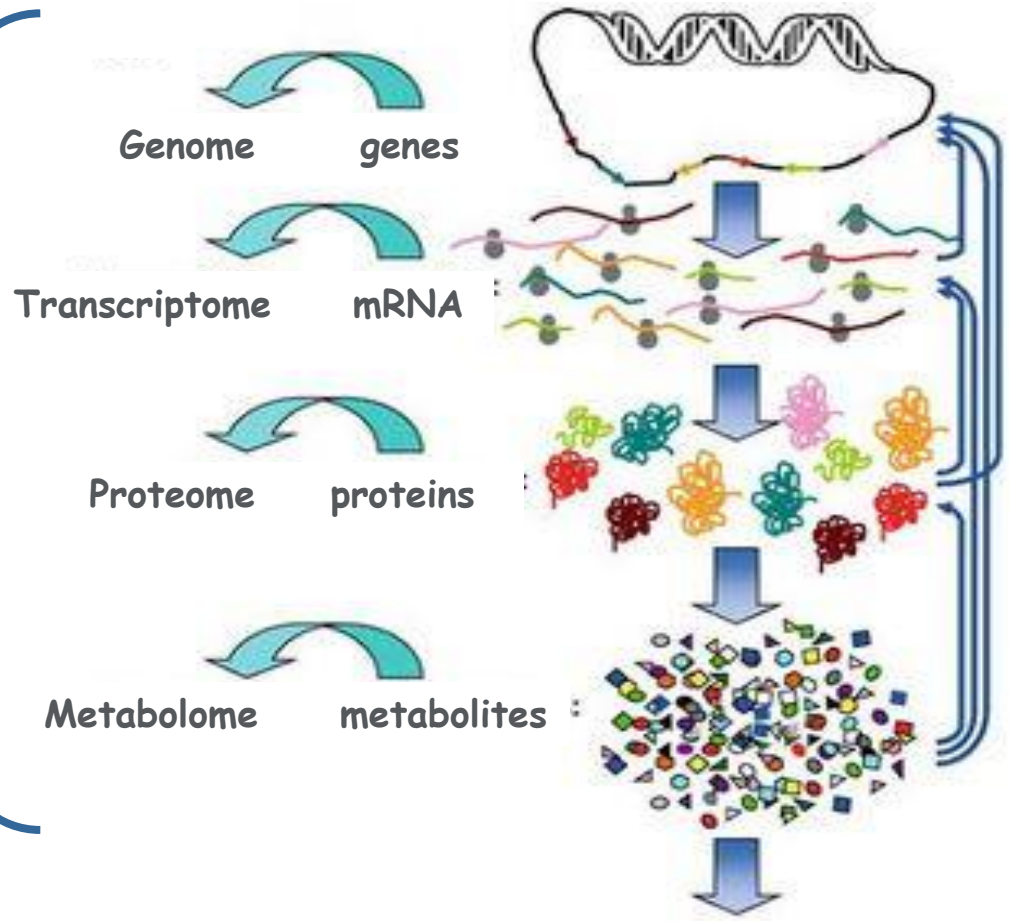
From Genotype to Phenotype

Johanna Vilkki
Green technology, Natural Resources Institute Finland

Luke
NATURAL RESOURCES
INSTITUTE FINLAND

Systems biology



Genotyping methodology
SNP TOOLS, WG SEQUENCING

Functional genomics tools
QPCR, EXPRESSION CHIPS, TRANSCRIPTOME SEQUENCING, RNAi

Proteomic tools
2D-GE, MS, CHROMATOGRAPHY, PROTEIN CHIPS

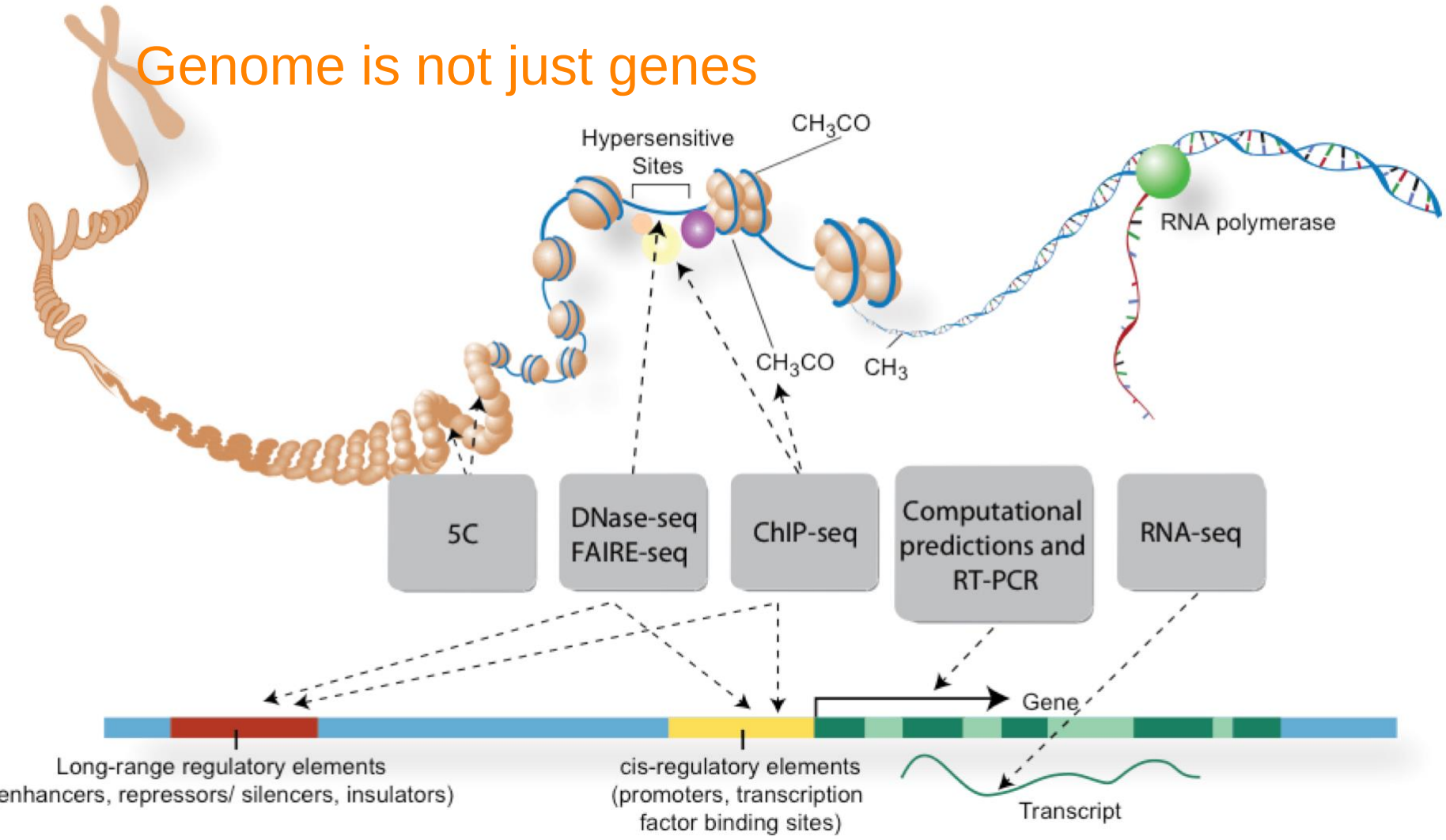
Metabolomic tools
MS, CHROMATOGRAPHY, NMR

