Abstract: Biostatistics and computational biology are increasingly facing the urgent challenge of efficiently dealing with a large amount of experimental data. In particular, high-throughput assays are transforming the study of biology, as they generate a rich, complex, and diverse collection of high-dimensional data sets. Through compelling statistical analysis, these large data sets lead to discoveries, advances and knowledge that were never accessible before, via compelling statistical analysis. Building such systematic knowledge is a cumulative process which requires analyses that integrate multiple sources, studies, and technologies. The increased availability of ensembles of studies on related clinical populations, technologies, and genomic features poses four categories of important multi-study statistical questions: 1) To what extent is biological signal reproducibly shared across different studies? 2) How can this global signal be extracted? 3) How can we detect and quantify local signals that may be masked by strong global signals? 4) How do these global and local signals manifest differently in different data types? 
We will answer these four questions by introducing novel classes of methodologies for the joint analysis of different studies. The goal is to separately identify and estimate 1) common factors reproduced across multiple studies, and 2) study-specific factors. We present different medical and biological applications. In all the cases, we clarify the benefits of a joint analysis compared to the standard methods. 
Our methods could accelerate the pace at which we can combine unsupervised analysis across different studies, and understand the cross-study reproducibility of signal in multivariate data.