Family-based analysis of genome-wide gene × gene interactions

Marit Ackermann

Biotec TU Dresden

July 9, 2009

Outline

Epistasis

- Epistasis: interaction between two or more genes
- known to be fundamental for the function of regulatory pathways in mammals
- implies its importance for the development of complex diseases such as cancer, Alzheimer‘s disease, diabetes

Traditional Approaches

- for yeast and worms large scale double knock-outs and knock-downs exist
- linkage and association studies in mammals concentrate on either single locus associations or interactions between few preselected loci
- major reasons: non-availability of large and suitable data for analysis of interaction effects, low power of the studies

Genome-Wide Screen in Mammals

- recent advances in biotechnology allow genome-wide genotyping of thousands of individuals
  $\longrightarrow$ can be used to **study epistatic effects over whole genome**
- genotyped individuals possibly related
  $\longrightarrow$ take population structure into account; even **make use of known relationships**

Outline

Outline

Method

- idea: two markers whose genotypes are correlated are likely to interact
- measure association via χ^2-test for contingency table

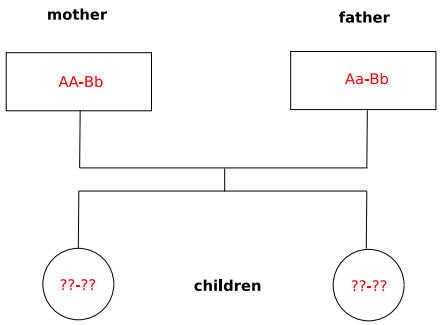
	BB	Bb	bb	
AA	n_{AABB}	n_{AABb}	n_{AAbb}	
Aa	n_{AaBB}	n_{AaBb}	n_{Aabb}	
aa	n_{aaBB}	n_{aaBb}	n_{aabb}	

Method

- idea: two markers whose genotypes are correlated are likely to interact
- measure association via χ^2-test for contingency table

	BB	Bb	bb
AA	n_{AABB}	n_{AABb}	n_{AAbb}
Aa	n_{AaBB}	n_{AaBb}	n_{Aabb}
aa	n_{aaBB}	n_{aaBb}	n_{aabb}

- make use of family information to avoid spurious findings:
  compare observed allele combination with what could have been inherited from parents
- additional correction for allelic drift



Problem

- extremely large number of interactions
  (example: 10,000 markers: $\sim 10^8$ interactions)
- leads to underpowered analysis, many false positive findings

Problem

- extremely large number of interactions
  (example: 10,000 markers: $\sim 10^8$ interactions)
- leads to underpowered analysis, many false positive findings
- need to complement with additional, **external information**

Outline

Databases

- use public knowledge about gene × gene interactions to confirm results;
  e.g. STRING: database of known and predicted physical and functional interactions
- include information from regulatory pathways

