Guidelines on combining genomic data across populations

R. Bonifazi[1], S. Aivazidou[1], A. Bowman[1], F. Macedo[2], R.F. Veerkamp[1], K. Quigley[3], T. Pabiou[3], R. Evans[3], H. Bunning[4], G.M. Tarekegn[5], P. Davoudi[4], K. Matthews[4], J. Layton[7], R.G.M.F. Damai[7], R. Finocchiaro[7], M. Cassandro[6,7,12], A. Quaglia[8], D. Vicario[9], L. Degano[9], M. Piazza[10], T. Roozen[2], M. Coffey[5], M. Burke[11] and J. Vandenplas[1]

[1] Animal Breeding and Genomics, Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, The Netherlands.

[2] Interbull Centre, Department of Animal Biosciences, SLU, Box 7023, S-75007 Uppsala, Sweden.

[3] Irish Cattle Breeding Federation, Link Rd Ballincollig, P31 D452 Cork, Ireland.

[4] Agriculture and Horticulture Development Board, Middlemarch Business Park, Siskin Parkway East, CV3 4PE Coventry, England.

[5] Scotland's Rural College, West Mains Road, EH9 3JG Edinburgh, Scotland, United Kingdom.

[6] National Federation of National Breeders Associations, XXIV Maggio 43, 00187 Roma, Italy.

[7] Associazione Nazionale Allevatori della Razza Frisona, Bruna e Jersey Italiana, Via Bergamo 292, 26100 Cremona, Italy.

[8] Associazione Nazionale Allevatori Bovini Italiani da Carne, Strada del Vio Viscioloso 21, 06132 Perugia, Italy.

[9] Associazione Nazionale Allevatori Bovini di Razza Pezzata Rossa Italiana, Via Ippolito Nievo 19, 33100 Udine, Italy.

[10] Associazione Nazionale degli Allevatori delle razze bovine Charolaise e Limousine Italiane, Via Ventiquattro Maggio 44/45, 00187 Roma, Italy.

[11] International Committee for Animal Recording, Arthur van Schendelstraat 550, 3511 MH Utrecht, The Netherlands.

[12] University of Padova, DAFNAE, Agripolis, viale dell'Università, 16 – 35020 – Legnaro Padova, Italy

Corresponding author: renzo.bonifazi@wur.nl

Abstract

Genomic selection has been widely implemented across several livestock species. Due to reduced genotyping costs, genomic data are being generated at increasing rates. Various genotyping arrays are available with single-nucleotide polymorphisms (SNPs) genotyping densities ranging from few hundred to hundreds of thousands of single-nucleotide polymorphisms (SNPs). Moreover, advancements in genome sequencing enable the generation of newer and more accurate reference genomes. Genotyping arrays are typically updated to remain close to newer reference genomes, thereby increasing the accuracy of genome-wide association studies and, potentially, genomic predictions. Thus, genomic data have been generated using different genotyping chips, i.e., arrays, and assemblies, creating the need to merge them into a unified map for downstream analyses. However, SNPs do not completely overlap across chips. The number of overlapping SNPs decreases even further when chips come from different manufacturers, from different organisations, which may use customised chips, or when the corresponding assembly versions differ. Moreover, even when chips refer to the same genome assembly, SNPs do not always point to the same physical positions, and metadata, such as SNP names, may differ. Therefore, merging genomic data across populations presents significant challenges which are further complicated by the discontinuation of public mapping resources. The aim of this study is to provide guidelines for combining SNP genotyping data collected across different genotyping chips and populations based on the experience of the M3GE project. More than 3.1 million individual genotypes from purebred and crossbred cattle were collected across six national breeding organisations from three countries: Ireland, Great Britain, and Italy. Genotypes were collected on 11 different genotyping chips with densities ranging from 30K to 777K SNPs. The corresponding assemblies were either UMD3.1 or ARS-UCD1.2. Genotypes were either in raw format or imputed to a custom set of SNPs as used in routine evaluations. Genotypes were collected through the Interbull Centre GenoEx-GDE database

following the Illumina A/B coding and counting the number of copies of the A allele. All genomic data were converted to the UMD3.1 assembly, validated for inconsistencies or errors in allele coding, physical positions on the genome, and allele frequency distributions, both within and across breeds. For maps alignment, we advise storing and exchanging maps together with their metadata (rsID, assembly version, and manifest), so that SNP positions can be linked or converted across chips and assemblies. For genotypes alignment, we advise using a consistent A/B allele coding and using allele frequencies as a diagnostic tool, combining correlation and root mean squared error and comparing each chip with at least two other chips. These alignment steps, together with the required file formats, are needed to ensure valid downstream analyses of combined genomic data.