Animal models
MOUSE MODELS, CLONES, CELL LINES

Rumen microbial ecology and function → **PHENOME**

Luke
NATURAL RESOURCES INSTITUTE FINLAND

Genome is not just genes



Genome
- sequence variation
- modification (epigenetic)

Environment
- nutrition
- dieseases
- stress

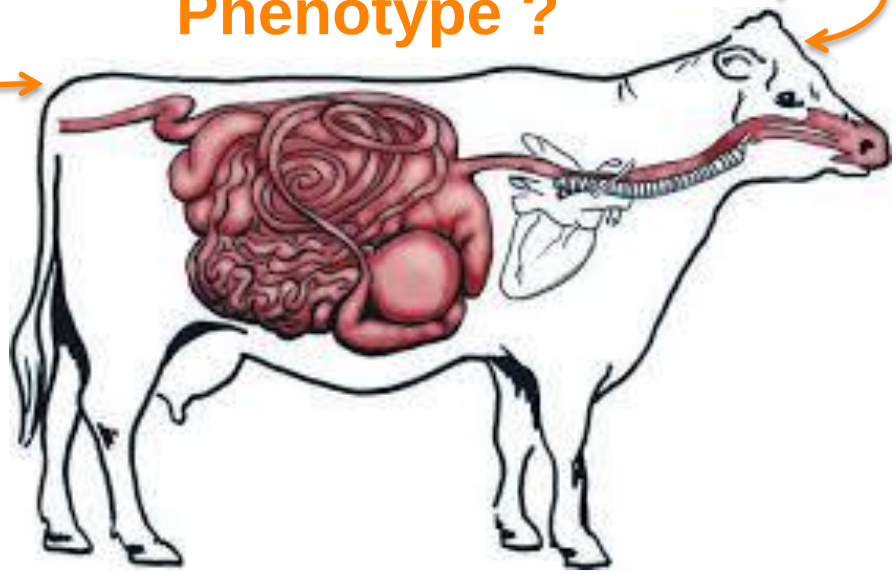
Genome expression
- transcription to mRNA
- translation to proteins
- protein modifications

Proteome

Metabolome

Rumen microbiome
- composition (taxa; genes)
- function (metatranscriptome, metaproteome)

Phenotype ?



Ruminomics questions from genome to phenome

1. Can we find genetic variation in the cow genome related to variation in the various efficiency or emission phenotypes or rumen microbiome composition?
2. If we can map variable loci (genes), how much of the variation do they explain and how are they connected through gene networks and biological pathways?
3. Does the host genotype determine some aspects of rumen microbial composition or function?
4. Does the rumen microbial composition affect host gene expression and physiology?
5. What is the effect of variation in nutrition and nutrients on the rumen microbiome and the various phenotypes

Ruminomics experiments aimed at finding answers

- 1000 cows: collection of very detailed phenotypes, microbiome metabarcoding and host genome genotypes (SNPs) to be included in genome wide association analysis (answering questions 1&2)
- Rumen digesta exchange experiments between varying genotypes: reindeer – cow and identical twin cows vs. unrelated cows to be analysed for effect on microbial communities as well as gene expression changes in rumen papillae, host liver and adipose tissues (answering questions 3&4)
- Detailed physiological studies with diets differing in carbohydrate, protein or lipid contents (answering question 5)

1000 cow experiment

- Standardized phenotype collection at several phenotype levels (e.g. methane emissions, feed intake, digestion efficiency, blood metabolites, milk fatty acids)
- Microbial community determined by 16S rRNA sequencing (specific primers for Archaea, Bacteria, Ciliates and Protozoa)
- Genotyping of 800 animals for a “discovery” stage and 200 animals in a “confirmation” stage. The discovery stage includes Holstein cows with Southern European feeding regime (forage component is mainly maize silage and hay); confirmation stage includes Nordic Red Cattle from Finland and Sweden with Northern European feeding regime where the forage component is mainly grass silage
- GeneSeek Genomic Profiler HD (76,883 SNPs)

