Grant agreement number: 101000226

H2020 – Research and Innovation Action

Deliverable 4.3

Map of inbreeding depression, deleterious variants and genetic load.

Due date of deliverable: M48 – May 2025

Actual submission date: M48 – May 2025

WP concerned	WP4	
Lead partner	**Gwendal Restoux** **Pascal Croiseau** **Thierry Tribout** **Anne Barbat** **Stephanie Minery**	**INRAE**
Author(s)	**Gwendal Restoux** **Pascal Croiseau** **Thierry Tribout** **Anne Barbat** **Stephanie Minery**	**INRAE**
	Jack Windig **Renzo Bonifazi** **Jeremie Vandenplas**	**WUR**
	Theodorus Meuwissen	**NMBU**
Version	V1	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°101000226.

Dissemination level	PU: Public

Call: H2020-SFS-2020-2

Topic: SFS-13-2020 ı Genome and epigenome enabled breeding in terrestrial livestock

Start date of the project: June 1st, 2021 **Duration:** 60 months

End date of the project: May 31st, 2026 **Project ID:** 101000226

Table of content

1. Summary

This deliverable presents a genome-wide analysis of inbreeding depression and genetic load in five local cattle breeds: Abondance, Tarentaise, Vosgienne (France), Meuse-Rhine-Yssel (Netherlands), and Norwegian Red. The objective was to identify genomic regions carrying deleterious recessive variants affecting traits such as fertility, health, and production.

Inbreeding depression arises from increased homozygosity of harmful alleles. While pedigree-based methods can limit the rise of inbreeding, they do not prevent the accumulation of rare, deleterious mutations. To address this, two complementary GWAS models were developed: a classical additive-dominance model (A_D), and a ROH-based model (ROH and A*_D*) that distinguishes homozygous genotypes occurring within runs of homozygosity (ROH) from those outside. Both models also include a pedigree-based inbreeding coefficient to separate local from global effects.

Data included ~50K SNP genotypes and phenotypes for production, fertility, and udder health traits. After quality control and applying a Bonferroni correction, 108 significant genomic regions were identified: 96 with the ROH model, 22 with A_D, and 18 with A*_D*. Most regions were method-specific, confirming the complementarity of the approaches.

Significant variability was observed between breeds and traits. More associations were detected in Abondance and Tarentaise, particularly for production traits, while few were identified in Vosgienne. The ROH model frequently identified regions where homozygosity due to autozygosity was associated with stronger inbreeding depression, supporting the idea that these regions are more likely to harbor recessive deleterious alleles. Including a global inbreeding coefficient helped ensure that these local effects were not confounded with overall inbreeding levels.

Each method has its own strengths. The A_D model leverages the full dataset and is efficient for detecting general recessive effects but lacks information on local autozygosity. The ROH-based approach targets local inbreeding but may have reduced power due to limited sample size in ROH segments. The A*_D* model increases specificity by focusing on recessive effects within ROH, though it uses fewer data points.

These findings can support more refined breeding strategies. Rather than relying solely on global inbreeding control, breeders can avoid matings that increase homozygosity in high-risk regions. This is particularly useful in small or local populations, where the impact of inbreeding is more pronounced and some deleterious variants may be breed-specific.

The results are consistent with previous RUMIGEN findings on the effects of genomic selection on diversity and inbreeding. Together, they illustrate how genomic selection, while promoting genetic gain, also influences the distribution and expression of genetic load. The identification of inbreeding-sensitive genomic regions provides valuable tools for designing sustainable selection programs that preserve genetic health alongside performance.

Teams involved:

The RUMIGEN partners involved and that contributed to this deliverable are: Institut national de la recherche agronomique (INRAE) with Pascal Croiseau (as Task Leader), Gwendal Restoux, Thierry Tribout, Anne Barbat and Stéphanie Minery, Wageningen Research (WR) with Jack J. Winding, Renzo Bonifazi and Jeremie Vandenplas, and Norwegian University of Life Sciences (NMBU) with Theodorus Meuwissen.

2. Introduction

Inbreeding depression remains a major concern in livestock breeding, as it leads to reduced fertility, productivity, and overall fitness. This reduction in performance is driven by the expression of deleterious recessive alleles when homozygosity increases due to mating between related individuals. This effect over the whole genome, known as genetic load, is difficult to tackle directly, as the causal mutations are often rare and difficult to detect using conventional genome-wide association study (GWAS) methods.

The control of inbreeding has been widely studied, particularly through the use of optimal contribution strategies that minimize pedigree-based coancestry while maintaining genetic gain (Meuwissen, 1997; Colleau et al., 2006). Although these strategies can limit the increase in inbreeding over time, they do not prevent the gradual accumulation of deleterious mutations. Empirical and theoretical studies suggest that such accumulation contributes to the long-term persistence of genetic load, even in populations under effective inbreeding control.

To quantify this load, Muller (1950) introduced the concept of lethal equivalents, defined as a group of deleterious alleles that would cause death if homozygous. In practice, the number of lethal equivalents is inferred from the slope of the regression of survival or fertility traits on individual inbreeding coefficients. While this approach offers a useful summary measure, it is limited to fitness-related traits and lacks resolution at the genomic level.

More recent studies have developed models to estimate individual genetic load more explicitly. Varona et al. (2019) proposed a pedigree-based framework that decomposes inbreeding into contributions from founders and non-founders, capturing the Mendelian transmission of coancestry. In this model, genetic load is treated as an additive effect transmitted through generations. In parallel, bioinformatic approaches aim to assess genetic load through predicted variant effects—e.g., missense or loss-of-function mutations—and conservation scores across species (e.g., GERP, VEP), focusing particularly on coding regions of the genome (Bertorelle et al., 2022).

Complementary to these approaches, runs of homozygosity (ROH) have emerged as a genomic signature of autozygosity—stretches of the genome inherited identically by descent. ROH-based methods have been proposed to study inbreeding depression more locally along the genome. Pryce et al. (2012) introduced a GWAS framework distinguishing a global inbreeding covariate and regional ROH effects, though their model did not consider dominance. Paul et al. (2022) and Doekes et al. (2018) applied models associating trait variation with ROH segments, but neither included global inbreeding covariates, potentially confounding local and genome-wide effects. Moreover, these models often ignore the recessive nature of deleterious alleles and the need to distinguish between ROH-mediated homozygosity and incidental cases of homozygosity.