Key words: genomic data, SNP chips, allele coding, cattle, multi-breed

Introduction

Genomic selection (Nejati-Javaremi et al., 1997; Meuwissen et al., 2001) has been routinely implemented in many livestock species (e.g., Wiggans et al., 2017; Georges et al., 2019; Song et al., 2023). With decreasing genotyping costs, the number of genotyped animals keeps increasing across all livestock species. Individual genotypes are being collected on different single nucleotide polymorphism (SNP) chips, i.e., genotyping arrays, with different densities, ranging from a few hundred to hundreds of thousands of SNPs. Furthermore, reference genome assemblies are updated as sequencing technologies improve. For instance, in cattle, the UMD3.1 assembly (Zimin et al., 2009) was followed by ARS-UCD1.2 (Rosen et al., 2020). Genotyping arrays are typically updated to remain close to newer reference genomes, thereby increasing the accuracy of genome-wide association studies and, potentially, genomic prediction. Therefore, genomic data available within and across organisations have been generated on different chips and refer to different assemblies. This heterogeneity creates the need to merge genomic data into a unified map for downstream analyses.

Combining genotypes collected on different chips requires identifying the same SNP across chips. However, chips only partly overlap. Higher-density chips do not necessarily include all SNPs present on lower-density chips (Nicolazzi et al., 2014). The number of overlapping SNPs decreases even further when chips come from different manufacturers, from different organisations, which may use customised chips, or when the corresponding assembly versions differ. Indeed, genomic data from different sources may differ in several ways. SNP's physical position may refer to different genome assemblies, the same SNP may have different names across genotyping chips even when mapped to the same assembly, and alleles may be coded using different formats and strands (e.g., A/B or A/C/G/T alleles, TOP/BOT or forward/reverse strands; Illumina, 2006; Nicolazzi et al., 2014). Even when two chips refer to the same assembly, the positions reported for the same SNP may not always agree due to manufacturers using different procedures to map probe sequences to the genome (e.g., Fadista & Bendixen, 2012; Schmitt et al., 2010). Errors in any of these elements do not prevent genotype files from being merged, and they may remain unnoticed until they affect downstream analyses, such as imputation and genomic predictions.

Public resources have been developed to link SNP names, positions, and allele coding across chips. For instance, SNPchiMp (Nicolazzi et al., 2014, 2015) collected information on commercial chips for six livestock species directly from the chip manufacturers and linked it to public databases, including SNP positions on the latest assembly available at the time. However, SNPchiMp

(https://webserver.ibba.cnr.it/SNPchimp/) is no longer updated and maintained.

Combining genotypes across organisations is needed, for instance, in projects requiring collaborations across countries, such as the M3GE project. The M3GE project aims to develop multi-trait, multi-breed, multi-country genomic evaluations for beef cattle, focusing on sustainability traits and small populations (Bonifazi et al., 2025). Within M3GE, genotypes of purebred and crossbred cattle were collected from national breeding organisations in three countries, on 11 different chips referring to two assemblies. Therefore, genomic data had to be combined across chips, assemblies, breeds, and countries before any downstream analysis.

The aim of this study is to provide guidelines for combining SNP genotyping data collected across different chips and multiple populations, based on the experience of the M3GE project. In particular, we describe the alignment of map information across chips and assemblies, the validation of genotype information using allele frequencies, and the resulting guidelines.

Materials and Methods

Data

Genotypes were collected within the M3GE project (Bonifazi et al., 2025) from six national breeding organisations in three countries: Ireland, Great Britain, and Italy. In total, about 3.1 million individual genotypes were collected on purebred beef cattle, beef × beef and beef × dairy crossbred animals, and purebred dairy AI selection candidates. Genotypes were available at medium (30K SNPS) to high density (777K SNPs), either in raw format or imputed to the custom set of SNPs used in the national routine evaluation. Genotypes were collected through the Interbull Centre GenoEx-GDE platform (Interbull Centre, 2025).

For each chip, a map file was collected with the SNP name, chromosome, and base-pair position, the rsID when available, and the genome assembly version. Genotypes were coded following the Illumina A/B allele coding (Illumina, 2006) and counting the number of copies of the A allele, as required by GenoEx-GDE, i.e., 0 = BB, 1 = AB, 2 = AA, and 5 = missing genotype. Table 1 summarises the files and information used in this study to combine genomic data across chips.

In total, 11 chips were collected, with SNP positions referring to either the UMD3.1 (Zimin et al., 2009) or the ARS-UCD1.2 (Rosen et al., 2020) assembly. Different breeds were genotyped on different chips, and the same chip could be used by more than one organisation. Chips and countries were anonymised (e.g., COU1_chip1), and the overlap between chips was computed using only autosomal SNP names.

To combine genomic data across chips, we ensured that i) maps were aligned across chips, and ii) genotypes were aligned with the corresponding map and allele coding was consistent across chips.

Maps alignment across chips

Two alternative approaches to align SNP information across assemblies and maps were investigated. The first approach converts positions between any two assemblies, whereas the second assigns positions on the UMD3.1 assembly only, as the SNP dictionary used was built from UMD3.1 positions.

