THE EFFECT OF USING DNA OBTAINED FROM BLOOD OF CATTLE WITH GENETIC CHIMERISM ON ILLUMINA'S BEADCHIP ASSAY PERFORMANCE*

Artur Gurgul[1]♦, Dominika Rubiś[2], Monika Bugno-Poniewierska[1]

[1]Laboratory of Genomics, [2]Department of Animal Cytogenetics and Molecular Genetics, National Research Institute of Animal Production, 32-083 Balice n. Kraków, Poland
♦Corresponding author: artur.gurgul@izoo.krakow.pl

Abstract
Blood cell chimerism is a common phenomenon occurring in cattle coming from double or multiple parturitions and can be observed as two DNA profiles present in blood of each of twin born animals. In the era of genomics, a large number of animals is being genotyped with high throughput genotyping methods, which are giving limited insight into the performance of single markers and rather only statistical description of the results is available for a common user. This hampers the detailed analysis of the results obtained and direct identification of the causes of poorer performance of some samples. In this study we describe the influence of analysis of DNA obtained from blood samples of cattle with genetic chimerism on basic parameters of Infinium technology-based Illumina's genotyping arrays. The results obtained may help to identify such samples, especially when no precise information about the animals' origin is available.

Key words: cattle, chimerism, Infinium, SNP, twins

Together with the development of marker assisted selection or genomic breeding value estimation methods, molecular genetics is becoming a more and more widely used tool in animal breeding. The genomic selection in farm animals needs advanced and high throughput DNA analysis methods, which allow for fast and accurate genotyping of a huge number of polymorphisms spanning the entire genome. These requirements are satisfied by whole-genome genotyping assays, working based on BeadChip technology and designed to capture the possibly highest percentage of genetic variation (Hayes et al., 2009; Szyda et al., 2011).

In different laboratories, various biological materials are used for DNA extraction and purification for microarray analysis, starting from hair bulbs, through ear punches to peripheral blood and mouth swabs. It is well known that hair bulbs (simi-

larly as ear punches and swabs) are difficult in handling and give rather poor quality of DNA isolates, which may significantly affect the BeadChip assay performance. One of the best raw materials for DNA isolation, which can be sampled intravitally and almost non-invasively is peripheral blood. In many cases DNA extracted from this tissue has high quality and needs no further purification and concentration. However, blood (lymphocytes) – as practically the only cattle tissue – may have features of genetic chimerism (Niku et al., 2007).

The genetic chimerism is a condition in which a single organism is composed of two or more different populations of genetically distinct cells originating from different zygotes (Anderson et al., 1951). In cattle the most common form of chimerism is blood chimerism. This phenomenon occurs in twin-born cattle in which shortly after the implantation, placental vascular anastomoses between the two fetuses usually (in about 90% of cases) occur (Komisarek and Dorynek, 2002). This allows for the exchange of haemopoietic stem cells between fetuses. The cells settle in bone marrow and produce blood cells throughout the animal's lifetime, causing the genetic chimerism in peripheral blood, which can be observed as two DNA profiles in blood of each of twin born animals. The share of the cells showing co-twin genetic profile differs greatly between the individuals and seems to be unrelated to any recognized mechanism (Basrur and Kanagawa, 1969).

Twinning in cattle is not a common phenomenon – on average only 4–5% of all births are double (Cady and Van Vleck, 1978; Nielen et al., 1989). Twin pregnancies may have some adverse impact on fertility and production traits, or may lead to the syndrome called freemartinism in cows (Vigier et al., 1988, 1991). Nevertheless, animals from double pregnancies can be found in breeding population. This is caused by the fact that not all twin-born animals are affected by the adverse influence of the exchange of hormones through the vascular anastomosis between the opposite gender fetuses (Buoen et al., 1992; Zhang et al., 1994; Fricke, 2001) or by the fact that the information about the double parturition can be withheld by the breeders not to lower the value of the animal. Due to that, the twin born animals may be included in the genomic breeding value estimation programmes and the genetic chimerism observed in their blood may cause problems with the quality and reliability of the SNPs genotyping results obtained. To date, no information is available allowing to characterize and distinguish the effect of the genetic chimerism from other or non-biological factors (like quality of the DNA obtained, defective laboratory work and many others) on the Illumina's BeadChip assay performance. This study is aimed to provide information on how mixed genetic profiles may impair the performance of the Infinium technology-based genotyping assays.

