



OPEN

Four novel candidate causal variants for deficient homozygous haplotypes in Holstein cattle

Irene M. Häfliger^{1✉}, Mirjam Spengeler², Franz R. Seefried² & Cord Drögemüller¹

Mendelian variants can determine both insemination success and neonatal survival and thus influence fertility and rearing success of cattle. We present 24 deficient homozygous haplotype regions in the Holstein population of Switzerland and provide an overview of the previously identified haplotypes in the global Holstein breed. This study encompasses massive genotyping, whole-genome sequencing (WGS) and phenotype association analyses. We performed haplotype screenings on almost 53 thousand genotyped animals including 114 k SNP data with two different approaches. We revealed significant haplotype associations to several survival, birth and fertility traits. Within haplotype regions, we mined WGS data of hundreds of bovine genomes for candidate causal variants, which were subsequently evaluated by using a custom genotyping array in several thousand breeding animals. With this approach, we confirmed the known deleterious *SMC2*:p.Phe1135Ser missense variant associated with Holstein haplotype (HH) 3. For two previously reported deficient homozygous haplotypes that show negative associations to female fertility traits, we propose candidate causative loss-of-function variants: the HH13-related *KIR2DS1*:p.Gln159* nonsense variant and the HH21-related *NOTCH3*:p.Cys44del deletion. In addition, we propose the *RIOX1*:p.Ala133_Glu142del deletion as well as the *PCDH15*:p.Leu867Val missense variant to explain the unexpected low number of homozygous haplotype carriers for HH25 and HH35, respectively. In conclusion, we demonstrate that with mining massive SNP data in combination with WGS data, we can map several haplotype regions and unravel novel recessive protein-changing variants segregating at frequencies of 1 to 5%. Our findings both confirm previously identified loci and expand the spectrum of undesired alleles impairing reproduction success in Holstein cattle, the world's most important dairy breed.

Holstein is by far the most popular breed of cattle in the world, bred and raised for its high milk yield¹. Especially with the improved reproductive technologies that arose in the last century, the genomic exchange over the world increased and led to the use of a limited number of prominent sires across the globe¹⁻³. For Holstein cattle the difference between the absolute population size and the effective population size differs highly in all parts of the world and the inbreeding is increasing after the recent implementation of genomic selection⁴. With the effective population size being much smaller, the decrease in genetic variation and the increase in inbreeding is accelerated². There has been a recognisable decline in reproductive performance in high producing dairy cattle^{5,6}, which could be shown to be due to artificial selection on production traits and negative hitchhiking effects³. Since fertility problems are one of the most common reasons for culling cattle^{7,8} and animals of the Holstein breed are more likely to be culled than animals of other breeds⁸, there is a high need to improve female fertility, especially in Holstein cattle.

Genetic analyses of female fertility range from genome-wide association studies (GWAS) identifying genomic regions associated with fertility traits⁹⁻¹², to the identification of genomic regions showing reduced homozygosity due to recessive mostly embryonic lethal variants¹³ and to the investigation of effects of specific variants in functional candidate genes^{14,15}. Due to the routinely used single nucleotide polymorphism (SNP) array genotyping implemented in breeding programs about a decade ago, comprehensive population-wide genomic data of the current breeding populations are available. VanRaden et al. were the first to propose to screen this kind of data to identify genomic regions with fewer homozygous animals than expect. In their article they listed five haplotypes deviating significantly from the Hardy-Weinberg equilibrium (HWE) and thereby potentially harbouring lethal variants in the American Holstein population (Table S1)¹³. Three of these haplotypes, namely Holstein haplotypes 1, 2 and 3 (HH1-HH3), showed significant negative associations to reproduction traits. Interestingly, two

¹Institute of Genetics, Vetsuisse Faculty, University of Bern, Bremgartenstrasse 109a, 3001 Bern, Switzerland. ²Qualitas AG, 6300 Zug, Switzerland. ✉email: irene.haeffliger@vetsuisse.unibe.ch

Haplotype name	Associated disorder	Gene	OMIA	Variant description					Publications	
				chr	Position ^a	Description	Transcript of interest ^b	Coding DNA change		Protein change
HH0	Brachyspina	<i>FANCI</i>	000151-9913	21	21,184,869–21,188,198	Gross deletion (frameshift)		3.3 kb deletion		19,20,27,85,86
HH1	Embryonic lethality	<i>APAF1</i>	000001-9913	5	62,810,245	SNV (non-sense)	NM_001191507.1	c.1702C>T	p.Gln568*	13,16,22,85
HH2	Embryonic lethality	<i>IFT80</i>	001823-9913	1	107,172,616	SNV (frameshift)	XM_024984168.1	c.747delT	p.Leu250fs	13,24,28,85
HH3	Abortion	<i>SMC2</i>	001,824-9913	8	93,753,358	SNV (mis-sense)	XM_015472668.2	c.3404T>C	p.Phe-1135Ser	13,18,19,85,87
HH4	Abortion	<i>GART</i>	001826-9913	1	1,997,582	SNV (mis-sense)	NM_001040473.2	c.869A>C	p.Asn290Thr	16,27,85
HH5	Abortion	<i>TFB1M</i>	001941-9913	9	93,223,651–93,370,998	Gross deletion		139 kb deletion		26,27,85
HH6	Embryonic lethality	<i>SDE2</i>	002149-9913	16	29,020,700	SNV (start-lost)	NM_001099065.2	c.2T>C	p.Met1?	27
HH7	Embryonic lethality	<i>CENPU</i>	001830-9913	27	15,123,636	5 bp deletion (splice site)	XM_002698654.5	c.15123637_15123640delTTACT		17
HHB	Bovine leukocyte adhesion deficiency (BLAD)	<i>ITGB2</i>	000595-9913	1	144,770,078	SNV (mis-sense)	NM_175781.1	c.383A>G	p.Asp128Gly	51,85
HHC	Complex vertebral malformation (CVM)	<i>SLC35A3</i>	001340-9913	3	43,261,945	SNV (mis-sense)	NM_001105386.1	c.538G>T	p.Val180Phe	21,85,88
HHD	Deficiency of uridine monophosphate synthase (DUMPS)	<i>UMPS</i>	000262-9913	1	69,151,931	SNV (non-sense)	NM_177508.1	c.1213C>T	p.Arg405*	85,89,90
CDH	Cholesterol deficiency; juvenile mortality	<i>APOB</i>	001965-9913	11	77,891,733	Large insertion (frameshift)		ERV insertion		26,29,30,53,85

Table 1. Previously identified lethal haplotype regions and their associated candidate causal variants in Holstein cattle. ^aAccording to the reference sequence ARS-UCD1.2³¹. ^bAccording to the NCBI Annotation Release 106⁹¹.

subsequently performed similar analyses in the French Holstein population revealed 14 additional haplotype regions while confirming the three previously described haplotypes HH1, 2 and 3 (Table S1)^{16,17}. Therefore, performing new analyses regularly as genotyping data are accumulating was recommended¹⁷. In two similar screenings of the Nordic Holstein population another nineteen haplotype regions in addition to the previously reported HH3 region and the *FANCI*-related brachyspina-associated haplotype HH0, originally designated as HBY, were identified by Sahana et al. and Wu et al. (Table S1)^{18,19}. In the Danish Holstein population a haplotype associated with a large genomic deletion representing a loss-of-function of the *FANCI* gene causing brachyspina²⁰, as well as a haplotype associated with a missense variant in the *SLC35A3* gene causing complex vertebral malformation²¹ were identified (Table 1, Table S1). The latter two variants were initially thought to cause congenital malformation, but it was shown that in homozygous state, both mutations typically cause death in utero. Therefore, the two associated haplotypes (HH0 and HHC) were included in the list of 43 known Holstein haplotypes with reduced homozygosity (Table S1).

So far, candidate causal variants for seven of these Holstein haplotypes were identified either by whole-genome or whole-exome sequencing affecting different genes of importance for reproduction and/or development: *APAF1* (HH1), *IFT80* (HH2), *SMC2* (HH3), *GART* (HH4), *TFB1M* (HH5), *SDE2* (HH6) and *CENPU* (HH7) (Table 1)^{16,17,22–28}. In addition, a haplotype associated with increased juvenile mortality due to cholesterol deficiency (CDH) in the German Holstein population²⁹ had been detected to be associated with a transposable element insertion in the *APOB* gene that causes the metabolic disorder (Table 1, Table S1)^{26,29,30}.