Individual variation in population

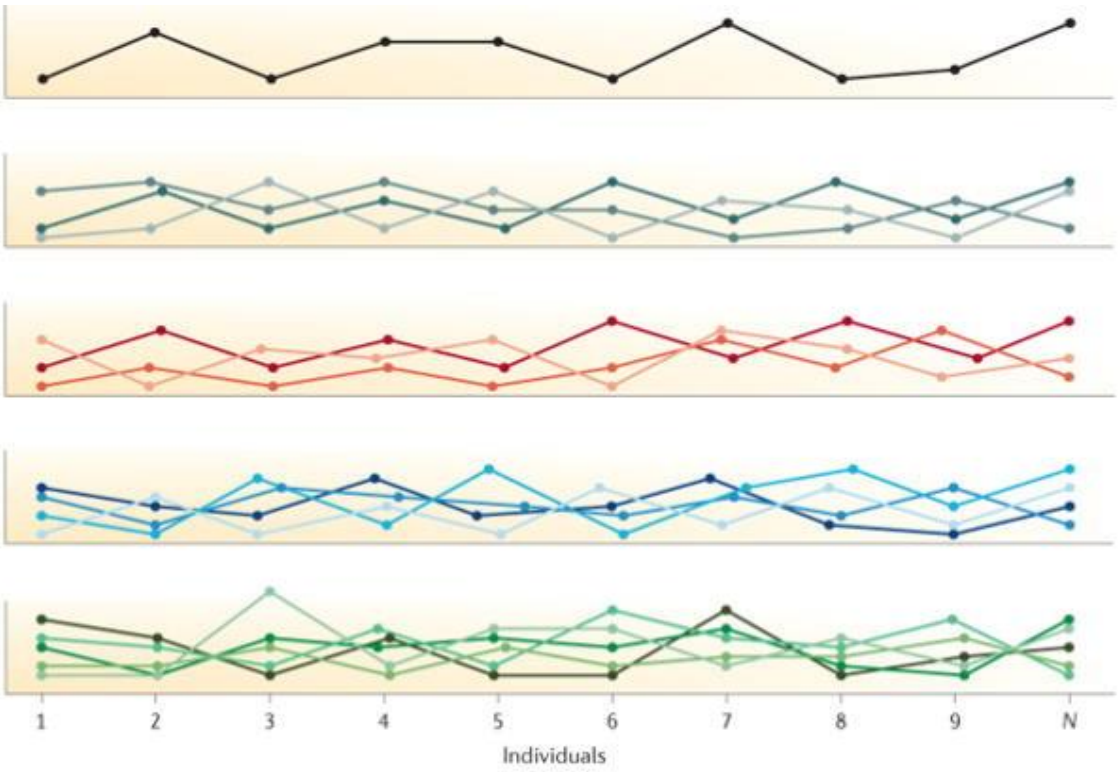
Animal phenotype

Microbial taxa comp.

Blood metabolites

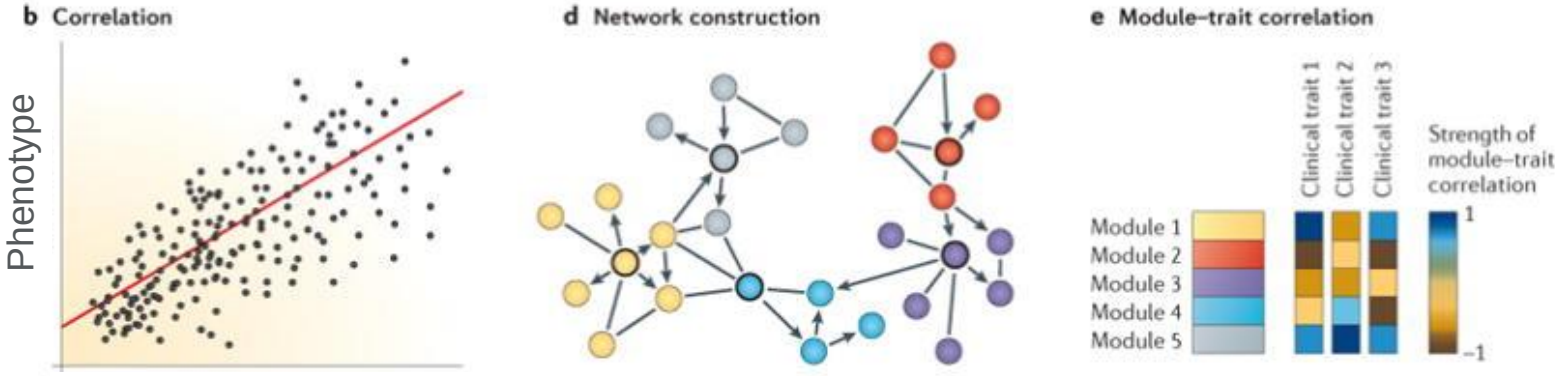
Milk FAs

Transcripts (/tissue)



Genome sequence (SNP)

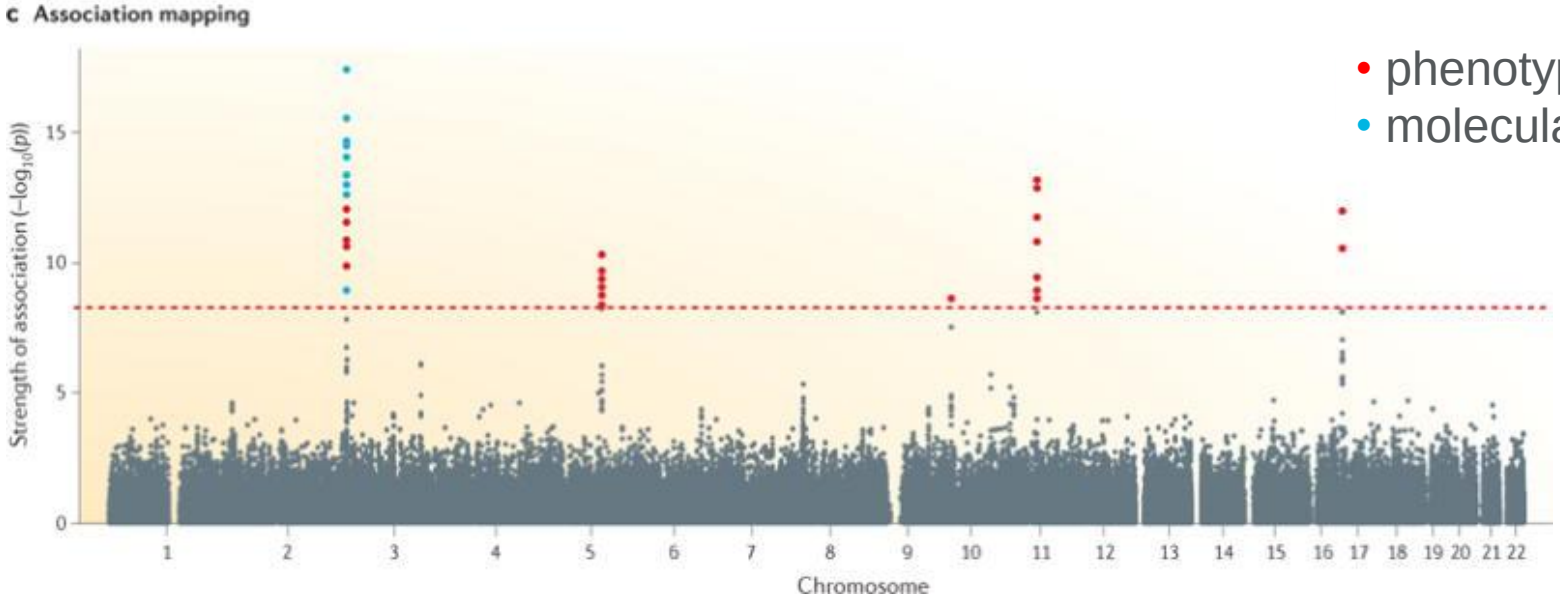
Luke
NATURAL RESOURCES INSTITUTE FINLAND



