



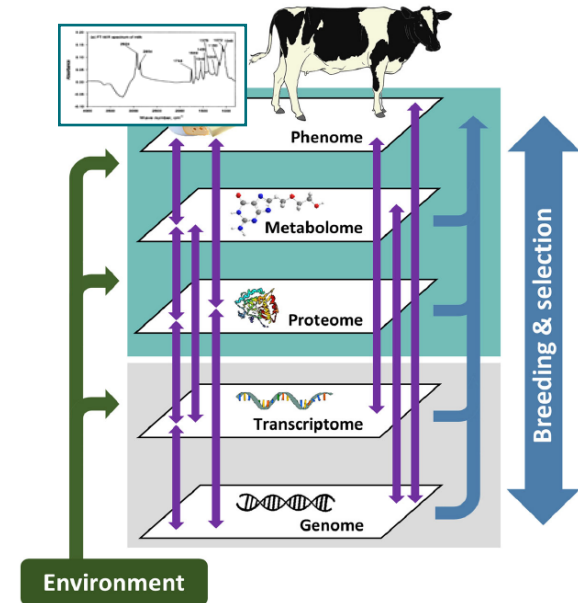
**Synergistic optimization:  
Utilizing genomic insights to enhance  
strategies for joint selective (deep) phenotyping  
and recursive creation of genomic reference populations**

N. Gengler, A. Markey, P. Lemal, Y. Chen, H. Wilmot, M. Schroyen  
ULiège-GxABT, Gembloux, Belgium

# Describing the Original Issue



- Creation of **intermediate phenotypes**
  - – **Omics data** → closer to physiological state

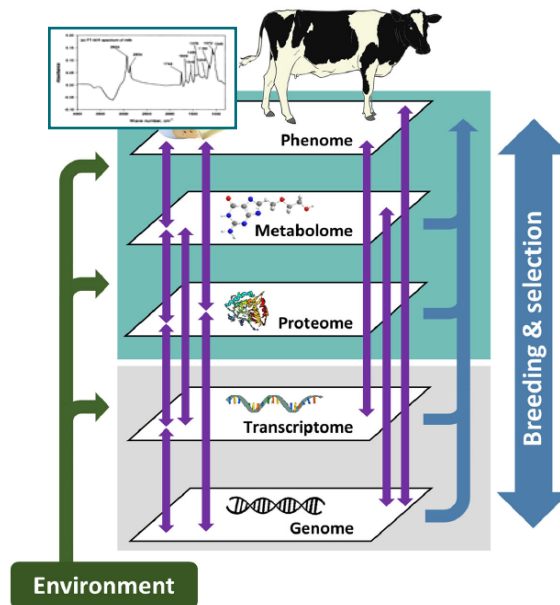


Modified from Tiplady et al. 2020  
<https://doi.org/10.1186/s40104-020-00445-2>

# Describing the Original Issue



- Creation of **intermediate phenotypes**
  - – **Omics data** → closer to physiological state
- Creation of these phenotypes
  - Often labelled “**high-throughput**”
  - Not necessarily **cheap** and **high-throughput**
  - There may be a **confusion**
  - **Deep phenotyping** maybe better word

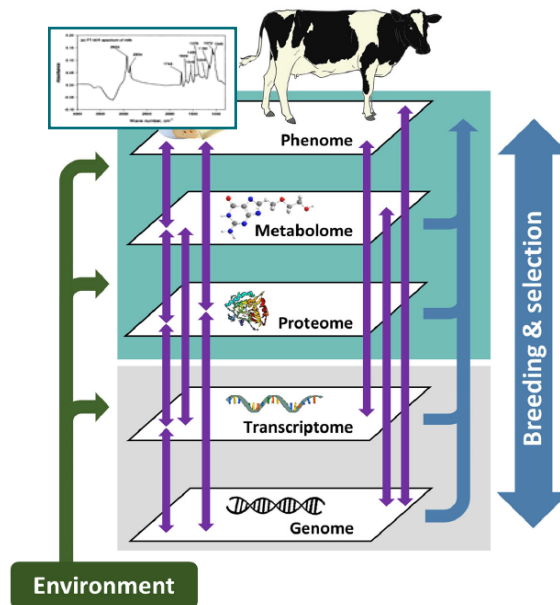


Modified from Tiplady et al. 2020  
<https://doi.org/10.1186/s40104-020-00445-2>

# Describing the Original Issue



- Creation of **intermediate phenotypes**
  - – **Omics data** → closer to physiological state
- Creation of these phenotypes
  - Often labelled “**high-throughput**”
  - Not necessarily **cheap** and **high-throughput**
  - There may be a **confusion**
  - **Deep phenotyping** maybe better word
- Today several **important keywords**
  - -Omics insights ← genomics
  - Selective (deep) phenotyping
  - Recursive creation of genomic reference populations

