

## Extracting Biological Meaning from High-Dimensional Datasets

John Quackenbush

*Dana-Farber Cancer Institute and the Harvard School of Public Health, U.S.A.*

The revolution of genomics has come not from the "completed" genome sequences of human, mouse, rat, and other species. Nor has it come from the preliminary catalogues of genes that have been produced in these species. Rather, the genomic revolution has been in the creation of technologies - transcriptomics, proteomics, metabolomics - that allow us to rapidly assemble data on large numbers of samples that provide information on the state of tens of thousands of biological entities. Although the gene-by-gene hypothesis testing approach remains the standard for dissecting biological function, 'omic technologies have become a standard laboratory tool for generating new, testable hypotheses. The challenge is now no longer generating the data, but rather in analyzing and interpreting it. Although new statistical and data mining techniques are being developed, they continue to wrestle with the problem of having far fewer samples than necessary to constrain the analysis. One way to deal with this problem is to use the existing body of biological data, including genotype, phenotype, the genome, its annotation and the vast body of biological literature. Through examples, we will demonstrate show how diverse datasets can be used in conjunction with computational tools to constrain 'omics datasets and extract meaningful results that reveal new features of the underlying biology.

[ John Quackenbush, 44 Binney Street, SM822 Boston, MA 02115-6084, U.S.A.; johnq@jimmy.harvard.edu]