Inferring causal gene-phenotype networks underlying complex traits using multi-omics data

Francisco Peñagaricano
Department of Animal Sciences
University of Florida

Introduction

- Genetic linkage and association studies have been successful in identifying genomic regions associated with phenotypic traits in livestock species
- Identification of the individual genes responsible for the phenotypic variation remains challenging
- These analyses do not provide in general any information about the molecular pathways underlying the phenotype under study

Objective

The objective of this research was to investigate gene-phenotype network reconstruction integrating phenotypic, genotypic, and transcriptomic data

1. We describe and perform multistep procedure for inferring causal networks integrating multiple layers of omics data
2. We show the construction of causal molecular networks underlying carcass fat deposition and muscularity in pigs

Methodology

Phenotypic, genotypic and transcriptomic data

- Data from a genetical genomics experiment with pigs that belong to an F_{2} Duroc x Pietrain population developed at Michigan State University
- Several phenotypes for carcass and meat quality traits
- Genotypes for 124 microsatellite markers spanning the whole genome
- Gene expression values of almost 20,000 transcripts measured in loin muscle tissue

Step 1: Exploring marginal associations between genotypes and phenotypic traits (pQTL mapping)

Step 2: Exploring marginal associations between genotypes and expression traits (eQTL mapping)

Step 3: Searching for genomic regions where significant pQTL co-localize with significant eQTL

Step 4: Inferring causal gene-phenotype networks using causal structure learning techniques
Methodology
1st step: perform a classical QTL mapping for several phenotypic traits
- carcass and meat quality traits that were measured on or were directly related to loin muscle
  \[ y_{ijkl} = \mu + s_i + g_j + c_k + e_{ijkl} \]
  - controlling for sex and contemporary group (exogenous variables)
  - genotypes:
    - Genotypes for 124 microsatellite markers spanning the whole genome
    - breed of origin probabilities across the genome of F2 animals were obtained assuming that the parental breeds were fixed for alternative QTL alleles
    - total 1279 putative QTL positions spanning the whole pig genome

Methodology
2nd step: perform a QTL mapping using the gene expression as a response variable (also known as eQTL mapping)
- gene expression data of almost 20,000 transcripts in longissimus dorsi muscle tissue
- controlling for dye and sex (as fixed effects) and array (as random effect)
- genotypes:
  - Genotypes for 124 microsatellite markers spanning the whole genome
  - breed of origin probabilities across the genome of F2 animals were obtained assuming that the parental breeds were fixed for alternative QTL alleles
  - total 1279 putative QTL positions spanning the whole pig genome

Methodology
3rd step: find genomic regions where significant pQTLs co-localize with significant eQTLs
- genomic region in chromosome 6

Methodology
Finally: the goal is to reconstruct gene-phenotype networks using causal structure learning algorithms integrating the phenotypic, genotypic, and transcriptomic information provided by this region
- The idea is to explore the universe of causal hypotheses in order to find a causal structure that is able to generate the observed pattern of conditional independencies among the variables
  - X1 and X2 are marginally independent
  - X1 and Y' and W are marginally independent
  - X2 and Y' are marginally dependent
  - Y and W are marginally dependent
  - Conditionally on Y, X1 and X2 are dependent
  - Conditionally on Y, X2/X1 and W are independent
Methodology

causal structure learning algorithms: the idea is to infer network structures underlying a set of correlated variables assuming that the conditional independencies in the joint probability distribution of these variables are compatible with the causal model

• causal structure learning algorithm: Inductive Causation (IC) algorithm
• the practical application of these algorithms involves performing a set of statistical decisions using conditional independent test: Fisher’s Z test

\[
Z_{X,Y,W} = \frac{1}{\sqrt{n - |W| - 3}} \log \frac{1 + \rho_{X,Y,W}}{1 - \rho_{X,Y,W}}
\]

Reconstructing gene-phenotype networks (1)

Reconstructing gene-phenotype networks (2)

Reconstructing gene-phenotype networks (3)

Conclusions

• We have detailed a multistep procedure for inferring causal networks integrating phenotypic, genotypic, and transcriptomic data
• We have applied this procedure for deciphering gene-phenotype networks underlying fat deposition and muscularity in pigs
• Our findings shed light on the mechanisms underlying the antagonist relationship between carcass fat deposition and meat lean content in pigs
• More generally, our results illustrate the potential of the procedure described here to unravel causal molecular networks underlying complex phenotypes in livestock species

Knowledge about gene-phenotype networks can be used to predict the behavior of complex systems

- the network predicts that modulation of the expression of ZNF24 should lead to changes in the expression of SSX2IP
- a recent study evaluated potential ZNF24 target genes; the authors overexpressed and silenced ZNF24 and then applied microarray assay to identify target genes
- the overexpression of ZNF24 significantly decreased the expression of SSX2IP while the silenced of ZNF24 resulted in a significant overexpression of SSX2IP
Acknowledgments
Bruno D Valente, Hasan Khatib, Guilherme JM Rosa
Juan Pedro Steibel, Ron O Bates, Cathy W Ernst

Contact Information:
E-mail: fpenagaricano@ufl.edu
Website: fpenagaricano-lab.org

Thanks for your attention!