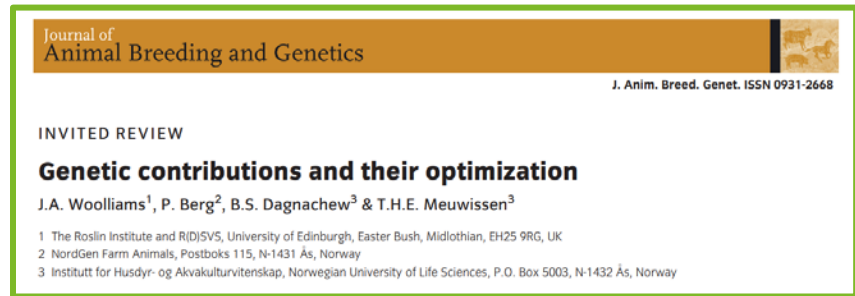
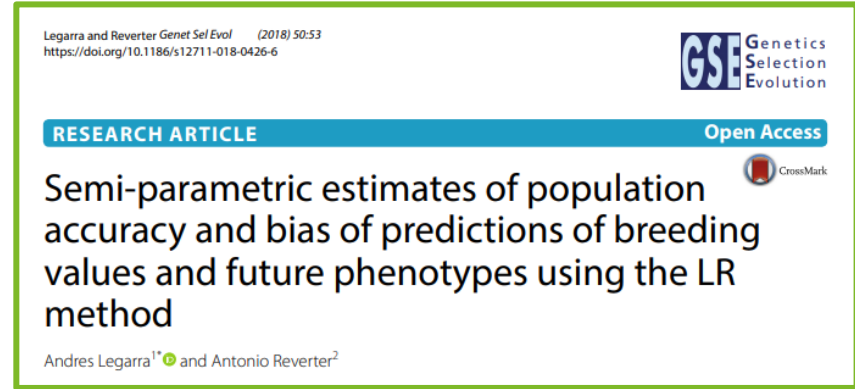


Modified from Tiplady et al. 2020  
<https://doi.org/10.1186/s40104-020-00445-2>

# Recursive Creation of Genomic Reference Populations



- **Adding** to an existing data **new data should not create bias**, this is related to the LR method →
- Genetic **“diversity”** as added animals should **represent the whole population** this is related to Optimum Contribution (OC) →



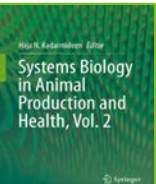
# Example Interest of –Omics → Context of Systems Biology



- Systems biology → **Holistic view**
  - **Interdisciplinary field** focusing on **complex interactions** within biological systems
- **Methods and Tools** driven
  - **High-Throughput** Technologies → **see last slide**
  - **Bioinformatics** Tools:
    - › Software and algorithms for analyzing and interpreting biological data
  - **Mathematical** Modeling:
    - › Mathematical frameworks → represent biological processes
- Examples often in **complex contexts** as diseases →

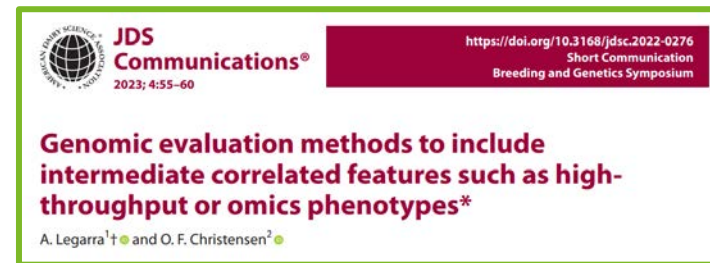
Schroyen, M., Liu, H., Tuggle, C.K. (2016).  
Applications of Systems Biology to Improve Pig Health

[https://doi.org/10.1007/978-3-319-43332-5\\_2](https://doi.org/10.1007/978-3-319-43332-5_2)