In the first approach, the maps referring to the ARS-UCD1.2 assembly were aligned to the UMD3.1 assembly. An in-house script was developed to convert map information between any two assemblies, hereafter referred to as "source" and "target" assemblies. This first approach is hereafter referred to as the "rsID-bridging" approach. First, variant call format (VCF) files corresponding to the source and target assemblies were downloaded from Ensembl (https://jun2026.archive.ensembl.org/Bos_taurus/Info/Index). Then, for each SNP, the chromosome and base-pair position in the map file were matched to the VCF file of the source assembly to retrieve the SNP identifier (rsID).

The rsID was then searched in the VCF file of the target assembly to retrieve the chromosome and base-pair position on the target assembly (Figure 1a).

In the second approach, hereafter called the "dictionary-based approach", SNP names were first harmonised across all 11 chips by capitalising all letters. Regardless of the assembly version each chip referred to, the corresponding SNP physical positions on the UMD3.1 assembly were then assigned from a SNP dictionary based on the commercial SNP names. The dictionary was built by combining two sources of SNP positions on UMD3.1: SNPchiMp v.3 (Nicolazzi et al., 2015), with ~1.9 million SNPs, and a SNP database provided by ANAPRI (Associazione Nazionale Allevatori Bovini di Razza Pezzata Rossa Italiana, Udine, Italy), with ~0.79 million SNPs. The two sources were merged into a dictionary of unique SNP names. When the two sources reported conflicting positions for the same SNP name, the position from SNPchiMp was retained. Figure 1b summarises the workflow.

Table 1. Files and information used to combine genomic data across chips[1].

File	Information	Notes
Map file (one per chip)	SNP name; chromosome; base-pair position; rsID (if available)	• Assembly version of the positions stored next to the corresponding map
Genotype file	Genotypes coded 0, 1, 2, or 5	• Raw or imputed genotypes • Number of copies of the A allele following the Illumina A/B coding: 0 = BB, 1 = AB, 2 = AA, 5 = missing (as for GenoEx-GDE) • SNP order as in the map file
Source and target assembly VCF	CHROM, POS, ID, REF, ALT, QUAL, FILTER, INFO	• CHROM, POS, and ID; used for the rsID bridging approach (Figure 1a)
SNP dictionary	SNP name; chromosome; position on UMD3.1	• Combined from SNPchiMp v.3 and the ANAPRI database; used for the dictionary-based approach (Figure 1b)

[1] CHROM = chromosome; POS = base-pair position; ID = SNP identifier (rsID); REF = reference allele; ALT = alternative allele; VCF = variant call format.

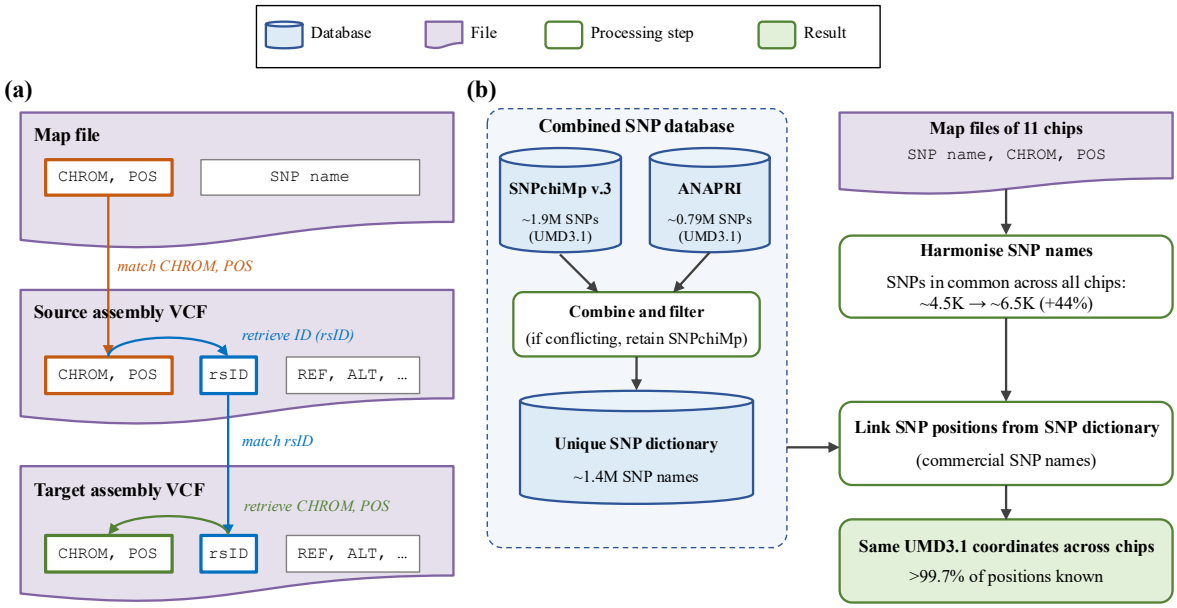


Figure 1. Workflow for the alignment of maps across chips: (a) conversion of positions between assemblies following the rsID-bridging approach; (b) harmonisation of SNP names and assignment of positions from a SNP dictionary combining SNPchiMp v.3 and the ANAPRI database. CHROM = chromosome; POS = base-pair position; ID = SNP identifier (rsID); VCF = variant call format.

