

RESEARCH ARTICLE

# Balancing selection on a recessive lethal deletion with pleiotropic effects on two neighboring genes in the porcine genome

Martijn F. L. Derks<sup>1\*</sup>, Marcos S. Lopes<sup>2,3</sup>, Mirte Bosse<sup>1</sup>, Ole Madsen<sup>1</sup>, Bert Dibbits<sup>1</sup>, Barbara Harlizius<sup>2</sup>, Martien A. M. Groenen<sup>1</sup>, Hendrik-Jan Megens<sup>1</sup>

**1** Wageningen University & Research, Animal Breeding and Genomics, Wageningen, The Netherlands, **2** Topigs Norsvin Research Center, Beuningen, the Netherlands, **3** Topigs Norsvin, Curitiba, Brazil

\* [martijn.derks@wur.nl](mailto:martijn.derks@wur.nl)



**OPEN ACCESS**

**Citation:** Derks MFL, Lopes MS, Bosse M, Madsen O, Dibbits B, Harlizius B, et al. (2018) Balancing selection on a recessive lethal deletion with pleiotropic effects on two neighboring genes in the porcine genome. *PLoS Genet* 14(9): e1007661. <https://doi.org/10.1371/journal.pgen.1007661>

**Editor:** Tosso Leeb, University of Bern, SWITZERLAND

**Received:** April 24, 2018

**Accepted:** August 27, 2018

**Published:** September 19, 2018

**Copyright:** © 2018 Derks et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Whole genome sequence data for carriers and non-carriers of the deletion in this Large White population are deposited in the European Nucleotide Archive under BioProject accession PRJEB1683. The raw-sequence data of carrier samples can be found under accessions: ERS177318, ERS177319, ERS177322, ERS177320. The RNA-sequencing data is deposited under BioProject PRJEB19268, the raw data of the carrier sample PigWUR166 is deposited in the sequence read archive under accession ERX187799. The RNA-sequencing data

## Abstract

Livestock populations can be used to study recessive defects caused by deleterious alleles. The frequency of deleterious alleles including recessive lethal alleles can stay at high or moderate frequency within a population, especially if recessive lethal alleles exhibit an advantage for favourable traits in heterozygotes. In this study, we report such a recessive lethal deletion of 212kb (del) within the *BBS9* gene in a breeding population of pigs. The deletion produces a truncated BBS9 protein expected to cause a complete loss-of-function, and we find a reduction of approximately 20% on the total number of piglets born from carrier by carrier matings. Homozygous del/del animals die mid- to late-gestation, as observed from high increase in numbers of mummified piglets resulting from carrier-by-carrier crosses. The moderate 10.8% carrier frequency (5.4% allele frequency) in this pig population suggests an advantage on a favourable trait in heterozygotes. Indeed, heterozygous carriers exhibit increased growth rate, an important selection trait in pig breeding. Increased growth and appetite together with a lower birth weight for carriers of the *BBS9* null allele in pigs is analogous to the phenotype described in human and mouse for (naturally occurring) *BBS9* null-mutants. We show that fetal death, however, is induced by reduced expression of the downstream *BMPER* gene, an essential gene for normal foetal development. In conclusion, this study describes a lethal 212kb deletion with pleiotropic effects on two different genes, one resulting in fetal death in homozygous state (*BMPER*), and the other increasing growth (*BBS9*) in heterozygous state. We provide strong evidence for balancing selection resulting in an unexpected high frequency of a lethal allele in the population. This study shows that the large amounts of genomic and phenotypic data routinely generated in modern commercial breeding programs deliver a powerful tool to monitor and control lethal alleles much more efficiently.

from the non-carrier Large White pig is available under BioProject: PRJNA311523, deposited in the sequence read archive under: SRR3160021. All genotype data from pig chromosome 18, for 23,722 pigs, are available from animalgenome.org (repository link: <https://www.animalgenome.org/repository/pub/WUR2018.0508/>) upon reasonable request, and subject to the following conditions: a) Data are not to be further distributed without consent of data owners. b) Data will be available only for research institutes and for research purposes.

**Funding:** This research was funded by the STW-Breed4Food Partnership, project number 14283: From sequence to phenotype: detecting deleterious variation by prediction of functionality. This study was financially supported by NWO-TTW and the Breed4Food partners Cobb Europe, CRV, Hendrix Genetics and Topigs-Norsvin. The use of the HPC cluster was made possible by CAT-AgroFood (Shared Research Facilities Wageningen UR). In addition, this study was supported by the IMAGE project (Horizon 2020, No. 677353). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** MSL and BH are employees of Topigs Norsvin Research Center, a research institute closely related to one of the funders (Topigs Norsvin). All authors declare that the results are presented in full and as such present no conflict of interest. The other Breed4Food partners Cobb Europe, CRV, Hendrix Genetics, declare to have no competing interests for this study.