Taken together, these findings indicate, that even though Holstein is an international breed characterized by a limited number of founding animals, the national subpopulations carry a different genetic load due to local selection decisions. Therefore, the aim of this study was to explore the accumulated massive genomic and phenotypic data of the local Swiss Holstein cattle population to identify genomic regions impairing fertility and rearing success. We applied a combination of methods including screening for missing homozygosity based on SNP-genotyped trios, trait association studies using estimated breeding values for female fertility traits and linkage disequilibrium analyses exploiting both SNP and whole-genome sequencing (WGS) data. The ultimate

Trait group	Trait sub-group	Trait	Description	Abbreviation
Fertility traits	Fertility	Non-return rate heifer	Heifers non-return rate after 56 days, binary	NRh
		Non-return rate cow	Cows non-return rate after 56 days, binary	NRc
		Interval first to last insemination heifer	Interval between first and last insemination for heifer, days	IFLh
		Interval first to last insemination cow	Interval between first and last insemination for cows, days	IFLc
		Interval calving to insemination	Interval from calving to first service, days	DFS
Birth traits	Birth history direct	Percentage normal births	Calving ease, scored between 1-without help to 5-dystocia	CEd
		Percentage live births	Percentage of calves born alive	SBd
		Birth weight	Weight of calve at birth, kg	BWd
		Gestation length	Days from successful insemination to birth	GLd
		Multiple birth	Percentage of multiple births	MBd
	Birth history maternal	Percentage normal births	Calving ease, scored between 1-without help to 5-dystocia	CEm
		Percentage live births	Percentage of calves born alive	SBm
		Birth weight	Weight of calve at birth, kg	BWm
		Gestation length	Days from successful insemination to birth	GLm
		Multiple birth	Percentage of multiple births	MBm
Survival traits	Rearing success	Survival period 1	Survival from day 3 up to 30th day of life	P1
		Survival heifer period 2	Survival of heifers from day 31 up to 458 days	P2h
		Survival bull period 2	Survival of young bulls from 31 up to 183 days	P2b

Table 2. Traits of interest.

goal was to pinpoint potentially causative protein-coding variants explaining reproductive failure due to recessive Mendelian disorders.

Materials and methods

The SNP array data of the Holstein population used was provided by the breeding association's swissherdbook and Holstein Switzerland. The quality-controlled (MAF > 0.01 and call rates > 0.9 per SNP and > 0.8 per animal) SNP data included 114 890 SNP combining a variety of SNP arrays with densities ranging from 3 to 150 k. SNP positions have been updated to the latest cattle reference sequence ARS-UCD1.2^{31,32}. The software Fimpute v2.2³³ was used to impute the data in order to correct for false genotypes as well as to increase the SNP data of lower density genotyped animals. The comprehensive data included 52,961 Holstein cattle. From these data, two analysis data sets were formed: the trio-based analysis, where the complete trio was genotyped (sire, dam and offspring), with 17,915 animals and the pgg-based analysis, where the dam is replaced by the maternal grandsire (sire, maternal grandsire and offspring), with 30,315 animals. While the trio-based dataset allows tracking the direct inheritance of alleles, the pgg-based dataset includes some uncertainty. Nevertheless, the pgg-based approach, where all male animals are genotyped, represents more accurately the common genotyping scheme of current breeding programs, as can be seen from the higher number of animals in the dataset. Both datasets include animals born in Switzerland after 2009 and their ancestors' genotypes.

The described data sets (trio and pgg) were used in the software snp1101³⁴ to identify haplotypes with a significant deviation from HWE, including a correction of false discovery rate according to Benjamini-Yekutieli³⁵. The chosen window size for the sliding-window approach was 50 SNP, based on the work of Hoff et al.³⁶. The snp1101 output provides a list with all significant haplotypes, the number of observed and expected homozygous carriers, as well as the allele frequency and the haplotype itself. For all regions of interest, where the observed number of homozygous animals were lower than the number of expected homozygous animals, we have chosen the most significant haplotype (lowest p-value) to reduce the number of false positive carriers. Further, we predicted the diplotypes for these haplotypes in the entire genotyped Holstein population and selected carrier animals for WGS.

Association studies were conducted using the predicted diplotypes and all female reproduction traits from the Swiss routine genetic evaluation system. In total 18 traits were analysed, while traits can be grouped into fertility, birth and survival traits (Table 2). The estimated breeding values (EBV) were deregressed and used as pseudo-phenotypes in linear mixed models (LMM) in the GCTA software^{37,38}. The genomic relationship matrix estimated using the method GRM of the GCTA software was included in the model to account for population structure. In addition, for the SNP data described above genome-wide association studies (GWAS) together with

deregressed EBVs as pseudo-phenotypes were performed. These single SNP regressions were implemented using the software `snp1101`³⁴, the genomic relationship matrix was calculated with the method proposed by VanRaden³⁹ and a Bonferroni-correction was applied ($p < 0.000044$).

Whole-genome sequencing (WGS) was performed in an attempt to sequence animals potentially carrying variants causing reduced homozygosity in the identified genomic regions. Therefore, 37 Holstein animals were selected based on their diplotype status for WGS. Additionally, 656 genomes from other projects were available, leading to 691 publicly available genomes including 244 purebred Holstein cattle (Table S2). Sequence data preparations was done according to previously described methods⁴⁰, with the only difference in the recalibration step, where the known variants from the 1000 Bull Genomes Project run 7 (BQSR file version 2) were used⁴¹. This resulted in a variant call format (VCF) file encompassing all single-nucleotide variants (SNV) and short indels from the 691 animals. The average read depth per genome were calculated using `covstat` from `goleft` v0.1.19⁴² and varied from $3.3\times$ to $72.8\times$, with an average of $18.7\times$ read depth (Table S1).

In order to design a custom SNP array, we selected candidate SNVs and short indels within the genomic regions of interest defined by the determined haplotype regions including plus/minus 2 Mb flanking sequence up- and downstream. Due to the variation in read depth, missing values and a single homozygous carrier animal were allowed, GATK recommended quality measures for hard filtering needed to be passed and at least a single Holstein animal needed to be carrier of a variant, but no more than 75% of all animals for it to be selected for the array design. Variants were included in a custom array using Axiom Microarray Genotyping Technology, designed under the umbrella of Swiss routine genomic systems. Due to the process of the design almost every second (44%) of the initially selected 465,768 variants could be designed and led to an array including 205,362 novel variants and 112,854 so called “routine variants” from previously applied arrays including common SNP markers considered for genomic selection. Subsequently, this custom array, called SWISScow array, was used routinely in the Swiss genomic breeding program. Therefore, comprehensive genotypes of 13,667 Swiss cattle are available, of which 5603 are Holstein.

Linkage disequilibrium (LD) (r^2) were estimated between diplotypes for haplotypes and both, custom array genotypes (a) and WGS-driven genotypes (b), in a merged dataset encompassing 5 603 (a) and 37 (b) animals, respectively. These analyses were conducted using `plink` v1.9⁴³.

For visualisation of the results from the different genome-wide analyses the R package `OmicCircos` was used⁴⁴. Figure 1 summarizes the results from the haplotype identification using the trio and `pgp` approach, indicates if there are linked marker on the custom array, if the haplotypes show associations to current traits of the breeding program, and if there are significant GWAS associations in the Swiss Holstein population for the trait groups fertility, birth and survival.

In order to improve the interpretation of candidate variants conservation scores `PhyloP` and `PhastCons` from the UCSC database were taken into account^{45,46}. Therefore, the tool `LiftOver` from UCSC tools was used to lift the bovine variants from the reference sequence ARS-UCD1.2³¹ to the human genome 38⁴⁷. For these positions, the conservation scores of 99 vertebrates could be extracted directly. Furthermore, for a more comprehensive prediction of the impact of protein-changing variants the tool `PROVEAN`⁴⁸ was applied. Lastly, the data provided by the 1000 Bull Genomes project run 8 was analysed for the distribution of candidate variants, as it is an international control cohort with a variety of breeds⁴¹.

Results

Detection of novel and previously described haplotypes. All identified haplotypes are described in Table 3, their statistical evidence is summarized in Table S3 and the two outer circles in Fig. 1 demonstrate their distribution across the 29 bovine autosomes. The search for haplotypes deviating from HWE led to the detection of 24 haplotype regions, of which five had been previously described (Table 3, Table S3). These identified haplotypes were named in accordance to 17 previously described Holstein haplotypes (Table S1) and designated as subsequent haplotypes (HH18–HH38) (Table 3, Table S3). The obtained results from the trio-based approach led to the detection of 10 haplotypes, of which 6 were never observed in homozygous state, although at least 10 were expected (Table 3). On the other hand, the `pgp`-based approach revealed 21 haplotypes, of which 4 never occurred in homozygous state. Combining the outcome of both approaches revealed 7 commonly detected haplotype regions. The most significant deficient homozygous haplotype is the previously described haplotype HH21 for which 157 homozygous carriers were expected, but none were observed (Table 3, Table S3)¹⁸. The identified haplotypes have an average length of 1.12 Mb and range from 0.24 up to 2.56 Mb and segregate at an average allele frequency of 0.024 ranging from 0.014 to 5.46 (Table S3).