Genotypes alignment with the map and allele coding

Genotype information was validated for agreement between genotypes and the corresponding map, and for consistency of allele coding across chips both within and across countries. Allele frequencies (AF) were used as a diagnostic tool. For each SNP in common between two chips, the AF of the A allele was computed as the sum of the genotype codes divided by twice the number of non-missing genotypes, using a custom AWK script to handle the large genotype files. For each pair of chips, the AF of overlapping SNPs were compared using the Pearson correlation (ρ) and the root mean squared error (RMSE), computed as: $\text{RMSE} = \sqrt{\frac{1}{m}\sum_{j=1}^{m}\left(p_{1\text{j}} - p_{2\text{j}}\right)^2}$, where m is the number of SNPs in common between the two chips, and $p_{1\text{j}}$ and $p_{2\text{j}}$ are the AF of SNP j on the first and on the second chip, respectively. No thresholds were set on ρ or RMSE, and each comparison was evaluated by inspecting the scatter plot of AF (made using "ggplot2" package in R) together with the corresponding ρ and RMSE.

Genomic information was compared at two levels: i) within-breed, i.e., using genotypes of the same breed, either within-country (when more than one chip for a country was available) or across-country; and ii) across-breed and across-country. In all comparisons, AF were computed using different animals since not enough individuals were genotyped for more than one chip at the same time to allow for meaningful comparisons. Finally, a stronger agreement between AF was expected within breed than across breeds, with across-breed comparisons expected to be noisier but still informative.

Results & Discussion

Overlap between chips

Table 2 shows the number of autosomal SNPs on each chip and in common between each pair of chips. The number of autosomal SNPs per chip ranged from 28,289 (COU1_chip2) to 667,044 (COU3_chip2). Across all chips, there were ~815K unique SNP names, of which ~160K were present on two or more chips, and only ~6.5K were in common across all chips. The overlap between pairs of chips varied widely. For instance, COU1_chip2 shared between 9,582 and 26,952 SNPs with the other chips, while COU2_chip3 and COU2_chip4 shared 48,862 SNPs. Most SNPs present on a single chip were on the highest-density chip: 535,040 SNPs on COU3_chip2.

Maps alignment across chips

Using the rsID-bridging approach, two chips that were based on the ARS-UCD1.2 assembly (COU1_chip3 and COU3_chip2) were converted to the UMD3.1 assembly. This approach matched 92.7% and 85.7% of the SNPs for COU1_chip3 and COU3_chip2, respectively. For these two chips, respectively, 1.8% and 8.4% of the SNPs were not matched in the source VCF (ARS-UCD1.2), and 5.5% and 5.9% of the SNPs were not matched in the target VCF (UMD3.1). The rsID-bridging approach is straightforward, fast, and flexible, and it works well for chips whose SNPs have a dbSNP rsID. However, this approach relies on the availability and consistency of rsIDs across assemblies, and it requires an exact match of chromosome and position in the source VCF. Coordinate liftover tools, such as UCSC liftOver (Hinrichs et al., 2006) and CrossMap (Zhao et al., 2014), convert positions between assemblies using a pre-computed alignment of the two assemblies and do not rely on rsIDs. Therefore, liftover could be used for the SNPs that remained unmatched. However, when applied to map files that contain only chromosome and position, liftover converts positions without checking alleles, and errors in the original positions are carried over to the target assembly.

In the dictionary-based approach, the combined SNP dictionary included a total of ~1.4 million unique SNP names. Of these, 776,922 had the same position in both the

Table 2. Number of autosomal SNPs on each chip (diagonal), number of SNPs in common between pairs of chips (below the diagonal), and number of SNPs present only on that chip (unique SNPs)[1].

Chip	COU1 chip1*	COU1 chip2*	COU1 chip3	COU1 chip4	COU2 chip1	COU2 chip2	COU2 chip3	COU2 chip4	COU2 chip5	COU3 chip1	COU3 chip2
COU1_chip1*	132,665										
COU1_chip2*	26,952	28,289									
COU1_chip3	60,874	14,812	64,888								
COU1_chip4	37,420	9,582	37,555	42,172							
COU2_chip1	34,670	11,716	23,964	14,091	41,855						
COU2_chip2	37,863	10,491	38,040	37,685	14,745	44,460					
COU2_chip3	39,280	11,016	38,940	38,640	15,154	44,165	48,862				
COU2_chip4	39,280	11,016	38,940	38,640	15,154	44,165	48,862	48,862			
COU2_chip5	37,863	10,491	38,040	37,685	14,745	44,460	44,165	44,165	51,828		
COU3_chip1	37,796	10,511	37,990	37,564	14,689	44,210	43,922	43,922	50,390	50,493	
COU3_chip2	116,260	24,961	60,118	35,716	37,420	38,064	41,877	41,877	38,064	38,065	667,044
Unique SNPs	491	139	44	127	-	-	-	-	1,188	-	535,040

[1] Chips are anonymised by country (COU1 to COU3) and chip number. Counts are based on autosomal SNP names; * Chip used by more than one national breeding organisation; - = none.

