A. Non-synonymous coding variants and variants within 5000bp of genes, or
B. Intergenic variants excluding regions close to genes (±5000bp).

Prediction accuracy for group A = 0.35 while B = 0.27. When sets A and B were combined, prediction accuracy = 0.39. The high LD in the cattle genome means that there is considerable overlap in the predictions from set A and B. However, the results do suggest that variants in and close to coding regions explain a large proportion of the trait variance but that intergenic regions are also important in regulating trait expression, in keeping with evidence from the human ENCODE project (Skipper *et al.* 2012). The fertility trait appeared to be highly polygenic, with an average of 3305 SNP effects fitted in the model. This is expected because the fertility trait was largely based on calving interval: a complex trait influenced by many factors such as cow energy balance, oocyte health and embryo development.

In Figure 1 we present some examples of QTL discovery from among the 50 most significant QTL regions occurring within 5000bp of a gene: first to demonstrate the advantages of our approach and second to illustrate the range of genetic factors that underpin the complex female fertility trait.

Several regions only showed strong evidence for QTL in the SEQ analysis, demonstrating the improved power of SEQ genotypes. One example (Fig. 1a) is a rare variant in the 3' UTR of the SCARA5 gene showing a strong probability of being either the causal variant or one in strong LD with a QTL. This SNP is likely a relatively recent mutation because it segregated only in the black and white Holsteins (MAF=0.08) and was not in strong LD with any other SNP. SCARA5 expression is upregulated in human endometrium tissue when an early embryo is present (Duncan *et al.* 2011), and was also found to be more highly expressed in bovine ovary tissue compared to 17 other tissues (Chamberlain *et al*. 2014). A second example (Fig 1b) is two SEQ variants in high LD (Holstein only, MAF=0.025). The highest probability variant lies between SMEK1 and CCDC88C gene, while the other is a missense mutation in SMEK1. Potentially either gene could be considered to be a good candidate. CCDC88C is a negative regulator in the Wnt signalling pathway that regulates embryo germ cell development (Enomoto *et al.* 2006). SMEK1 has been demonstrated to regulate hepatic gluconeogenesis in mice (Yoon *et al.* 2010) and also appears to regulate the differentiation of embryonic stem cells (Lyu *et al.* 2011). In an analysis of the same data for milk traits (results not shown), these same mutations have a strong association with milk yield and the allele that increased milk yield reduced fertility.

A further region that showed a strong association for fertility and milk yield is between the GC and NPFFR2 genes (Fig. 1c). There is strong LD across this region and the association was spread across several variants. Again both genes are potentially good candidates: GC encodes Vitamin D transporter and disruption of the Vitamin D pathway affects oestrogen biosynthesis, while NPFFR2 interacts with kisspeptin which plays a key role in neuroendocrine regulation of reproduction (Matzuk and Lamb 2008). A number of regions on the X chromosome showed several strong QTL signals including SNP very close to KAL1 (Fig 1d) and UBE2A. In humans several mutations in KAL1 are responsible for “Kallmann syndrome”, affecting the embryonic migratory pathway of neurons that synthesize gonadotropin-releasing hormone (Hardelin *et al.* 1992). This results in impaired gonad development in males and females. Mutations in UBE2A have been shown to be associated with maternal effects on early embryo survival (Matzuk and Lamb 2008).

The validity of our results are dependent on the accuracy of imputation and reference genome annotation, neither of which is perfect. However, this study demonstrates that imputed sequence genotypes with Bayesian analysis improved the accuracy of genomic prediction and the QTL discovery highlighted a broad range of genetic factors potentially affecting dairy cow fertility.