# Example Interest of –Omics → Context of Genetic Evaluations

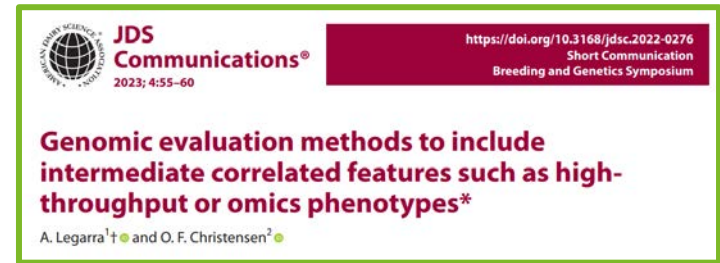
- **Past studies** largely focused on ways to integrate to gain **new knowledge**
- But also, to support more reliable **genomic evaluations**





# Example Interest of –Omics → Context of Genomics

- **Past studies** largely focused on ways to integrate to gain **new knowledge**
- But also, to support more reliable **genomic evaluations**



- With some **interesting methodological** ideas ← **missing –omics data**
  - **Predicting individual contributions** of –omics to phenotypes
  - Inferring **missing individual contributions** (e.g., similar to single-step GBLUP)

# Deep Phenotyping (-Omics and Beyond)



- Challenging in **real-life situations** for several reasons:
  - Biological variability
  - Data complexity and volume
  - Technical and analytical challenges
  - Computational and infrastructural limitations
- Fortunately:
  - **Advancements in technology and computational methods**
  - **Collaborative efforts** (e.g., EU Cost Action EU-LI-PHE <https://eu-li-phe.eu/>)
- Often still **requires massive investments**:
  - **Limiting** effective and widespread use under field and experimental conditions
  - **Sampling needed!**



## Extensions of This Issue

- Problem extends to other **simpler “complex phenotype”-taking** situation as
  - **Limited available facilities** → feed intake, methane emissions
  - **Expensive** molecular or wet-chemistry **analysis**
  - General use of **expensive sensors** → **on which animals?**
    - › **Digital phenotypes** → presentation later by Ching-Yi Chen
    - › **Uncertainty of identification** → presentation later by Juan P. Steibel



## Extensions of This Issue

- Problem extends to other **simpler “complex phenotype”-taking** situation as
  - **Limited available facilities** → feed intake, methane emissions
  - **Expensive** molecular or wet-chemistry **analysis**
  - General use of **expensive sensors** → **on which animals?**
  - Situations where producers want to **predict performances** of actual animals of the next generation **but need validations**



## Extensions of This Issue

- Problem extends to other **simpler “complex phenotype”-taking** situation as
  - **Limited available facilities** → feed intake, methane emissions
  - **Expensive** molecular or wet-chemistry **analysis**
  - General use of **expensive sensors** → **on which animals?**
  - Situations where producers want to **predict performances** of actual animals of the next generation **but need validations**
- Even with **all the possible investments**
  - Many phenotyping efforts **limited by poor choice of animals!**



# Extensions of This Issue

- Problem extends to other **simpler “complex phenotype”-taking** situation as
  - **Limited available facilities** → feed intake, methane emissions
  - **Expensive** molecular or wet-chemistry **analysis**
  - General use of **expensive sensors** → **on which animals?**
  - Situations where producers want to **predict performances** of actual animals of the next generation **but need validations**
- Even with **all the possible investments**
  - Many phenotyping efforts **limited by poor choice of animals!**
- Also: directly linked to **creation of genomic reference populations**



# Deep Phenotyping Few in a Given Population

- Many **sampling methods**, some examples:
  - **Random sampling**
    - › Unbiased but high sampling variance
  - **Stratified random sampling**
    - › Stratified random sampling decreases the standard error of our sample estimators vs. simple random sampling
  - **Systematic sampling**
    - › E.g., ordered (randomly) than  $n^{\text{th}}$  member chosen
  - **Cluster sampling**
    - › Population divided into non-overlapping groups or clusters
    - › Choosing cluster than individuals inside clusters
  - ....