SNPchiMp and ANAPRI databases, 612,189 were only present in the SNPchiMp database, 10,224 were only present in the ANAPRI database, and 1,568 had conflicting positions between the two databases. For these ~1.5K SNPs, the SNPchiMp database information was retained. Finally, a total of 46,721 SNP names had no chromosome and base-pair position in either database. After harmonising SNP names, the number of autosomal SNPs in common across all 11 chips increased from ~4.5K to ~6.5K, corresponding to an increase of 44%. SNP positions on UMD3.1 were then assigned to each chip using the built SNP dictionary by matching the SNP names. On each chip, the SNP position was known for more than 99.7% of the SNPs (Figure 1b). The main advantage of the dictionary-based approach is that the same SNP name receives the same coordinates on all chips.

Genotypes alignment with the map and allele coding

Comparing AF between chips revealed different types of misalignment. Figure 2 shows an example of an allele coding switch. The AF of chip 1 were negatively correlated with those of chip 2 for the same breed within country (Figure 2a), and with those of chip 3 for the same breed in another country (Figure 2b). In contrast, the AF of chip 2 and chip 3 agreed (Figure 2d). A negative correlation, especially when as strong as that observed in Figure 2a indicates a complete switch of the allele coding (i.e., the A allele has been counted in one chip, while the B allele was counted in another chip). However, it is not possible from a single comparison to determine which chip had the allele coding switch; comparing three chips allowed us to attribute the switch to chip 1. The issue was solved by recoding the genotypes of chip 1, i.e., 0 became 2 and vice versa. The across-breed comparison between chip 1 and chip 3 (Figure 2c) showed the same negative trend, although with a larger spread of points around the regression line, as expected for AF of different breeds. Thus, across-breed comparisons were noisier but still useful to confirm allele coding switches.

Figure 3a shows the comparison between two chips of the same breed within country, in which chip 2 was misaligned. The AF of the two chips were not correlated (ρ = 0.01), and the RMSE was 0.34. Importantly, the marginal distribution of chip 2 AF was closer to a normal distribution than a uniform distribution. Such a pattern indicated a potential misalignment between the genotypes and their corresponding map for chip 2. However, AF checks alone cannot indicate whether the error lies in the map

or in the genotypes. In this case, both map and genotypes were flagged and re-sent by the data provider.

Figure 3b shows a less evident case of misalignment, in which the maps of chip 1 and chip 2 were offset by one SNP, such that the AF of a SNP was assigned to the next SNP. Despite the shift, ρ remained high (0.90), likely because the shift led to comparing AF of adjacent SNPs, which, due to linkage disequilibrium, may have similar AF. However, the RMSE was 0.12, with many SNPs deviating from the diagonal (Figure 3b).