Outline

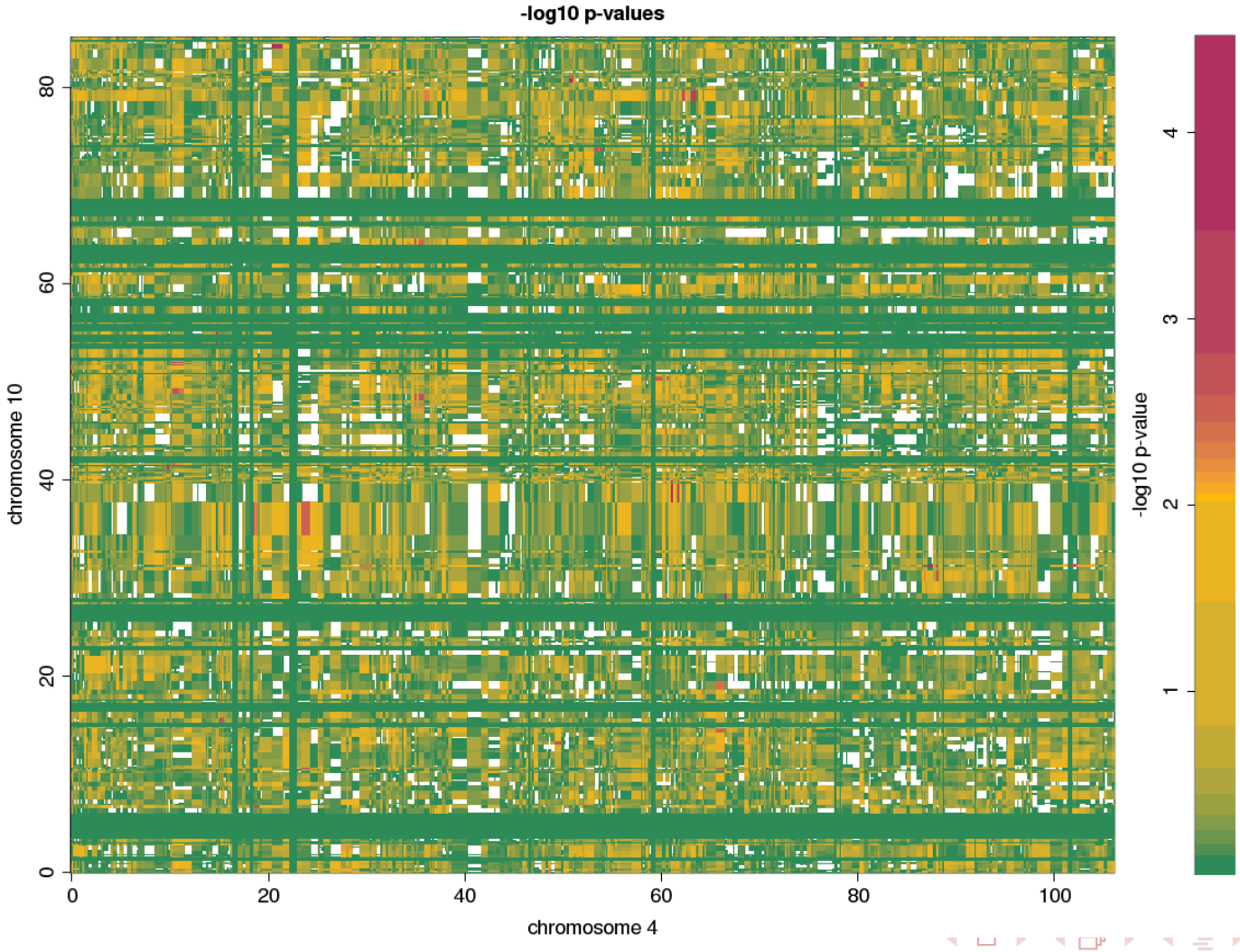
Outline

Data

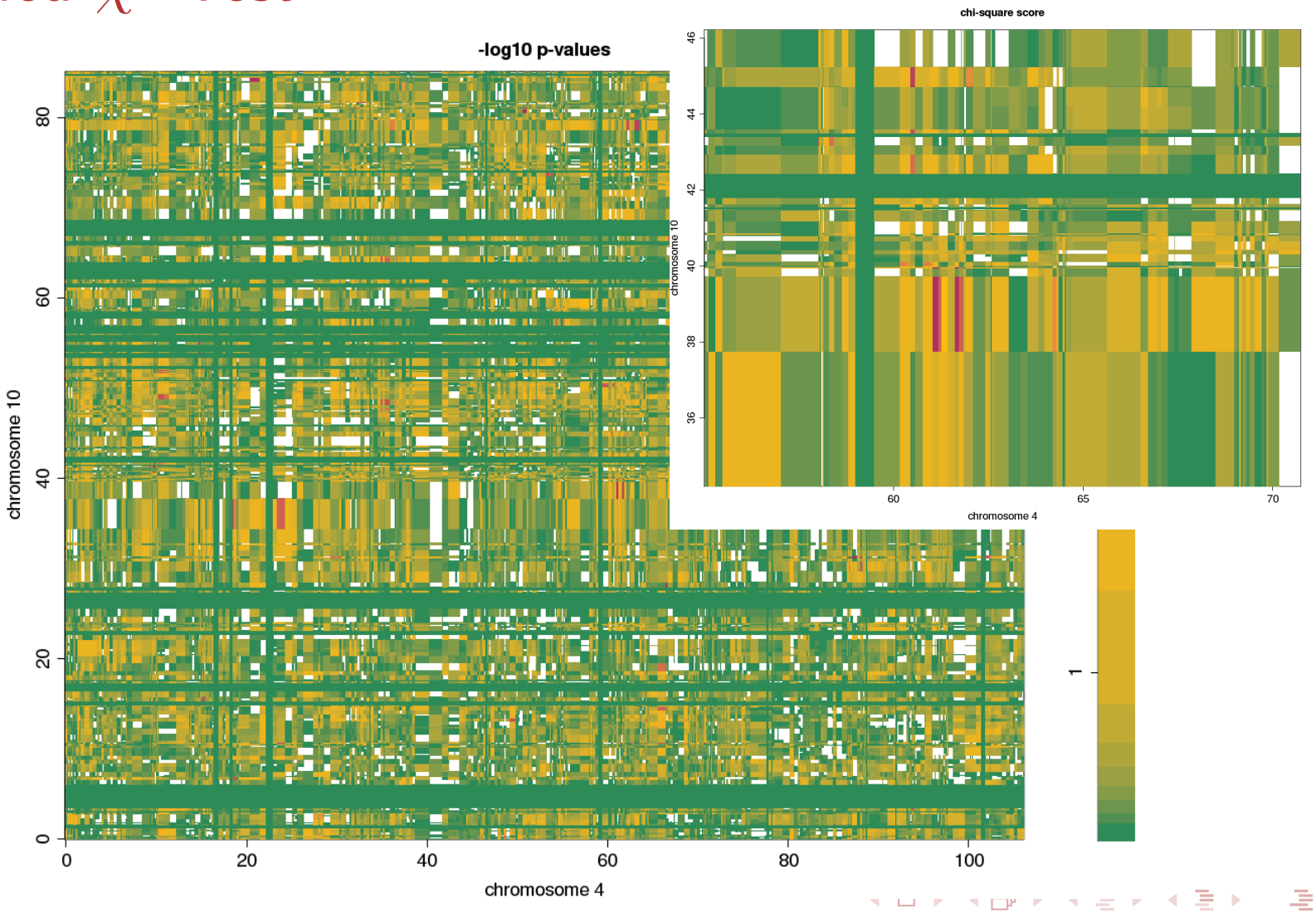
- Solberg, L.C. et al. (2006). A protocol for high-throughput phenotyping, suitable for quantitative trait analysis in mice. *Mammalian Genome*, **17**, 129-146.

- genotype data from more than 2000 outbred mice consisting of $\sim 12,000$ markers
- only consider interactions on two different chromosomes