# Deep Phenotyping Few in a Given Population

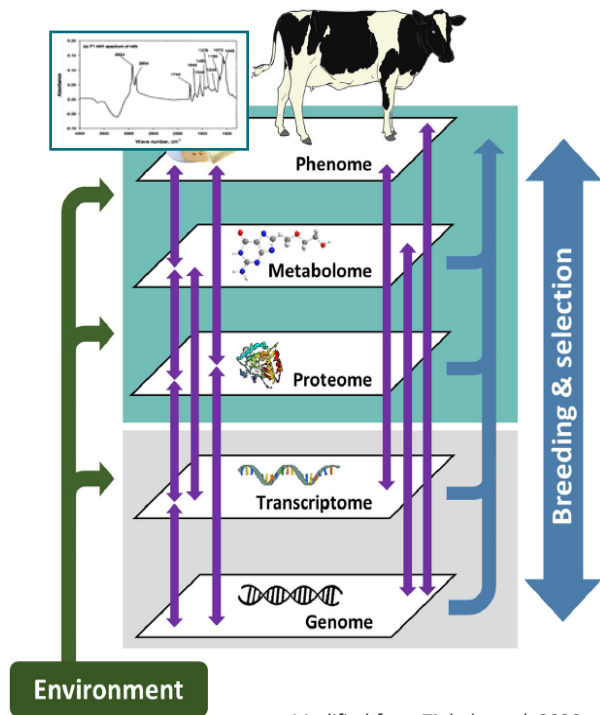
- Many **sampling methods**, some examples:
  - **Random sampling**
    - › Unbiased but high sampling variance
  - **Stratified random sampling**
    - › Stratified random sampling decreases the standard error of our sample estimators vs. simple random sampling
  - **Systematic sampling**
    - › E.g., ordered (randomly) than  $n^{\text{th}}$  member chosen
  - **Cluster sampling**
    - › Population divided into non-overlapping groups or clusters
    - › Choosing cluster than individuals inside clusters
- What is lacking → **use of genomic knowledge**



# Genome ↔ Phenome (G2P)

- **Some facts:**
  - **Genotyping** in many cases currently **much cheaper** than **phenotyping**
  - And can be performed at **early stages before phenotyping**
- Important question:
  - **Can genomics help organize animal phenotyping in an efficient manner?**
- Therefore, the main **objective** of this presentation:
  - **Initiate discussions** how genomics could help optimize selective phenotyping
  - Some **practical examples** from the research in our group
- Explains the first part of the title:  
**“Utilizing genomic insights to enhance selective phenotyping strategies”**

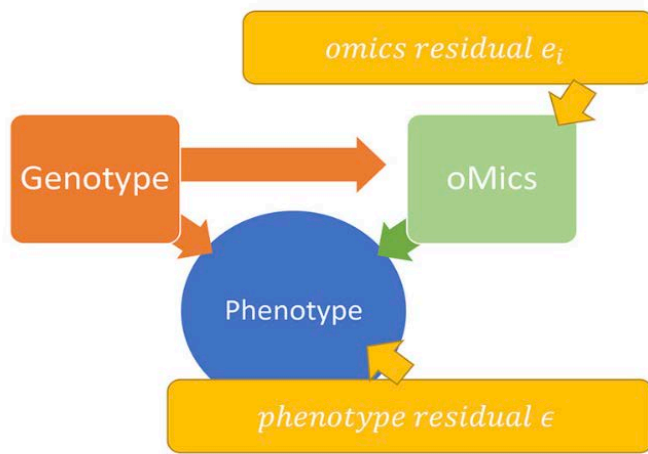
# Genomics Insights ⇔ Other -Omics



Modified from Tiplady et al. 2020  
<https://doi.org/10.1186/s40104-020-00445-2>

- Question already asked before:
  - **Can we predict individual contributions of –omics to phenotypes?**
- Answer:
  - If genetic variation explains –omics contribution to phenotypes  
→ **yes, it should be possible**

# Genomics Insights ↔ Other -Omes



$$y = X\beta + ZM\alpha + Z_r\alpha_r + \epsilon$$

$$m_i = \tilde{X}\tilde{\beta}_i + \tilde{Z}g_i + e_i,$$

Models for prediction of breeding value using Omics

Modified from Legarra and Christensen 2023  
<https://doi.org/10.3168/jdsc.2022-0276>

- Question already asked before:
  - Can we predict individual contributions of –omics to phenotypes?
- Answer:
  - If genetic variation explains –omics contribution to phenotypes  
→ yes, it should be possible
- Open possibilities to select animals for (deep) phenotyping

