

# VALIDATION IN GENOMICS

T. Hastie

*Stanford University, Stanford, CA, USA*

<sup>†</sup> E-mail: [hastie@stanford.edu](mailto:hastie@stanford.edu)

High throughput genomic experiments gather measurements on large numbers of genes, typically for relatively small numbers of samples. The literature abounds with novel methods for analyzing such data, identifying genes in causative pathways, building prognostic signatures, and the like. Never before have statistical methods been so prone to the hazards of overfitting and over-eager interpretation. In this talk we discuss some specific examples of these pitfalls. We then suggest ways to robustify data analysis, provide proper validation, and hence protect the researchers from reporting spurious results.