Modified χ^2-Test

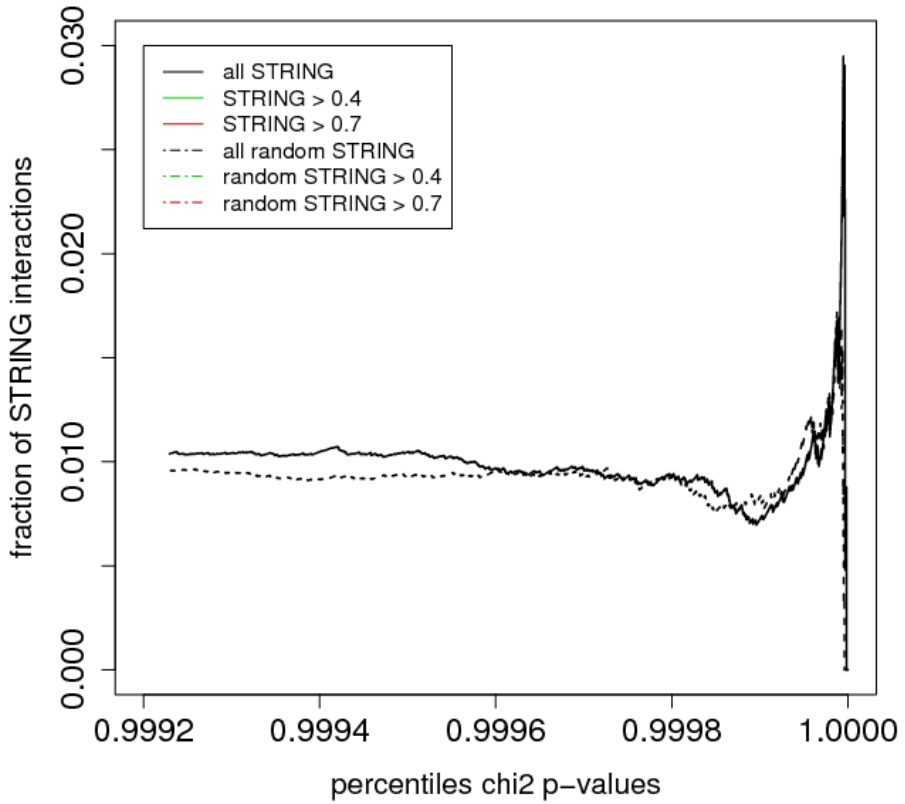


Modified χ^2-Test



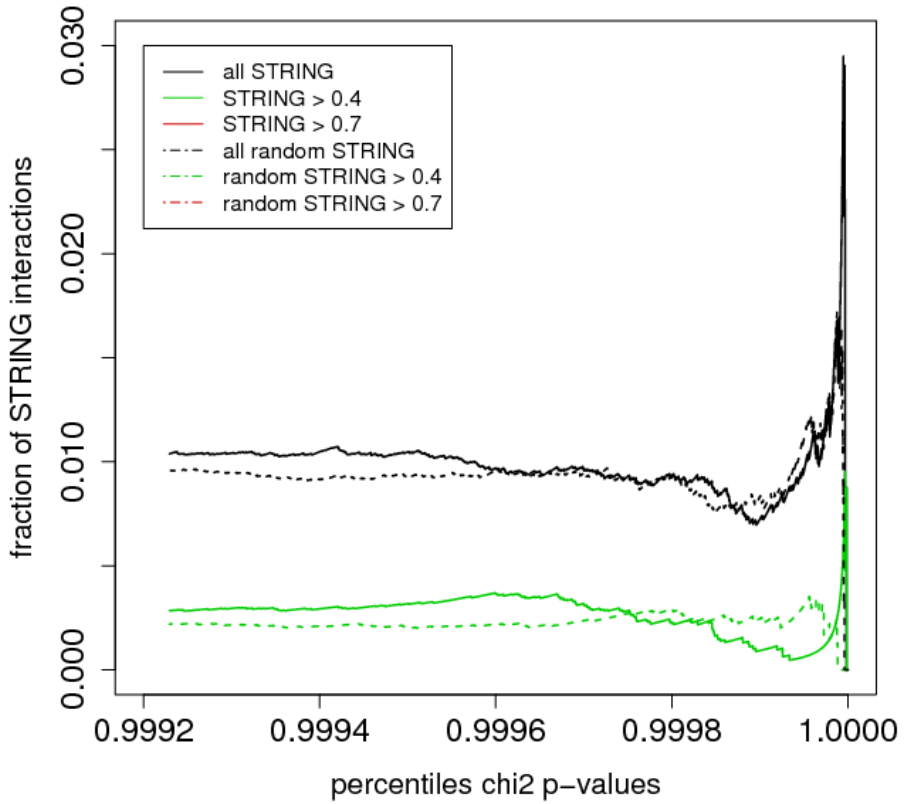
Confirmation with STRING

- fraction of SNP pairs with a low χ^2 p-value that lie close to interacting genes
- proportion of confirmed interactions should increase with increasing χ^2 score