To address these limitations, we developed a novel GWAS model designed to identify genomic regions involved in genetic load by explicitly modelling both dominance and inbreeding effects. Our model introduces an innovative structure in which the genotype of each variant is classified based on whether it occurs within a ROH, thereby capturing homozygosity by descent (HBD). Importantly, we also incorporate global inbreeding coefficients derived from pedigree information, ensuring that the ROH-based effects are statistically independent from overall inbreeding. This combined modelling of additive, dominance, and ROH-specific effects offers a more accurate framework to detect deleterious recessive alleles and better understand the genetic architecture of inbreeding depression.

This methodology was applied to production and health traits in heifers from five local cattle breeds: Abondance, Tarentaise, and Vosgienne in France; Dutch Meuse-Rhine-Yssel (MRY) in the

Netherlands; and Norwegian Red in Norway. The study aimed to (i) evaluate the contribution of ROH in GWAS models for detecting genetic load, (ii) assess the added value of distinguishing between genotypes inside and outside ROH, and (iii) identify candidate genomic regions potentially contributing to inbreeding depression across different breeds and environments.

3. Materials and Methods

Populations and phenotypes

This study included genotypic and phenotypic data from five local dairy or dual-purpose cattle breeds originating from France, the Netherlands, and Norway. These breeds represent contrasting situations in terms of population size, selection pressure, management practices, and implementation of genomic selection.

- MRY (Meuse-Rhine-Yssel) is a Dutch dual-purpose breed originating from central Netherlands. Genomic selection began in 2018, and animals were genotyped at ~50K SNPs.
- Norwegian Red Cattle (NRF) is the main dairy breed in Norway. Genotypes were originally obtained at high density (777K SNPs), but a subset of 50K SNPs was retained for analysis. NRF had the largest effective population size among all five breeds.
- Abondance (ABO) is the fourth largest French dairy breed, adapted to alpine environments, producing ~6,000 kg of milk per year. Genomic selection began in 2014.
- Tarentaise (TAR) is another alpine French dairy breed, similar in management to ABO, but with a smaller population and average milk production of ~4,800 kg/year. Genomic selection began in 2014.
- Vosgienne (VOS) is a small French dual-purpose breed from the Vosges massif, producing ~4,200 kg of milk annually. It is the only French breed in which all animals are genotyped. Genomic selection also started in 2014.

For the French and NRF breeds, phenotypic records considered were production and udder health traits whereas for the MRY breed, the traits considered were production and fertility. Phenotypes were pre-adjusted for fixed effects such as herd, year, and season. For the French and NRF breeds, the response variable used was the yield deviation (YD), associated with an appropriate reliability weight for the French and NRF breeds. In contrast, YD values were used without applying individual reliability weights for MRY. As each animal had a single phenotypic record, the information content was comparable across individuals. Furthermore, variance components were re-estimated based on YD values, such that the residual variance implicitly captured individual-level variation. The resulting estimated breeding values and variances were highly consistent with those obtained using original phenotypes, supporting the use of unweighted YD values in this context.

Genotyping and ROH detection

Animals were genotyped with 50K SNP chips (or imputed to this density when needed). Genotypes were phased and imputed using breed-specific reference panels. Runs of homozygosity (ROH) were detected using PLINK. A ROH was defined as a continuous stretch of at least 15 homozygous SNPs spanning a minimum length of 1000 kb, with a maximum inter-SNP distance of 150 kb, a minimum SNP density of one SNP per 75 kb, and no heterozygous calls allowed within the ROH.

Pedigree-based inbreeding

Pedigree records were used to compute the pedigree-based inbreeding coefficient (F_PED) for each individual. This allowed us to include a measure of global inbreeding that is independent from genomic information. F_PED was calculated using the tabular method and incorporated in the model as a covariate.

Statistical model

Genome-wide associations studies were performed, considering 2 models.

The aim of the first model was to search for variants whose allele with an unfavourable effect on the phenotype is recessively determined. For this purpose, the response variable yi for individual i and variant k was modelled as:

$$y_i = \mu + F_{ped_i} f + u_i + g_{ik} a_k + h_{ik} d_k + e_i \ , \qquad \text{(M1)}$$

where

- μ is the overall mean;
- f is the pedigree inbreeding regression coefficient whose covariate is F_{ped_i}, aimed to correct data for global inbreeding effect;
- u_i is the additive breeding value of animal i, calculated as $\sum_j (x_{ij}\alpha_j)$ over a set of around 15,000 SNPs with MAF greater than 0.10 and evenly distributed over the 29 autosomes, x_{ij} being the centered genotype of animal i at marker j, and α_j being the additive random effect of marker j with MAF p_j ($\alpha_j \sim N(0, {\sigma_g^2}/{\sum_j 2p_j(1-p_j)})$;
- e_i is the residual ($e_i \sim N(0, \sigma_e^2)$;
- a_k and d_k are, respectively, the additive and dominance fixed effects of variant k;
- g_{ik} is number of alleles 2 of individual i at variant k (0,1,or 2);
- h_{ik} = 0 if g_{ik} = 0 or 2, and h_{ik} = 1 if g_{ik} = 1.

The second model was designed to estimate the effect of local inbreeding by considering the effect on the phenotype whether or not a variant is in a ROH segment. To do this, we estimated for each variant the effect of homozygous genotypes 11 in a ROH segment (11R) or not (11), homozygous 22 in a ROH segment (22R) or not (22), in deviation from the effect of heterozygous genotype (12). Here, the response variable yi for individual i and variant k was then modelled as:

$$y_i = \mu + F_{ped_i} f + u_i + c_{1_{ik}} E_{11/12_k} + c_{2_{ik}} E_{11R/12_k} + c_{3_{ik}} E_{22/12_k} + c_{4_{ik}} E_{22R/12_k} + e_i \ , \qquad \text{(M2)}$$

where

- μ, F_{ped_i}, f, u_i and e_i are the same as in first model;
- $E_{11/12_k}$, $E_{11R/12_k}$, $E_{22/12_k}$ and $E_{22R/12_k}$ are, respectively, the differences in effect between genotypes 11 and 12, between genotypes 11R and 12, between genotypes 22 and 12, and between genotypes 22R and 12;.
- $c_{1_{ik}}$, $c_{2_{ik}}$, $c_{3_{ik}}$ and $c_{4_{ik}}$ are variables (0 or 1) indicating the individual's genotype and ROH status at variant k. For example, [$c_{1_{ik}}$, $c_{2_{ik}}$, $c_{3_{ik}}$, $c_{4_{ik}}$]=[0,1,0,0] if individual i has genotype 11R at variant k, and [$c_{1_{ik}}$, $c_{2_{ik}}$, $c_{3_{ik}}$, $c_{4_{ik}}$]=[0,0,0,0] if individual i is heterozygous.