Measured molecular trait

Genetic interaction network -> modules of co-regulated genes; hub genes

e.g. principal components in a module

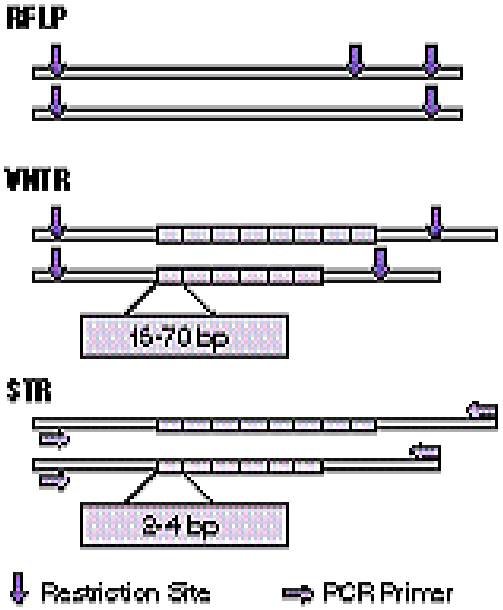


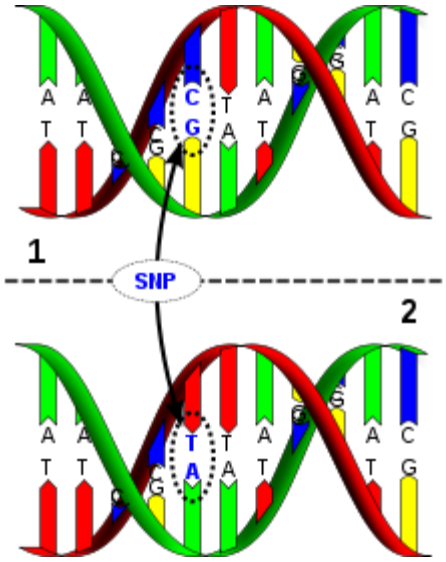
- phenotype
- molecular trait

Causal relationship?

From: Civelek & Lusis, 2014 Nat Rev Genet

DNA markers for detecting sequence variation





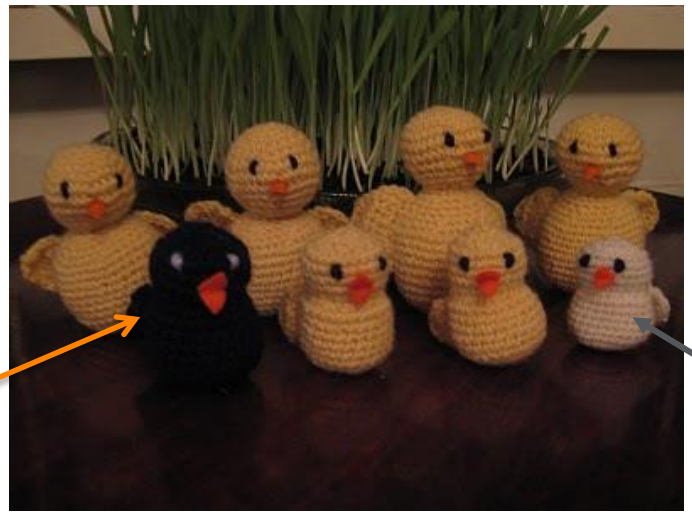
CNV

microsatellite

”SNP”

Detecting association

AA



GG

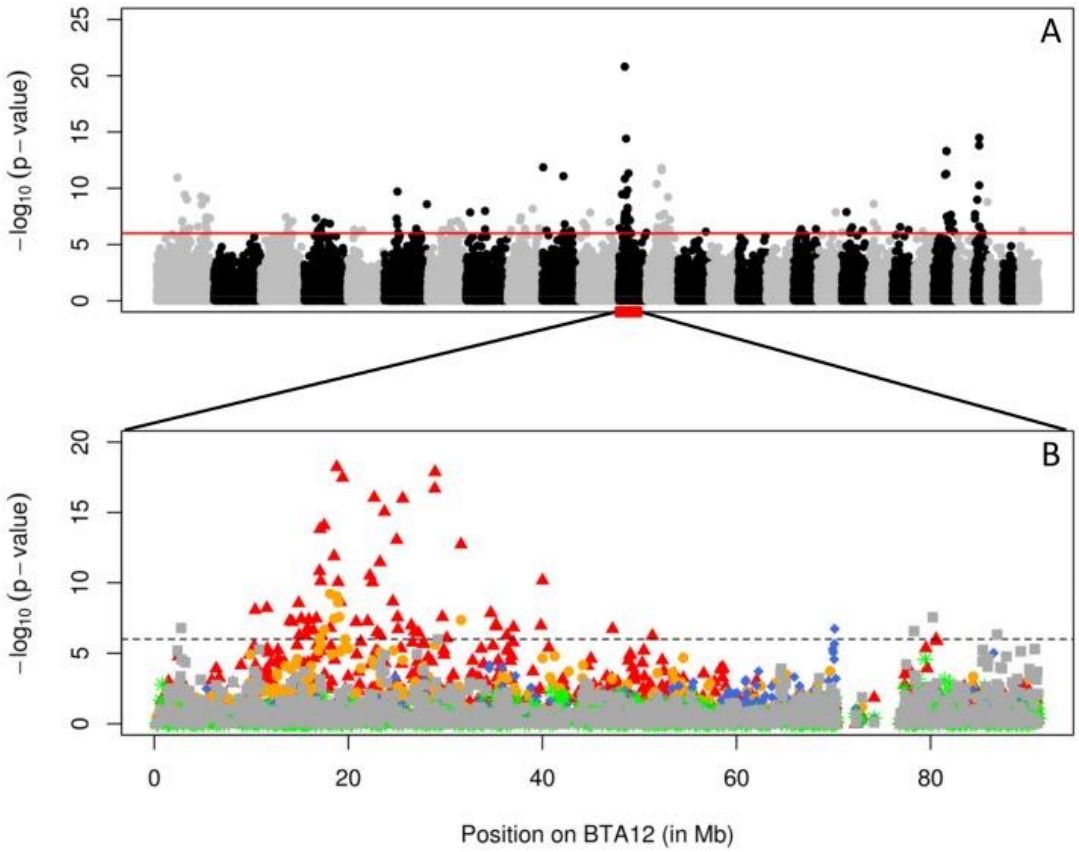
AA

Individuals differing for the phenotype
(hundreds or thousands)