# For the Transcriptome: A Recent Paper (2023)



Cell Genomics

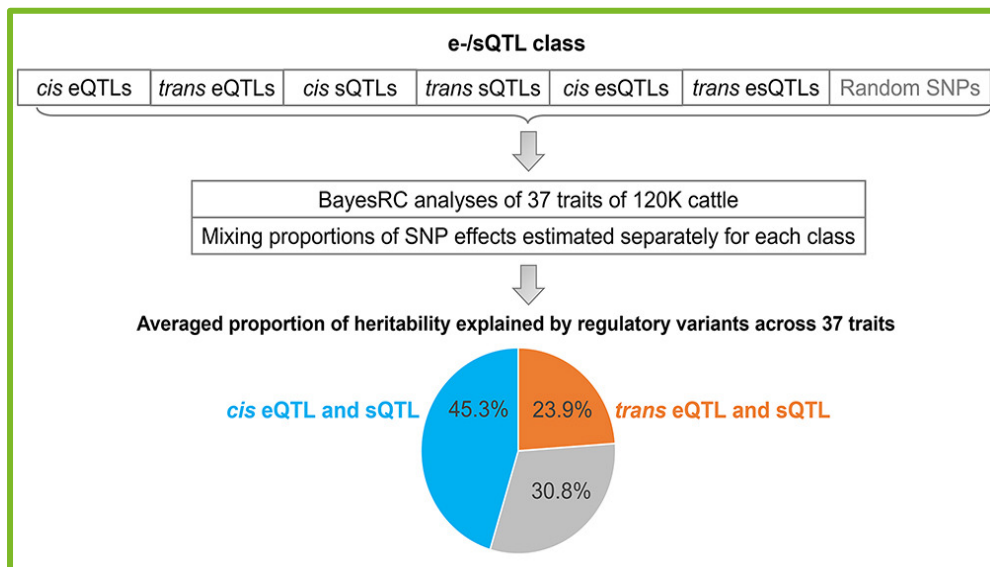
<https://doi.org/10.1016/j.xgen.2023.100385>



## Article

### Gene expression and RNA splicing explain large proportions of the heritability for complex traits in cattle

Ruidong Xiang,<sup>1,2,3,12,\*</sup> Lingzhao Fang,<sup>4,5</sup> Shuli Liu,<sup>6</sup> Iona M. Macleod,<sup>2</sup> Zhiqian Liu,<sup>2</sup> Edmond J. Breen,<sup>2</sup> Yahui Gao,<sup>7</sup> George E. Liu,<sup>7</sup> Albert Tenesa,<sup>4,8</sup> CattleGTEx Consortium, Brett A. Mason,<sup>2</sup> Amanda J. Chamberlain,<sup>2,11</sup> Naomi R. Wray,<sup>9,10</sup> and Michael E. Goddard<sup>1,2</sup>





# For the Proteome and the Metabolome

- Both also under direct and/or indirect **genetic control**
- **Proteome** → Protein QTL (pQTL) → protein levels
  - Example of Casein QTL → Casein variants → cheese yield ↗



# For the Proteome and the Metabolome

- Both also under direct and/or indirect **genetic control**
- **Proteome** → Protein QTL (pQTL) → protein levels
  - Example of Casein QTL → Casein variants → cheese yield ↗
- **Metabolome** → Metabolite QTL (mQTL) → metabolite levels
  - Linking genetic variations to specific metabolic pathways, physiological traits



# Example Selective Phenotyping Strategy → Discussions



1. **Objective** → getting n “new” phenotypes
2. Using **genomics**:
  - Basic principle:
    - › Selecting “new” animals based on predicted phenotypes (or GEBV) of the trait to phenotype
  - Proposal:
    - › Systematic sampling of animals sorted by predicted phenotypes (or GEBV)



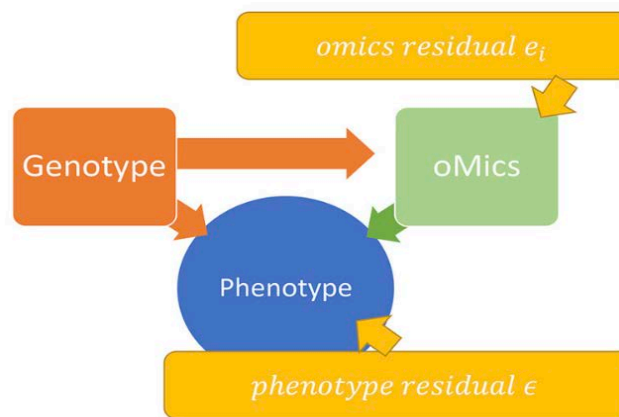
# Example Selective Phenotyping Strategy → Discussions

1. **Objective** → getting n “new” phenotypes

2. Using **genomics**:

- Basic principle:
  - › Selecting “new” animals based on predicted phenotypes (or GEBV) of the trait to phenotype
- Proposal:
  - › Systematic sampling of animals sorted by predicted phenotypes (or GEBV)

3. Extension to **intermediate phenotypes** →



$$y = X\beta + ZM\alpha + Z_r\alpha_r + \epsilon$$

$$m_i = \tilde{X}\tilde{\beta}_i + \tilde{Z}\mathbf{g}_i + e_i,$$

Models for prediction of breeding value using Omics

NB: GEBV vs.  $\hat{y}$  or  $\hat{m}$  ? ← **depends on situation**

Modified from Legarra and Christensen 2023  
<https://doi.org/10.3168/jdsc.2022-0276>