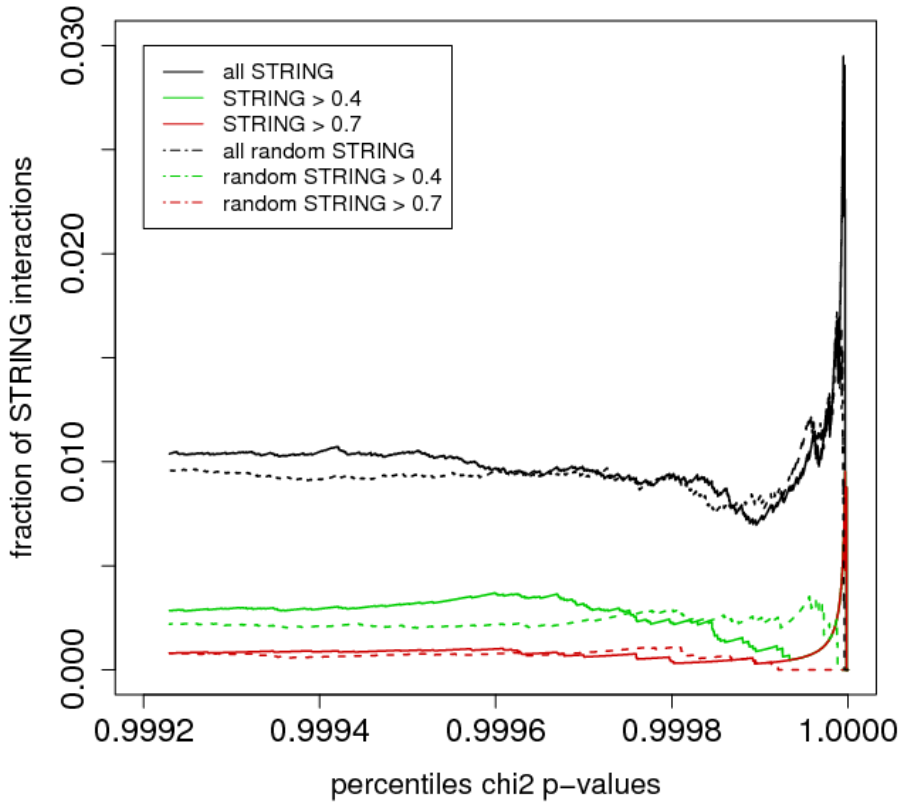
Confirmation with STRING

- fraction of SNP pairs with a low χ^2 p-value that lie close to interacting genes
- proportion of confirmed interactions should increase with increasing χ^2 score



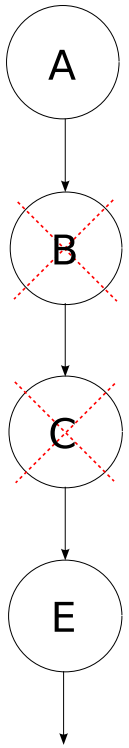
Confirmation with STRING

- fraction of SNP pairs with a low χ^2 p-value that lie close to interacting genes
- proportion of confirmed interactions should increase with increasing χ^2 score



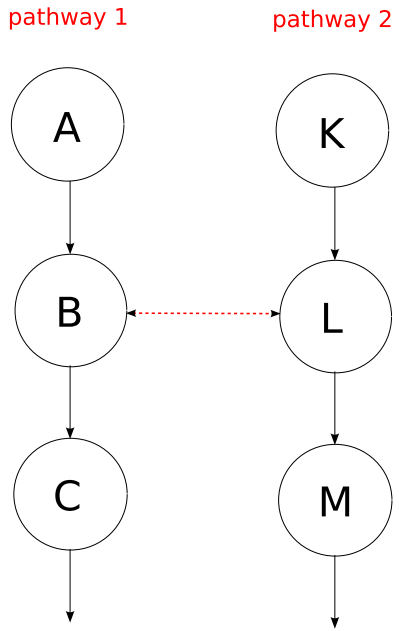
Incorporating Pathway Information

- interactions in one pathway can be crucial, e.g. when signal weakened by two consecutive dysfunctional pathway members



Incorporating Pathway Information

- interactions in one pathway can be crucial, e.g. when signal weakened by two consecutive dysfunctional pathway members
- interactions between pathways indicate common endpoint

