Yeast GENOMIC EXPRESSION ANALYSES

Notes and Tips by Audrey Gasch
As discussed at the 2000 Yeast Genetics Meeting in Seattle, Washington

I. Experimental setup

A. Choosing the experimental parameters

1. Choosing the strain background: Many labs routinely work with their favorite strains, but it is worthwhile to consider which strain is best for your genomic studies, because different strains can have very different genotypes and phenotypes

a. Be aware of the differences between common lab strains (eg. S288C vs W303, etc.) and how those differences might affect the experimental results

b. The mating type can affect phenotype, including gene expression

c. Best to use strains with minimal auxotrophic markers to simplify analysis of the results

d. Mutant strains should be in an isogenic background as the wild-type control

2. Choosing the experimental conditions

a. Determine the optimal growth conditions for your needs. Genomic expression can be affected by growth conditions, including media (rich versus minimal), culture system (shaker flasks versus chemostat), temperature, etc.

b. Investigate the optimal "dosage" of the stimulus: Many genomic studies observe the response of cells to various stimuli. It is worth investigating the dosage of the applied stimulus to pick conditions that will give meaningful results

1) Lethal conditions may result in data that are difficult to interpret, while mild conditions may not provoke a detectible genomic expression program

c. Investigate the timing of the cellular response.

1) Determine the appropriate time points that will reveal the genomic
expression response … some responses occur and subside within 15 minutes, other responses peak hours after the stimulus.

2) Timecourse experiments yield a higher level of detail than single-
timepoint experiments, including temporal information

B. Planning the experiment: ONE VARIABLE ONLY!

As in any experiment, it is very important to insure that there is only One variable in genomic expression experiments. Often, experimental variables that are overlooked can provoke substantial changes in genomic expression and confound analysis of the results.

1. Hypothetical and real examples of multiple variable experiments:
   a. ** Diauxic shift during experiment: Likely the most common oversight is the progression of the cells through diauxic shift, when the cells become limited for glucose and alter their metabolism accordingly. The expression of thousands of genes is altered during this phase of growth. The timing of diauxic shift is dependent on the culture conditions (strain, media, growth temperature, aeration, environmental stress) so it is very important to know when diauxic shift occurs under your conditions and avoid it (see more below).
   b. Pleiotropic drugs will result in pleiotropic cellular effects and thus genomic expression
      1) The "DNA damaging agent" Methyl-methane sulfonate (MMS) methylates many cellular targets in addition to DNA
      2) High sodium – alters ionic strength, osmotic strength, as well as Na+ concentration
   c. Experiments with extensive cell handling: avoid unnecessary cell handling (such as washing cells during collection), or account for cell handing in a control experiment
      1) Changes in culture aeration can lead to hypoxia
   d. Drugs suspended in a carrier solution: add carrier alone in mock control

C. Choose the reference for microarrays: The main goal in choosing a reference is to insure significant hybridization signal in every spot on the arrays so that the Ratio of R/G signal in each spot can be quantified … therefore the identity of the reference is somewhat arbitrary; the data can be subsequently mathematically transformed to reveal the biologically-relevant ratios.

1. Example reference samples
   a. Genomic DNA
   b. An arbitrary RNA reference pool
   c. Time-zero RNA, taken just before beginning the experiment
   d. A pool of all of the RNA samples recovered from an experiment
      RNA taken from the control sample
2. Regardless of which reference is used, be sure to use the identical reference on all arrays in a given timecourse so as to compare the timepoints to eachother
3. Mathematical transformation example:

   a. for a timecourse in which genomic DNA is used as the reference, there will be one array for each time point INCLUDING the time = 0 sample
   b. Red/Green signal in each spot = signal from time point RNA sample/signal from genomic DNA

c. To transform the data, divide the R/G ratio measured for each gene on the t>0 arrays by the corresponding R/G ratio measured on the t = 0 array to cancel the "genomic DNA" denomenator:

$(R/G_{t>0}$ array) / $(R/G_{t=0}$ array) = $(RNA_{t>0}$ / Genomic DNA) / $(RNA_{t=0}$ / Genomic DNA) =

$\mathbf{RNA_{t>0}/RNA_{t=0}}$

II. Execution of the experiment

A. Before beginning sample collection, allow the subculture to recover from stationary phase

1. At least 2 doublings (not absolute time, as different strains have different growth rates)

B. Begin the experiment at a cell density that will avoid diauxic shift at end of the experiment

1. Be aware that timing of diauxic shift is condition-specific (strain, media, temperature, aeration, environmental stress)
2. For long experiments use a chemostat to maintain culture conditions

C. Record ALL possible details – you'll be glad you did

1. Examples:
   a. OD_{600}, cell number, cell volume over time
   b. Cell viability through experiment
   c. Cell morphology: take periodic photographs of the cells to characterize cell shape, cell cycle arrest
   d. Nutrient concentrations (glucose, NH_4, ethanol, etc.): Always freeze a small aliquot of the culture media to measure such things later
   e. If possible, measure drug concentrations in the culture during experiment
2. Record any anomalies
3. Record anything you can think of … you many not know what details will be valuable until After you see the results

III. Sample collection

A. Be as controlled as possible! Collect all samples as identically as possible

B. Collect by centrifugation (3-5 min at ~3g) or filtration (<1 min by filtering culture over a sterile 0.45 um filter and collecting entire filter)

C. Again, few variables

1. Collect cells ~experimental temperature to avoid temperature shock
2. Collect cells as quickly as possible
3. Do not wash or handle cells unnecessarily

D. Example problems:

1. washing cells induces many variables and can induce the stress response (induces changes in nutrients, osmolarity, ionic strength, pH, aeration, etc.)
2. collecting cells on ice can induce cold shock
3. lengthy collection can induce hypoxia
4. variable collection time is Very Bad!

V. Microarray hybridization

A. Again, as controlled as possible

1. For optimal comparison within a timecourse, perform sample collection, RNA preparation and labeling, and especially hybridizations together for consistency

B. Always be consistent with labeling and hybridizations (time, temperature) to improve reproducibility

VI. Duplicate experiments

A. Duplicate experiments as identically as possible

1. Maintain as many experimental details (such as starting OD_{600}, culture shaker speed, timing of experiment, etc.) as possible

2. The most common differences between experiments seem to be differences in the expression of metabolic genes

V. The practical art of data Analysis

A. Remember all of the experimental details when analyzing the data

1. Be aware of:

a. Pleiotropic conditions during the experiment

b. Potential diauxic shift problems

c. Strain background (auxotrophic markers, mating type)

d. Cell cycle progression or arrest during the experiment

e. Secondary effects of the primary stimulus

B. Keep an open mind when interpreting the results!

C. To identify responses that are specific to your conditions, compare the data to other datasets of unrelated conditions … remember that many of the observed responses may not be specific to your conditions.

D. Remember that whatever analytical method used to analyze the data (hierarchical clustering, Self Organizing Maps, K-means clustering, Singular Value Decomposition), these are methods of *Organizing* the data

1. Which analysis method is the best: NONE!

a. Different methods have different strengths and weaknesses which make each method suited to different analytical problems

b. Often the most thorough analysis involves multiple permutations of multiple computational methods

c. A given cluster is not necessarily "THE" answer