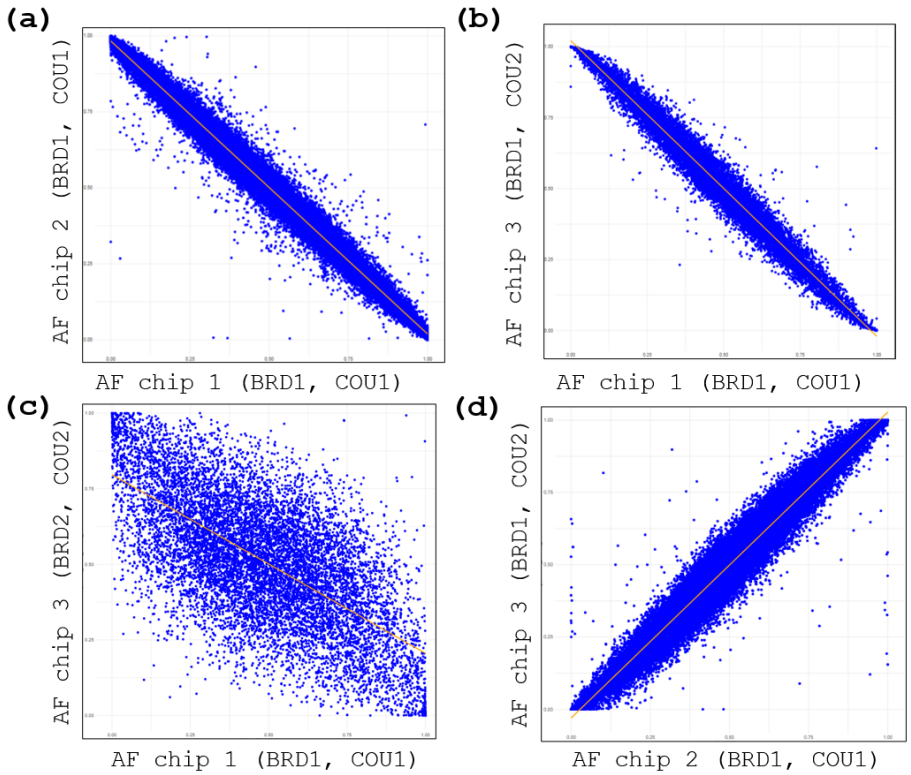


Figure 2. Scatterplot of allele frequencies (AF) for SNPs in common between pairs of chips. (a) chip 1 vs chip 2, same breed (BRD1) and country (COU1); (b) chip 1 vs chip 3, same breed (BRD1) in different countries (COU1 and COU2); (c) chip 1 vs chip 3, different breeds (BRD1 and BRD2) and countries; (d) chip 2 vs chip 3, same breed (BRD1) in different countries.

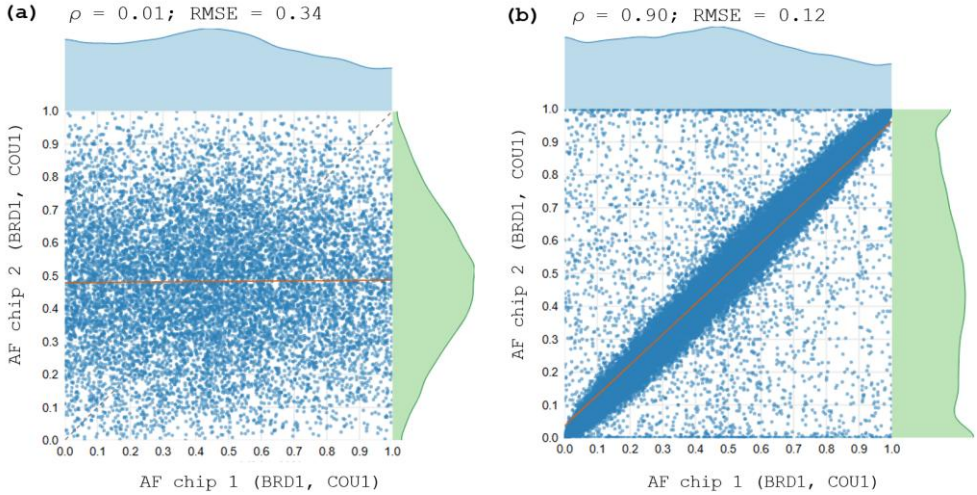


Figure 3. Scatterplot of allele frequencies (AF) of SNPs in common between two chips of the same breed (BRD1) and country (COU1): (a) example of misaligned chip 2; (b) maps of both chips offset by one SNP relative to the genotypes. Marginal distributions of AF for chip 1 and chip 2 are shown above and to the right of each panel, respectively. ρ = Pearson correlation; RMSE = root mean squared error.

Table 3. Guidelines for combining genomic data across populations[1].

Step	Guideline	Reason
Maps alignment	Use public resources when possible (e.g., SNPchiMp) to assign positions from commercial SNP names	Same coordinates across chips; >99.7% of positions known (Figure 1b)
	Store the rsID and the assembly version together with the map	Needed to convert positions between assemblies (Figure 1a)
	Keep complete manifest files of each chip	Manifest information on older chips is not always available or in the public domain
	Add support for metadata (rsID, assembly version, manifest) to GenoEx	These metadata are needed to link and convert positions across chips and assemblies (Figure 1)
Genotypes alignment	Use a consistent Illumina A/B coding, i.e., always count the same allele (A or B). A/B alleles are designated from the actual or contextual sequence of each SNP (TOP/BOT strand) and therefore remain the same even if genome assemblies or the submitted strand change (Illumina, 2006)	An allele coding switch reverses the AF of a chip (Figure 2)
	Use AF checks as a diagnostic tool, combining both ρ and RMSE	A shift by one SNP still gave $\rho = 0.90$, while RMSE = 0.12 (Figure 3b)
	Compare each chip with at least two other chips, using either genotypes from the same breed (within or across countries) or, if not available, from different, preferably related, breeds	Needed to identify which chip may cause a misalignment (Figure 2)
	Automate AF checks (ρ and RMSE) at submission, using already-approved genotypes that act as a reference, if available	Issues can be detected before genotypes are combined

[1] AF = allele frequency; ρ = Pearson correlation; RMSE = root mean squared error.

The RMSE indicated that the alignment between map and genotype was not correct. Once correctly aligned, ρ was 0.99 and RMSE was 0.04. Thus, ρ and RMSE should be used together to detect misalignments, as ρ alone may not be sufficient.

