

5th Annual Joint Bioinformatics Workshop

July 19, 2005
2229 Seamans Center
University of Iowa

by

Bartley Brown
University of Iowa

Microarray Probe Selection and Probe Set Annotation

ABSTRACT

The current large-scale sequence-based gene discovery program at the University of Iowa provides a unique opportunity for microarray expression profiling.

In order to select cDNA probes to be printed on microarray slides one must identify a reasonably sized subset of clones from a much larger initial set and from a variety of sources. In this selection process the annotation of clones is a sine qua non. The type and extent of annotation required is different for each potential probe set. The annotation may include “primary” information concerning sequencing such as library, tissue, library tag, cluster size, rearray verification, polyA signal, double banding. So-called “secondary” annotation can also be important in the probe selection process. This includes Unigene cluster, accession number, go-terms, best NT hit, etc.

In order to store this wide variety of information we have adapted our database schema so that it utilizes tables with generic fields. Instead of “number in cluster”, or “accession number” our fields are “id”, “type”, and “value”. With this table structure and our especially designed web application we can quickly and easily make sequence annotation available to the user.