Then, estimates of the $E_{11/12_k}$, $E_{11R/12_k}$, $E_{22/12_k}$, $E_{22R/12_k}$ and their respective prediction error (co)-variances (PEV) were combined to obtain estimates of the effect of being in a ROH versus not being in a ROH while being homozygous 11 (ROH11) or homozygous 22 (ROH22) at variant k and their PEV then P-values:

$\widehat{ROH}11_k = \hat{E}_{11R/12_k} - \hat{E}_{11/12_k}$, and $\widehat{ROH}22_k = \hat{E}_{22R/12_k} - \hat{E}_{22/12_k}$.

A drawback of the M1 model is that if the linkage disequilibrium between a marker and the causal mutation being sought is low, then individuals with genotypes 11, 12 and 22 at the marker are each a mixture, in varying proportions, of genotypes aa, Aa and AA at the causal mutation, reducing detection power. We therefore defined an additive effect (a*) and a dominance effect (d*) and their respective P-values by considering only the contrasts at a k variant between homozygous individuals 11 and 22 in a ROH and heterozygous individuals, based on the assumption that in a ROH segment the LD between marker and causal mutation is stronger:

$\hat{a}_k^* = 0.5 * \hat{E}_{22R/12_k} - 0.5 * \hat{E}_{11R/12_k} = \frac{1}{2}[(2\hat{a}_k^* + 0\hat{d}_k^*) - (\hat{a}_k^* + \hat{d}_k^*)] - \frac{1}{2}[(0\hat{a}_k^* + 0\hat{d}_k^*) - (\hat{a}_k^* + \hat{d}_k^*)]$;

$\hat{d}_k^* = -0.5 * \hat{E}_{22R/12_k} - 0.5 * \hat{E}_{11R/12_k} = -\frac{1}{2}[(2\hat{a}_k^* + 0\hat{d}_k^*) - (\hat{a}_k^* + \hat{d}_k^*)] - \frac{1}{2}[(0\hat{a}_k^* + 0\hat{d}_k^*) - (\hat{a}_k^* + \hat{d}_k^*)]$.

The PEV of $\hat{E}_{11R/12_k}$ and $\hat{E}_{22R/12_k}$ were combined accordingly to obtain $\hat{a}_k^*$ and $\hat{d}_k^*$ PEV and then P-values.

To define statistical significance, we applied a Bonferroni correction based on the number of SNPs tested. A corrected p-value threshold of 0.05 corresponds to a raw p-value of 10^{-6}, or a –Log(p-value) threshold of 6. This threshold was used uniformly across all models (ROH, A_D, A*_D*) and for all traits and breeds.

Characterization of variants of interest:

Model M1: a variant k was considered significant if [-LPV($\hat{a}_k$)>Thr & -LPV($\hat{d}_k$)>Thr & $\hat{d}_k$ is favourable] ; the idea is to look for variants with recessive determinism and whose recessive allele is unfavorable for the trait, and this approach will be referred as “A_D” below.

Model M2: Two approaches were considered. A variant k was considered significant either if:

- [-LPV($\widehat{ROH}11_k$)>Thr & $\widehat{ROH}11_k$ is unfavourable] or [-LPV($\widehat{ROH}22_k$)>Thr & $\widehat{ROH}2_k$ is unfavourable] ; the idea here is to look for unfavourable effect of ROH (*i.e.* local inbreeding), and this approach will be referred as “ROH” below;

  or

- [-LPV($\hat{a}_k^*$)>Thr & -LPV($\hat{d}_k^*$)>Thr & $\hat{d}_k^*$ is favourable]. Again, the idea is to look for variants with recessive determinism and whose recessive allele is unfavourable for the trait, and this approach will be referred as “A*_D*” below.

Quality control and significance threshold definition

After merging the SNP genotypes within each breed, quality control procedures were applied to filter variants and individuals prior to association testing. Only autosomal SNPs were retained. For each of the three French breeds (Abondance, Tarentaise, Vosgienne) and for the NRF breed, genotyping was performed using the 50K SNP chip, and quality filtering included the removal of SNPs with a call rate below 95%, minor allele frequency (MAF) below 1%, or Hardy-Weinberg equilibrium p-value below 10^{-6}. Individuals with a genotype call rate below 90% were also excluded. After filtering, the number of SNPs retained for analysis ranged from 45,265 (Vosgienne) to 49,424 (Abondance) for the French breeds and is about 45,255 for NRF.

For the MRY population, genotypes were imputed at 44,645 autosomal SNP for 4,645 purebred animals (pedigree-based MRY composition ≥87.5%). The SNP dataset was checked for duplicated

markers (1 SNP removed), and quality control was also conducted resulting in a final dataset of 40,959 SNPs (Table 1).

Definition of genomic regions

Genomic regions associated with inbreeding depression were defined based on the positions of detected SNPs. SNPs located less than 1 Mb apart were grouped into the same region, using a ±1 Mb window around each SNP. This grouping was performed separately for each trait. Afterwards, regions identified for different traits were compared based on their genomic positions, and any overlapping regions were merged and considered as a single, unified region. This approach allowed the identification of genomic segments potentially involved in inbreeding depression, both within and across traits.

4. Results

Population structure and quality control

Table 1 presents the number of SNPs retained after quality control and the number of genotyped animals analyzed per trait and breed. Quality control included filtering SNPs with minor allele frequency below 0.01, call rate below 95%, and Hardy-Weinberg equilibrium p-value below 10^{-6}. After these filters, the number of SNPs retained per breed ranged from 40,959 (MRY) to 49,424 (Abondance).

The number of animals with available phenotypes varied substantially across breeds and traits. For instance, more than 14,000 genotyped animals were available for production traits and somatic cell score in Abondance, compared to fewer than 3,500 in the Vosgienne breed. In the MRY breed, 3,022 animals were used for milk, fat, and protein yield analyses, and 2,603 for conception rate.

Breed	Nb SNP	Number of genotyped animals with phenotype considered in the GWAS			
		MY,FY, PY [1]	FC, PC,	SCS [2]	CR [3]
MRY	40,959	3022	-	-	2603
Norwegian Red	45255	6580	-	6576	-
Abondance	49,424	14,241	14,241	14,927	-
Tarentaise	46,758	7095	7095	7604	-
Vosgienne	45,265	3361	3361	3396	-

[1] MY = Milk Yield, FC = Fat Content, PC = Protein Content, FY = Fat Yield, PY = Protein Yield
[2] SCS = Somatic Cell Score [3] CR = Conception rate (success/failure) at the first insemination

Table 1. Summary of genotyped animals and retained SNPs per breed and trait after quality control.