Material and methods

Samples, genotyping and quality parameters evaluation

The test material was genomic DNA isolated from full blood of 22 unrelated cattle of the Polish Holstein-Friesian breed (Holstein – HO), both females and males, which were registered as coming from double calving (cattle with chimerism – CH

group). As a control, DNA from a single parent of each of the studied animals was used (CTR group). The blood for DNA isolation was obtained from the DNA Bank held by the National Research Institute of Animal Production, so no Local Ethics Committee permission was needed for this study. The mixed genetic profiles in the blood of the twin-born animals were confirmed with use of 12 standard microsatellite markers recommended by ISAG for parentage testing in cattle (Radko et al., 2010). Subsequently, all the samples were genotyped for about 7K SNPs with use of Illumina's BovineLD BeadChip v1.1 assay according to standard Infinium Ultra protocol. The results were analysed in GenomeStudio software, which also generated all basic statistics. As an additional control (referred as CTRad group), 22 samples obtained from HO cattle genotyped with use of Bovine SNP50 v2 assay were used. Only SNPs common for SNP50 and LD assays were evaluated in this group (6854 out of 6909 assayed in LD array). The samples showed lowered quality parameters (like call rates and gencall scores), but they were confirmed not to come from double parturitions and not having mixed genetic profiles as shown by microsatellite analysis. The quality parameters of the samples were lowered for unknown reasons, probably by the other or non-biological factors like DNA quality, defective laboratory work or other random factors.

Hardy-Weinberg equilibrium was analysed by using Fisher's exact test. The expected genotypic counts under Hardy-Weinberg equilibrium were calculated and evaluated for heterozygotes excess/deficiency.

B allele frequency (BAF) and log R ratio

B allele frequency (BAF) was calculated using GenomeStudio software. BAF may be defined as a normalized measure of the allelic intensity ratio of the two possible alleles. Theta intensity values were calibrated to generate BAF using canonical genotype clusters, produced by analysis of large sample sets by Illumina Company. BAF is a value between 0 and 1 and represents the proportion contributed by one SNP allele (B) to the total number of alleles: BAF is an estimate of $N_B/(N_A + N_B)$, where N_A and N_B are the number of A and B alleles, respectively (Staaf et al., 2008). This value is normalized, so that it is zero if theta is less than or equal to the AA cluster's theta mean, 0.5 if it is equal to the AB cluster's theta mean, or 1 if it is equal to or greater than the BB cluster's theta mean. The BAF distribution was analysed in all three groups separately.

The log R ratio parameter, referring to the absolute SNP intensity (R), was calculated in GenomeStudio software as $(\log_2(R_{observed}/R_{expected}))$, where R expected is interpolated from the observed allelic ratio with respect to the canonical genotype clusters.

Results

Changes in the values of basic quality parameters in samples with blood chimerism

The first step in assessing the quality of the results obtained from Illumina's Infinium assay is the revision of control dashboard and basic parameters like call rate

(the percentage of markers called in respect to the number of markers assayed) and 10%GC (10th percentile of GenCall score over all SNP) for all samples. The analysis of these parameters in most cases is sufficient for identification of problematic samples, but it does not give the precise information about the source of the issue. In this experiment, the control dashboard for all the analysis did not show any problems with the assays, which suggests that all steps of array preparations like single base extension, target removal or staining were impeccable. When analysing call rates and 10%GC scores we found that all samples obtained from cattle with blood chimerism (CH group) showed some deviations from the level of parameters observed for control group. The mean call rate for control samples was 0.996 (±0.002) and the 10%GC score ranged from 0.807 to 0.8161 with mean value of 0.813 (±0.003). For samples with genetic chimerism the values were clearly lower and developed as 0.847 (±0.056) – mean call rate and 0.625 (±0.127) – mean 10%GC with minimum and maximum values of 0.408 and 0.813, respectively. The additional control samples showed quality parameters of intermediate values when compared to both remaining groups. Although the mean call rate for this group was rather close to that observed in control samples (0.963±0.079), the mean 10%GC was low (0.627±0.0718) with minimum and maximum values of 0.317 and 0.665 and was strongly similar to that observed in CH samples.