Association studies indicate phenotypic relevance of deficient homozygous haplotypes. Significant associations ($p < 0.05$) were detected for 21 haplotypes distributed across 17 chromosomes (Fig. 1, Table 3). Nevertheless, three of the detected deficient homozygous haplotypes show no significant associations to any of the 18 considered traits (Table S4). Most of the deficient homozygous haplotypes are associated with various fertility- and birth-related traits, whereas only two detected haplotypes show significant association to two traits of calf survival (Table 3). For example, HH37 shows significant associations to totally six different analysed fertility- and birth-related traits (Table 3). Detailed results from the association studies including all 24 detected deficient homozygous haplotypes and 18 traits of interest are summarised in Table S4. The by far strongest association was detected for the HH5 haplotype to the trait NRh (heifers non-return rate after 56 days) ($p = 0.00006$), which is in accordance with previously reported findings that a gross genomic deletion on chromosome 9 encompassing the entire *TFB1M* gene causes embryonic lethality²⁶. Another example of a deficient homozygous haplotype showing a negative association to non-return rate in heifers is HH24 ($p = 0.02$), beside further significant associations of that haplotype to three birth-related traits (Table 3). The fertility-related trait

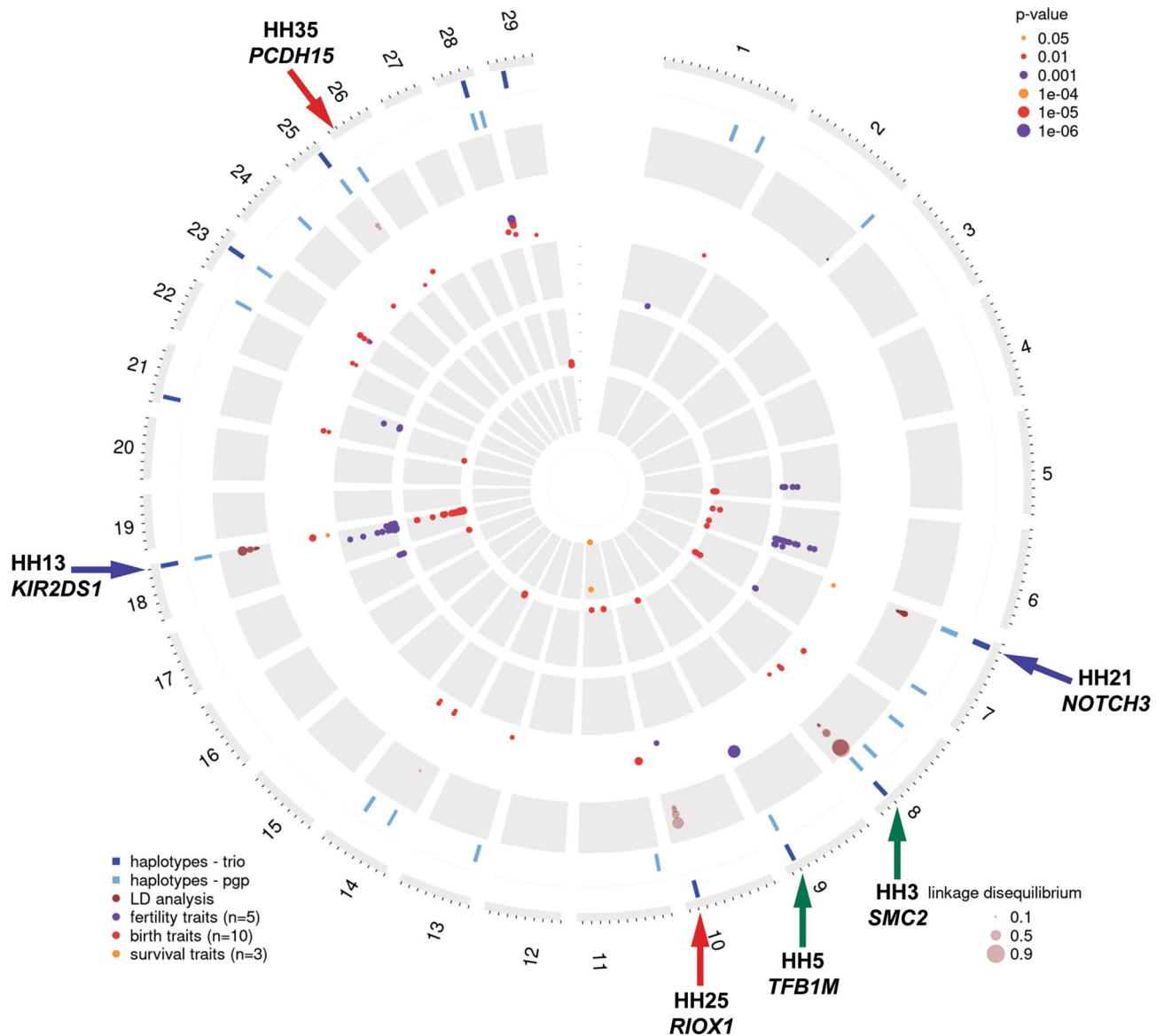


Figure 1. Summary of the SNP and WGS data analyses of the Swiss Holstein population. This plot visualises all the comprehensive genomic analysis. The most outer circles show the haplotypes per chromosome with reduced homozygosity for the trio approach in dark blue and the pgp approach in light blue. LD (r^2) between haplotypes and markers of the custom SNP array is indicated in the circle with the brown dots. Note the dot size correlates with the LD extent. The fourth circle shows significant results from the haplotype to phenotype association analyses. Note that the three evaluated trait groups are represented by different colours and the dot size correlates with the significance values. The three inner circles present the significant ($p < 0.0000043$) GWAS results across the fertility (purple), birth (red) and survival (yellow) traits. Scales are based on the $-\log_{10}(p\text{-value})$. Note the green arrows indicating previously identified haplotypes and candidate causal variants. The blue arrows indicate the previously identified haplotypes HH21 and HH13^{13,16,18} and the herein described candidate causative variants in the genes *NOTCH3* and *KIR2DS1*, respectively. Finally, the red arrows indicating the newly identified haplotypes HH25 and HH35 harbouring most likely causative variants in the genes *RIOX1* and *PCDH15*.

interval DFS (calving to first insemination) is negatively associated with the HH25 haplotype and the trait IFLc (interval fist to last insemination) with the HH32 haplotype whereas HH32 and HH30 show a reducing effect on the trait BWm (birth weight) (Table S4). The herein described deficient homozygous haplotype HH21 most likely corresponds to the previously described haplotypes 175.5 (VanRaden et al.¹³) and 07-126¹⁸ and shows a negative association to the survival-related traits P2b (survival from day 3 up to 30 days) (Table 3). The also previously published deficient homozygous haplotype HH3 shows significant association to the birth-related trait GLm (gestation length) (Table 3), which can be explained with the already known *SMC2*-related abortion of homozygotes²⁴. Finally, for the further two haplotypes with putative known causal variants (see below) HH35

Name ^a	chr	Pos in Mb ^b	Analysis	Homozygous haplotype carrier			Allele frequency %	Known and novel associated gene ^c	Associated traits		
				observed	expected	deficient			Fertility	Birth	Survival
HH18	1	10.302–104.376	pgp	7	28	75%	2.39				
HH19 ^d	1	139.874–140.643	pgp	12	47	74%	3.00		GLd		
HH20	2	134.477–135.148	pgp	1	14	93%	1.66				
HH21 ^e	7	7.872–10.433	trio/pgp	0	157	100%	5.46	NOTCH3			P2b
HH22	7	93.582–94.642	pgp	2	16	88%	1.73				
HH23	8	15.562–16.863	pgp	5	25	80%	2.20			MBm	
HH24	8	66.699–68.006	pgp	2	59	97%	3.36		NRh	BWm, CEm, MBm	
HH3 ^f	8	90.959–92.085	trio/pgp	0	10	100%	1.38	SMC2	GLm		
HH5 ^e	9	90.928–91.824	trio/pgp	0	30	100%	2.39	TFB1M	NRh		
HH25	10	86.876–87.773	trio	5	20	75%	1.96	RIOX1	DFS		
HH26	11	4.531–5.362	pgp	7	34	79%	2.56		GLd		
HH27	13	7.083–8.297	pgp	2	17	88%	1.82		GLm		
HH28	14	24.591–24.827	pgp	3	20	85%	1.98			SBd, SBm	
HH29	14	58.128–59.238	pgp	14	52	73%	3.14			SBm, CEm	
HH13 ^b	18	60.932–62.101	trio	1	17	94%	1.81	KIR2DS1		BWd	P1
		62.070–63.045	pgp	2	22	91%	2.06				
HH30	21	7.844–8.671	trio	1	24	96%	2.13			BWd, CE _d , CE _m	
HH31	22	59.811–60.704	pgp	3	17	82%	1.80		GLd, GLm		
HH32	23	32.773–33.726	pgp	1	13	92%	1.57			BWd, CEm	
		33.545–34.948	trio	0	13	100%	1.55		GLd, IFL _c	CE _m	
HH33	24	44.693–45.678	pgp	14	42	67%	2.82			SBd	
HH34	25	36.162–37.330	trio/pgp	0	18	100%	1.86		GLm		
HH35	26	3.359–4.235	pgp	7	31	77%	2.44	PCDH15		SBm	
HH36	28	28.541–29.736	pgp	2	15	87%	1.71			CE _m	
		29.731–30.877	trio	2	26	92%	2.25				
HH37	28	40.348–41.271	pgp	6	66	91%	3.54		GLd, GLm, IFL _h	CE _m , SBd, MBm	
HH38	29	14.243–15.317	trio	0	14	1.00	1.66			CE _m	