Number and distribution of significant variants

The results from the GWAS show clear differences between methods and a strong heterogeneity across breeds and traits. As illustrated in the Figure 1, models ROH and A_D (corresponding to the M1 model) detect a much larger number of significant SNPs (–Log(p-value) > 6) than the A*_D*

model (M2). Across all breeds and traits combined, ROH and A_D identify 831 and 893 significant SNPs, respectively, whereas A*_D* detects only 43. This discrepancy may reflect the more conservative nature of the M2 model or lower statistical power due to a limited number of “usable” data (i.e. contrasts where computed only considering SNP within ROH). In contrast, M1 and ROH are more sensitive to detecting regions associated with inbreeding depression and additive or dominance effects.

The Figure 2 highlights the marked variation between breeds and traits. For example, the A_D model identifies 65 significant SNPs for milk yield in Abondance, and 63 for protein content in Tarentaise, while only 2 SNPs are detected for somatic cell count in Vosgienne. Likewise, the ROH model finds 60 SNPs associated with protein content in Tarentaise, but none for some traits in other breeds. These disparities likely reflect differences in trait architecture, selection history, and population size, which impact both the presence and detectability of deleterious variation.

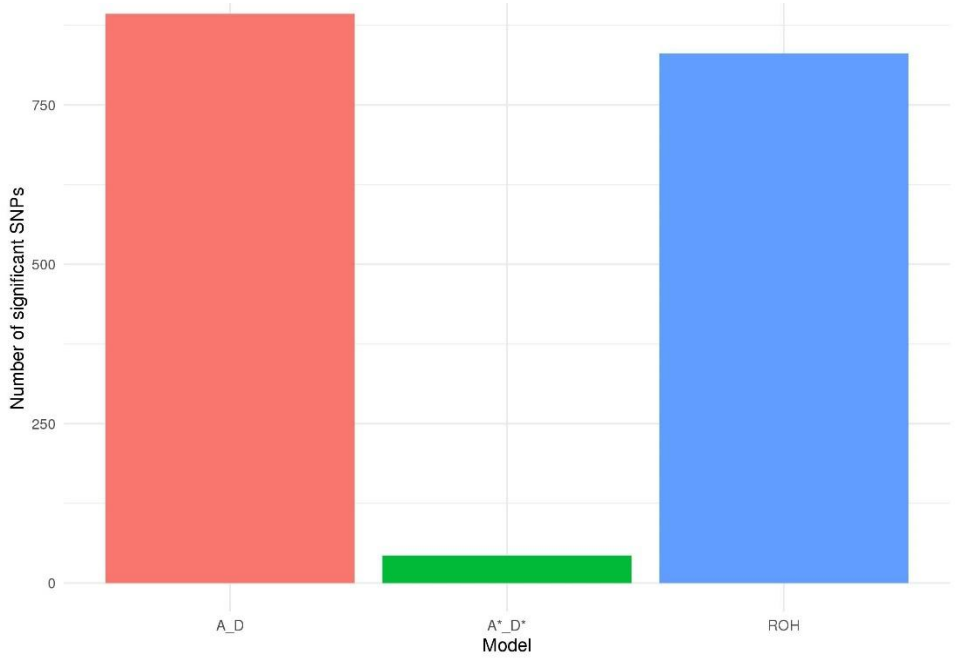

Figure 1: Total number of significant SNPs detected (–Log(p-value) > 6) by model, all traits and breeds combined

