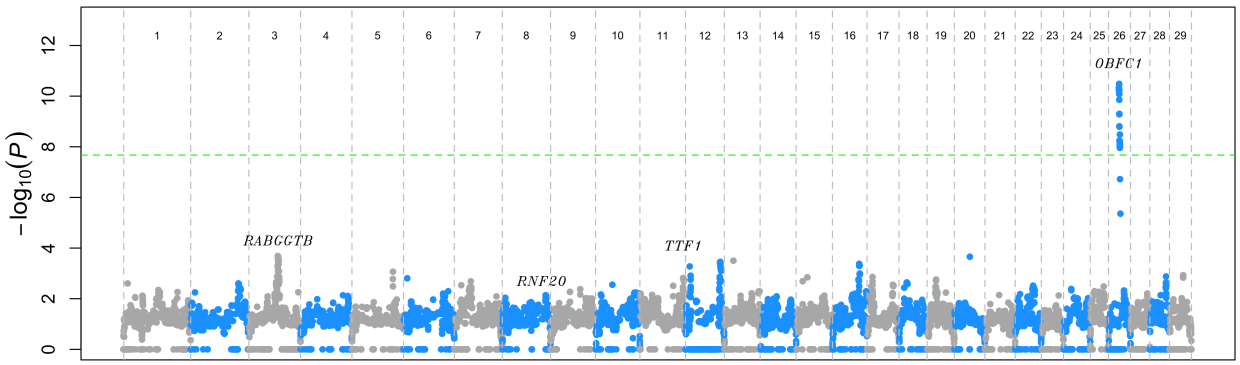
Supplemental Material S4: Haplotype-based genome scan for EL mutations.

We performed a haplotype-based scan to identify regions with haplotypes with significant depletion in homozygotes. We used the data set described previously[1]. It consists in a dairy cattle population from New-Zealand (NZ; mainly Holstein, Jersey and crossbred individuals) including 58,369 individuals genotyped on either Illumina Bovine 50K (v1 and v2) or Illumina BovineHD arrays. We kept markers common to the three arrays and mapping to bovine autosomes (using UMD 3.1 Bovine Reference genome assembly). After checking for parentage errors, we removed markers with a call rate < 95%, generating more than 10 Mendelian inconsistencies, which were monomorphic or strongly deviating from Hardy-Weinberg proportions (p < 1e-8). In addition, we removed 35 small segments that are associated with errors in the genome build[1]. The final data set contained 37,769 SNPs. Remaining Mendelian inconsistencies were erased (removing genotypes in either the offspring, the parent or both).

Haplotypes were first reconstructed based on familial information using LINKPHASE3[1]. The partial haplotypes were further phased (some markers remain unphased) using LD information with DAGPHASE[2] and Beagle[3]. Beagle automatically clusters haplotypes at each marker position based on local similarity using variable length Markov chains as previously described[4].

Regions with putative EL mutations were identified by testing for deviation from Hardy–Weinberg (HW) equilibrium separately for each haplotype cluster (individuals can either carry 0, 1 or 2 copies of a given haplotype cluster). We only considered significant p-values when reflecting a depletion in homozygotes, and when the number of homozygotes for the haplotype numbered < 10.

The results are illustrated in the accompanying Manhattan plot. The genome-wide significance (indicated by green dashed line) was set at p=2.13 e-08 corresponding to Bonferroni corrected p of 0.05 for 2,349,367 tests (performed at 37,769 marker positions each with on average 62 haplotype clusters). Only one region on BTA26 (corresponding to the *OBFC1* EL) showed a genome-wide significant depletion in homozygosity (p = 3.3 e-11). At the most significant position, the haplotype cluster driving the signal had a frequency of 2.7% in the population, while no homozygote individuals were observed. The haplotype-based approach gave no significant signal for the three other ELs detected in the NZ population (*RABGGTB, RNF20* and *TTF1*).



1. Druet T, Georges M. LINKPHASE3: an improved pedigree-based phasing algorithm robust to

genotyping and map errors. *Bioinformatics* 31, 1677-9 (2015).

2. Druet T, Georges M. A hidden markov model combining linkage and linkage disequilibrium information for haplotype reconstruction and quantitative trait locus fine mapping. *Genetics* 184, 789-98 (2010).

3. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet* 84, 210-23 (2009).

4. Browning SR. Multilocus association mapping using variable-length Markov chains. *Am J Hum Genet* 78, 903–913 (2006).