Alleles present in different phenotype groups (frequency)

Luke
NATURAL RESOURCES INSTITUTE FINLAND

”Manhattan plot” across the genome (GWAS)



Open questions 1

- Difficult to pinpoint causative SNPs, or genes
- Most SNPs in (human) GWAS studies map outside protein-coding regions
- Cattle functional regulatory elements have not been well annotated

www.faang.org

FAANG
Functional Annotation of Animal Genomes

A coordinated international action to accelerate genome to phenome

FAANG aims to:
Standardize core assays and experimental protocols
Coordinate and facilitate data sharing
Establish an infrastructure for analysis of these data
Provide high quality functional annotation of animal genomes

Open questions 2

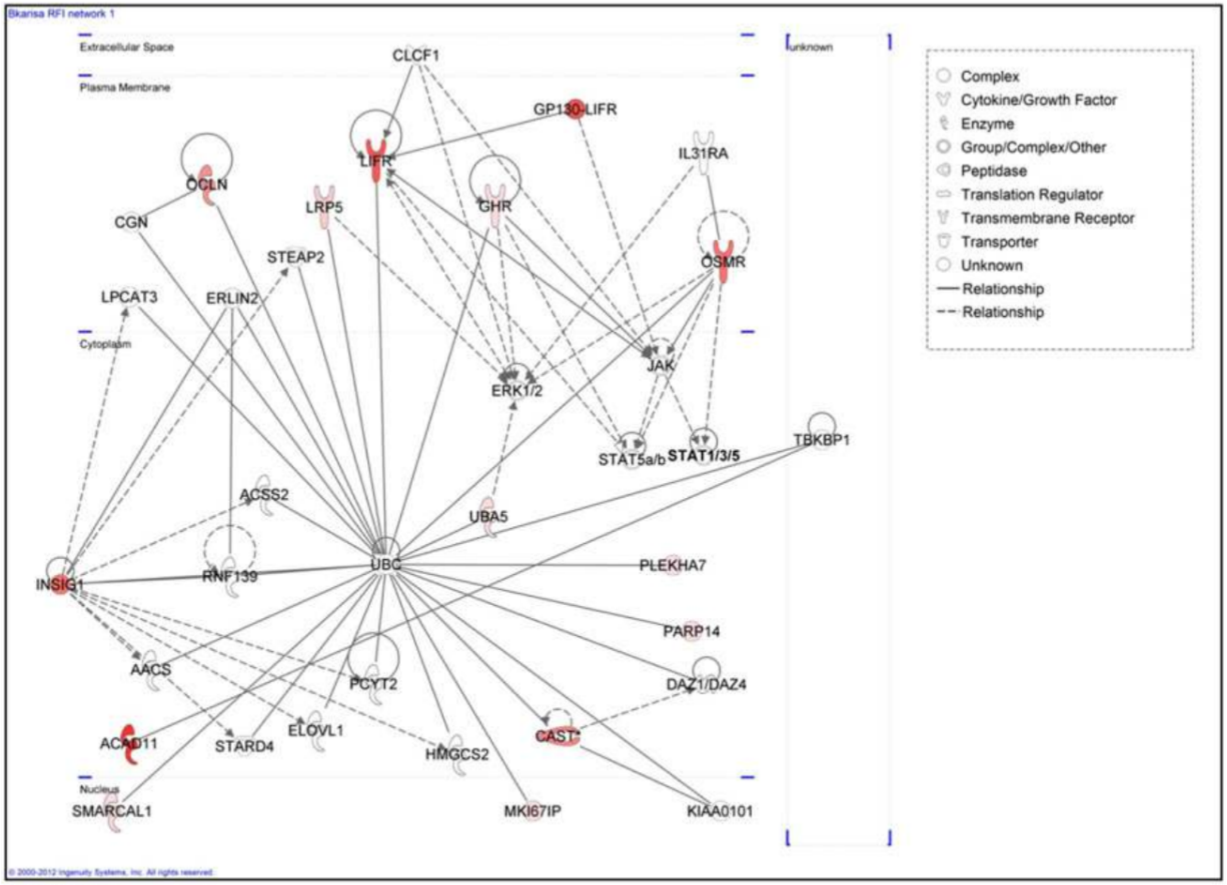
- Usually significant associations are able to explain only a minor part of heritable variation in a complex trait phenotype - “missing heritability”
- Individual variants have small effects
- Epistasis – interactions between genes
- Epigenetics – modification of gene expression without sequence variation, parent-of origin effects

From associations to causal variants to pathways

- Positional candidate genes from associated regions
- Prioritization according to function (inferred from GO terms) associated with trait under study
- E.g. carbohydrate, protein, lipid metabolism or growth, milk production
- Sequence variations at prioritized candidate genes (whole genome sequence data available/ 1000 Bulls Genome consortium)
- Prioritization of sequence variations according to predicted effect (nonsynonymous, missense, splice site)
- Repeat association analysis – construct networks of significantly associated genes