When analysing Hardy-Weinberg equilibrium (HWE) for all studied markers we presumed that CH group should be characterized by an increased number of heterozygotes. The assumption was based on the fact that mixed genetic profile should show increased number of heterozygous calls arising from manifestation of opposite homozygotes present in original genetic profiles of each of the cell lines. The size of this phenomenon should be correlated with the share of the cells showing co-twin genetic profile, and should only result in heterozygous calls when about half of the cells will be showing genetic profile of the co-twin. In the subset of control animals, the genotype distribution of 399 SNPs deviated from Hardy-Weinberg equilibrium (HWE) at the significance level of 0.05. Of these markers, 210 (53%) showed significant overrepresentation of heterozygotes. In the group of twins, 1028 markers deviated from HWE and as expected, as much as 941 (92%) showed significant heterozygotes excess. In CTRad group the deviation from HWE was observed for 600 (8.7%) markers of which only 176 (29%) showed heterozygote excess, which is clearly less than in CH group. Between the parent-child pairs in CH an CTR groups, we found no inconsistencies in Mendelian inheritance across all called SNPs.

B allele frequency (BAF) and log R ratio

It was shown that by the analysis of pooled DNA samples on genotyping microarrays it is possible to establish the share and frequency of separate alleles within the pool (Meaburn et al., 2005; Peiris et al., 2011). This approach is currently based on analysis of B allele frequency parameter (BAF), which reflects the intensity of each allele in the studied material (Peiris et al., 2011). On the basis of that result we assumed that presence of mixed genetic profiles in the studied samples, caused by genetic chimerism, will be the most clearly visible in deviations of BAF values. The mean, median and standard deviation of BAF for separate groups was similar and

did not allow distinguishing the mixed genetic profiles. Based on data obtained from CTR group, we established that most of BAF values in our samples clustered in the ranges of 0–0.125, 0.375–0.725 and 0.925–1 (Figure 1). These intervals were set as standard BAF values for samples analysed in our laboratory and number of out-laying observations was calculated for each group separately. For CTR group only 0.49% of all BAF observations outlaid specified intervals. In CH group, as much as 20.89% of markers showed BAF values in inter-clusters regions. The CTRad group was characterized by intermediate percentage of outlying observations, which deve-loped as 2.92. The devastations in CH group were clearly visible in BAF plots across all markers. In most cases of chimerism, BAF values formed additional clusters of observations in the ranges between the expected intervals (Figure 1).
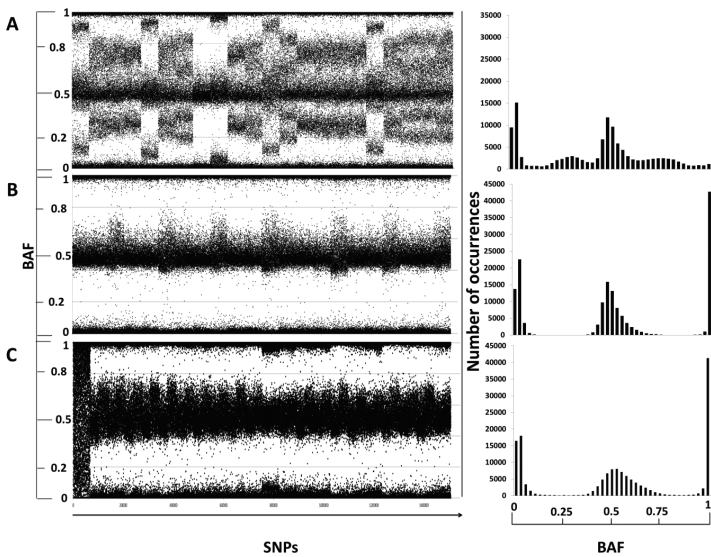
Clear split in BAF values is visible for most samples with mixed genetic profiles (A).
Deviating observations are forming additional clusters of observations in between-expected-clusters
ranges. No such tendency is observed in both control (B) and additional control (C) samples.

Figure 1. B allele frequency (BAF) plots (left side) and BAF histograms (right side) for samples with blood chimerism (A), control group (B) and additional control group of poor performing samples (C)

Another parameter which is being disturbed when low quality or inappropriate concentrations of DNA are being used for microarray analysis is log R ratio. Geno-typing results of common samples showing log R ratio standard deviation higher than 0.3 are considered to be of poor quality, especially for copy number variation analysis, which uses intensity parameters to evaluate the number of copies of specific genomic regions (Metzger et al., 2013). In CH and CTR groups, the standard devia-tion for log R ratio developed as 0.23 and 0.27, respectively. Only in CTRad group this parameter exceeded the assumed threshold and developed as 0.35. The plots of log R ratio across all samples and all markers are shown in Figure 2.
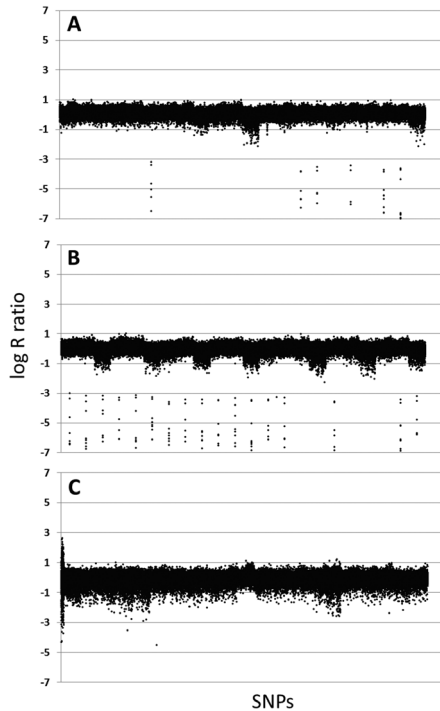
Similar distribution of LRR values was observed for both chimeric and control samples (A and B). Higher deviations in LRR were observed for additional control group – showing rather loss of absolute intensity for separate markers.

