GENOTYPE PANEL REQUIREMENTS FOR INCLUSION INTO BREEDPLAN SINGLE STEP EVALUATIONS

N.K. Connors and M.H. Ferdosi

Animal Genetics & Breeding Unit*, University of New England, Armidale, NSW, 2351 Australia

SUMMARY

Increased demand for genomic data driven by the transition of BREEDPLAN to single-step genomic BLUP, has seen an increase in the numbers of genotyping providers and SNP panels. To assess the suitability of new panels, AGBU has developed a standardised procedure to ensure high quality genomic data for inclusion into BREEDPLAN.

INTRODUCTION

The transition of Australia's national beef genetic evaluation, BREEDPLAN, to Single-Step Genomic Best Linear Unbiased Prediction (ssGBLUP) in 2017 (Johnston *et al.* 2018) has driven increased demand for genomic data. The resultant growth in genotyping has stimulated the rapid introduction of new commercial genotyping companies and new single nucleotide polymorphism (SNP) panels. Given that the end use for the majority of the beef industry's genotyping is incorporation into BREEDPLAN, it is necessary to ensure that SNP panels offered to breeders/breed societies are compatible with the genomic pipeline quality control (QC) requirements (Connors *et al.* 2017) used for building the genomic relationship matrix for BREEDPLAN single-step evaluations. These QC checks can only be performed if the genotypes are compatible and as such, new genotypes and new panels require analysis prior to inclusion into the BREEDPLAN genomic pipeline. The Animal Genetics and Breeding Unit (AGBU) has developed a set of industry standards for genotype panels, along with a process of analysing new SNP panel products and validating their compatibility for the BREEDPLAN genomic pipeline. This paper describes the analyses and validation process for new SNP panels and their genotypes.

MATERIALS AND METHODS

The fundamental requirements for BREEDPLAN genomics compatibility are consistent format, SNP quality and informativity. These requirements ensure accuracy and consistency across all genomic records included in the national genomic evaluations. These standards are currently in consultation with Meat and Livestock Australia, industry partners and genotype providers. Firstly, genotypes are required to be sent in a particular format derived from Illumina Genotyping Exports with specifications designated by AGBU. This consistent format ensures all genotypes are processed in the correct order, enables automated analysis of the new SNP panels, and ensures genomic pipeline QC can be performed. Analyses of new SNP panels requires access to the panel map file, which provides SNP names, SNP location (chromosome and base pair position), and allele coding (e.g. manufacture strand/customer strand) in a particular genome assembly. The map file must also be in a particular consistent format as designated by AGBU. The map file provides an overview of the panel, including SNP density and chromosome coverage.

Analysis of SNPs on new panels enables a comparison with existing panels, and more importantly, the consensus panels developed at AGBU (AGBU 6k, AGBU 150k) used for the BREEDPLAN genomic pipeline to combine SNP data from panels of varying densities. AGBU 6k consists of a set of approximately 6000 SNPs common across SNP panels and is available in the public domain (Boichard

* A joint venture of NSW Department of Primary Industries and the University of New England

et al. 2012); AGBU 150k is a set of approximately 150,000 SNPs used to format all genotypes to a common consensus panel (Connors *et al.* 2017). Analysis of the new panel map file in relation to the consensus panels indicates its compatibility with the BREEDPLAN genomic pipeline. For example, genotypes from panels missing the common 6000 SNPs (AGBU 6k) cannot be compared with existing genotypes and will be incompatible with BREEDPLAN.

AGBU has developed automated reporting processes to analyse new SNP panels, providing a detailed breakdown including the number of SNPs (overall and per chromosome), minimum and maximum distance between SNPs, first and last position of SNPs, mean and standard deviation of positions per chromosome. A common SNPs analysis is performed by creating a SNP overlap matrix between the new panel and a subset of the most informative existing panels, including the AGBU consensus panels. These statistics are presented in an automated report in both tabular and graphical representations, enabling rapid identification of potential issues. Automated reports can be made available to relevant industry and commercial bodies considering the use of the analysed new SNP panel.