Example of gene network associated with RFI

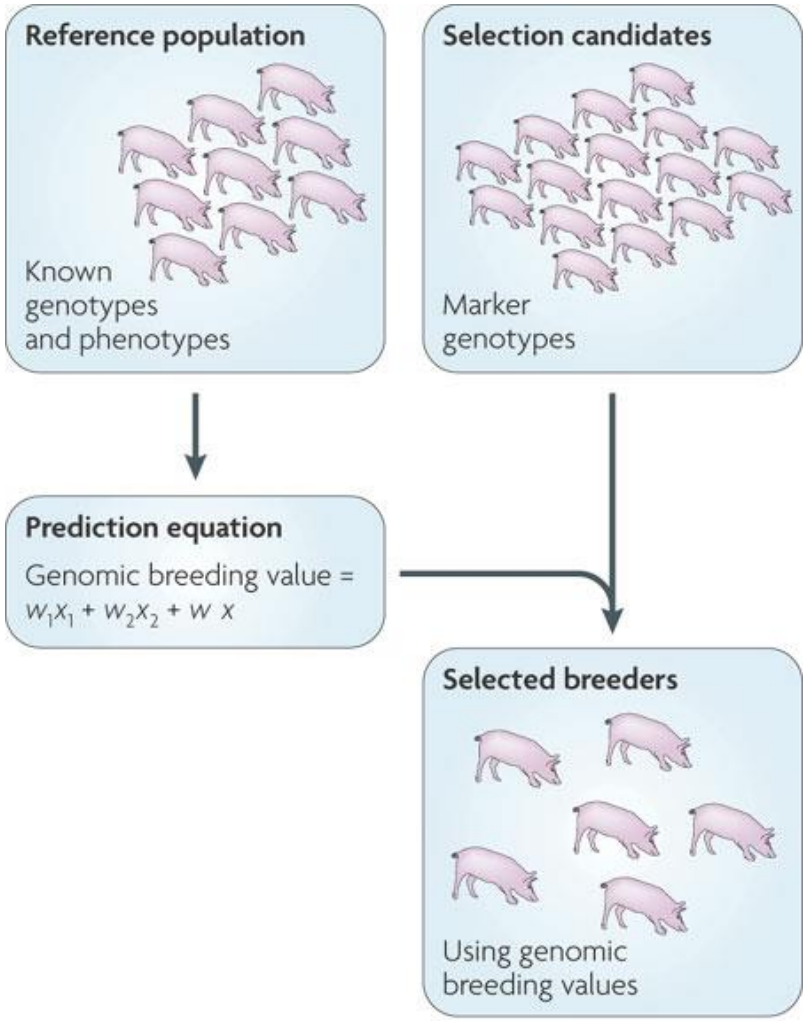
Karisa et al. 2014; Animal Sci J



- 24 genes associated with RFI in beef cattle
- metabolites associated with RFI

- Ingenuity Pathway Analysis (IPA)
- reveals important hub genes (UBC, INSIG) and other genes involved with the pathways
- metabolic networks

Genomic prediction



- densely situated markers (e.g. 50 000 SNPs)
- simultaneous genome-wide estimation of effects requires large reference population (thousands of animals)
- genomic breeding value (GBV)

Goddard & Hayes, 2009 Nature Reviews | Genetics

Accuracies of genomic predictions

- Rely on linkage disequilibrium (LD) between causal variant and closest SNP
- Realized accuracies usually low across populations
- Feed efficiency and feed intake traits challenging, because difficulties in getting phenotype measurements from large numbers of individuals

- Including biological background information and causal variants would improve predictions and reduce dependence on LD


Don’t blame me – I have the best GEBV for methane emissions!

Twin cow experiment

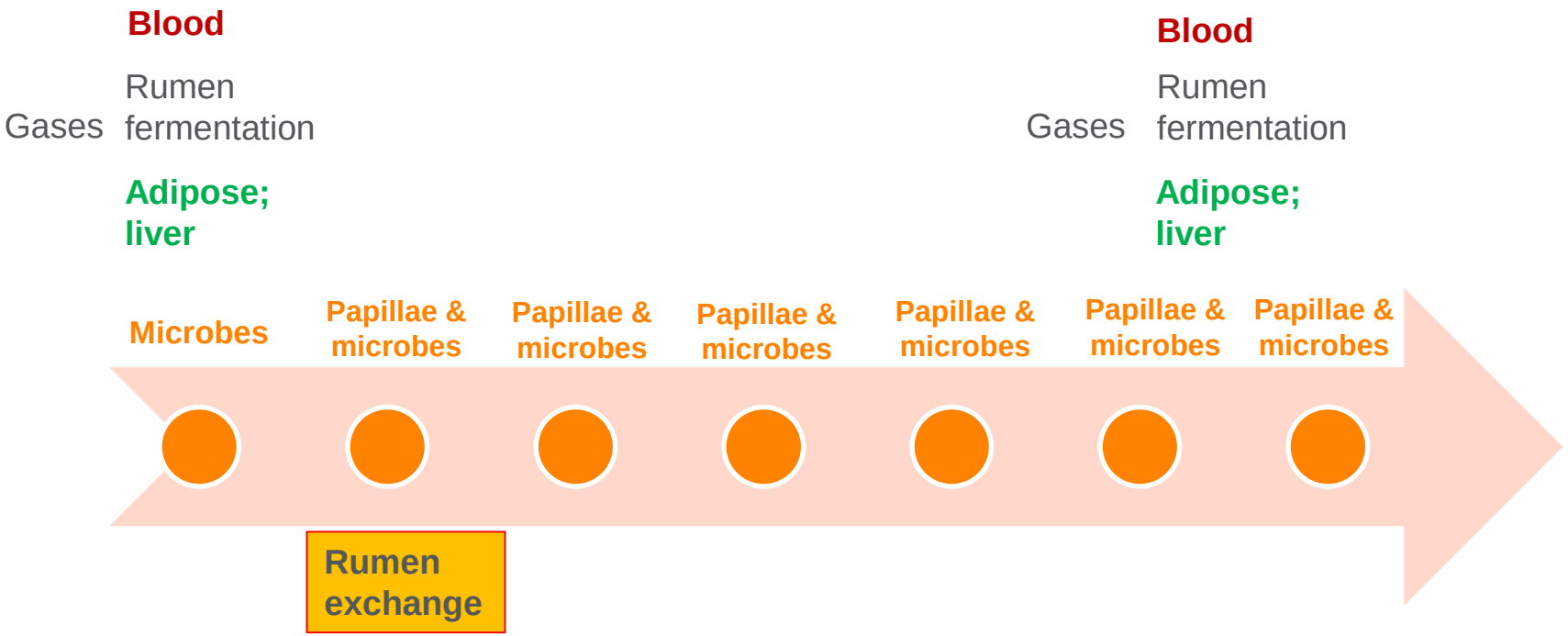
- two pairs of identical twins produced by embryo splitting
- 4 genetically diverse cows
- fitted with rumen cannula
- same diet
- mid lactation
- rumen content exchange

Twin cow experiment - transcriptome studies

- Changes in microbial population may affect rumen epithelium function and thereby nutrient absorption kinetics
- These changes may in turn be reflected in the associated gene networks in the liver and adipose tissue

- The main objective to study the changes in the rumen epithelium, liver and adipose transcriptomes due to total exchange of rumen content -> change in microbial population

