# Summary

In this thesis we developed statistical methods for the analysis of high dimensional data. We particularly focussed on high dimensional networks reconstruction. In genomics, the identification of gene regulatory networks is crucial for understanding gene function, and hence important for both treatment and prediction of diseases. High dimensional networks reconstruction is a very challenging task since the number of possible graphs grows exponentially with the number of variables (e.g. genes). However, some of the relationships between these variables may be known from the literature. For instance, the current beliefs on interactions among genes is condensed in repositories like KEGG and Reactome. We introduce a framework which allows the incorporation of such prior information in the reconstruction in a soft manner such that it informs the analysis if correct, but can be overruled if completely incompatible with the data. We also treat the subjects of genetic association studies (eQTL mapping) and data integration.

In chapter 1 we introduce a new global-local shrinkage ridge-type prior for undirected networks reconstruction based on SEMs with posterior edge selection. The proposed approach is computationally fast and outperforms known competitors such as the *graphical lasso*.

In chapter 2 we extend chapter 1 to include prior information in reconstructing undirected networks. The incorporation of the prior knowledge is done in a soft manner allowing the data at hand to overrule the prior information if not relevant. Furthermore, the proposed method is able to explicitly estimate the agreement of the prior knowledge with the data at hand which is a novelty in incorporating prior information in network inference.

In chapter 3 we introduce a framework for simultaneously analysing multiple related high dimensional and complex datasets. Such analyses include gene regulatory network reconstruction, genetic association studies (e.g. eQTL mapping) and data integration in genomics, to name but a few. To enable the analysis for small $n$ relative to large $p$, we introduce the *horseshoe* prior which allows for sparsity; a desired property for the analysis of such data. We illustrate the approach by two applications, namely: to the reconstruction of gene regulatory networks and to eQTL mapping.

In chapter 4 we explore several approaches to reconstruct gene regulatory networks from combining observational (*in vivo*) and time-course cell line (*in vitro*) gene expression data. The dynamics of the human cell are assumed to obey a first-order vector autoregression VAR(1) model and it is investigated how the underlying model parameters can be efficiently learned using the two types of datasets. We see in an application to real data that reconstruction of the conditional independence graph by borrowing information from the cell line data improves significantly. Moreover, our newly proposed strategies to learn the VAR(1) model parameters are able to indicate preserved transcriptional dynamics between the *in vitro* and *in vivo* environments.