RESULTS AND DISCUSSION

To date, AGBU has analysed more than 50 different SNP panels, enabling a detailed understanding of the requirements for BREEDPLAN compatibility. Based on this experience, AGBU has formulated a set of requirements for new panels to ensure compatibility for the genomic pipeline and inclusion into BREEDPLAN single-step evaluations, forming the basis of new panel assessment Levels 1-3.

Currently Level 1 assessments are enforced for BREEDPLAN inclusion, such that genotypes from a new panel not meeting these requirements will be excluded from the evaluation until such time as the requirements are met. Levels 2 and 3 assessments, along with the formation of reference populations, are proposed for future implementation and are currently being negotiated with industry bodies to determine funding structures and accountability.

Level 1 assessment – Panel format, quality and content. New panel inclusion requirements are as follows:

- Provide AGBU with the aim and/or target breed/s of a newly developed panel;
- Genotypes and map files must be formatted as per specified guidelines (as supplied by AGBU);
- The genome assembly used for the panel map file must be provided;
- Panel must contain at least 90 percent of SNPs in the AGBU 6k consensus panel;
- Panel must have at least 10000 SNPs in common with the autosomal region of the AGBU 150k consensus panel;
- There must be at least 200 SNPs on the non-autosomal region of the X chromosome;
- Confidential or patented SNPs and any restrictions regarding their use must be provided;
- The raw genotypes are to be supplied in Illumina AB format (guidelines as supplied by AGBU) and must not be imputed or manipulated (e.g. removal of chromosome X, Y or mitochondrial SNPs);
- Genotypes must possess GenCall (GC) scores. If GC scores are not provided, a statement of quality assurance is required. AGBU will not accept any responsibility of quality regarding issues related to SNPs without GC scores;
- Companies which do not use GC as quality control (e.g. Affymetrix) must provide AGBU with the formula to convert their QC to GC. Additionally, they should provide a statement regarding the concordance of their panels with Illumina panels;
- SNP names should not contain a prefix or suffix. If prefixes or suffixes are present, specific recommendations to AGBU must be included on the SNPs involved;
- There must be no SNP duplication in the map file; i.e. SNPs with same position and chromosome but different name, or same SNP name and different positions. If important SNPs must be tested on the

panel more than once, they must have a suffix 'dup' and the reason for duplication must be provided.

In addition to meeting the above criteria for panel inclusion, AGBU also recommends the following:

- At least 20 individuals should be genotyped with both an existing panel of equal or higher density and the new panel, to check quality and concordance;
- Suggested inclusion of SNPs in mitochondrial regions on new panel;
- Suggested inclusion of SNPs on the non-autosomal region of Y chromosome on new panel.

These BREEDPLAN inclusion requirements and recommended features are communicated to genotyping companies in relation to the design of new panels, and/or in response to inclusion/exclusion of genotypes from new panels.

Level 2 assessment – Heterozygosity. The Level 2 heterozygosity assessment is proposed for implementation in future, dependent on ongoing negotiations with industry and commercial bodies. This assessment is not currently enforced for new panels to be included into BREEDPLAN. Genotypes for at least 200 individuals per breed are required. Individuals should be purebred according to software used for the BREEDPLAN genomic pipeline (Boerner 2017; Boerner and Wittenburg 2018) and therefore representative of the breed population, forming a reference dataset for that breed. Individuals can be genotyped either with the new panel, or with a panel of high enough density (e.g. Illumina Bovine HD (777k)) such that there are two panels (one of which is the new panel) sharing at least 95 percent of SNPs in common for comparison. The heterozygosity and allele frequencies of the genotypes are assessed to check for the following requirements:

- The heterozygosity over all SNPs for the entire breed's population must be greater than 30 percent;
- The heterozygosity for pure breed animals must be less than 50 percent;
- Major allele frequencies for each individual must not exceed 75 percent;
- The histogram for SNPs with minor allele frequency (MAF) above 0.05 should be reasonably distributed.

