CDCB Policy

USE OF NATIONAL COOPERATOR DATABASE FOR SNP ARRAY VALIDATION AND SNP INFORMATION DISCLOSURE

This policy for regulating the use of the national cooperator database for the purposes of SNP array validation and disclosure of SNP information with certified genomic laboratories was approved by the CDCB Board of Directors on 08/23/2019.

1. POLICY STATEMENT

The Council on Dairy Cattle Breeding (CDCB) performs a validation process for all SNP arrays included in the CDCB collaborators database. The purpose of the validation procedure is to verify the concordance and overall quality of the SNP calls produced by a new commercial array (or technology). Once validated, SNP genotypes produced with the validated SNP array can be routinely submitted to the CDCB and animals genotyped with such array can be included in the genomic evaluation. Not all the SNPs of any given SNP array are used by CDCB, as CDCB selects a set of SNPs for performing its evaluations and services. CDCB discloses the updated list of selected SNPs with all certified genomic laboratories as an incentive to keep the commercial SNP arrays available as relevant as possible for the dairy industry. Since both the SNP array validation process and the inclusion of new SNPs in the evaluation set rely on the national cooperator database, this policy addresses the rules, costs associated and information required for such purposes.

2. SNP ARRAY VALIDATION

In order to be included in the CDCB collaborators database, SNP arrays are required to undergo a validation process.

The validation process has been an informal collaboration between the CDCB and the interested laboratory. However, a formal procedure is now required as the frequency of such validation requests has increased greatly, which implies substantial allocation of time and resources by CDCB. Furthermore, the validation procedures rely on the data stored in the national cooperator database.

The new SNP array validation process starts once the respective certified genomic laboratory has completed all the following steps[1]:

- Completed and signed “CDCB SNP array validation” form: https://redmine.uscdcb.com/documents/163

[1] This procedure is required for SNP arrays provided by CDCB approved genotyping laboratories. New SNP arrays used for international genotyping exchanges (e.g. InterGenomics, CDDR exchanges) will undergo validation entirely under CDCB responsibility and cost management.

- Submission of at least 50 genotyped samples using the SNP array up for validation.
    - At least 20% of the animals genotyped with the new SNP array provided to CDCB during the validation process are required to be previously genotyped with a different SNP array and stored in the CDCB collaborator database. Alternatively, 25% of the animals genotyped with the new SNP array provided to CDCB during the validation process are required to be *progeny* of animals previously genotyped with a different SNP array and stored in the CDCB collaborator database.
    - New genotypic information is to be shared in the accepted genotype file formats for approved genotyping laboratories, as specified in https://redmine.uscdcb.com/projects/cdcb-customer-service/wiki/CDCB_Accepted_genotype_file_formats#GENOTYPE-FILES-FOR-APPROVED-GENOTYPING-LABORATORIES
- Payment of "SNP array validation" fee.

SNP array validation process

Once the completed "CDCB SNP array validation" form and the minimum number of animal genotypes has been received, CDCB staff will analyze the data received and provide feedback on performance of the SNP array. Initial phases of the validation can start before the payment of the "SNP array validation fee", but payment must be received by the CDCB before the first full test run is performed.

The SNP array validation fee gives approved genotyping laboratories up to 3 (three) full test runs of the data submitted in a period of 30 (thirty) days. SNP array validation is to be finalized within these test runs and time frame. Should further test runs or more time be required to validate the SNP array, an additional SNP array validation fee will need to be paid by the genotyping laboratory, which will provide for up to 3 (three) further full test runs and 30 (thirty) more days to finalize the validation.

Once SNP array validation process is completed, the CDCB staff will issue a letter of compliance to the approved genotyping laboratory. The CDCB staff will have up to 30 (thirty) days after the letter of compliance is issued to finalize the implementation on production tables and start accepting genotypes on the newly validated SNP array routinely.

At implementation, all genotype exchange partners will receive notification of the newly validated SNP array name and characteristics (SNP counts, SNP names, storage sequence, etc). Note that the CDCB also publicly displays the names and basic characteristics of all SNP arrays used in genomic evaluations.

SNP array validation service cost

The "new SNP array validation" form includes a declaration of agreement to pay the validation service fee. Such validation is different for the different circumstances. Fee information is provided here: https://redmine.uscdcb.com/projects/cdcb-customer-service/wiki/CDCB_SNP_array_validation_costs

3. DISCLOSING INFORMATION ON SNPs USED IN GENOMIC EVALUATIONS

The total number of SNPs used for genomic services by the CDCB varies across time. The SNPs in use are selected for different purposes, e.g. sex determination, breed conflicts pre-screening, etc (henceforth named 'QC SNPs'). However, most of the SNPs used for genomic services are selected for their association with one or more traits under selection ('prediction SNP', which may include a subset of QC SNPs). These SNPs are usually identified in public research or in projects performed or supported by CDCB. Before their implementation, AGIL and CDCB staff relies on the national cooperator database to validate trait associations (if applicable), effects and assess general SNP performance.

There is an established practice of sharing SNP name subsets with CDCB collaborator genotyping laboratories because it's in CDCB interest that genotyping laboratories developing SNP arrays include as many of the prediction SNPs as possible to reduce imputation burden and increase accuracy of evaluations. Furthermore, there is interest in including as large a number of QC SNPs as possible, to improve quality checks on genotyped animals.

It's CDCB intention to continue this successful collaboration for both QC and associated SNPs **only to CDCB approved genotyping labs and international genetic evaluation centers** for the purposes described above. Collaboration and sharing of such information with researchers will be examined and decided upon on a case-by-case basis.

No information about which trait a SNP subset under investigation is associated to will be shared with any collaborator or party on SNPs under testing and unpublished. Prioritization of SNPs may be disclosed by CDCB without references to which traits the priority is referred to. In order to prevent a competitive advantage, **any SNP information released by CDCB about SNPs under investigation or testing will be shared to all approved genotyping laboratories.**

For no reason and under no circumstances SNP effects on any trait will be disclosed by CDCB to any party.