Experimental Design

Todd Lowe
Bio 210
Jan 10, 2003

Experimental Design

What types of experiments can you do with microarrays?

Changes / Differences in

- mRNA expression
  - Single stage of growth, expression profile not changing; comparison between two different environmental responses or organisms with different genetic backgrounds
  - Time course, specifically measuring dynamic changes within the same organism
- Genomic DNA
  - Single nucleotide changes
  - Multiple, complex changes

Type I vs. Type II

- Type I: simplest type, two samples, directly compared on one spotted slide by ratio of hybridization
  - “Reference” or “wild-type”, generally green, “Experimental” or “mutant” in red
- Type II: Multiple comparisons, such as time course experiments
  - Multiple experimental samples hyb’d against a common reference (i.e. time zero); allows direct comparison of multiple samples

Choice of "Reference"

- For type II experiments, "reference" sample does not have to be related to experiment
- To be a good reference, simply need non-zero signal at maximum possible genes or elements being measured
- A "zero" or very small reference denominator causes large fluxuations in ratio of

  experimental/reference   intensities

References

- Common choice for references:
  - Time zero of time course experiment
  - Cancer cell lines (one or more), which commonly express more genes than normal cells b/c of rapid growth program
  - Genomic DNA

Reference Comparison

	Pros	Cons
Time Zero mRNA	Convenient; Easy to interpret ratios	May be difficult to make in large enough quant.; not for cross-lab comp
Cancer cell lines mRNA	Good diversity of mRNAs; can be made in quant. so used as a common ref across labs	Unique pool of mRNAs; once used, must make a new pool that will be different; labor intensive
Genomic DNA	Convenient, plentiful, fairly reproducible	Small “denominator” for highly expressed mRNAs

Types of mRNA Expression Studies

- Cyclical cell events: Cell cycle, circadian gene expression (in fruit flies)
- Metabolism – changes in transcription due to changes in supplied nutrients & environ.
  - Ex: diauxic shift, DeRisi *et al.*, shift from fermentation (anaerobic) to respiration (aerobic) metabolism
- Developmental stages
  - Ex: sporulation in yeast, Chu *et al.*
  - Ex: several studies in *C. elegans*

Probing Cellular Response

- Response to environmental stress
  - Heat or cold shock
  - DNA damage (UV, gamma radiation, etc.)
  - Starvation (amino acids, nitrogen source)
  - Osmotic shock
  - pH
  - Hydrogen peroxide

Studying Changes in Genomes with Arrays

- Changes in DNA copy number
  - Hughes et al., (2000) study showed extra copies of chromosomes in yeast strains
  - Some deletion mutants survived by keeping extra copy of chromosome with deleted gene
  - mRNA expression pattern showed similar increase in all transcripts on duplicated chromosome

Genome Changes in Cancer

- Pollack et al. (1999) used similar idea to compare cancer v. normal human cells
- Used genomic DNA directly (not cDNA) to show changes in gene copy number in cancerous cell

Tracking Adaptive Evolution

- Ferea et al. (1999) used arrays to track changes in genome due to natural selection
- "Wild-type" yeast grown in glucose-poor media for 250 generations
- Resulting yeast changed in "adaptive shifts" to low glucose-environment
- Used arrays to find which genes & pathways were affected: glycolysis genes down-regulated, TCA cycle genes up-regulated (energy-related pathways were adjusted to use glucose more efficiently)
- Results indicate only a small number of mutations (~5) in control genes responsible for adaptations

Annotating a Genome

- Shoemaker *et al.* (2001) used “exon-tiling” Affy arrays to improve genome annotation
- Looked for co-expression of neighboring exons to define correct gene structures
- For several small regions, tiled *all* sequence to define exon structure, based mRNA hybridization signal
- Very expensive, limited in same way EST verification of gene structure

SNP / Haplotype Mapping

- SNP = single nucleotide polymorphism
- Estimated 5.3 million “common” SNPs that can be found in 10-50% of all people (approx. one every 600 bp)
- Finding all these common genetic differences between people gives stable landmarks in “blocks” of DNA that are co-inherited with particular alleles that may be associated with a disease
- Perlegen effectively “resequenced” 20 copies of chromosome 21 to look for all common SNPs

Spellman et al. Cell Cycle Paper

Biological Motivation

- To identify all the genes in yeast that are cell-cycle regulated

Array Protocol

- Spotted DNA arrays
- full length PCR products of all 6000+ yeast protein-coding genes
- RNA labelling: total RNA (15ug) + oligo dT making labelled cDNA
- Hyb’d with arrays for 4-6 hours at 65 C

Experimental Design

- Yeast cells were synchronized by three different methods:
- Block & release:
  - Alpha factor – pheromone that makes cells of opposite mating type (MATa) arrest in G1 phase
  - Cdc15 synchronization (cdc15-2 mutants grow at 23 C, arrest at 37 C at telophase)
- Sort by size ~ size correlates to point in cell cycle
  - Elutriation – centrifugal separation of asynchronous population to obtain just G1 cells (of same size)

Exp. Design (cont.)