Importantly, all SNPs included in BREEDPLAN single-step evaluations will require at least 1000 genotypes with that SNP (i.e. 1000 individuals), to ensure high imputation accuracy. If less than 1000 genotypes are available (e.g. because they're included in a new SNP panel) these SNPs will not be used until they can be further validated. Thus the new panels require more stringent assessment of those SNPs in common with other panels.

Level 3 assessment – Imputation accuracy. The Level 3 imputation accuracy assessment is proposed for implementation in future, dependent on ongoing negotiations with industry and commercial bodies. This assessment is not currently enforced for new panels to be included into BREEDPLAN. The aim of this assessment is to investigate how well genotypes from existing panels can be imputed to the new panel, and vice versa. Genotypes of at least 2000 individuals with high density (e.g. 777k) per breed are required. Individuals should be suitably representative of the average breed population, such that they form a reference dataset for that breed. This assessment will be performed by extracting a subset of SNPs based on the new panel's map file from 1000 individuals and imputing up to 777k, to other panels, and to SNPs used for building the GRM. The remaining 1000 individuals will be used as a reference set for imputation. Furthermore, if the new panel is of higher density (i.e. >50k), imputation accuracy of low density panels (i.e. <20K) will also be assessed.

Reference populations. Testing SNP panels for heterozygosity and imputation accuracy requires reference populations. The creation of reference populations is proposed for implementation in the near future to ensure ongoing robustness of genomics in BREEDPLAN. Genotypes for at least 200 individuals per breed are required. Individuals should be purebred, according to BREEDPLAN

genomic pipeline requirements (Boerner 2017; Connors *et al.* 2017) and thus suitably representative of the breed's population. Reference individuals can be used to test concordance between panels, and ongoing research needs for BREEDPLAN single-step evaluations.

These reference populations should be dynamic, continually supplemented and updated yearly, to ensure genetic trends in the population are captured. The significant cost involved in establishing reference populations and maintaining them requires strategic negotiations with appropriate industry bodies and commercial parties to determine funding structures and accountability. These strategic discussions are currently underway and will determine the expected time for implementation of assessment Levels 2-3 and reference populations.

CONCLUSIONS

This paper describes industry standards developed by AGBU for the inclusion of genotypes from new SNP panels into the BREEDPLAN single step evaluations. These requirements ensure accuracy and consistency across genomic records from various different genotyping platforms. Increasing numbers of new SNP panels are being introduced, requiring SNP panel analyses to determine BREEDPLAN compatibility. AGBU has developed an automated analysis process, which provides reports on new SNP panels for interested commercial and industry partners. This assessment process benefits industry by ensuring animals are not genotyped with panels that are not compatible with BREEDPLAN, and ensures genotyping quality is maintained.

ACKNOWLEDGEMENTS

This research includes data generated by the Beef CRC. This research is supported by Meat and Livestock Australia (MLA) project L.GEN.0174.

REFERENCES

Boerner V. (2017) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **22:** 97.

Boerner V., Wittenburg D. (2018) *Frontiers in Genetics* **9:** 185.

Boichard D., Chung H., Dassonneville R., David X., Eggen A., Fritz S., Gietzen K.J., Hayes B.J., Lawley C.T., Sonstegard T.S., Van Tassell C.P., VanRaden P.M., Viaud-Martinez K.A., Wiggans G.R., for the Bovine L.D.C. (2012) *PLOS ONE* **7:** e34130.

Connors N.K., Cook J., Girard C.J., Tier B., Gore K.P., Johnston D.J., Ferdosi M.H. (2017) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **22:** 317.

Johnston D.J., Ferdosi M.H., Connors N.K., Boerner V., Cook J., Girard C., Swan A., Tier B. (2018) *Proc. World Congr. Genet. App. Livest. Breed.* . **11:** 269.