Sampling



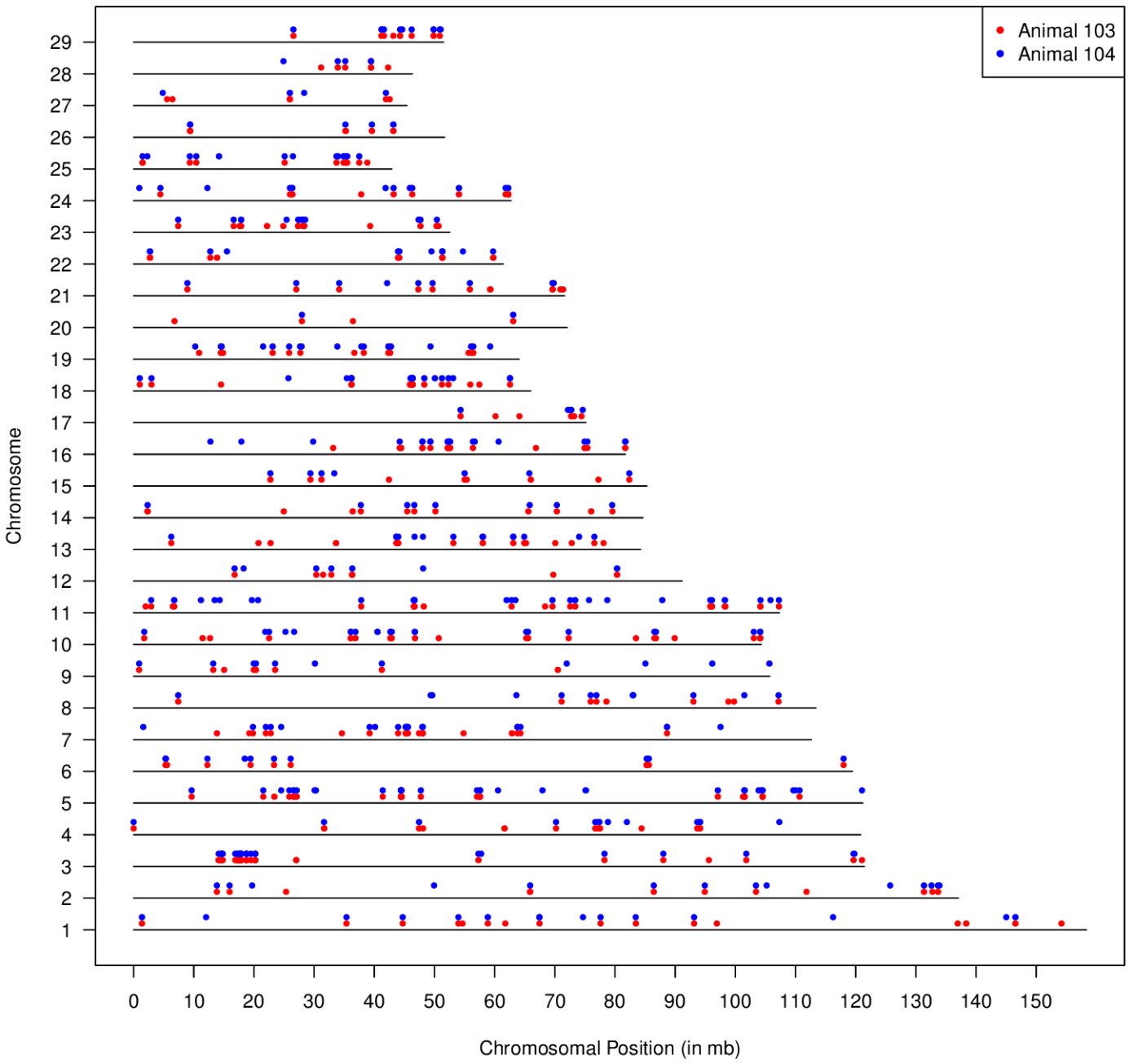
RNA sequencing (transcriptome)

- Adipose & liver: 150 bp paired-end mRNA sequencing Illumina TruSeq/HiSeq3000;
- 33M reads/sample
- Papillae: 150 bp paired-end mRNA sequencing and 50 bp single-end miRNA sequencing, Illumina TruSeq

Papillae transcriptome

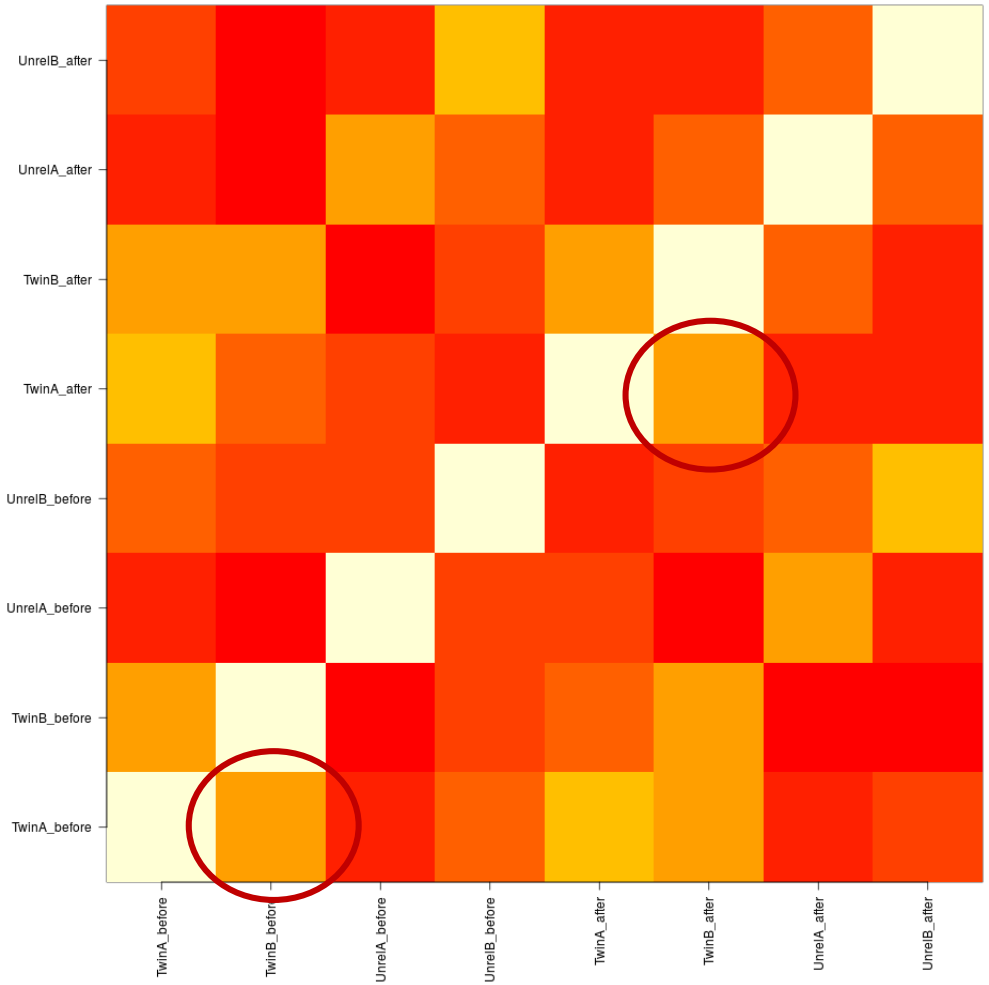
- the papillae have a distinctive expression profile
- approx 2500 genes expressed with 5RPKM cutoff; 80% of reads come from top 5% genes
- Kegg pathways for these genes: Oxidative phosphorylation, nitrogen metabolism, PPAR signaling pathway, Ubiquitin mediated proteolysis, Protein processing in endoplasmic reticulum, Cholesterol biosynthesis