Table 3. Haplotypes indicating reduced homozygosity in Swiss Holstein. *BWd* weight of calve at birth in kg, direct trait, *BWm* weight of calve at birth in kg, maternal trait, *CEd* direct calving ease, scored between 1–without help to 5–dystocia, *CEm* maternal calving ease, scored between 1–without help to 5–dystocia, *DFS* interval from calving to first service in days, *GLd* days from successful insemination to birth, direct trait, *GLm* days from successful insemination to birth, maternal trait, *IFLc* interval between first and last insemination for cows in days, *IFLh* interval between first and last insemination for heifer in days, *MBm* percentage of multiple births, *NRh* heifers non–return rate after 56 days, binary, *P1* survival from day 3 up to 30th day of life, *P2b* survival of young bulls from 31 up to 183 days, *SBd* percentage of calves born alive, direct trait, *SBm* percentage of calves born alive, maternal trait. ^aHH meaning Holstein haplotype. ^bAccording to the reference sequence ARS-UCD1.2³¹. ^cAccording to NCBI Annotation Release 106⁹¹. ^dHaplotype co-localises with previously described haplotype HHB by Shuster et al. and Cole et al.^{51,85}. ^eHaplotype described before as 175.5 and 07–126 by VanRaden et al. and Sahana et al.^{13,18}. ^fHaplotype previously described by VanRaden et al., McClure et al., Sahana et al. and Wu et al.^{13,18,19,87}. ^gHaplotype previously described by Schütz et al. and Fritz et al.^{26,27}. ^hHaplotype previously described by Fritz et al.¹⁶.

shows a negative effect for the birth-related trait SBm (live births) while HH13 is associated to BWm (birth weight) as well as the survival related trait P1 (survival from day 3 up to 30th day of life) (Table 3).

In addition, the GWAS results are visualised using Manhattan plots in the supplementary material (Supplementary Figs. S1–S18). The genome-wide significant ($p < 0.0000043$) results are collected in Table S5 and in summary displayed in the three inner circles of Fig. 1. Among the five studied fertility-related traits we identified clear associations to a locus on chromosome 18 for the three traits DFS (calving to insemination), NRh (non–return rate for heifer) and IFLc (interval between first and last insemination for cows), that co-localize with the region of the herein detected deficient homozygous HH13 (Table S5). In addition, at the same locus interesting associations to five different birth-related traits were observed (Table S5). At the genome region of HH28 on chromosome 14, a GWAS hit to the two birth-related traits direct traits BWd (birth weight) and CE_d (percentage normal births) was observed (Table S5). No co-localization between GWAS signals of the three studied survival traits and the herein detected deficient homozygous haplotypes could be identified. Nonetheless, a GWAS hit on chromosome 11 for the calf survival-related trait P2h (survival of heifers from 31 up to 458 days) was observed at the identical genome region on chromosome 11 of the previously described CDH haplotype containing a deleterious *APOB* loss-of-function variant causing rearing loss (Table S1)²⁹.

Haplotype	Gene	OMIM	Variant description				
			Genomic position ^a	Description	Transcript ^b	Coding DNA change	Protein change
HH21 ^c	<i>NOTCH3</i>	600276	chr7:7913459	3 bp deletion (inframe deletion)	XM_003586246.3	c.129_131delTTG	p.Cys44del
HH3 ^d	<i>SMC2</i> ^e	605576	chr8:93753358	SNV (missense)	XM_015472668.2	c.3404T>C	p.Phe1135Ser
HH25	<i>RIOX1</i>	611919	chr10:84938370	30 bp deletion (inframe deletion)	NM_001099702.1	c.396_425delGGCGCAGACCCC GGCGGCACGCTTGGTGA	p.Ala133_Glu142del
HH13 ^f	<i>KIR2DS1</i>	604952	chr18:62758881	SNV (nonsense)	NM_001097567.1	c.475C>T	p.Gln159*
HH35	<i>PCDH15</i>	605514	chr26:5325675	SNV (missense)	XM_015460562.2	c.2599C>G	p.Leu867Val

Table 4. Candidate variants potentially explaining the identified haplotype regions. ^aAccording to the reference sequence ARS-UCD1.2³¹. ^bAccording to NCBI Annotation Release 106⁹¹. ^cHaplotype described before as 175.5 and 07-126 by VanRaden et al. and Sahana et al., respectively^{13,18}. ^dHaplotype previously described by VanRaden et al., McClure et al., Sahana et al. and Wu et al.^{13,18,19,87}. ^eVariant previously detected by McClure et al.²⁴. ^fHaplotype previously described by Fritz et al.¹⁶.

Identification of four novel potentially causal variants. The generated comprehensive WGS data was mined in genome regions containing deficient homozygous haplotype for variants potentially causing pre- or post-natal lethality or sublethal conditions. By filtering for variants indicating an obvious depletion in homozygosity and showing a strong LD with the haplotype, we successfully uncovered five protein-changing DNA variants as candidates (Table 4). This short list of possibly causal variants includes the previously described missense mutation in the *SMC2* gene associated with HH3-related embryonic lethality²⁴ (Table S6). In addition, we detected four novel variants linked to different deficient homozygous haplotypes affecting these four genes: *NOTCH3* (HH21), *RIOX1* (HH25), *KIR2DS1* (HH13), and *PCDH15* (HH35) (Table 4). Unfortunately, only two of these five variants, those affecting *SMC2* and *KIR2DS1*, could be successfully added to the customized SWISS-cow array and thereby genotyped in several thousand animals of the current Swiss dairy population. In accordance to the results of the haplotype analysis for HH3 and HH13, no single homozygous carrier of the *SMC2* and *KIR2DS1* variants was observed neither in the current population of more than 14 thousand Swiss dairy cattle nor in any other breed of cattle included in the 1000 Bull Genomes project (Table S6).

As reported earlier by Ref.²⁴ the *SMC2* missense variant p.Phe1135Ser affects an evolutionary conserved residue of the C-terminal P-loop-nucleoside triphosphate hydrolase domain classified as a deleterious change according to in silico predictions. In the studied Swiss Holstein population this variant occurs a low frequency (1.3%) in high LD ($r^2 = 0.85$) with the HH3 haplotype on chromosome 8 at ~94 Mb (Table S6). The *SMC2* variant was absent from all other studied breeds (Table S6).

The detected stop-gain variant in the *killer cell immunoglobulin like receptor, two Ig domains and short cytoplasmic tail 1* (*KIR2DS1*) gene showing missing homozygosity is located on chromosome 18 at ~62 Mb in LD ($r^2 = 0.49$) with the HH13 haplotype. The detected *KIR2DS1* variant located in exon 4 leads to the premature termination of the protein sequence at codon 159 (p.Gln159*) truncating the protein by almost two third including functionally important domains (Fig. 2A). It segregates within the Swiss Holstein population at an allele frequency of 5% and shows a highly significant deviation from HWE ($p = 0.000007$) based on a Chi-square test (Table S6). Interestingly the *KIR2DS1* variant obviously occurs in various other breeds beside Holstein cattle (Table S6).

Further, we propose three coding variants in *NOTCH3*, *RIOX1*, and *PCDH15* as candidate causal variants that were so far not evaluated in the broader population due to limitations in the design process of the customized SWISS-cow array.

Firstly, we identified a disruptive in-frame deletion in the *notch receptor 3* (*NOTCH3*) gene moderately linked ($r^2 = 0.49$) with the completely deficient homozygous haplotype HH21 on chromosome 7 at ~9 Mb (Table 4). This protein-changing variant knocks-out a highly conserved cysteine residue (p.Cys44del) of an epidermal growth factor-like repeats (EGFRs) that comprise the extracellular domain of the NOTCH3 receptor. The deletion results in a loss of a cysteine residue in the first of these numerous EGFRs (Fig. 2B) and is predicted to be deleterious (Table S6). This variant occurs at a low allele frequency of 0.014 and was never observed in homozygous state in any of the more than 4000 bovine genomes with available WGS data (Table S6). The studied *NOTCH3* variant appears to be very rare, and among the 4109 animals of the 1000 Bull Genomes project, it was found in only three animals of different breeds (Table S6).

Secondly, another disruptive in-frame deletion encompassing 30 nucleotides of the open reading frame of the *ribosomal oxygenase 1* (*RIOX1*) gene was detected that occurs perfectly linked ($r^2 = 1$) with the HH25 haplotype on chromosome 10 at ~87 Mb (Table S6). Interestingly, within our complete WGS data a single homozygous animal was detected, whereas 16 homozygotes were present in the 1000 Bull Genomes variant catalogue. By visual inspection of the mapping quality of the reads around this variant, we noticed that the mapping algorithm (Supplementary Fig. S19) did not splice many of the reads covering the 30 bp deletion properly. Therefore, we assume that these homozygotes most likely represent false positives and are actually only heterozygous carriers of the *RIOX1* variant. Obviously, the *RIOX1* deletion occurs in many breeds other than Holstein (Table S6). The disruptive in-frame deletion affects the N-terminal region of the RIOX1 protein (p.Ala133_Glu142del) (Fig. 2C).

