

Integrating Biology and Statistics: Gene Set Methods

BIOS 691-003

Winter/Spring 2010

Philosophical Overture

- Integrating biology and statistics
- Gene sets: genes whose protein products collaborate on a well-defined function
 - Vague!
- Hard to define ‘function’ or draw boundary on ‘gene sets’
- Statistical methods often ad-hoc
- Be *skeptical*... but optimistic

Historical Motivations

- Too many genes are significant
 - Researchers used to generate a list by p-value and comb for genes that work together
 - First pathway tools automated this process
- Patterns may be more significant than any individual gene
 - e.g. if most genes in glycogen biosynthesis are up, but none is significant individually (after multiple-comparisons adjustment)
 - We can infer that glycogen is being made