Summary

Genetical genomics is an interdisciplinary field concerned with the consequences of natural genetic variation on multiple molecular traits (mRNA expression levels, or protein and metabolite abundance). The goal of genetical genomics is to contribute to the establishment of informative molecular models explaining how DNA variation leads to observable phenotypic differences such as the emergence of a disease. Genetical genomics aims to become a genetic approach to systems biology and can therefore be referred to as a 'systems genetics' approach.

In this thesis, we introduce the principles of genetical genomics for a general readership (**Chapter 1**). The applicability of genetical genomics is illustrated with a screen of gene expression in hematopoietic cells from a population of inbred mice (**Chapter 2**). The results of this experiment demonstrate that the genetic variants controlling gene expression levels are highly sensitive to the differentiation state of the cells.

A computational protocol for mapping of genetic variants underlying variation in gene expression traits in inbred populations is fully developed in **Chapter 3**, addressing both theoretical and practical aspects of the implementation of the genetical genomics approach.

Such genetic variants are referred to as eQTL (expression quantitative trait loci). **Chapter 4** is concerned with a particularly controversial issue in genetical genomics: the relevance of eQTL hotspots on the genome. Those hotspot regions harbor genetic variation that seems to affect the expression of a (very) large number of genes (sometimes thousands). They could therefore reveal the presence of major biological regulators. However, because of limitations of the statistical methods commonly used, some studies have questioned the biological significance of hotspots. Here, we propose a permutation strategy that allows us to discard numerous hotspots due to statistical artifacts induced by widespread coexpression.

In **Chapter 5** we present DifCoEx a new bioinformatics method for differential coexpression analysis. We illustrate the use of DiffCoEx by applying it to the analysis of a publicly available microarray study of the effect of carcinogenic products on mutant tumor-prone rats. We show that the method is able to reveal meaningful groups of genes which do not show differential expression patterns, but are differentially correlated.

Chapter 6 and **Chapter 7** are concerned with statistical inference of causal relationships between phenotypes using genetic data. This topic has great significance for the of field biomedical research because it has been presented as a way to identify drug targets for the treatment of diseases and metabolic conditions with complex genetic inheritance. In **Chapter 6**, we review the different methods that have been used in genetics studies to try to connect phenotypes in functional networks. Subsequently, in **Chapter 7** we focus more specifically on a popular method

based on co-mapping of phenotypes and we delineate the critical conditions required for its proper use. In particular, we show that this method cannot produce reliable results within the settings of most current genetic experiments (including genetical genomics) because of the limited population sizes, the limited effect size of most genetic variants and the omnipresence of noise in high-throughput biological technologies.

The last part of this thesis is devoted to a discussion of when and how genetical genomics can successfully contribute to a systems genetics approach of biology (**Chapter 8**).