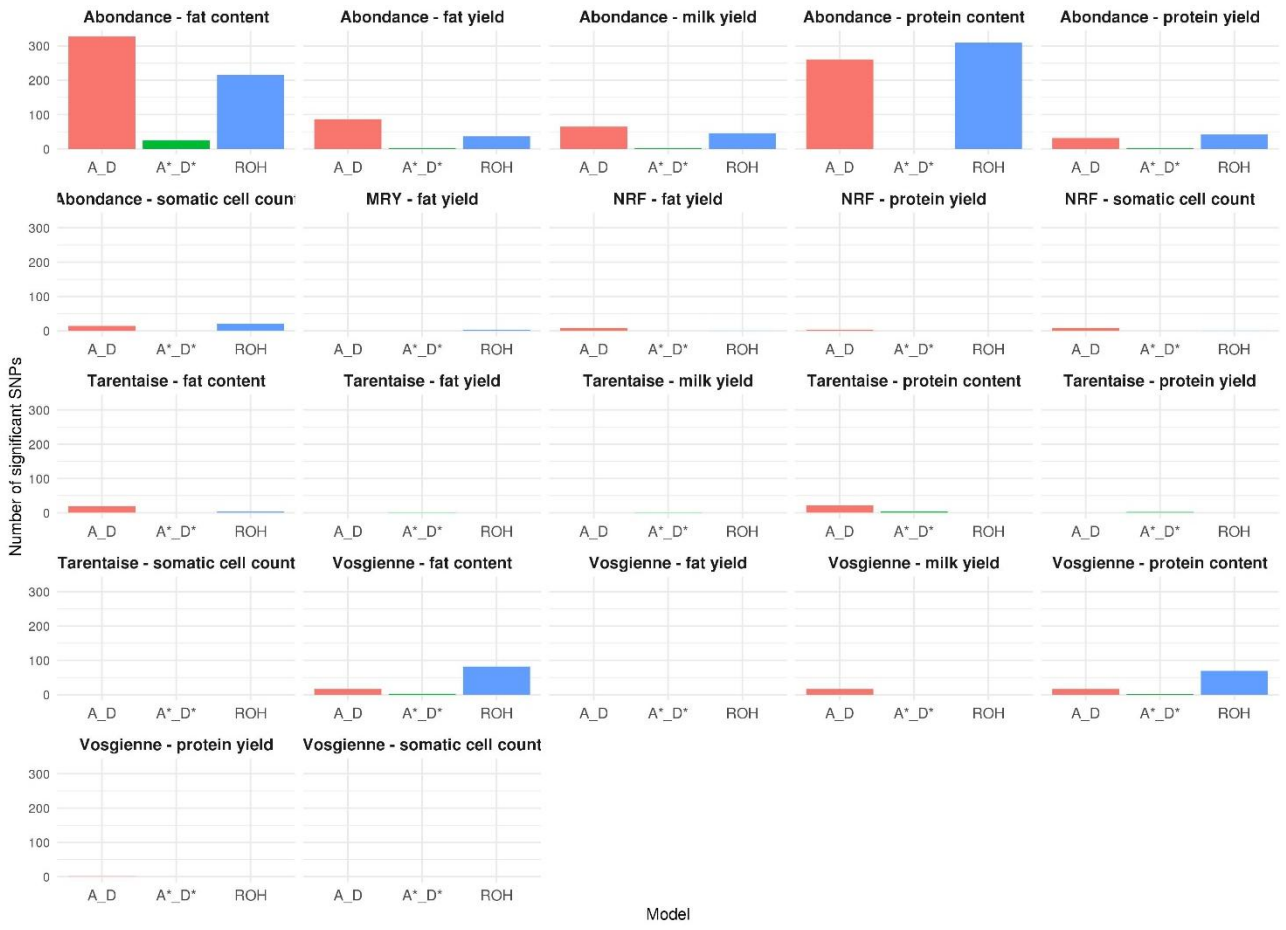

Figure 2: Number of significant SNPs detected (–Log(p-value) > 6) by trait and breed across models

Genomic regions and overlap between methods

Variants were considered significant when the negative $\log_{10}$(p-value) exceeded 6, which corresponds approximately to a 5% genome-wide significance threshold after Bonferroni correction, given the number of variants tested per breed. The genomic positions of all detected regions across the three methods (ROH, A_D, A*_D*) for the three French breeds (Abondance, Tarentaise, Vosgienne), the Dutch breed MRY and for the seven studied traits (milk yield, fat yield, protein yield, fat content, protein content, somatic cell score and conception rate) are detailed in Annex Table 1. To facilitate the interpretation of these results, a summary of the total number of detected regions by breed, trait, and method is provided in Table 2. For each combination, the table also reports how many regions are shared with at least one other method, indicating the level of concordance across analytical strategies.

A total of 108 regions were identified across all traits, breeds, and methods. Among these, 96 regions were detected by the ROH model, 22 by A_D, and 18 by A*_D*. The vast majority of regions (92 out of 108) were specific to a single method, highlighting the capacity of each model to capture distinct facets of the genetic load. Only 16 regions were shared between two methods, and 3 were detected simultaneously by all three approaches.

This analysis confirms that most signals are method-specific. However, overlap was observed primarily between models A_D and A*_D*, as well as between A*_D* and ROH, supporting the idea that these approaches are complementary rather than redundant. The presence of shared regions, particularly in known functional areas (e.g., chromosomes 1, 5 or 14), strengthens the confidence in these loci as candidates for inbreeding depression effects.

Method	Race	Trait	Regions	Shared
A_D	Abondance	fat content	2	2
A_D	Abondance	fat yield	2	1
A_D	Abondance	milk yield	2	1
A_D	Abondance	protein yield	2	1
A_D	Tarentaise	fat yield	1	0
A_D	Tarentaise	milk yield	1	0
A_D	Tarentaise	protein content	2	1
A_D	Tarentaise	protein yield	2	0
A_D	Vosgienne	fat content	2	1
A_D	Vosgienne	protein content	1	1
A*_D*	Abondance	fat content	4	4
A*_D*	Abondance	fat yield	2	1
A*_D*	Abondance	milk yield	3	1
A*_D*	Abondance	protein content	2	2
A*_D*	Abondance	protein yield	2	1
A*_D*	Tarentaise	protein content	1	1
A*_D*	Vosgienne	fat content	4	3
A*_D*	Vosgienne	protein content	3	3
ROH	Abondance	fat content	37	4
ROH	Abondance	fat yield	9	0
ROH	Abondance	milk yield	7	0
ROH	Abondance	protein content	17	2
ROH	Abondance	protein yield	4	1
ROH	Abondance	SCC	1	0
ROH	Vosgienne	fat content	5	2
ROH	Vosgienne	protein content	5	2
ROH	MRY	fat yield	1	0
ROH	NRF	fat yield	1	0
ROH	NRF	somatic cell score	2	0

Table 2: Summary of genomic regions detected by three methods (ROH, A_D, and A*_D*) across five breeds (Abondance, Tarentaise, Vosgienne, MRY and NRF) and for the 7 studied traits. The table reports, for each combination of method, breed, and trait, the total number of detected regions and the count of regions shared with at least one other method.

method	number of regions	mean size of the regions (Mb)	mean number of SNP per regions
ROH	96	0.61	7.01
A_D	22	0.76	5.09
A*_D*	18	0.08	2.5

Table 3: Summary statistics of genomic regions identified by method.

The average size and number of significant SNPs per region are summarized in Table 3. The ROH method identified the highest number of regions (96), with an average size of 0.61 Mb and 7.01 significant SNPs per region. Regions detected by A_D (22 regions) were slightly larger on average (0.76 Mb), but contained fewer significant SNPs (5.09 per region). In contrast, A*_D* identified similar number of regions than A*_D* model (18), but with an average size of only 0.08 Mb and around 2.5 significant SNPs per region.

5. Discussion

This study provides a comprehensive genome-wide analysis of inbreeding depression and genetic load in five local cattle populations by applying two distinct GWAS models, each designed to capture different dimensions of deleterious variation. Rather than relying on a single statistical framework, we considered 3 approaches:

- a classical additive/dominance model (A_D) that identifies SNPs whose unfavourable effect is recessive;
- a ROH-based approach (ROH) that targets local inbreeding effects by comparing homozygous genotypes occurring within versus outside ROH segments;
- and an integrated recessive-ROH approach (A*_D*) that specifically seeks recessive effects expressed only within ROH, reflecting regions with high identity-by-descent.

Each approach incorporates different assumptions with their advantages and their weakness. The A_D method (classical dominance model) is straightforward to take advantage of the entire genotyped sample to estimate the recessive effect of a variant: the test provides high statistical power to detect SNPs where the recessive allele is detrimental. However, it does not incorporate information from runs of homozygosity (ROH), which limits its ability to precisely identify genetic load when the causal variant is rare and present only in autozygous segments.

The ROH method focuses instead on local inbreeding effects by comparing homozygotes within ROH to those outside ROH, without modelling classical additive or dominance effects. It benefits from the ability of ROH to capture long IBD segments, which increases its potential to detect regions harbouring rare deleterious mutations. However, the number of homozygous individuals within a given ROH segment is often limited, reducing statistical power and widening confidence intervals.

The A*_D* method combines recessive modelling with ROH status by explicitly modelling contrasts between homozygotes in ROH and heterozygotes, while adjusting for overall inbreeding. This improves specificity for detecting genetic load carried by rare variants embedded in autozygous haplotypes. The main drawback of this approach lies in its model complexity and reduced effective sample size (only individuals within ROH contribute to the test).

