

Statistical Approaches to Microarray Data Analysis

or

Primer on Statistical Analysis of Gene Expression

Brian S. Yandell
Statistics, Horticulture, Biometry

www.stat.wisc.edu/~yandell/statgen

©ZB Zeng & BS Yandell array.overview.1 March 24, 2001

Microarray Overview

- what is a microarray?
spotted arrays & oligonucleotides
- image analysis
separating signal from noise
- informatics
blasting gene accession numbers
mining literature for function
building & interpreting patterns
- experimental design
comparing conditions
replication & factorial design
- data analysis
clustering & classification
hypothesis testing or Bayesian inference

©ZB Zeng & BS Yandell array.overview.2 March 24, 2001

What is a Microarray?

goal: highly specific binding of mRNA
– what proteins are actively being expressed?
– compare tissues under different conditions
– identify patterns of differential expression

spotted arrays
– highly specific complementary DNA (cDNA)
– spots on glass (nylon) slide as probes
– two mRNA samples with different dyes
– hundreds or thousands of spots per slide

oligonucleotide arrays (Affymetrix©)
– 20 sets of 25-mer probes per gene
– pairs of positive match and mis-match
– careful, secret selection of 25-mers
– tens of thousands of genes per chip

©ZB Zeng & BS Yandell array.overview.3 March 24, 2001

Image Analysis

- creating & using array
– making spots on array
– isolating mRNA samples
– binding to spots
– reading signal
– adjusting for background
- image analysis
Yang Buckley Dudoit Speed (2000)
- combining oligonucleotide probe pairs
Li Wong (2001) PNAS 98: 31-36
Schadt Li Su Wong (2001) J Cell Biochem 80:
192-202

©ZB Zeng & BS Yandell array.overview.4 March 24, 2001

Microarray Informatics

- gene has GenBank or other accession number
BLAST search: www.ncbi.nlm.nih.gov
NCGR: www.ncgr.org
- mine available literature for function
- doublecheck with Northern, RT-PCR or RPA(?)
- can be slow: how to speed up?
BioPERL Project (www.bioperl.org)
R Project (www.r-project.org)
- when to use function information? (Gould)
prefilter: assign genes to groups for analysis
postfilter: detected functional groupings

©ZB Zeng & BS Yandell array.overview.5 March 24, 2001

Microarray Experiments: Design

- design experiment
genes, conditions, replicates(!)
multiple conditions on one chip?
multiple chips per gene array?
- collect tissue under different conditions
extract mRNA
hybridize mRNA to chip(s)
organize microarray data
- protocol to detect patterns
prescreen genes for low abundance?
tools to find putative differential expression
postscreen genes for known function?

©ZB Zeng & BS Yandell array.overview.6 March 24, 2001

Microarray Experiments: Analysis

- compare conditions
with or without replication
find patterns of differential expression
- linear models (anova)
formal tests
assumptions and models
parametric or non-parametric
log-normal, gamma, poisson
- data reduction
clustering, classification
principal components (PCA or SVD)
fold change vs. other patterns?

©ZB Zeng & BS Yandell array.overview.7 March 24, 2001

Future of Microarray Studies

- clinical trials
repeated measures on cases
multiple tissue samples per visit
- QTL studies in experimental crosses
expression profile as huge phenotype
epistasis & transcription factors
recombinant inbred & congenic lines
- environmental studies
- proteomics, metabolomics & cell models
protein chips already in advanced research
in situ views of cell activity
in silico simulations of function

©ZB Zeng & BS Yandell array.overview.8 March 24, 2001