Making vast existing data resources available for genomic imprinting analyses

I. Blunk1,2, M. Mayer2, H. Hamann3 & N. Reinsch2

1 University of Rostock, Faculty of Agricultural and Environmental Sciences, Justus-von-Liebig-Weg 6, 18059 Rostock, Germany
2 Leibniz Institute for Farm Animal Biology (FBN), Institute for Genetics and Biometry, Wilhelm-Stahl-Allee 2, Dummerstorf 18196, Germany
3 State-Office for Geo-information and Rural Development Baden-Wuerttemberg, Geodata-Center, Stuttgarter Straße 161, Kornwestheim 70806, Germany
reinsch@fbn-dummerstorf.de

Summary

Alleles of imprinted loci are fully or partially inactivated through epigenetic mechanisms, depending on their parental origin. Their effects on phenotypes contribute to the broader class of parent-of-origin effects (POEs). Standard methodology for mapping imprinted loci in association or linkage analyses requires both phenotypes and ordered (phased) marker genotypes to be known from the same individuals. Therefore, large numbers of already recorded phenotypes from ungenotyped progeny as well as available genotypes from parents without own phenotypes remain unused for such genomic imprinting analyses. Both by theoretical considerations and simulation studies we were able to prove that the above mentioned limitations can be overcome. Consequently, in association studies phenotypic information from e.g. ungenotyped progeny can be related to unphased marker genotypes of parents. This can be achieved by first using all available phenotypic information in estimating POEs from population data with a suitable mixed model and later employing these ePOEs as dependent variables in an association analyses after matching ePOEs with unphased genotypes. The theoretical foundation is that at a single imprinted biallelic locus the regression of the true POE – defined as the difference of a genotype’s transmitting abilities under a male and a female expression pattern – on the simple gene count at this locus is exactly equal to the imprinting effect. As a demonstration of the practical applicability of our approach the ePOEs for up to 1,857 Brown Swiss sires were used to map imprinted quantitative trait loci expressed in their fattening progeny. For the trait net body weight gain a genome-wide significant signal was found on BTA11. Chromosome-wide significant signals were also found for muscularity and fatness traits. Thus, ePOEs provide a new opportunity to cost-effectively detect imprinted loci without genotyping of progeny and allows the exploitation of data, which otherwise would remain unused for imprinting analyses.

Keywords: epigenetics, genomic imprinting, genome wide association study

Introduction

Paternal and maternal gametes are differently expressed in case the corresponding gene is imprinted. The effects of imprinted genes contribute to the epigenetic part of the genetic
variation and are attributed to the family of parent-of-origin effects (POEs). When imprinted loci were investigated in earlier studies, these studies presupposed that the parental origin of the alleles (phase) of heterozygote genotypes can be established. A difference between their average phenotypes would indicate the presence of an imprinted quantitative trait locus (iQTL). However, phase of marker genotypes is not always certain and it is required that the individuals with genotypes also deliver the phenotypes, although traits of interest are frequently measured in progeny without genotypes (e.g., fattening progeny). Based on the assumption that the resulting regression coefficient is the imprinting effect, we propose to regress estimated POEs (ePOEs) of parents on their unphased gene counts to map iQTLs expressed in their progeny. This would relieve genome scans in imprinting analyses from the necessity of phased genotypes to be available for all animals with phenotypes. Firstly, this proposal was examined in theory and was secondly applied to Brown Swiss cattle slaughterhouse data.

Material and methods

Simulation study

A single two-generation population was simulated. In analysis A, the general ability to map iQTLs by regressing the ancestor’s ePOEs as dependent variables on their gene counts was examined. In analysis B, the results were compared to the case when transmitting abilities (TAs) were used. The ePOEs and TAs were estimated for 100 animals in a founder generation. These animals were inter-mated in a cross-classified manner leading to a second generation with 10,000 full-sib families and 3 descendants per family. Thus, 30,000 progeny provided records, which were used as observations to estimate the original parental ePOEs and TAs via a special type of imprinting model (Blunk et al., 2017). The records were created by adding the additive and imprinting effects according to the simulated QTL genotypes. A residual variance of two led to a heritability of 0.40. Fifteen mutually independent markers were simulated, where five were in linkage disequilibrium (LD) with five QTLs (LD of 0.1 in all cases), respectively. The QTL near marker two with a minor allele frequency (MAF) of 0.4 was biparentally expressed contributing 30% to the total additive genetic variance of 1.35. A purely paternally (maternally) expressed QTL nearby locus marker six (10) with a MAF of 0.5 (0.4) contributed 5% (5%) to the total additive genetic variance and 25% (25%) to a total imprinting variance of 0.27. QTLs near markers 14 and 18 (MAF = 0.5, respectively) were contrarily partially imprinted and contributed 30% to the additive genetic variance and 25% to the imprinting variance. The imprinting variance accounted for 20% of the total additive genetic variance. For each analysis 100 replications were performed and the results averaged.

