RNA & Microarrays

- RNA can be analyzed by Microarray expressions.
- Microarray expression have a high dimensionality of profile data.
- There is currently with no standard experimental platform.
- This complicated the development of web-accessible databases and analytical tools needed for RNA research.

Public Expression Profiling Resource

- PEPR == Public Expression Profiling Resource
- PEPR provides centralized Affymetrix expression profiling data to the public research community
- There is no need for the public resaercher to have their own Affymetrix software to use PEPR datbase
- PERP claims to be the largest public resource having quality control and standard operating procedures (QC/SOP).
- All future implementations will include these QC/SOP criteria.
- Currently there are 1426 profiles available to the public

Tools from PEPR

- Oracle based web solutions:
  - Single Gene Query Tool
  - PEPR project/file search.

    Goal: To query PEPR for all profiles matching the desired queries.
  - Functional Clusters

    Goal: To provide rapid download of data within a "project"
  - Export of Microarray profile data
- Back end solutions:
  - Direct deposit of data to NBCI GEO database

    Goal: To provide automated link to move profile data to GEO

Affymetrix

- Commercial firm for delivering GeneChips
- Products of Affymetrix : GeneChip Arrays, Assays & Reagents, Instruments and Software
- Microarray Suite (MAS 5.0)

How Does It Work

The GeneChip® Expression Assay
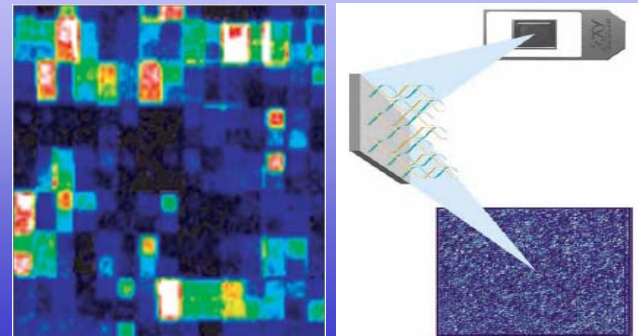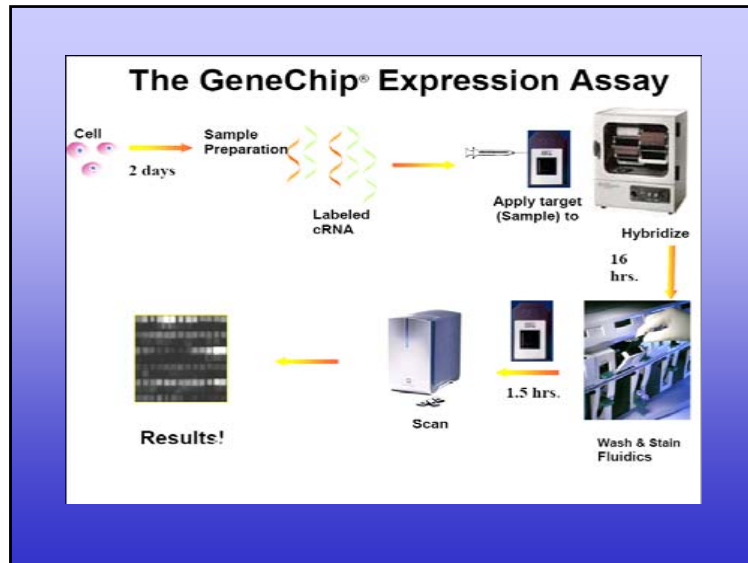
Single Gene Query Tool

- SGQT == Single Gene Query Tool
- Currently implemented for time series only
- Goals is to provide:
  - dynamic queries
  - any gene in the genome
  - graphical visualizations
  - spreadsheet downloads

How does the public SGQT work functionally

- Simple project query interface.
- Gene search function supporting ambiguous queries
- Graphical visualization showing all replicates
- Average generated on the fly
- Mouse-over information for each replicate and average
- Access to additional database resources with a mouse click
- Able to download all data in the graph for Spreadsheet use

Access

- Trough URL or using the menu from the main web site "Data" "Search" "Single Gene Query Tool"
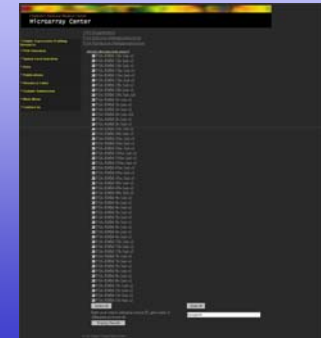- User selects species and "time series"

Project Selection

- User selects the list of implemented projects that are shown

Profile Selection

- User select the expanded list of profiles of the selected project
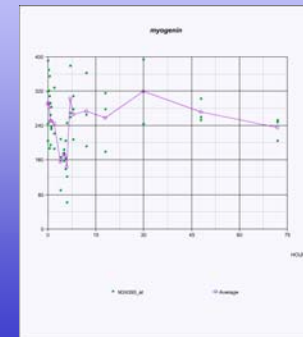- User types in the requested Gene (search function supports ambiguous queries)

Gene Selection

- User select the required gene from the resulting drop-down menu
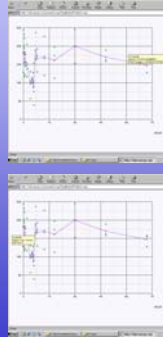

Graphical visualization

- Showing all replicates, with average generated on the fly
- Able to download for all data in the graph for use with a spreadsheet

- Mouse-over for each replicate and average, showing data associated with that data point

Additional resources

- Clickable data points for access to additional database resources for the gene under study

  - Gene Name: myogenin
  - Unigene: Rn.9465
  - Genbank: M24393
  - LocusLink: 29148
  - Affymetrix: M24393_at
  - Download the original dataset: PGA-RMM-Ctr-4aA-s2

Conclusion

- User needs skills in microarray experiments to use all the available tooling
- SGQT is usable without experience in interpretation of microarray data
- Some of the tools are in the background
- PERP is trying to have reliable project/data using QC/SOP available for the public

Questions?

The End.

References

Websites used:
- www.affymetrix.com
- microarray.cnmcresearch.org

Additional documentation used:
- pepr2004nar.pdf

Images used from Affymetrix (Must be used according to the licensee agreement):
- an_actual_array_image.pdf
- chip_probes_output.pdf
- gene_expression_assay.pdf