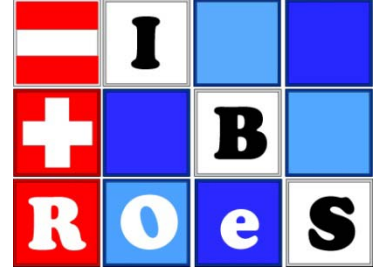


How the levee bears out against the flood



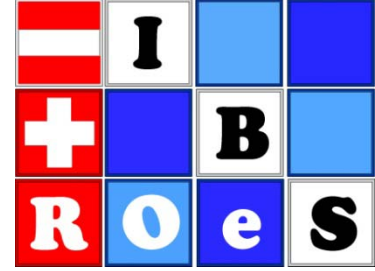
Georg Heinze

Section for Clinical Biometrics

Center for Medical Statistics, Informatics and Intelligent Systems (CeMSIIS)

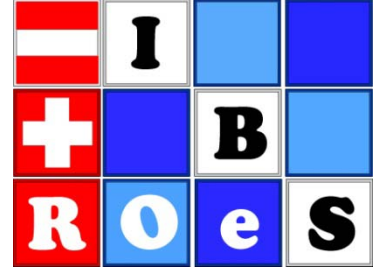
Medical University of Vienna

How the levee bears out against the flood



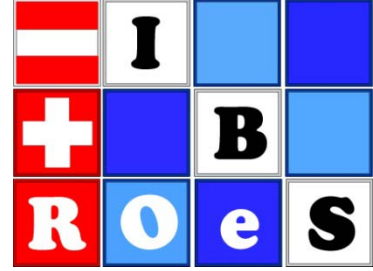
- *The data flood:*
 - Many simultaneous tests
 - New hypotheses beyond $\mu_1 = \mu_2$
 - Prediction with $\gg n$ explanatory variables
 - A variety of new platforms: genomics, methylation, copy numbers, transcriptomics, NGSseq, proteomics, metabolomics, miRNA
 - Impossibility to check assumptions
- *The levees:*
 - FDR, cost-effective sequential designs
 - New test statistics (moderated t, LIM, adaptive trimmed t, ...)
 - Regularized estimation,
 - cross-validated tuning,
 - double-cross-validated internal validation
 - (do not optimize and validate with the same data!)
 - Adopt methods to platforms,
 - Keep analysis pipelines traceable
 - Robust methods that do not make strong assumptions (c')

How the levee bears out against the flood



- Lists of differential expressed genes: Please don't take them for granted; use concepts like **locFDR** or **Odds of differential expression** to weight genes in further gene enrichment analysis
- Statistical models: serve as a (simplified) description of reality. They should help us to deduce general principles in biological associations. Incorporate existing knowledge (eg, confounders, pathways) where possible, Don't let machines do this artistic task alone!

How the levee bears out against the flood



- Why care about omics?
- They may provide useful answers to understand:

Why patients are different,

Why some patients have better prognosis than others,

The etiology of diseases, and How to improve prognosis,

How to find new (perhaps personalized) therapy targets.