Brown Swiss data

Parental ePOEs and their reliabilities were generated in a previous analysis of slaughterhouse data measured in Brown Swiss fattening-progeny (Blunk et al., 2017). The ePOEs contributed a significant proportion of the imprinting variance to the total genetic variance in the net BW gain (g/d), carcass musculature and carcass fatness. Musculature was described using five monetary grades, whereas fatness was categorized by scores from one (lean) to five (very fat). Both traits were analyzed via a linear (muscleL; fatL) and threshold (muscleT; fatT) imprinting model. Prior to the genome scan, the ePOEs were parent-average corrected, de-regressed and weighted according to methods in Garrick et al. (2009) for breeding values.
The average of de-regressed reliabilities of ePOEs was 0.15. For 1,857 sires, 37,443 genotypes were available for the trait net BW gain and 37,433 genotypes were available for 1,831 sires for the muscularity and fatness traits. In the genome scan the ePOEs were regressed on the gene count of each marker representing un-phased genotypic information. The marker effects were individually tested conducting a conditional Wald F-statistic via ASReml 3.0 (Gilmour et al. 2009). The genetic variance due to relationships between sires was captured by including the inverse of the additive genomic relationship matrix (VanRaden, 2008). All genomic information was obtained from the UCSC Genome Browser (http://genome.ucsc.edu) based on the Bos Taurus UMD3.1.1/bosTaur8 (assembly date: Dec. 2009). Pairwise LD between single nucleotide polymorphisms (SNPs) was specified as $r^2$ using Haplovew 4.2 (Barrett et al., 2005). The imprinting status of genes was derived from the Geneimprint database (http://www.geneimprint.org; R. L. Jirtle)

**Results and discussion**

**Simulated data**

In analysis A, it was investigated whether ePOEs are suitable to detect iQTLs. At the two simulated markers representing fully imprinted QTLs, the $p$-values were $p = 0.012$ and $p = 0.025$, which indicated the presence of iQTLs, whereas the effect estimated at marker locus two (linked to a Mendelian QTL) did not significantly deviate from zero with $p = 0.459$. The signals at the marker loci linked to partial iQTLs remained visible with $p$-values of 0.010, respectively (Fig. 1A). Thus, analysis A provides strong evidence that ePOEs are suitable to map iQTLs.

To examine whether TAs can be used to detect iQTLs, TAs of parents were regressed on their gene counts in analysis B. With $p$-values ranging from 0.001 to 0.002, the marker effects due to Mendelian and partially imprinted loci showed clear deviations from zero. The effects estimated for the two fully imprinted QTLs did not deviate from zero with $p$-values of 0.328 and 0.369 (Fig. 1B). Thus, TAs were not suitable to detect fully imprinted loci.

**Practical data**

A SNP on BTA11 was associated with ePOEs estimated in net BW gain at the 5% genome-wide significance level (Fig. 2). The type I error rate was controlled using the concept of the false discovery rate (Benjamini and Hochberg, 1995). The SNP can be assigned to the receptor accessory protein 1 gene (*REEP1*), whose imprinting status in cattle is unknown. Imumorin et al. (2011) identified bovine iQTLs containing orthologs of imprinted genes in mice and human. One iQTL with an effect on weaning weight was located on BTA11. However, the iQTL is located about 25Mb away from the detected SNP locus. The human ortholog of *REEP1* is related to the hereditary spastic paraplegia, a syndrome characterized by progressive lower-limb spastic paralysis (Züchner et al., 2006). In mice, the expression of *Reep1* was suggested to regulate the adipogenesis in white adipose tissue (Renvoisé et al., 2016). With regard to muscleL, two 5% chromosome-wide significant SNPs were found on BTA7 and BTA18 (Fig. 2). The presence of imprinted genes on BTA7 is unknown but four imprinted genes were reported for BTA18. However, these genes are located far away from the SNP found in our study. Moreover, the SNP could not be reproduced when the ePOEs estimated in muscleT were analyzed. Instead, a 5% chromosome-wide significant SNP was found for this trait in a non-coding area on BTA7 (Fig. 2). However, no relationship can be
assumed to the SNP found on BTA7 when ePOEs on muscleL were analyzed because this marker is distantly located. With regard to fatL, no significant SNPs were found. In contrast, when fatT was analyzed, 45% chromosome-wide significant SNPs were found on BTA5 (Fig. 2). The SNPs with the strongest signals were located within or close to the trophinin associated protein gene (TROAP) and the protein lifeguard 2 gene (FAIM2). So far, no imprinted loci on BTA5 are known in cattle. However, the bovine orthologue of the imprinted Sodium-coupled neutral amino acid transporter 4 gene (SLC38A4) in mice is located on BTA5 close to TROAP and FAIM2. This may suggest an imprinting cluster on BTA5. With regard to fatT, another 5% chromosome-wide significant SNP was found on BTA9 (Fig. 2). The SNP lies between the imprinted PLAG1 like zinc finger 1 gene and the imprinted insulin like growth factor 2 receptor gene. LD ($r^2 \leq 0.03$) suggested full independence. Another 5% chromosome-wide significant SNP was found on BTA29 (Fig. 2). It is, however, located in a great distance from the imprinted Insulin-like growth factor 2 gene and the H19 gene.

Conclusion

Our study suggests new genome regions in cattle possibly being subject to genomic imprinting. Moreover, our method made it possible to use phenotypic data and genotypes, which would otherwise remain unused for imprinting analyses.

In general, as ePOEs were used as pseudo-phenotypes, it remains unclear whether their effects actually arose from genomic imprinting or another parent-specific genetic phenomenon (e.g., maternal genetic effects).

List of References

Figures

**Figure 1.** Average (100 repetitions) $-\log_{10} p$-values generated by regressing the ancestor’s estimated parent-of-origin effects A) and their transmitting abilities B) on their own gene counts at 15 marker loci in simulated data sets.
Figure 2. Marker loci in relation to their $-\log_{10} p$-values generated by regressing estimated parent-of-origin effects in Brown Swiss slaughter data on unphased genotypes. Red line = 5% genome-wide significant; dashed line = 10% genome-wide significant; blue diamonds = 5% chromosome-wide significant markers.