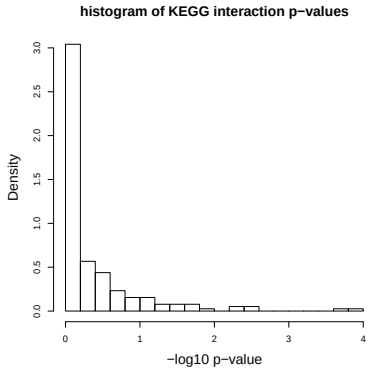


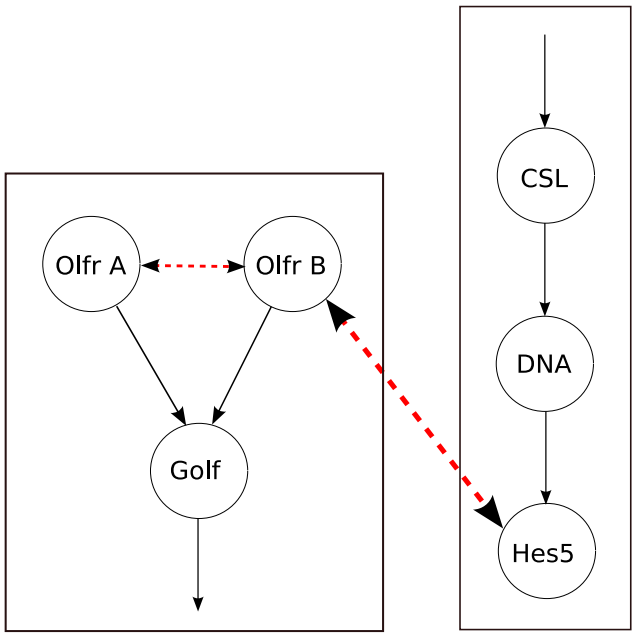
Example: KEGG Pathway

KEGG: database of signaling and metabolic pathways

Example: KEGG Pathway

KEGG: database of signaling and metabolic pathways





indicates importance of olfactory receptors in embryonic development and interplay with notch

Outline

- we propose a new approach to infer epistatic interactions in mammals

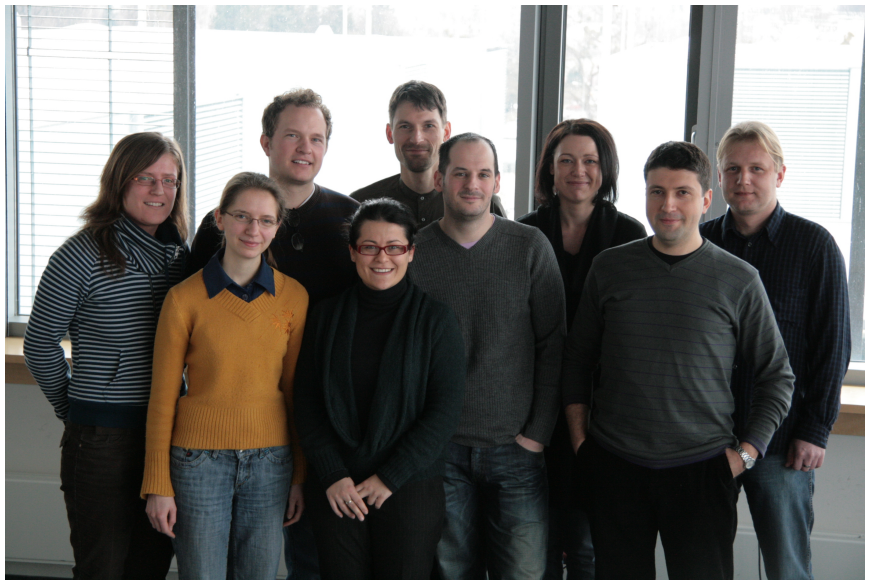
- we propose a new approach to infer epistatic interactions in mammals
- works on a genome-wide scale

- we propose a new approach to infer epistatic interactions in mammals
- works on a genome-wide scale
- population structure explicitly taken into account

- we propose a new approach to infer epistatic interactions in mammals
- works on a genome-wide scale
- population structure explicitly taken into account
- other counfounding factors readily included

- we propose a new approach to infer epistatic interactions in mammals
- works on a genome-wide scale
- population structure explicitly taken into account
- other counfounding factors readily included
- data integration from different sources increases power and facilitates biological interpretation of results

Acknowledgements



- Dr. Andreas Beyer
- my colleagues in the Cellular Networks and Systems Biology group

- Klaus Tschira Foundation for funding

KTS
Klaus Tschira Stiftung
gemeinnützige GmbH