Lastly, we identified a missense variant in the *protocadherin related 15* (*PCDH15*) gene that is perfectly linked ($r^2 = 1$) with the HH35 haplotype on chromosome 26 at ~4 Mb (Table S6). Interestingly, for this variant, as well as for the HH35 haplotype we observed some homozygous carriers (Table 3, Table S6). In the 1000 Bull Genomes

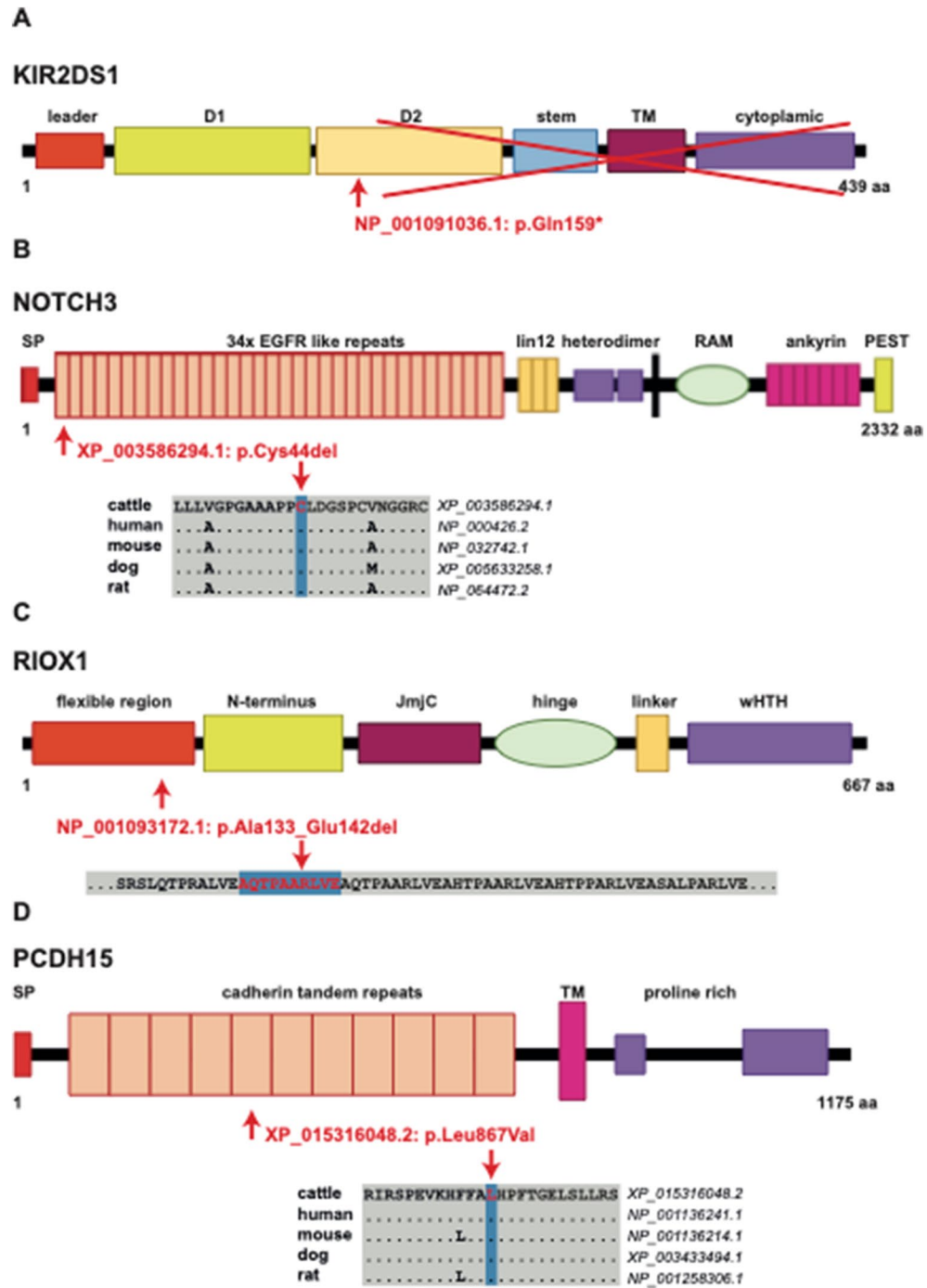


Figure 2. Features of the four novel candidate causal variants. Note the predicted protein changes of the detected variants indicated in red. ((A) *KIR2DS1*:p.Gln159^{*82}, (B) *NOTCH3*:p.Cys44del⁷⁶, (C) *RIOX1*:p.Ala133_Glu142del⁸³, (D) *PCDH15*:p.Leu867Val⁸⁴).

Project variant catalogue all identified variant carriers were indicated as Holstein, Ayrshire and Norwegian Red (Table S6). Nevertheless, the amino acid exchange (p.Leu867Val) affects a highly conserved residue of a cadherin tandem repeat domain (Fig. 2D) although the computer predicted consequence was inconclusive (Table S6).

Discussion

After giving an overview of the known haplotypes with missing homozygosity in the worldwide Holstein breed, we have conducted a comprehensive survey of genomic and phenotypic data of the Swiss Holstein cattle population to search for further deficient homozygous haplotypes. We focussed on the detection of hidden recessive variants and found five candidates that cause the observed missing homozygosity. These variants can lead to the exclusion from the population due to natural selection processes, e.g., embryonic lethality, abortions and

stillbirth, or due to artificial selection against animals that show illthrift or other non-lethal more subtle conditions, e.g., metabolic disorders that are often not diagnosed.

In previous studies performed in Holstein cattle a simple allele frequency-based approach was applied to identify haplotypes showing a depletion of homozygous animals by applying the assumption of random mating and a Chi-squared test^{13,18,49}. Other studies accounted for the mating status by included the haplotype state of the sire and the maternal grandsire^{16,50}. In comparison, the herein presented study uses two different approaches, a trio- and pgp-approach, that include genotyped trios and applies a Fisher exact test. This had the advantage that we were able to detect haplotypes that segregate at a lower allele frequency. The pgp-based approach includes almost double the data from the trio-based approach. This is explained by the genotyping management, in which male breeding animals are usually genotyped, but by far not all females. Therefore, we identified almost twice the number of haplotypes with the pgp-approach. Nevertheless, the applied trio-approach is assumed to be more powerful as the parental haplotype inheritance can be traced directly.

Introducing this novel approach, we were able to verify our outcomes with the results of earlier studies in different Holstein cattle populations. With the applied methods, we were able to detect four haplotype regions, namely HH3, HH5, HH13 and HH21 that were previously described to occur in several subpopulations of Holstein cattle^{13,16,18,26,51}. Another herein detected haplotype (HH19) co-localizes with the genome regions of the known deficient homozygous haplotype HHB on the proximal end of chromosome 1. However, it obviously differs to the BLAD-associated HHB haplotype as we confirmed by using the customized SWISScow array that the causative *ITGB2* variant does not segregate anymore in the Swiss Holstein population (results not shown) probably due to the strict exclusion of carriers in the early 1990s.

Furthermore, by applying the WGS data screen, we confirmed the known Holstein breed-specific HH3-linked missense variant in *SMC2* on chromosome 8 that also in our data showed complete missing homozygosity²⁴. Interestingly, for this deleterious variant we could only detect a single significant phenotypic effect in our additive models for gestation length based on the haplotype, but a suggestive negative effect on non-return rate in heifer. Negative effects on the number of inseminations in heifers within that genomic region were shown in Canadian Holstein cattle⁵². Nevertheless, these findings support a purely recessive embryonic lethal effect of the *SMC2* variant that allows merely the indirect detection in non-return rate²⁴. For HH5, we were unfortunately unable to confirm the association with the most likely causative gross deletion encompassing the *TFB1M* gene as we focused in the scope of this project on SNV and short indel genotyping.

Interestingly, we found a co-localization of a GWAS peak for survival during the first 458 days of life in the region of the previously reported CDH haplotype²⁹. The CDH haplotype was shown to be linked with a pathogenic endogenous retroviral insertion into the *APOB* gene and thereby leading to a subvital malabsorption of cholesterol, provoking increased juvenile mortality^{26,29,30,53}. This haplotype was difficult to detect due to it segregating in homozygous state in the adult population, as the LD between the CDH haplotype and the *APOB* variant was not perfect as the ancestral version of the haplotype still occurred^{29,30}. Initially, cholesterol deficiency was thought to be a recessive or codominant inherited disorder, what led to the identification of the responsible locus by applying a case-control based GWAS²⁹. Further studies examining heterozygous carriers showed that the *APOB*-associated disorder is not a simple Mendelian disease, as heterozygous animals can also show severe clinical signs of cholesterol deficiency⁵⁴. The identified GWAS peak in the current study underlines these findings with an additive model. For a monogenic recessive lethal disorder, one would not expect to see any effect within an additive model, as all the homozygous affected animals would be more or less excluded from the genotyped population and it can be assumed that these heterozygous animals are phenotypically normal like the homozygotes. It is in line with recent findings in cattle that additive models are shown to have their limitations and non-additive models should be taken into account for further phenotypic evaluations of recessive disorders⁵⁵.