- high number of expressed regions, where so far no known bovine transcripts have been annotated
- characterizing these regions would be highly interesting for understanding the responses in the papillae (and to the annotation of the bovine genome)



Luke
NATURAL RESOURCES
INSTITUTE FINLAND

Spearman correlations – liver gene expression finding genes contributing to differences in gene expression in twins after rumen content exchange



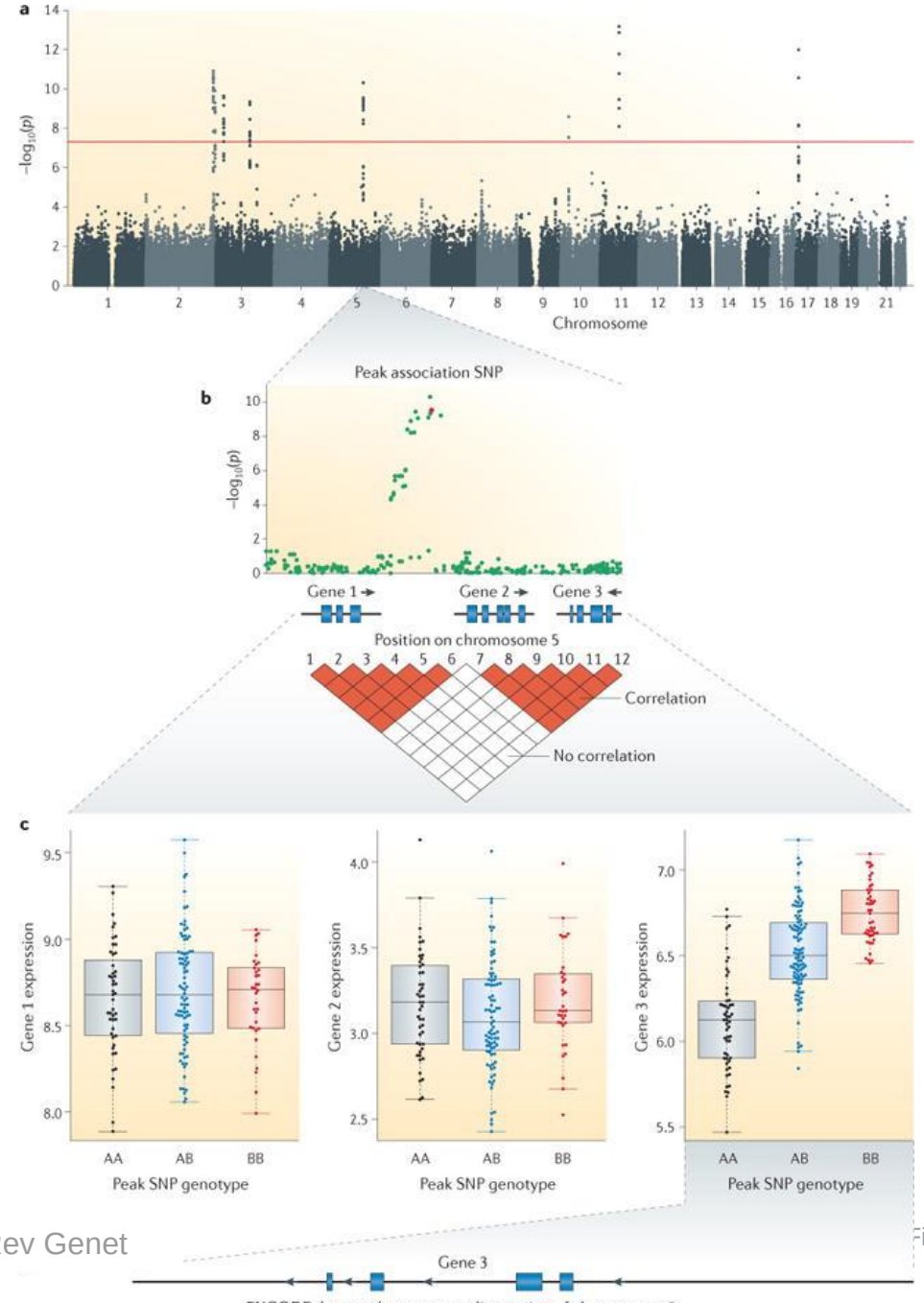
Luke
NATURAL RESOURCES INSTITUTE FINLAND

Outcome

- Genome-wide association results may indicate
  - genes/loci with considerable effect on the phenotypes
  - through which biological pathways or processes the host may regulate rumen microbiome composition
  - which pathways are associated with the specific phenotypes and the relationships between them
- Correlations between multilevel phenotypes and associated genetic loci (possibly causal mutations) will aid in developing genomic prediction equations for efficiency traits
- Transcriptome studies will provide understanding of molecular mechanisms regulating changes in rumen epithelial function in response to variation in microbial composition

Thank you!

RUMINOMICS



From: Civelek & Lusis, 2014 Nat Rev Genet

Finland

Luke
NATURAL RESOURCES
INSTITUTE FINLAND