- Samples collected
  - for 2.2 hours (alpha factor) every 7 min
  - for 5 hours (cdc15) every 10 min
  - for 6.5 hours (elutriation) every 30 min

- Average yeast cell cycle ~ 90 minutes in optimal growth conditions

Experimental Pitfalls

- Each synchronization method may induce transcription specific to method (i.e. artificial spikes for some genes)
  - Alpha factor release causes mating pathway expression
  - Cdc15-2 temperature shift can cause heat shock response

Computational Analysis

- Fourier transform of all genes, giving a vector which reflects the phase & period of the $\log_2$-ratio of intensities, measured over course of experiment
- Vectors for each gene from cdc15, alpha factor, and Cho *et al.* cdc28 experiment added together after synchronizing the phase
- Genes “known” to be expressed at only one point in cell cycle chosen to define reference point

Computational Analysis

- Vectors for these known reference genes were compared to all other gene vectors using Pearson correlation (+1 = perfect match, 0 no correlation, -1 = perfect negative relationship)

(Aside for biologists: to brush up on your stats, see

http://davidmlane.com/hyperstat/)

- All genes ranked by aggregate score, and 800 had a score greater than 91% of known cell cycle regulated genes (arbitrary cutoff)

Spellman-table3.htm

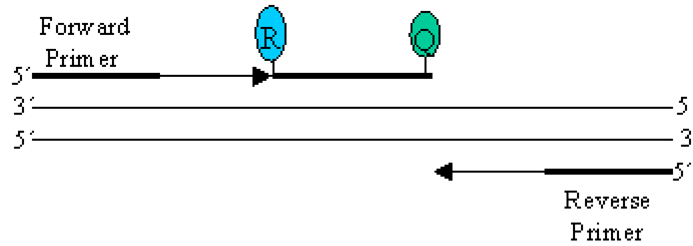
“Novel” Results

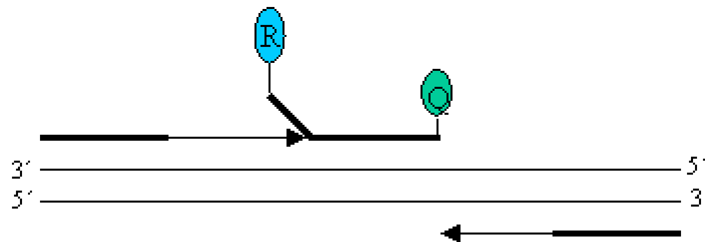
- Approx. 800 genes appear to be cell-cycle controlled, up from 104 previously known
- Thus, over 10% of all genes in yeast oscillate expression with cell cycle, a much larger number than previously thought

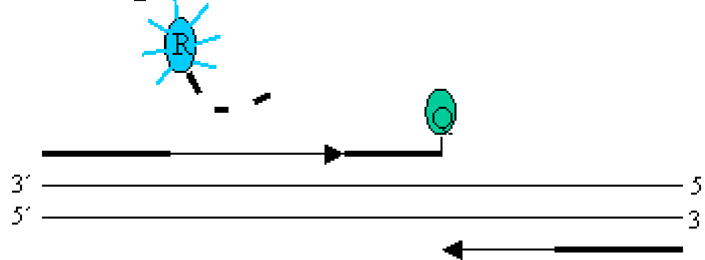
Verification of Array Measurements

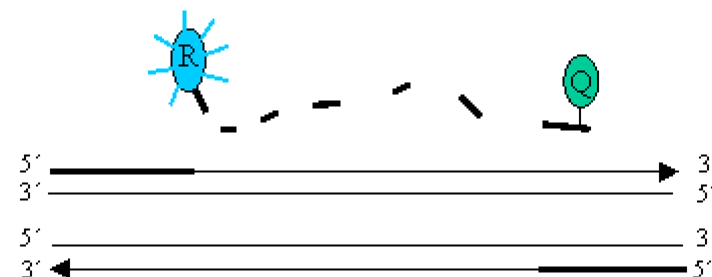
- TAQman assay – highly accurate quantitation of cDNA using RT-PCR and flourescent dyes

TAQman Assay









- More PCR product, more "R" flourescent dye released from "Q" quenching dye
- Highly accurate for measuring small amounts of mRNA

Further Experiments

- Cln3 and Clb2 experiments – both are cell-cycle regulator proteins
- Cells were arrested in cell cycle, and samples with and w/o Cln3 or Clb2 were measured for increased expression of genes under their control

Additional Computational Analysis

- Promotor analysis: frequency of DNA binding sites 700bp upstream of 800 genes

Assessment

- Reality check: data points were shuffled by time point and gene & scored in same way
- Only 24 "genes" had scores exceeding empirically chosen threshold (false pos rate ~3%)
- Shuffled by gene only: still <10% exceeding threshold