# Example Selective Phenotyping Strategy → Discussions



1. **Objective** → getting n “new” phenotypes
2. Using **genomics**:
  - Basic principle:
    - › Selecting “new” animals based on predicted phenotypes (or GEBV)
  - Proposal:
    - › Systematic sampling on animals sorted by predicted phenotypes (or GEBV)
3. Extension to **intermediate phenotypes**

# Example Selective Phenotyping Strategy → Discussions



1. **Objective** → getting  $n$  “new” phenotypes
2. Using **genomics**:
  - Basic principle:
    - › Selecting “new” animals based on predicted phenotypes (or GEBV)
  - Proposal:
    - › Systematic sampling on animals sorted by predicted phenotypes (or GEBV)
3. Extension to **intermediate phenotypes**
4. Required: **maintaining genetic diversity**
  - › Minimizing average coancestry  $C = \mathbf{c}'\mathbf{H}\mathbf{c}/2$  under the restriction that  $\mathbf{c}'\mathbf{c} = n$  where  $\mathbf{H}$  = combined relationship matrix,  $\mathbf{c}$  = column vector of 0 and 1
5. Needed **optimizing** 2. (3.) and 4. jointly
  - Idea → adding condition  $\mathbf{c}'\mathbf{g} = \text{average}(\mathbf{g})$  where  $\mathbf{g}$  = vector of GEBV

## Several Practical Example Directly From My Group



- Examples not (yet) linked to **deep phenotyping**
- But nice examples....  showing that the **presented issues appear very quickly!**

# Pauline Lemal: Finding Animals Worth Equipping Sensors



- Studying Heat Stress (HS) distinguishing through **objective measures**  
e.g., sensors

- Interest of using sensors to assess behavior under HS was already shown →



- Problem:
  - **Given the costs of sensors which herds / animals to equip?**
- Solution:
  - Predict **HS susceptibility based on available data**, e.g., MIR-based biomarkers
  - Select **most variable herds**
  - Adopt appropriate **sampling strategies inside herds**



# Yansen Chen: CH<sub>4</sub> and Thousand MIR-Traits Models

➤ **Not only in -omics  
potentially many traits** →

➤ Yansen's context



- **Focal traits** → direct observed CH<sub>4</sub> → breeding objective traits
- **Secondary traits** → routinely available MIR spectral traits → index traits

➤ Problem

- Correlated MIR-predicted CH<sub>4</sub> not direct trait but **measuring expensive / difficult**
- **Need to detect relevant animals to be directly phenotyped for CH<sub>4</sub>**

➤ Solution

- Using appropriate model to **predict GEBV for CH<sub>4</sub>**
- **Selecting variable herds** for sniffer/GF installation ← **genetic proximity** to all cows



# Alice Markey: Ascertaining Boar Taint from Fat Analysis



- Interplay between
  - Human nose scores (HNS) → **standard** for boar taint detection
  - Needed for **validation** → scatole (SCA) and androstenone (AND) in fat
- Problem
  - Alice's context: **PI boars are tested batch by batch** using X animals (PI x LR)
  - HNS → **easy** to obtain
  - But SCA and AND in fat → **expensive**, often only partial phenotyping possible
- Solution
  - Have HNS on all X animals (all genotyped at birth) and fat samples
  - Compute GEBV for SCA and AND in a 3-trait model with HNS
  - For each new batch → based on GEBV for SCA and AND  
**using rankings, respecting genetic diversity → select samples to analyze**