We propose to add four novel candidate causative variants to the list of Holstein haplotypes with known cause for missing homozygosity. We emphasize to highlight two variants in *KIR2DS1* and *NOTCH3* linked to the abovementioned confirmed deficient homozygous haplotypes HH13 and HH21, respectively. Both detected variants segregate in Holstein cattle but, in contrast to the HH3-linked *SMC2* variant, represent derived alleles that predates establishing of modern breed as they rarely occur also in unrelated breeds of cattle.

Interestingly, in human *KIR2DS1* is reported to be associated with reproduction, namely placentation success^{56,57}. *KIR2DS1* is an activator receptor, belongs to the KIR family of natural killer cell Ig-like receptors and plays a vital role in the immune system (OMIM #604952). Regarding the involvement in disorders, *KIR2DS1* is known to be associated in human with an autoimmune disease called psoriasis vulgaris^{58,59} and leukaemia⁶⁰⁻⁶². In later studies reproductive failure and foetal growth restriction were associated with maternal *KIR2DS1* and foetal leukocyte antigen C2 (HLA-C2) imbalance, however, being complicated by the interference of the maternal and the embryonic genotype^{56,63-66}. Nonetheless, the function of *KIR2DS1* was described as protective against pregnancy loss^{66,67}. Moreover a study showed the importance of *KIR2DS1*/HLA-C2 in response to viral placental infections⁶⁸. The herein detected bovine variant most likely represents a true loss-of-function mutation of *KIR2DS1*, as possibly nonsense-mediated decay selectively recognizes and degrades mRNAs whose open reading frame is truncated by a premature translation termination codon. Therefore, this variant is supposed to lead to pregnancy losses of homozygous embryos. Furthermore, as the function of *KIR2DS1* is sensitive to its expression⁶⁸, we speculate that also a negative effect in heterozygous carriers might exist. This hypothesis is supported by GWAS hits in the genome region of HH13 for several fertility-related traits (interval calving to insemination, non-return rate heifer and interval first and last insemination), as well as birth-related traits (percentage normal births, percentage live births, birth weight as well as gestation length for maternal and direct traits). Interestingly, major QTLs for fertility and birth traits were mapped at 60 Mb on chromosome 18 in the German Holstein population⁶⁹.

The *NOTCH3* gene belongs to the mammalian Notch family that includes four proteins, which are transmembrane developmental signalling receptors⁷⁰⁻⁷². The Notch signalling pathway has important roles in the

development of organs, regulation of cell death, cell survival and cell abundance^{70–72}. *NOTCH3* is known to be associated with fatal dominant inherited disorders such as myofibromatosis, cerebral arteriopathy with sub-cortical infarcts and leukoencephalopathy (CADASIL) and lateral meningocele syndrome (OMIM # 600276), as well as different forms of cancer⁷³. In *Notch3*-deficient mice, it was initially shown that *Notch3* seems to be non-essential for proper embryonic development and fertility⁷⁴, however, post-natal pathology in *Notch3*-null mice revealed structural defects in the smooth muscle cells, arterial differentiation and the arterial morphology⁷⁵. The *NOTCH3* variant linked to the HH21 haplotype is predicted to delete a highly conserved cysteine residue in the first of 34 EGFR modules that are characterised by six highly conserved cysteine's that are essential for the stability of each EGFR-module fold by their three disulphide bonds⁷⁶. EGFR1–3 had been shown to be critical for Notch pathway activation⁷⁷. Interestingly, the most significantly associated trait with HH21 was survival of young bulls from 31 up to 183 days, which indicates a possibly subvital condition caused by the *NOTCH3* variant. So far, the proposed variant was never observed in homozygous state in any of more than 4000 cattle genomes, while a population evaluation in Swiss Holstein is still pending to prove if there are indeed 100% no homozygous carriers as to be assumed from the haplotype analysis.

Lastly, we would like to propose two candidate causal variants in *RIOX1*, and *PCDH15*, that show perfect LD to the deficient homozygous haplotypes HH25 and HH35, respectively. *RIOX1* is an oxygenase that can act as both a histone lysine demethylase and a ribosomal histidine hydroxylase and plays a central role in histone code, chromatin organisation and DNA transcription regulation (OMIM #611919). Due to the basic importance of these processes for DNA and cell function, we speculate that a disruption of the protein by a motif of 10 amino acids might lead to disturbance of the meiosis or to the inactivation of cell differentiation in very early stages of the embryogenesis. This hypothesis would be supported by the haplotype association for HH25 that indicates a prolonged interval calving to first insemination (DFS). Here we also would like to address the fact that such small deletions affecting repetitive DNA motifs clearly show up the limitations of the applied short-read methods for whole-genome sequencing and subsequent indel calling (Supplementary Fig. S19). Nevertheless, because the variant *RIOX1* allele obviously occurs in a variety of breeds, this variant might represent a quite old mutation that arose already before modern breed formation that could also be of less impact. If the proposed variant is indeed reduced in number of homozygous carriers needs to be evaluated by further genotyping.

Finally, we propose a missense variant in *PCDH15* as strong causal candidate for the depletion in homozygosity of HH35. The occurrence of this bovine variant is limited to Holstein and animals of populations with documented introgression of Holstein cattle. It affects an evolutionary conserved residue of a functionally important domain and the affected gene is a member of the cadherin superfamily⁷⁸. It encodes an integral membrane protein that is known to be vital for the maintenance of normal retinal and cochlear function (OMIM #605514). In men, the affected gene *PCDH15* is associated with recessively inherited forms of deafness and neurosensory Usher syndrome⁷⁹. The HH35 haplotype is not completely deficient homozygous indicating no embryonic lethal effect of the associated *PCDH15* variant. Therefore, we speculate that possibly a subtle disorder leads to the social exclusion of young individuals, probably due to their abnormal behaviour as they have a different perception of the environment than unaffected animals. Alternatively, we hypothesize that possibly not all biological functions of *PCDH15* are known and more severe conditions could arise leading to pre- or neonatal lethality and thereby explaining the lack of homozygotes. Nevertheless, we suggest also for this newly described variant in *PCDH15* to follow up by extended genotyping and propose a detailed clinical examination of homozygous living animals to evaluate their individual health status.

At this point we would like to lay out some limitations of the presented study. The imputation step, which is based on pedigree information and allele frequencies, can affect the detection of decreased homozygosity. Therefore, imputation has limited power for recently arisen variants due to the existence of ancestral and derived versions of identical haplotypes. This could potentially lead to the false negative detection of recent haplotype regions and the false positive detection of haplotype regions due to naturally over-represented alleles. Furthermore, the accuracy of imputation depends on the SNP density of the genotype array and number of genotypes available per array⁸⁰. During the design of the custom genotyping array, it was not possible to include all variants of interest. Mainly due to technical reasons, but also due to the more complicated types of variants (e.g. deletions). The manufacturer did not recommend those variants since in silico predicted genotyping quality was expected to be limited. Nevertheless, as shown for the HH3 and HH13 associated variants in *SMC2* and *KIR2DS1*, the custom array was a very efficient tool to monitor several thousands of animals. The displayed results illustrate that even though Holstein is an international breed local selection strategies affect the genomic structure of its subpopulations. From a genome-wide perspective, it might be possible that we have overlooked additional haplotype regions due to the restricted sliding-window length of 50 SNPs, considering that there is a correlation between window size and the detected homozygosity³⁶. In addition, the restriction for protein-changing SNVs and indels limits the identification for candidate causal variants. We have not considered non-coding regulatory variants or larger structural variants that have a pathogenic impact, which we do not know yet. An example for this is the HH5 associated large deletion including the gene *TFB1M*²⁶, for which we could confirm the haplotype region in our data. Furthermore, it was shown that graph based reference sequences can improve the annotation and could potentially lead to the detection of further protein-changing variants⁸¹. In order to improve the understanding of the described candidate variants, further studies are recommended to evaluate all homozygous haplotype and variant carriers. As the example of CDH in Holstein cattle has taught, the oversimplification of the depletion of homozygosity by assuming the cause to be a monogenetic recessive disorder might be misleading.

In conclusion, we demonstrate that with mining massive SNP data in combination with WGS data, we mapped 19 novel and 5 previously described haplotype regions and unravel novel recessive protein-changing variants in *NOTCH3*, *RIOX1*, *KIR2DS1* and *PCDH15* segregating at low to moderate frequencies. In addition, our findings confirm previously identified loci such as *SMC2* for HH3. The phenotypic associations support additive effects in the described haplotype regions and variants. Taken together this study expands the spectrum of undesired alleles

impairing reproduction success in Holstein cattle, the world's most important dairy breed, enabling improved DNA-based selection in order to reduce reproductive failure and animal loss.