Interestingly, the ROH model was particularly successful in detecting regions where the homozygosity due to ROH segments was associated with stronger inbreeding depression than regular homozygosity. This observation supports the hypothesis that deleterious recessive alleles tend to be more harmful when embedded in ROH, likely due to their identity-by-descent. A similar

conclusion was reached by Bosse et al. (2019), who showed that ROH segments can carry a disproportionate load of deleterious mutations.

There are at least two plausible explanations for this phenomenon. First, the rapid inbreeding that leads to ROH formation may cause recent, strongly deleterious mutations to become homozygous more quickly than would occur under equilibrium conditions. Second, ROHs are more likely to reflect true autozygosity, thus increasing the chance that causal QTL alleles are also homozygous. In contrast, homozygous genotypes outside ROHs may arise by chance or due to incomplete linkage with causal mutations, and therefore may not consistently capture the underlying QTL effect.

These considerations reinforce the importance of distinguishing ROH-based homozygosity in the detection of genetic load. Our results suggest that the recessive load in ROH segments is more severe, which has significant implications for breeding strategies aiming to reduce inbreeding depression. Identifying and managing these high-risk ROH regions may offer a more efficient strategy to limit the impact of harmful recessive alleles.

Unlike the approach proposed by Pryce et al. (2014), our model explicitly includes a global inbreeding coefficient as a covariate. This adjustment is essential to avoid confounding local effects captured through ROH with general effects related to genome-wide inbreeding. Indeed, individuals with high overall inbreeding tend to have more and longer ROH segments, which could lead to spurious associations if not properly controlled. By incorporating a global inbreeding coefficient, our framework ensures that the detected ROH effects truly reflect local autozygosity rather than the background inbreeding level. This enhances the specificity of our results and limits false positives due to general inbreeding depression.

The complementarity of these approaches supports their combined use to obtain a comprehensive map of genetic load.

Among the 108 genomic regions identified, 92 were detected by a single model only, confirming that each approach reveals distinct facets of the genetic load. The ROH model, by design, captures broader regions enriched in homozygosity by descent. In contrast, the A*_D* model, though more conservative, isolates recessive effects likely linked to rare deleterious mutations. The A_D model bridges between the two, revealing recessive alleles expressed more generally, regardless of local autozygosity.

These patterns are also reflected in the characteristics of the regions detected: ROH regions tended to be longer and include more significant SNPs, whereas A_D and A*_D* typically identified smaller, more focused regions with fewer significant markers per region (Table 3).

This plurality of models strengthens the robustness of our findings and highlights the value of using GWAS strategies to dissect complex genomic phenomena like inbreeding depression. Importantly, several of the genomic regions identified by our models correspond to well-known loci associated with economically important traits. For instance, we consistently detected a strong signal near the DGAT1 gene on chromosome 14, a locus already known to influence milk yield and fat content, but also for chromosome 1 and 5. The recurrence of such regions across breeds and traits strengthens the biological plausibility of our findings and confirms the ability of our models to accurately capture deleterious recessive variants involved in inbreeding depression. These concordant signals provide an internal validation of our approach and support its relevance for practical applications in breed management and selection programs

Beyond the immediate interest of identifying genomic regions associated with inbreeding depression, our results open important perspectives for the future management of genetic load in livestock populations. Until now, strategies to control inbreeding have primarily focused on limiting the increase of pedigree- or genome-based inbreeding coefficients. However, the identification of

specific regions carrying deleterious recessive variants offers a more targeted approach: rather than only reducing global inbreeding, future breeding programs could actively avoid at-risk matings that would lead to homozygosity in these high-risk regions. This finer level of management, based on the actual genetic architecture of the load, would allow for more sustainable selection schemes, balancing genetic gain with long-term population fitness. Moreover, in the context of crossbreeding or conservation, knowing which regions are breed-specific or shared becomes crucial to prevent the accumulation or reintroduction of deleterious variants.

6. Conclusion

This study introduces an innovative framework to map the genetic load associated with inbreeding depression in five local cattle populations, by combining three complementary approaches (A_D, ROH, and A*_D*) based on two GWAS models. By integrating additive and dominance effects, ROH information, and a global inbreeding covariate, we provide a detailed and robust view of genomic regions likely to harbour deleterious recessive variants.

Our findings show that most detected signals are specific to a single method, emphasizing the complementarity of the approaches and the risk of overlooking important components of the genetic load when relying on a single model. The consistent detection of well-known loci, such as DGAT1, further supports the biological relevance and validity of our approach.

Looking forward, this comprehensive mapping will enable more refined genetic management strategies. Rather than focusing solely on controlling overall inbreeding, future breeding programs can use this information to avoid at-risk matings that may increase homozygosity in critical regions, while still promoting genetic gain. Furthermore, understanding which regions are breed-specific or shared will be essential for managing crossbreeding and conservation programs, by minimizing the introduction or accumulation of deleterious variants.

In summary, this methodology paves the way for more proactive and precise genomic stewardship, balancing short-term selection objectives with the long-term health and sustainability of livestock populations. It provides a solid foundation for future developments, especially in integrating genetic load information into selection schemes and population management strategies (see WP8).

Finally, the present results align closely with those reported in RUMIGEN Deliverable 4.1, which assessed the effects of genomic selection on genetic gain, diversity and drift in the same local breeds. The integration of homozygosity mapping and the detection of inbreeding-sensitive regions, as done in this deliverable, is highly complementary to D4.1's findings on changes in population structure, genetic trends, and inbreeding rates—particularly those based on F_ROH. These combined insights highlight how genomic selection, while boosting genetic gain, may also alter the dynamics of inbreeding and genetic load, especially in small populations.

In this context, the monitoring of ROH-based inbreeding becomes especially relevant. As shown here, ROH segments are more strongly associated with inbreeding depression, likely because they reveal truly homozygous deleterious QTLs. Meanwhile, D4.1 documented breed-specific patterns in inbreeding trends, with genomic selection leading to either stabilization or increase in F_ROH depending on the breed. These patterns confirm that the design of breeding programs must be tailored to population size, structure, and breeding objectives. Together, the deliverables emphasize the need for genomic management strategies that balance selection intensity with preservation of genetic diversity and long-term viability of local breeds.