# Recursive Creation of Genomic Reference Populations

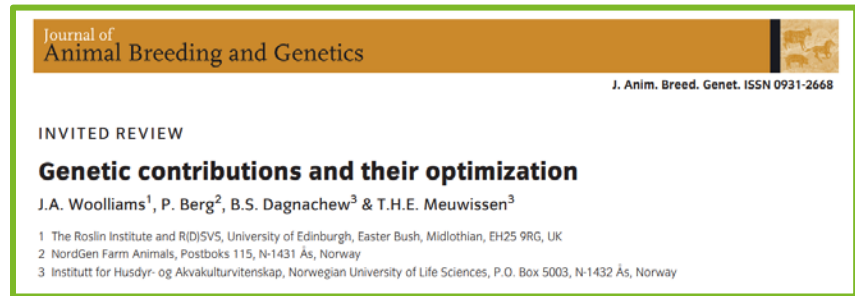
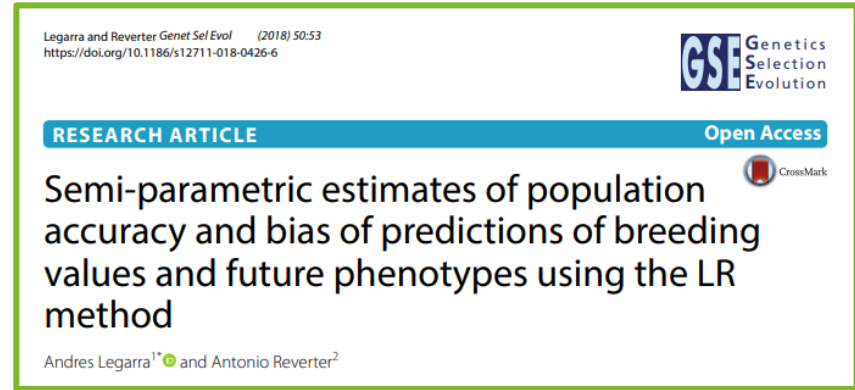


- Context
  - Animal breeding → recursive process → **Alice's case with each batch**
  - **Recursive creation of a pig reference population**
- Obviously
  - (Newly) phenotyped animals → added to reference population
- Explains also this part of the title:
  - **....selective (deep) phenotyping and recursive creation of genomic reference populations....**

# Recursive Creation of Genomic Reference Populations



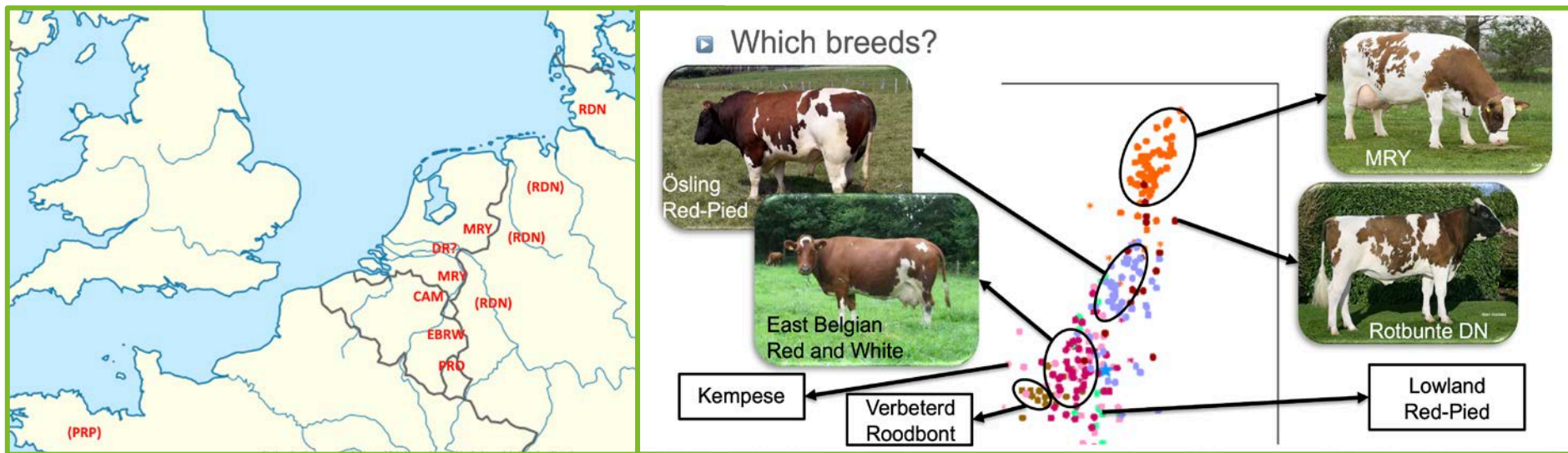
- **Adding** to an existing data **new data should not create bias**, this is related to the LR method →
- Genetic **“diversity”** as added animals should **represent the whole population** this is related to Optimum Contribution (OC) →



# Underlying Hypothesis: Genetic Homogeneity of Population



- Old problem in genomic selection for **smaller breeds**
- Example of the **Red&White (Red-Pied)** non-Holstein breeds in NWEurope