Data availability

The whole-genome sequencing data is stored in the European Nucleotide Archive and can be found with the sample IDs available in Table S2. For access to the SNP genotyping data, we ask interested people to contact the authors or the breeding associations directly.

Received: 4 October 2021; Accepted: 7 March 2022

Published online: 31 March 2022

References

- Hansen, L. B. Consequences of selection for milk yield from a Geneticist's viewpoint. *J. Dairy Sci.* **83**, 1145–1150 (2000).
- Zenger, K. R., Khatkar, M. S., Cavanagh, J. A. L., Hawken, R. J. & Raadsma, H. W. Genome-wide genetic diversity of Holstein Friesian cattle reveals new insights into Australian and global population variability, including impact of selection. *Anim. Genet.* **38**, 7–14 (2007).
- Ma, L. *et al.* Genome changes due to artificial selection in U.S. Holstein cattle. *BMC Genomics* **20**, 1–14 (2019).
- Makanjuola, B. O. *et al.* Effect of genomic selection on rate of inbreeding and coancestry and effective population size of Holstein and Jersey cattle populations. *J. Dairy Sci.* **103**, 5183–5199 (2020).
- Pryce, J. E. *et al.* World trends in dairy cow fertility. In *World Congress of Genetics Applied to Livestock Production*, Vol. 10, 6 (2014).
- Walsh, S. W., Williams, E. J. & Evans, A. C. O. A review of the causes of poor fertility in high milk producing dairy cows. *Anim. Reprod. Sci.* **123**, 127–138 (2011).
- Seegers, H., Beaudeau, F., Fourichon, C. & Bareille, N. Reasons for culling in French Holstein cows. *Prev. Vet. Med.* **36**, 257–271 (1998).
- Rilanto, T. *et al.* Culling reasons and risk factors in Estonian dairy cows. *BMC Vet. Res.* **16**, 173 (2020).
- Mao, X. *et al.* Genome-wide association mapping for dominance effects in female fertility using real and simulated data from Danish Holstein cattle. *Sci. Rep.* **10**, 1–9 (2020).
- Cai, Z., Guldbbrandsen, B., Lund, M. S. & Sahana, G. Prioritizing candidate genes for fertility in dairy cows using gene-based analysis, functional annotation and differential gene expression. *BMC Genomics* **20**, 1–9 (2019).
- Jiang, J. *et al.* A large-scale genome-wide association study in U.S. Holstein cattle. *Front. Genet.* <https://doi.org/10.3389/fgene.2019.00412> (2019).
- Minozzi, G. *et al.* Genome wide analysis of fertility and production traits in Italian Holstein cattle. *PLoS ONE* **8**, 1–10 (2013).
- VanRaden, P. M., Olson, K. M., Null, D. J. & Hutchison, J. L. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *J. Dairy Sci.* **94**, 6153–6161 (2011).
- Puckowska, P., Borowska, A., Szwaczkowski, T., Oleński, K. & Kamiński, S. Effects of a novel missense polymorphism within the SIGLEC5 gene on fertility traits in Holstein–Friesian cattle. *Reprod. Domest. Anim.* **54**, 1163–1168 (2019).
- Taylor, J., Schnabel, R. & Sutovsky, P. Review: Genomics of bull fertility. *Animal* **12**, s172–s183 (2018).
- Fritz, S. *et al.* Detection of haplotypes associated with prenatal death in dairy cattle and identification of deleterious mutations in GART, SHBG and SLC37A2. *PLoS ONE* **8**, e65550 (2013).
- Hozé, C. *et al.* A splice site mutation in CENPU is associated with recessive embryonic lethality in Holstein cattle. *J. Dairy Sci.* **103**, 607–612 (2020).
- Sahana, G., Nielsen, U. S., Aamand, G. P., Lund, M. S. & Guldbbrandsen, B. Novel harmful recessive haplotypes identified for fertility traits in Nordic Holstein cattle. *PLoS ONE* **8**, 8–12 (2013).
- Wu, X., Mesbah-Uddin, M., Guldbbrandsen, B., Lund, M. S. & Sahana, G. Haplotypes responsible for early embryonic lethality detected in Nordic Holsteins. *J. Dairy Sci.* **102**, 11116–11123 (2019).
- Charlier, C. *et al.* A deletion in the bovine FANCI gene compromises fertility by causing fetal death and brachyspina. *PLoS ONE* **7**, e43085 (2012).
- Thomsen, B. *et al.* A missense mutation in the bovine SLC35A3 gene, encoding a UDP-N-acetylglucosamine transporter, causes complex vertebral malformation. *Genome Res.* **16**, 97–105 (2006).
- Adams, H. A. *et al.* Identification of a nonsense mutation in APAF1 that is likely causal for a decrease in reproductive efficiency in Holstein dairy cattle. *J. Dairy Sci.* **99**, 6693–6701 (2016).
- Daetwyler, H. D. *et al.* Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.* **46**, 858–865 (2014).
- McClure, M. C. *et al.* Bovine exome sequence analysis and targeted SNP genotyping of recessive fertility defects BH1, HH2, and HH3 reveal a putative Causative mutation in SMC2 for HH3. *PLoS ONE* **9**, e2769 (2014).
- Cooper, T. A., Wiggans, G. R., Null, D. J., Hutchison, J. L. & Cole, J. B. Genomic evaluation, breed identification, and discovery of a haplotype affecting fertility for Ayrshire dairy cattle. *J. Dairy Sci.* **97**, 3878–3882 (2014).
- Schütz, E. *et al.* The Holstein Friesian lethal haplotype 5 (HH5) results from a complete deletion of TBF1M and cholesterol deficiency (CDH) from an ERV-(LTR) insertion into the coding region of APOB. *PLoS ONE* **11**, e0154602 (2016).
- Fritz, S. *et al.* An initiator codon mutation in SDE2 causes recessive embryonic lethality in Holstein cattle. *J. Dairy Sci.* **101**, 6220–6231 (2018).
- Ortega, M. S. *et al.* Truncation of IFT80 causes early embryonic loss in cattle. *BioRxiv* **2021**, 99. <https://doi.org/10.1101/2021.07.02.450952> (2021).
- Kipp, S. *et al.* Identification of a haplotype associated with cholesterol deficiency and increased juvenile mortality in Holstein cattle. *J. Dairy Sci.* **99**, 8915–8931 (2016).
- Menzi, F. *et al.* A transposable element insertion in APOB causes cholesterol deficiency in Holstein cattle. *Anim. Genet.* **47**, 253–257 (2016).
- National Center for Biotechnology Information. *ARS-UCD1.2 Assembly* (National Center for Biotechnology Information, 2018).
- Rosen, B. D. *et al.* De novo assembly of the cattle reference genome with single-molecule sequencing. *Gigascience* **9**, 021 (2020).
- Sargolzaei, M., Chesnais, J. P. & Schenkel, F. S. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* **15**, 478 (2014).
- Sargolzaei, M. *SNP1101 User's Guide. Version 1.0* (Guelph HiggsGene Solution Inc., 2014).
- Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188 (2001).
- Hoff, J. L., Decker, J. E., Schnabel, R. D. & Taylor, J. F. Candidate lethal haplotypes and causal mutations in Angus cattle. *BMC Genomics* **18**, 799 (2017).
- Garrick, D. J., Taylor, J. F. & Fernando, R. L. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet. Sel. Evol.* **41**, 55 (2009).

38. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
39. VanRaden, P. M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **91**, 4414–4423 (2008).
40. Häfliger, I. M. *et al.* Identification of small and large genomic candidate variants in bovine pulmonary hypoplasia and anasarca syndrome. *Anim. Genet.* **51**, 382–390 (2020).
41. Hayes, B. J. & Daetwyler, H. D. 1000 Bull Genomes Project to map simple and complex genetic traits in cattle: Applications and outcomes. *Annu. Rev. Anim. Biosci.* **7**, 89–102 (2019).
42. Brentp. *goleft*. *GitHub* (2018). <https://github.com/brentp/goleft>. Accessed 4 Oct 2021.
43. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
44. Hu, Y. *et al.* OmicCircos: A simple-to-use R package for the circular visualization of multidimensional Omics data. *Cancer Inform.* **13**, 13–20 (2014).
45. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
46. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
47. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
48. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE* **7**, e46688 (2012).
49. Wu, X., Mesbah-Uddin, M., Gulbrandtsen, B., Lund, M. S. & Sahana, G. Novel haplotypes responsible for prenatal death in Nordic Red and Danish Jersey cattle. *J. Dairy Sci.* **103**, 4570–4578 (2020).
50. Derks, M. F. L. *et al.* A systematic survey to identify lethal recessive variation in highly managed pig populations. *BMC Genomics* **18**, 858 (2017).
51. Shuster, D. E., Kehrli, M. E. Jr., Ackermann, M. R. & Gilbert, R. O. Identification and prevalence of a genetic defect that causes leukocyte adhesion deficiency in Holstein cattle. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9225–9229 (1992).
52. Mankanjuola, B. O. *et al.* Identification of unique ROH regions with unfavorable effects on production and fertility traits in Canadian Holsteins. *Genet. Sel. Evol.* **53**, 68 (2021).
53. Artesi, M. *et al.* Pooled CRISPR Inverse PCR sequencing (PCIP-seq): Simultaneous sequencing of retroviral insertion points and the integrated provirus with long reads. *BioRxiv* **2019**, 558130 (2019).
54. Häfliger, I. M. *et al.* APOB-associated cholesterol deficiency in Holstein cattle is not a simple recessive disease. *Anim. Genet.* **50**, 372–375 (2019).
55. Reynolds, E. G. M. *et al.* Non-additive association analysis using proxy phenotypes identifies novel cattle syndromes. *Nat. Genet.* **53**, 949–954 (2021).
56. Xiong, S. *et al.* Maternal uterine NK cell-activating receptor KIR2DS1 enhances placentation. *J. Clin. Investig.* **123**, 4264–4272 (2013).
57. Faridi, R. M. & Agrawal, S. Killer immunoglobulin-like receptors (KIRs) and HLA-C allorecognition patterns implicative of dominant activation of natural killer cells contribute to recurrent miscarriages. *Hum. Reprod.* **26**, 491–497 (2011).
58. Łuszczek, W. *et al.* Gene for the activating natural killer cell receptor, KIR2DS1, is associated with susceptibility to psoriasis vulgaris. *Hum. Immunol.* **65**, 758–766 (2004).
59. Suzuki, Y. *et al.* Genetic polymorphisms of killer cell immunoglobulin-like receptors are associated with susceptibility to *Psoriasis vulgaris*. *J. Investig. Dermatol.* **122**, 1133–1136 (2004).
60. Marin, D. *et al.* KIR2DS1 genotype predicts for complete cytogenetic response and survival in newly diagnosed chronic myeloid leukemia patients treated with imatinib. *Leukemia* **26**, 296–302 (2012).
61. La Nasa, G. *et al.* Homozygosity for killer immunoglobulin-like receptor haplotype A predicts complete molecular response to treatment with tyrosine kinase inhibitors in chronic myeloid leukemia patients. *Exp. Hematol.* **41**, 424–431 (2013).
62. Naumova, E. *et al.* Genetic polymorphism of NK receptors and their ligands in melanoma patients: Prevalence of inhibitory over activating signals. *Cancer Immunol. Immunother.* **54**, 172–178 (2005).
63. Dambaeva, S. V. *et al.* Recurrent pregnancy loss in women with killer cell immunoglobulin-like receptor KIR2DS1 is associated with an increased HLA-C2 allelic frequency. *Am. J. Reprod. Immunol.* **75**, 94–103 (2016).
64. Colucci, F. The role of KIR and HLA interactions in pregnancy complications. *Immunogenetics* **69**, 557–565 (2017).
65. Yang, X. & Meng, T. Killer-cell immunoglobulin-like receptor/human leukocyte antigen-C combination and ‘great obstetrical syndromes’ (review). *Exp. Ther. Med.* **22**, 1178 (2021).
66. Hiby, S. E. *et al.* Maternal activating KIRs protect against human reproductive failure mediated by fetal HLA-C2. *J. Clin. Investig.* **120**, 4102–4110 (2010).
67. Parham, P. & Guethlein, L. A. Pregnancy immunogenetics: NK cell education in the womb? *J. Clin. Investig.* **120**, 3801–3804 (2010).
68. Crespo, Á. C., Strominger, J. L. & Tilburgs, T. Expression of KIR2DS1 by decidual natural killer cells increases their ability to control placental HCMV infection. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 15072–15077 (2016).
69. Müller, M. P. *et al.* Genome-wide mapping of 10 calving and fertility traits in Holstein dairy cattle with special regard to chromosome 18. *J. Dairy Sci.* **100**, 1987–2006 (2017).
70. Kopan, R. & Ilagan, M. X. G. The canonical Notch signaling pathway: Unfolding the activation mechanism. *Cell* **137**, 216–233 (2009).
71. Bray, S. J. Notch signalling in context. *Nat. Rev. Mol. Cell Biol.* **17**, 722–735 (2016).
72. Baron, M. Combining genetic and biophysical approaches to probe the structure and function relationships of the notch receptor. *Mol. Membr. Biol.* **34**, 33–49 (2017).
73. Bellavia, D. *et al.* Notch3: From subtle structural differences to functional diversity. *Oncogene* **27**, 5092–5098 (2008).
74. Krebs, L. T. *et al.* Characterization of Notch3-deficient mice: Normal embryonic development and absence of genetic interactions with a Notch1 mutation. *Genesis* **37**, 139–143 (2003).
75. Domenga, V. *et al.* Notch3 is required for arterial identity and maturation of vascular smooth muscle cells. *Genes Dev.* **18**, 2730–2735 (2004).
76. Hosseini-Alghaderi, S. & Baron, M. Notch3 in development, health and disease. *Biomolecules* **10**, 485 (2020).
77. Kershaw, N. J. *et al.* Notch ligand delta-like1: X-ray crystal structure and binding affinity. *Biochem. J.* **468**, 159–166 (2015).
78. Pepermans, E. & Petit, C. The tip-link molecular complex of the auditory mechano-electrical transduction machinery. *Hear. Res.* **330**, 10–17 (2015).
79. Keats, B. J. B. & Savas, S. Genetic heterogeneity in Usher syndrome. *Am. J. Med. Genet. A* **130**, 13–16 (2004).
80. VanRaden, P. M. *et al.* Genomic imputation and evaluation using high-density Holstein genotypes. *J. Dairy Sci.* **96**, 668–678 (2013).
81. Crysanto, D., Leonard, A. S., Fang, Z.-H. & Pausch, H. Novel functional sequences uncovered through a bovine multi-assembly graph. *Proc. Natl. Acad. Sci.* **118**, e2101056118 (2021).
82. Robinson, J., Waller, M. J., Stoehr, P. & Marsh, S. G. E. IPD—The immuno polymorphism database. *Nucleic Acids Res.* **33**, D523–D526 (2005).
83. Chowdhury, R. *et al.* Ribosomal oxygenases are structurally conserved from prokaryotes to humans. *Nature* **510**, 422–426 (2014).

84. Rouget-Quermalet, V. *et al.* Protocadherin 15 (PCDH15): A new secreted isoform and a potential marker for NK/T cell lymphomas. *Oncogene* **25**, 2807–2811 (2006).
85. Cole, J. B. *et al.* Haplotype tests for recessive disorders that affect fertility and other traits. *Animal Improvement Program Research Report* (2018). https://www.aipl.arsusda.gov/reference/recessive_haplotypes_ARR-G3.html. Accessed 4 Oct 2021.
86. Agerholm, J. S., McEvoy, F. & Arnbjerg, J. Brachyspina syndrome in a Holstein calf. *J. Vet. Diagnostic Investig.* **18**, 418–422 (2006).
87. McClure, M. *et al.* Fine mapping for weaver syndrome in Brown Swiss cattle and the identification of 41 concordant mutations across NRCAM, PNPLA8 and CTTNBP2. *PLoS ONE* **8**, e59251 (2013).
88. Agerholm, J. S., Bendixen, C., Andersen, O. & Arnbjerg, J. Complex vertebral malformation in Holstein calves. *J. Vet. Diagnostic Investig.* **13**, 283–289 (2001).
89. Shanks, R. D., Dombrowski, D. B., Harpestad, G. W. & Robinson, J. L. Inheritance of UMP synthase in dairy cattle. *J. Hered.* **75**, 337–340 (1984).
90. Schwenger, B., Schöber, S. & Simon, D. DUMPS cattle carry a point mutation in the uridine monophosphate synthase gene. *Genomics* **16**, 241–244 (1993).
91. National Center for Biotechnology Information. *NCBI Bos Taurus Annotation Release 106* (National Center for Biotechnology Information, 2018).

Acknowledgements

The authors are grateful to the Swiss cattle breeding associations (swissherdbook, Holstein Switzerland) for providing genomic and phenotypic data. We thank Nathalie Besuchet-Schmutz for expert technical assistance. We thank the Interfaculty Bioinformatics Unit of the University of Bern for providing high performance computing infrastructure. In addition, we acknowledge Christy Vander Jagt, Amanda Chamberlain and Hans Daetwyler of the 1000 Bull Genomes Project for providing the variant catalogue. The Swiss National Science Foundation, Grant Number 172911, funds this study.

Author contributions

I.M.H. performed the whole-genome sequencing data preparation and analysis. Further, I.M.H. interpreted all collective results, drafted this manuscript with all tables and figures. C.D. acquired the whole-genome sequencing data and drafted this manuscript. M.S. and F.R.S. performed the SNP genotyping data preparation and analysis, including the GWAS and haplotype associations. C.D. and F.R.S. designed and managed the project. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-09403-6>.

Correspondence and requests for materials should be addressed to I.M.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022