Figure 2. Plot for log R ratio (LRR) for samples with chimerism (A), control group (B) and additional control group of poor performing samples (C)

Discussion

In high throughput genotyping laboratories the quality of the results obtained and their reliability is one of the most important issues. All factors affecting performance of the applied methods should be promptly identified and eliminated to minimize the financial outlays associated with inadequate results. The accurate identification of these issues is not always straightforward and may take some time to pinpoint the source of the problem. In this study we attempt to analyse the impaired functioning of Infinium based genotyping assays (Illumina), resulting from analysis of samples with mixed genetic profiles. Although the results of our study will not allow for the discrimination of cases of samples contaminated with DNA of other individuals from cases of genetic chimerism, it may help to recognize the issue and allow paying special attention to the material used or procedures applied in the laboratory.

Based on the results obtained, we found that basic quality parameters used for evaluation of Infinium assay performance does not allow for reliable identification of the source of impaired results obtained during genotyping. Both call rates and quality scores for the group with blood chimerism were comparable with those obtained for the additional control group of poor performing samples. The only difference observed was that the CH samples were characterized by slightly lower call rates at the similar level of gencall scores. More suggestive results were obtained from HWE analysis, which indicated that the higher percentage of SNPs shows excessive heterozygous calls in CH samples than in both remaining groups.

Based on previous reports (Craig et al., 2009; Conlin et al., 2010; Holl et al., 2013) we assumed that the mixed genetic profiles present in one sample should be easily recognized by the analysis of intensity parameters, which will deviate from the expected values. This will be especially visible in the B allele frequency distribution across genome. For instance, if two genetic profiles (with equal share of different cell lines) will be present in one sample there is possibility that two different genotype intensities will be analysed at a single locus. Assuming that in one of the cell lines present in chimeric blood AA genotype will be observed and AB in the other line, the allelic intensity expressed by BAF should be around 0.33, i.e. outside the expected BAF ranges. These phenomena will be seen for several SNPs across the genome and will result in splits in BAF plots. However, the share of cell lines with different genetic profiles in the blood of animals with blood chimerism is variable between individuals, which hampers the calculation of BAF ranges characteristic for mixed genetic profiles. Interestingly, in a previous report, the share of different cell lines in the samples with chromosomal chimerism was estimated by analysis of size of BAF and log R ratio shifts (Conlin et al., 2010).

It can be concluded that analysis of BAF plots or histograms in most cases is sufficient to identify samples showing mixed genetic profiles arising from blood chimerism or substantial contaminations with other samples. The visible split in BAF plots was observed for most of the studied samples with blood chimerism, but not in other poor performing samples or good quality samples. We suggest that the revision of BAF histograms can be a good indicator of mixed DNA profiles present in the samples and can help to identify at least some of the possible causes of inferior performance of Infinium assay.

References

A n d e r s o n D., B i l l i n g h a m R.E., L a m p k i n G.H., M a d a w a r P.B. (1951). The use of skin-grafting to distinguish between monozygotic and dizygotic twins in cattle. Heredity, 5: 379–397.

B a s r u r P.K., K a n a g a w a H. (1969). Parallelism in chimeric ratios in heterosexual cattle twins. Genetics, 63: 419–425.

B u o e n L.C., Z h a n g T.Q., V e b e r A.F., R u t h G.R. (1992). Non-freemartin rate in Holstein heterosexual twins. Am. Assoc. Bov. Pract. Confr., 1: 300–303.

C a d y R.A., Va n V l e c k L.D. (1978). Factors affecting twinning and effects of twinning in Holstein dairy cattle. J. Anim. Sci., 46: 950–956.