7. References

Ablondi, M., Sabbioni, A., Stocco, G., Cipolat-Gotet, C., Dadousis, C., Kaam, J. T. van, Finocchiaro, R., & Summer, A. (2022). Genetic Diversity in the Italian Holstein Dairy Cattle Based on Pedigree and SNP Data Prior and After Genomic Selection. *Frontiers in Veterinary Science*, *8*, 773985. https://doi.org/10.3389/fvets.2021.773985

Bertorelle, G., Raffini, F., Bosse, M., Bortoluzzi, C., Iannucci, A., Trucchi, E., Morales, H.E., & van Oosterhout, C. Genetic load: genomic estimates and applications in non-model animals. *Nat Rev Genet* **23**, 492–503 (2022). https://doi.org/10.1038/s41576-022-00448-x

Biscarini F, Cozzi P, Gaspa G, & Marras G. (2019). *detectRUNS: Detect Runs of Homozygosity and Runs of Heterozygosity in Diploid Genomes. R package version 0.9.6.* [Computer software]. https://CRAN.R-project.org/package=detectRUNS

Bosse, M., Megens, H. J., Derks, M. F., de Cara, Á. M., & Groenen, M. A. (2019). Deleterious alleles in the context of domestication, inbreeding, and selection. Evolutionary applications, 12(1), 6-17.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 7. https://doi.org/10.1186/s13742-015-0047-8

Colleau, J. J., & Moureaux, S. (2006). Optimizing management of kinship and inbreeding coefficients in dairy cattle selection.

CRV. (n.d.). *GENOOMTESTEN BIJ MRIJ EEN FEIT*. Retrieved November 14, 2022, from https://www.cooperatie-crv.nl/aeu-blog/genoomtesten-bij-mrij/

Daetwyler, H. d., Villanueva, B., Bijma, P., & Woolliams, J. a. (2007). Inbreeding in genome-wide selection. *Journal of Animal Breeding and Genetics*, *124*(6), 369–376. https://doi.org/10.1111/j.1439-0388.2007.00693.x

Doekes, H. P., Veerkamp, R. F., Bijma, P., Hiemstra, S. J., & Windig, J. J. (2018). Trends in genome-wide and region-specific genetic diversity in the Dutch-Flemish Holstein–Friesian breeding program from 1986 to 2015. *Genetics Selection Evolution*, *50*(1), 15. https://doi.org/10.1186/s12711-018-0385-y

Doublet, A.-C., Croiseau, P., Fritz, S., Michenet, A., Hozé, C., Danchin-Burge, C., Laloë, D., & Restoux, G. (2019). The impact of genomic selection on genetic diversity and genetic gain in three French dairy cattle breeds. *Genetics Selection Evolution*, *51*(1), Article 1. https://doi.org/10.1186/s12711-019-0495-1

Forutan, M., Ansari Mahyari, S., Baes, C., Melzer, N., Schenkel, F. S., & Sargolzaei, M. (2018). Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. *BMC Genomics*, *19*(1), Article 1. https://doi.org/10.1186/s12864-018-4453-z

Lillehammer, M., Meuwissen, T. H. E., & Sonesson, A. K. (2011). A comparison of dairy cattle breeding designs that use genomic selection. *Journal of Dairy Science*, *94*(1), 493–500. https://doi.org/10.3168/jds.2010-3518

Lozada-Soto, E. A., Parker Gaddis, K. L., Tiezzi, F., Jiang, J., Ma, L., Toghiani, S., VanRaden, P. M., & Maltecca, C. (2023). Inbreeding depression for producer-recorded udder, metabolic, and reproductive diseases in US dairy cattle. *Journal of Dairy Science*, S0022030223008275. https://doi.org/10.3168/jds.2023-23909

Makanjuola, B. O., Miglior, F., Abdalla, E. A., Maltecca, C., Schenkel, F. S., & Baes, C. F. (2020). Effect of genomic selection on rate of inbreeding and coancestry and effective population size of Holstein and Jersey cattle populations. *Journal of Dairy Science*, *103*(6), 5183–5199. https://doi.org/10.3168/jds.2019-18013

Meuwissen, T. H. E. (1997). Maximizing the response of selection with a predefined rate of inbreeding. Journal of animal science, 75(4), 934-940.

Meuwissen, T. H. E., & Luo, Z. (1992). Computing inbreeding coefficients in large populations. *Genetics Selection Evolution*, *24*(4), Article 4. https://doi.org/10.1186/1297-9686-24-4-305

Paul, K., d'Ambrosio, J., & Phocas, F. (2022). Temporal and region-specific variations in genome-wide inbreeding effects on female size and reproduction traits of rainbow trout. *Evolutionary Applications*, *15*(4), 645-662.

Pérez-Enciso, M. (1995). Use of the uncertain relationship matrix to compute effective population size. *Journal of Animal Breeding and Genetics*, *112*(1–6), 327–332. https://doi.org/10.1111/j.1439-0388.1995.tb00574.x

Pryce, J. E., & Daetwyler, H. D. (2012). Designing dairy cattle breeding schemes under genomic selection: A review of international research. *Animal Production Science*, *52*(3), 107. https://doi.org/10.1071/AN11098

Sargolzaei, M., Iwaisaki, H., & Colleau, J.-J. (2005). A fast algorithm for computing inbreeding coefficients in large populations. *Journal of Animal Breeding and Genetics*, *122*(5), 325–331. https://doi.org/10.1111/j.1439-0388.2005.00538.x

Sarviaho, K., Uimari, P., & Martikainen, K. (2023). Estimating inbreeding rate and effective population size in the Finnish Ayrshire population in the era of genomic selection. *Journal of Animal Breeding and Genetics*, *140*(3), 343–353. https://doi.org/10.1111/jbg.12762

Scott, B. A., Haile-Mariam, M., Cocks, B. G., & Pryce, J. E. (2021). How genomic selection has increased rates of genetic gain and inbreeding in the Australian national herd, genomic information nucleus, and bulls. *Journal of Dairy Science*, *104*(11), 11832–11849. https://doi.org/10.3168/jds.2021-20326

Stoop, M., Veninga, G., de Jong, G., & Reinhardt, F. (2017). Genomics in small populations: The MRY breed. *Interbull Bull*, *51*, 26–28.

Topolski, P., & Jagusiak, W. (2020). Inbreeding in a Population of Polish Holstein-Friesian Young Bulls Before and After Genomic Selection. *Annals of Animal Science*, *20*(1), 71–83. https://doi.org/10.2478/aoas-2019-0065

Varona, L., Altarriba, J., Moreno, C., Martínez-Castillero, M., & Casellas, J. (2019). A multivariate analysis with direct additive and inbreeding depression load effects. Genetics Selection Evolution, 51, 1-12.

