Heritable variation in gene expression is the key contributor of phenotypic change and genetic diversity in Holstein cattle.

After 15 years of genomic selection, we now know a lot more molecular biology.

Dr. Tom Lawlor
Holstein Association USA
Integrating molecular biology with Fisher’s infinitesimal model.

SNP effects

General population

94%

Single nucleotide polymorphism (SNP)

6%
Genome changes in Holsteins is very dynamic

Rapid changes in allele frequency

Subpopulations constantly forming, mixing and reforming.

Elite Cows – born 2015
Contemporary Cows born around 2000
Bulls born around 1980
Cows with genetics from 1964
RNA → Protein → Metabolites → Phenotype

What’s new in molecular biology

Gene Regulatory Networks
Coordinated response with redundancy

Gene expression
Major driver of genetic change
Gene expression and RNA splicing explain large proportions of the heritability for complex traits in cattle.

Published: October 11, 2023
https://doi.org/10.1016/j.xgen.2023.10038

Averaged proportion of heritability explained by regulatory variants across 37 traits:

- cis eQTL and sQTL: 45.3%
- trans eQTL and sQTL: 23.9%
- NO regulatory control: 30.8%

Variants contributing to heritability:

- regulatory control: 69%
- NO regulatory control: 31%
Gene expression

*Regulation* – when, where and how much of a protein

*Exon splicing* – what type of protein is produced

**Regulation**

- **UP** regulated
- **DOWN** regulated

**Exon Splicing**

- Process involving the joining of exons to produce mRNA.
- Alternative splicing can produce different proteins.

**Different proteins**

- Protein 1
- Protein 2
Large number of possible interactions

*Cis* – nearby

*Trans* – long distance

These loops allow interactions between different regions of DNA

Long range inter-chromosomal interactions
Genetic Redundancy:
only a subset of all these interactions is needed
Hundredsof
millions
of
combinations

Exon splicing
provides more
variation

Wide variety of genetic
material to select from

Population
differentiation

Parents of the next
generation will have a **unique genotype**
Different subpopulations will have different genetic architecture depending on animal’s ancestry.

Different transcripts lead to different gene regulatory networks.
These three breeds are genetically different from each other.

\[ F_{st} \] measures allele frequency differences between populations.

- \[ F_{st} = 0 \] same breed
- \[ F_{st} = 0.07 \] different dairy breeds
- \[ F_{st} = 0.15 \] dairy breed compared to a beef breed
Although centered on the same physiological pathways, set of differentiated genes were almost not overlapping among the breeds. This suggests a kind of plasticity in the genome allowing different solutions to respond to a similar breeding goal.

Each of the breeds has found its own pathway to a genetic solution wired differently.

Connections between genes are different in the different breeds.
The genome-to-phenotype map has a hierarchical organization.

- **DNA sequence**
- **Proteins**
- **Biological pathway**
- **Metabolites**
- **Phenotype**

Different subpopulations can use alternative pathways to achieve the same phenotypic response.
Evolutionary advantages to having redundancy and hierarchical organization
Improved or novel functions can evolve gradually, rather than in a single step.

**Modularity**: components may be separated and recombined, often with the benefit of flexibility and variety in use.

**Coordinated response with redundancy**

- **similar** - compensates for mistakes in other pathways
- **improved** – better use of a similar pathway
- **new** - modified pathway leading to new phenotype (evolution)
Gametic disequilibrium

Different gametes with different combinations of alleles leads to different genetic solutions (pathways) being favored in different subpopulations.

Gamete $A_b c$ is desirable in Family 1
Gamete $a_B C$ is desirable in Family 2

The two subpopulations become genetically different

Rapid changes in genetic architecture are seen in nature:
with selection and population division

The genetics of different subpopulations change differently
even when they’re selected for the same objective
Rapid evolutionary changes in gene expression in response to climate fluctuations

"Biggest heritable change in response to drought was changes in gene expression"

Different subpopulations took different pathways in response to the same drought conditions.
A single population was divided into ten sub-populations.
Then exposed to higher temperatures.

Differences in the genetics of the original founders steers the subpopulation towards using a different set of SNPs to achieve the same phenotypic goal.
Different genotypes leads to different transcripts leading to different proteins leading to different interactions leading to different networks leading to different pathways leading to similar metabolites being utilized to produce the same phenotype.

Lai et al, 2023
By understanding how genetic variation is created and organized, along with the proper management of our population structure, we can achieve rapid genetic progress and maintain genetic diversity.
The average effect of SNP $i$ in a population:

$$\alpha_i = a_i + (1 - 2p_i)d_i + \sum \alpha_{ij}^i$$

- **Main effects**
  - Depends upon frequency of the SNP $p_i$
  - Gene action (additive and dominance) $a_i$ and $d_i$

- **Interactions**
  - Interactions with other SNPs $\alpha_{ij}^i$

We don’t ignore interactions, we averaged them out.
Without epistasis, consistency across families gives us a high SNP effect.

Assuming phenotypic value is related to change in SNP frequency.

With epistasis, the inconsistency across families gives us a much lower SNP effect.

Epistasis appears to be small because we’re looking at the “average effect” across all subpopulations.
Problem with pooling data from multiple subpopulations together

Lose information on interactions

Analyzing Data

In a pooled data set.

We limit our selection to the SNPs associated with core genes or cis-regulation.

The conundrum:
More observations leads to a more accurate prediction.

By pooling, I get a more accurate prediction on a LIMITED set of SNPs.

Separate analysis, I get a less accurate prediction on a LARGER set of SNPs.
Better use of the available genetic variation
Multiple breeding populations, each seeking its own path to obtain a similar goal.

Wright’s *Shifting Balance Theory*

Create large family units.
Each with its own reference population.
Select best within-family animals.

Evaluate animals from all breeding organizations together on a national scale.
Evaluate all animals from all countries for the global population.
2022 Within-stud selection

F_{st} = 0.04

U.S. Average

F_{st} = 0.00

concentrating on their own families

F_{st} = 0.02

German bred

F_{st} = 0.02

F_{st} = 0.03

ST genetics

F_{st} measures genetic distance from overall population

F_{st} = 0.06
Any Questions