High-Throughput Sequencing Course
Time Course Hypotheses

Biostatistics and Bioinformatics

Summer 2019
So far we have considered comparing mean abundance level at a single time-point.

Example: Let $\mu_0$ and $\mu_1$ denote the mean mRNA abundance level for the untreated and treated group.

- $H_0 : \mu_0 = \mu_1$ (there is no treatment effect)
- $H_1 : \mu_0 \neq \mu_1$ (there is a treatment effect)

What may be of interest is to identify genes for which the mRNA abundance level varies over time.

We will consider the one-sample and two-sample time-course hypotheses.
One-Sample Problem: No Time Course Effect

There is no time-course effect: The mean level is constant over time
**ONE SAMPLE PROBLEM: TIME COURSE EFFECT**

There is a time course effect: The mean level varies over time
Let $\mu(t)$ denote the mean mRNA abundance level at time $t > 0$

If the mean level is constant over time, there is no time effect

$H_0 : \mu(t) = c$ for all $t$ for some constant $c$

$H_1 : \mu(s) \neq c$ for some $t$
Two-sample Problem: Time Course Effect?

There is no time-course effect within each condition, while there is a treatment effect. Is this interesting?
TWO-SAMPLE PROBLEM: TIME COURSE EFFECT?

There is a time-course effect within each condition but not time-course effect across conditions. Is this interesting?
**Two-sample Problem: Time Course Effect**

There is a time-course effect for the treated group only. Is this interesting?
**Time-course Hypothesis: Two-Sample**

- Let $\mu_0(t)$ denote the mean mRNA abundance level at time $t > 0$ for the *untreated* group
- Let $\mu_1(t)$ denote the mean mRNA abundance level at time $t > 0$ for the *treated* group
- $H_0 : \mu_0(t) = \mu_1(t)$ for all $t$
- $H_0 : \mu_0(t) \neq \mu_1(t)$ for some $t$
TWO-SAMPLE PROBLEM: TIME COURSE EFFECT?

There is a time-course effect within each condition and a phase shift with respect to treatment. Is this interesting?
TWO-SAMPLE PROBLEM: TIME COURSE EFFECT?

There is a time-course effect within each condition and a vertical shift with respect to treatment. Is this interesting?
**Standard Analysis (Not Recommended)**

- For each gene, do a two-sample t-test at each time point
- Declare a time-course if any of the $P$-values are “significant
- To make things worse: Use the $P$-values to describe the time-course
- This approach ignores multiple testing aspect (not only due to genes but also due to multiple timepoints within each gene)
- This analysis would only be appropriate if one time-point is identified upfront
- What is the point of a time-course experiment if only one timepoint is of interest?
Previously, we have modeled the mean abundance level at a single time point as
\[ Y = \mu + \epsilon \]
You can model the expression level at time \( t \) as
\[ Y(t) = \mu(t) + \epsilon(t) \]
The challenge here is that \( \mu(t) \) is an unknown function of time.
Methods using this type of model use various approaches for estimating \( \mu(t) \).