# Hélène Wilmot: Reference Population in R&W DP Cattle

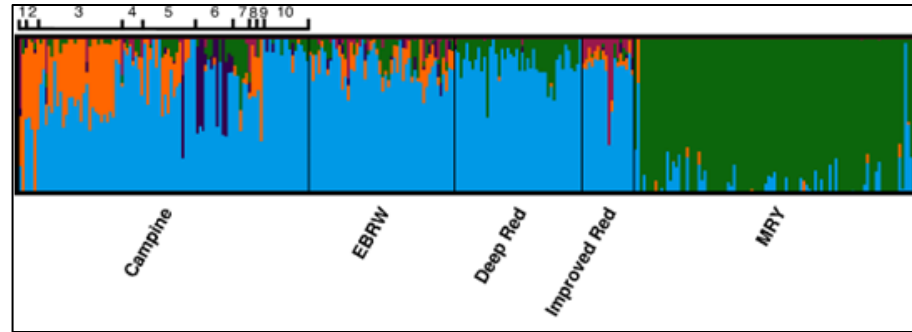


- Supervising ongoing MSc Thesis in Collaboration with WUR
- Hélène's context:
  - **Red&White (Red-Pied) non-Holstein breeds in NWEurope**
  - Breeds are Dual-Purpose and Red&White, share a common history **but are not the same....**

# Red&White (Red-Pied) DP Non-Holstein Breeds



- Genomic differences



- Different breeding objectives



- Even preferred colors

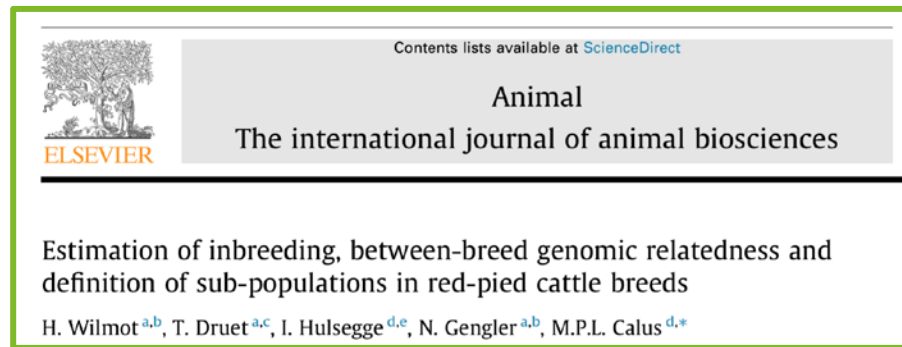




# Hélène Wilmot: Reference Population in R&W DP Cattle



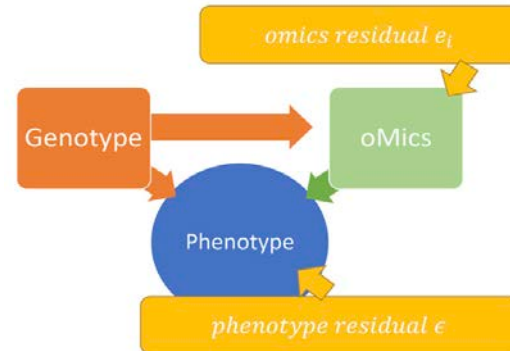
- Supervising ongoing MSc Thesis in Collaboration with WUR
- Hélène's context:
  - **Red&White (Red-Pied) non-Holstein breeds in NWEurope**
  - Breed are Dual-Purpose and Red&White, share a common history **but are not the same....**
  - Main breed of interest **East-Belgian Red&White (EBRW)**
- Needed:
  - Deriving a strategy to add **reference animals (sires) from sister breeds**
  - Without diluting EBRW specificity  
→ **sub-population structure**



# Synergistic Optimization of Selective (Deep) Phenotyping and Recursive Creation of Genomic Reference Populations



- $1 + 1 = 3$  😊 !
- But also: **Better genotyping**  $\leftrightarrow$  **Better phenotyping**
- + Extra benefits  
→ **Predicting future records**  
(cf Donagh Berry's talk)



Modified from Legarra and Christensen 2023  
<https://doi.org/10.3168/jdsc.2022-0276>



**I want to get your opinions....**

# Acknowledgements



- Support of the Futurospectre partnership
  - AWE – Comité du Lait – CRA-W – ULiège-GxABT
- CECI Consortium for computational resources 
- Service Public de Wallonie 
- National Fund for Scientific Research 
- Support by different other Regional and European Projects



**ASAS-CSAS-WSASAS Annual Meeting**  
Calgary, AB, Canada • July 21-25, 2024

# EAAP/ASAS Speaker Exchange Program

**[nicolas.gengler@uliege.be](mailto:nicolas.gengler@uliege.be)**