Guidelines

Table 3 summarises our guidelines for combining genomic data across populations, with the reason for each guideline based on the results of this study. The guidelines are grouped into maps alignment and genotypes alignment. For maps alignment, the main guideline is that map information should be stored and exchanged together with its metadata, i.e., rsID, assembly version, and manifest, such that positions can be linked or converted across chips and assemblies. For genotypes alignment, the main guideline is to use AF as a diagnostic tool, comparing chips for which AF are expected to agree (i.e., the same or closely related breeds), with checks done on submission based on both ρ and RMSE.

Implications

These guidelines apply whenever genotypes collected by different organisations are to be combined. This is the case for many collaborative projects, such as the development of international genomic evaluations (e.g., M3GE (Bonifazi et al., 2025), InterGenomics (Jorjani et al., 2011; Savoia, 2021) SNP-MACE (Goddard et al., 2019)), studies on relationships among different populations (Vandewalle et al., 2026), and the exchange of genotypes through GenoEx for genomic evaluations or parentage verification (Interbull Centre, 2025). Public resources, such as SNPchiMp (Nicolazzi et al., 2015) and the SNP coordinates made available by Schnabel (Schnabel, 2018), were of great help in this study. However, these resources are

either no longer updated or incomplete. Therefore, new and maintained tools are needed to link (commercial) SNP names, SNP positions, and allele coding across chips and assemblies.

Conclusions

Combining genomic data collected on different chips and assemblies across populations is challenging and prone to mistakes, which may go unnoticed after merging. In the M3GE project, harmonising SNP names and assigning positions from a common SNP dictionary provided the same coordinates across 11 chips and increased the number of SNPs in common across all chips by 44%. Comparing AF of overlapping SNPs between chips detected allele coding switches and misalignments between maps and genotypes, and highlighted that, next to visual inspections, both ρ and RMSE should be considered. Public resources were useful for this study but are outdated. New and updated tools are needed to reliably combine genomic data across populations.

Acknowledgments

This study was financially supported by the Dutch Ministry of Economic Affairs (TKI Agri & Food Project LWV23062) and the M3GE partners ICAR (Utrecht, the Netherlands), SLU (Uppsala, Sweden), AHDB (Coventry, United Kingdom), SRUC (Scotland, United Kingdom), ICBF (Cork, Ireland), and FedANA (Rome, Italy), including its member organisations ANABIC (Associazione Nazionale Allevatori Bovini Italiani da Carne, Perugia, Italy), ANACLI (Associazione Nazionale degli Allevatori delle razze bovine Charolaise e Limousine Italiane, Rome, Italy), ANAFIBJ (Associazione Nazionale Allevatori della Razza Frisona, Bruna e Jersey Italiana, Cremona, Italy), and ANAPRI (Associazione Nazionale Allevatori Bovini di Razza Pezzata Rossa Italiana, Udine, Italy) for providing the data. We also acknowledge the Interbull Centre (Uppsala, Sweden) for providing the infrastructure to perform the analyses.

References

Bonifazi, R., S. Aivazidou, R.F. Veerkamp, F. Macedo, R. Evans, T. Pabiou, K. Quigley, K. Matthews, H. Bunning, G.M. Tarekegn, M. Cassandro, R. Finocchiaro, J. Layton, R.G.M.F. Damai, A. Quaglia, D. Vicario, L. Degano, M. Piazza, F. Tiezzi, T. Roozen, M. Coffey, M. Burke, and J. Vandenplas. 2025. The M3GE project: developing beef cattle multi-trait multi-breed multi-country genomic evaluations for sustainability traits and small populations. *Interbull Bulletin* 61:231–241.

Fadista, J., and C. Bendixen. 2012. Genomic Position Mapping Discrepancies of Commercial SNP Chips. PLoS ONE 7:e31025. https://doi.org/10.1371/journal.pone.0031025.

Georges, M., C. Charlier, and B. Hayes. 2019. Harnessing genomic information for livestock improvement. *Nat. Rev. Genet.* 20:135–156. https://doi.org/10.1038/s41576-018-0082-2.

Goddard, M.E., A. Jighly, H. Benhajali, H. Jorjani, and Z. Liu. 2019. SNPMace – A meta-analysis to estimate SNP effects by combining results from multiple countries. *Interbull Bulletin* 54:1–6.

Hinrichs, A.S., D. Karolchik, R. Baertsch, G.P. Barber, G. Bejerano, H. Clawson, M. Diekhans, T.S. Furey, R.A. Harte, F. Hsu, J. Hillman-Jackson, R.M. Kuhn, J.S. Pedersen, A. Pohl, B.J. Raney, K.R. Rosenbloom, A. Siepel, K.E. Smith, C.W. Sugnet, A. Sultan-Qurraie, D.J. Thomas, H. Trumbower, R.J. Weber, M. Weirauch, A.S. Zweig, D. Haussler, and W.J. Kent. 2006. The UCSC Genome Browser Database: update 2006. Nucleic Acids Res 34:D590–D598. https://doi.org/10.1093/nar/gkj144.