C o n l i n L.K., T h i e l B.D., B o n n e m a n n C.G., M e d n e L., E r n s t L.M., Z a c k a i E.H., D e a r d o r f f M.A., K r a n t z I.D., H a k o n a r s o n H., S p i n n e r N.B. (2010). Mechanisms of mosaicism, chimerism and uniparental disomy identified by single nucleotide polymorphism array analysis. Hum. Mol. Genet., 19: 1263–1275.

C r a i g D.W., M i l l i s M.P., D i S t e f a n o J.K. (2009). Genome-wide SNP genotyping study using pooled DNA to identify candidate markers mediating susceptibility to end-stage renal disease attributed to Type 1 diabetes. Diabet. Med., 26: 1090–1098.

F r i c k e P.M. (2001). Review: Twinning in Dairy Cattle. Prof. Anim. Sci., 17: 61–67.

H a y e s B.J., B o w m a n P.J., C h a m b e r l a i n A.C., V e r b y l a K., G o d d a r d M.E. (2009). Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet. Sel. Evol., 41: 51.

H o l l H.M., L e a r T.L., N o l e n - W a l s t o n R.D., S l a c k J., B r o o k s S.A. (2013). Detection of two equine trisomies using SNP-CGH. Mamm. Genome, 24: 252–256.

K o m i s a r e k J., D o r y n e k Z.J. (2002). Genetic aspects of twinning in cattle. J. Appl. Genet., 43: 55–68.

M e a b u r n E., B u t c h e r L.M., L i u L., F e r n a n d e s C., H a n s e n V., A l - C h a l a b i A., P l o m i n R., C r a i g I., S c h a l k w y k L.C. (2005). Genotyping DNA pools on microarrays: tackling the QTL problem of large samples and large numbers of SNPs. BMC Genomics, 6: 52.

M e t z g e r J., P h i l i p p U., L o p e s M.S., d a C a m a r a M a c h a d o A., F e l i c e t t i M., S i l v e s t r e l l i M., D i s t l O. (2013). Analysis of copy number variants by three detection algorithms and their association with body size in horses. BMC Genomics, 14: 487.

N i e l e n M., S c h u k k e n Y.H., S c h o l l D.T., W i l b r i n k H.J., B r a n d A. (1989). Twinning in dairy cattle: a study of risk factors and effects. Theriogenology, 32: 845–862.

N i k u M., P e s s a - M o r i k a w a T., T a p o n e n J., I i v a n a i n e n A. (2007). Direct observation of hematopoietic progenitor chimerism in fetal freemartin cattle. BMC Vet. Res., 3: 29.

P e i r i s B.L., R a l p h J., L a m o n t S.J., D e k k e r s J.C. (2011) Predicting allele frequencies in DNA pools using high density SNP genotyping data. Anim. Genet., 42: 113–116.

R a d k o A., S ł o t a E., M a r c z y ń s k a J. (2010). Usefulness of a supplementary set of microsatellite DNA markers for parentage testing in cattle. Pol. J. Vet. Sci., 13: 313–318.

S t a a f J., V a l l o n - C h r i s t e r s s o n J., L i n d g r e n D., J u l i u s s o n G., R o s e n q u i s t R., H ö g l u n d M., B o r g A., R i n g n é r M. (2008). Normalization of Illumina Infinium whole-genome SNP data improves copy number estimates and allelic intensity ratios. BMC Bioinformatics, 9: 409.

S z y d a J., Ż a r n e c k i A., S u c h o c k i T., K a m i ń s k i S. (2011). Fitting and validating the genomic evaluation model to Polish Holstein-Friesian cattle. J. Appl. Genet., 52: 363–366.

V i g i e r B., W a t r i n F., M a g r e S., T r a n D., G a r r i g o u O., F o r e s t M.G., J o s s o N. (1988). Anti-mullerian hormone and freemartinism: inhibition of germ cell development and induction of seminiferous cord-like structures in rat fetal ovaries exposed *in vitro* to purified bovine AMH. Reprod. Nutr. Dev., 28: 1113–1128.

V i g i e r B., M a g r e S., C h a r p e n t i e r G., B e z a r d J., J o s s o N. (1991). Anti-mullerian hormone and natural and experimental freemartin effect. Bull. Assoc. Anat., 75: 29–32.

Z h a n g T.Q., B u o e n L.C., S e g u i n B.E., R u t h G.R., W e b e r A.F. (1994). Diagnosis of freemartinism in cattle: the need for clinical and cytogenetic evaluation. J. Am. Vet. Med. Assoc., 204: 1672–1675.