Windig, J. J., & Hulsegge, I. (2021). Retriever and Pointer: Software to Evaluate Inbreeding and Genetic Management in Captive Populations. *Animals*, *11*(5), 1332. https://doi.org/10.3390/ani11051332

8. Annexe

A.

BTA	start (Mb)	end (Mb)	Trait	ROH	A*_D*	A_D
1	90.24	90.29	protein content	x	-	-
1	151.12	151.87	protein content	x	-	-
1	153.39	153.89	protein content	x	-	-
1	155.39	157.12	protein content	x	-	-
3	102.88	102.88	milk yield	x	-	-
3	112.62	112.62	fat yield	x	-	-
3	114.87	114.87	fat yield	x	-	-
3	116.03	116.03	fat yield	x	-	-
3	102.78	102.78	protein content	x	-	-
4	66.50	66.50	protein content	x	-	-
5	91.73	91.73	milk yield	x	-	-
5	90.69	90.69	fat yield	x	-	-
5	92.29	92.53	fat yield	x	-	-
5	39.69	40.72	fat content	x	-	-
5	75.80	75.91	fat content	x	-	-
5	84.74	85.15	fat content	x	-	-
5	86.09	89.37	fat content	x	x	x
5	90.61	92.53	fat content	x	x	-
5	93.32	97.12	fat content	x	x	x
5	104.21	106.78	fat content	x	x	-
5	28.50	28.50	protein content	x	-	-
5	39.69	40.52	protein content	x	-	-
5	96.18	97.06	protein content	x	-	-
6	22.17	23.50	somatic cell score	x	-	-
6	64.44	64.44	milk yield	x	-	-
6	54.23	54.23	fat content	x	-	-
6	85.45	85.72	protein content	x	-	-
6	86.75	87.40	protein content	x	-	-
7	5.78	5.78	milk yield	-	x	-
7	5.78	5.78	fat yield	-	x	-
7	5.78	5.91	protein yield	-	x	-
10	91.77	92.33	milk yield	x	-	-
10	91.72	91.89	fat yield	x	-	-
10	91.77	91.77	protein yield	x	-	-
10	35.99	36.26	fat content	x	-	-
10	77.85	78.56	fat content	x	-	-
10	80.76	80.79	fat content	x	-	-
10	86.12	86.65	fat content	x	-	-
11	103.67	104.16	milk yield	x	-	-
11	102.91	104.71	protein yield	x	-	-
14	11.37	12.27	milk yield	x	-	-
14	22.67	22.87	fat yield	x	-	-
14	24.38	24.38	fat yield	x	-	-
14	0.17	0.17	fat content	x	-	-
14	1.35	1.35	fat content	x	-	-
14	2.70	3.44	fat content	x	-	-
14	5.85	5.85	fat content	x	-	-
14	45.50	45.62	fat content	x	-	-
14	51.50	51.50	fat content	x	-	-
14	54.68	54.68	fat content	x	-	-
14	61.17	62.36	fat content	x	-	-
14	78.28	78.28	fat content	x	-	-
14	45.53	47.31	protein content	x	x	-
14	48.44	55.72	protein content	x	-	-
14	56.86	58.24	protein content	x	-	-
14	59.79	65.05	protein content	x	x	-
14	66.16	68.91	protein content	x	-	-
14	71.08	71.08	protein content	x	-	-
15	74.16	76.69	milk yield	-	x	x
15	77.34	77.34	milk yield	-	-	x
15	75.30	76.69	fat yield	-	x	x
15	77.34	77.34	fat yield	-	-	x
15	74.16	76.69	protein yield	x	x	x
15	77.34	77.34	protein yield	-	-	x
16	59.27	59.27	fat content	x	-	-
16	60.32	60.32	fat content	x	-	-
16	64.92	64.92	fat content	x	-	-
16	56.78	56.94	protein content	x	-	-
16	59.27	59.27	protein content	x	-	-
16	62.58	63.23	protein content	x	-	-
17	5.34	5.34	fat yield	x	-	-
19	28.99	28.99	fat content	x	-	-
19	54.86	54.86	protein content	x	-	-
20	58.97	58.97	fat content	x	-	-
20	32.96	32.96	protein content	x	-	-
20	69.47	69.47	protein content	x	-	-
21	22.79	23.24	milk yield	x	-	-
24	42.09	43.48	fat content	x	-	-
24	3.57	3.65	protein content	x	-	-
26	46.60	46.68	fat content	x	-	-
26	48.43	48.43	protein content	x	-	-
27	36.30	36.30	milk yield	-	x	-
27	32.97	32.97	protein yield	x	-	-
29	9.76	9.82	protein content	x	-	-

B.

BTA	start (Mb)	end (Mb)	Trait	ROH	A*_D*	A_D
1	72.10	72.24	protein content	-	x	x
3	19.62	19.62	milk yield	-	-	x
3	19.62	19.62	fat yield	-	-	x
3	19.62	19.62	protein yield	-	-	x
3	19.62	19.62	protein content	-	-	x
16	15.48	15.48	protein yield	-	-	x

C.

BTA	start (Mb)	end (Mb)	Trait	ROH	A*_D*	A_D
1	62.21	63.43	fat content	x	x	-
1	65.19	65.19	fat content	-	x	-
1	65.33	65.54	fat content	x	-	-
1	66.71	66.73	fat content	-	x	x
1	68.29	69.34	fat content	x	x	-
1	72.00	72.68	fat content	x	-	-
1	74.31	74.31	fat content	-	-	x
1	62.21	63.43	protein content	x	x	-
1	65.54	65.54	protein content	x	-	-
1	66.52	66.73	protein content	-	x	x
1	66.80	66.89	protein content	x	-	-
1	68.29	69.34	protein content	x	x	-
1	71.76	72.88	protein content	x	-	-
14	1.04	2.33	fat content	x	-	-

D.

BTA	start (Mb)	end (Mb)	Trait	ROH	A*_D*	A_D
1	74.84	76.60	fat yield	x	-	-

E.

BTA	start (Mb)	end (Mb)	Trait	ROH	A*_D*	A_D
11	70.89	70.89	somatic cell score	x	-	-
20	64.30	64.30	somatic cell score	x	-	-
10	77.13	77.13	fat yield	x	-	-

Table 1: Genomic regions significantly detected across three methods (ROH, A_D, and A*_D*) in the 5 breeds (Abondance [A], Tarentaise [B], Vosgienne [C], MRY [D] and NRF [E]) and for the seven studied traits (milk yield, fat yield, protein yield, fat content, protein content, somatic cell score and conception rate).