Illumina. 2006. Technical Note: "TOP/BOT" Strand and "A/B" Allele. Accessed July 26, 2026. https://www.illumina.com/documents/products/technotes/technote_topbot.pdf.

Interbull Centre. 2025. GenoEx - International Genotype Exchange Platform for Animal

Breeding. Accessed July 24, 2026. https://genoex.org/.

Jorjani, H., J. Jakobsen, M.A. Nilforooshan, E. Hjerpe, B. Zumbach, and V. Palucci. 2011. Genomic Evaluation of BSW Populations InterGenomics: Results and Deliverables. *Interbull Bulletin*. 43:5–8.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Nejati-Javaremi, A., C. Smith, and J.P. Gibson. 1997. Effect of total allelic relationship on accuracy of evaluation and response to selection. *J. Animal Sci* 75:1738. https://doi.org/10.2527/1997.7571738x.

Nicolazzi, E.L., M. Picciolini, F. Strozzi, R.D. Schnabel, C. Lawley, A. Pirani, F. Brew, and A. Stella. 2014. SNPchiMp: A database to disentangle the SNPchip jungle in bovine livestock. *BMC Genom.* 15:1–6. https://doi.org/10.1186/1471-2164-15-123

Nicolazzi, E.L., A. Caprera, N. Nazzicari, P. Cozzi, F. Strozzi, C. Lawley, A. Pirani, C. Soans, F. Brew, H. Jorjani, G. Evans, B. Simpson, G. Tosser-Klopp, R. Brauning, J.L. Williams, and A. Stella. 2015. SNPchiMp v.3: integrating and standardizing single nucleotide polymorphism data for livestock species. *BMC Genom.* 16:283. https://doi.org/10.1186/s12864-015-1497-1.

Rosen, B.D., D.M. Bickhart, R.D. Schnabel, S. Koren, C.G. Elsik, E. Tseng, T.N. Rowan, W.Y. Low, A. Zimin, C. Couldrey, R. Hall, W. Li, A. Rhie, J. Ghurye, S.D. McKay, F. Thibaud-Nissen, J. Hoffman, B.M. Murdoch, W.M. Snelling, T.G. McDaneld, J.A. Hammond, J.C. Schwartz, W. Nandolo, D.E. Hagen, C. Dreischer, S.J. Schultheiss, S.G. Schroeder, A.M. Phillippy, J.B. Cole, C.P. Van Tassell, G. Liu, T.P.L. Smith, and J.F. Medrano. 2020. De novo assembly of the cattle reference genome with single-molecule sequencing. Gigascience 9:giaa021. https://doi.org/10.1093/gigascience/giaa021

Savoia, S. 2021. Interbull genomic evaluation of small Holstein populations: InterGenomics-Holstein (IG-HOL). *Interbull Bulletin* 56:180–187.

Schmitt, A.O., R.H. Bortfeldt, and G.A. Brockmann. 2010. Tracking chromosomal positions of oligomers - a case study with Illumina's BovineSNP50 beadchip. *BMC Genom.* 11:80. https://doi.org/10.1186/1471-2164-11-80

Schnabel, R.D. 2018. ARS1.2 coordinates for 25 bovine SNP chips available. Accessed. https://www.animalgenome.org/repository/cattle/UMC_bovine_coordinates/.

Song, H., T. Dong, X. Yan, W. Wang, Z. Tian, A. Sun, Y. Dong, H. Zhu, and H. Hu. 2023. Genomic selection and its research progress in aquaculture breeding. *Reviews in Aquaculture* 15:274–291. https://doi.org/10.1111/raq.12716.

Vandewalle, M., R. Bonifazi, J. Vandenplas, S. Hiemstra, G. De Jong, D. Hinrichs, N. Gengler, and H. Wilmot. 2026. Unravelling the Genetic Structure of Local and Mainstream Red-Pied Cattle Breeds Using Genomics and Extended Pedigree Analysis. *J. Anim. Breed. Genet.* 143:705–718. https://doi.org/10.1111/jbg.70054.

Wiggans, G.R., J.B. Cole, S.M. Hubbard, and T.S. Sonstegard. 2017. Genomic Selection in Dairy Cattle: The USDA Experience. *Annu. Rev. Ani. Biosci.* 5:309–327. https://doi.org/10.1146/annurev-animal-021815-111422.

Zhao, H., Z. Sun, J. Wang, H. Huang, J.-P. Kocher, and L. Wang. 2014. CrossMap: a versatile tool for coordinate conversion between genome assemblies. Bioinformatics 30:1006–1007. https://doi.org/10.1093/bioinformatics/btt730.

Zimin, A.V., A.L. Delcher, L. Florea, D.R. Kelley, M.C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C.P. Van Tassell, T.S. Sonstegard, G. Marçais, M. Roberts, P. Subramanian, J.A. Yorke, and S.L. Salzberg. 2009. A whole-genome assembly of the domestic cow, Bos taurus. *Genome Biol.* 10:R42. https://doi.org/10.1186/gb-2009-10-4-r42.