## Author summary

We report a large deletion within the *BBS9* gene that induces late fetal mortality in homozygous affected animals in a commercial pig population. This late fetal mortality causes the fetus to become encapsulated and desiccated during the remaining time of the pregnancy, a process called mummification. The unusually high carrier frequency for this lethal deletion (10.8%) likely results from its strong positive association with growth rate in heterozygous individuals, an important selection trait in the pig breeding industry. Interestingly, we show that the positive effect on growth is induced by a heterozygous loss-of-function of the *BBS9* gene, associated with obesity in human and mouse. However, late fetal mortality is induced by insufficient expression of the *BMPER* gene located directly downstream of the deletion which affects its regulatory elements required for gene expression. Together, our study shows an unique example of allelic pleiotropy in which one allele (deletion) is responsible for both increased growth and late fetal mortality by affecting two different genes.

## Introduction

Domesticated animals are excellent models to study the effect of inbreeding on fitness, and the role of selection in inbreeding depression. Breeding of domesticated animals increases inbreeding by applying artificial insemination that allows breeding populations to be sired by a small number of elite males. The frequency of deleterious alleles including recessive lethal alleles can rise in populations as a consequence of drift due to small effective population size, but also due to selection [1]. Inherited defects usually derive from unique “founder” mutations [2]. Especially in cattle breeds, several high frequency lethal alleles have been described [3, 4] reaching carrier frequencies up to 32% [5], that can be traced back to prime bulls that were used extensively in the past decades. However, the effect of individual sires on the population depends on the breeding goal and the structure of the breeding program. In cattle breeding, the genetic contribution of a single bull can be extreme, producing up to hundreds of thousands of daughters. In pig breeding, however, drift effects are expected to be less severe because recessive lethal alleles from founder boars are less likely to rise in frequency very rapidly, because of a lower male selection intensity compared to cattle breeding [6].

The role of random drift and/or selection in increasing the frequency of deleterious variants is complex. When effective population size is small, drift effects can result in less effective selection [7]. Interestingly, the number of lethal variants found at relatively high frequency in commercial pig populations appears to be low [8–10]. The relative paucity of high-frequency deleterious alleles in pig and chicken, species that generally show a more gender-balanced selection [6, 11], and larger effective population size compared to cattle breeds [12, 13], raises the question why still some alleles rise to moderate frequency despite having a very clear adverse effect. Heterozygote advantage for traits selected in commercial populations provides a tantalizing alternative hypothesis [14]. In cattle, various instances of balancing selection have been described, driving deleterious alleles to higher population frequencies [5, 15]. In pigs, similar observations were made involving a transposable element (L1) insertion with positive effect on litter size, but negative consequences for boar fertility [16].

In a previous study we identified various recessive lethal alleles in three pig breeds [8], but the majority of these lethal alleles were found at low frequencies. One recessive lethal haplotype, however, was found at moderate frequency (~9% carrier frequency) causing a significant increase in foetal mortality at mid- to late-gestation and resulting in a high fraction of

mummified piglets in a Large White commercial population. The strong deleterious nature of the allele and the high frequency suggests a factor other than drift driving this haplotype to high frequency.

In this study, we report evidence of balancing selection on a recessive lethal 212kb deletion within the *BBS9* gene with antagonistic effects on fertility and growth. The allele affects fertility by causing early fetal death in homozygous progeny, resulting in mummified piglets. The same allele increases growth rate and feed intake for carrier animals compared to non-carrier animals. We propose that the deletion is maintained at moderate frequency in the Large White breed because of its association with this positive effect, despite it being lethal in homozygous state.

## Results

### A haplotype inducing foetal lethality segregates at moderate frequency in a Large White pig population

Genomic loci that harbour recessive lethal alleles can be identified by searching for haplotypes showing reduced or missing homozygosity. In this study, we analysed a previously identified recessive lethal haplotype on pig chromosome 18 (SSC18: 39.25–40.1 Mb) using 23,722 Large White animals from a single purebred sow line genotyped on the Porcine50K SNPchip (Sscrofa11.1 build). The haplotype frequency is estimated at 5.4% (10.8% carrier frequency, Table 1), showing that the haplotype is segregating at moderate frequency in this Large White population. In total, we expect 55 homozygote carriers for the SSC18 haplotype within the population. However, no homozygous del/del animals were observed, supporting that all copies of the haplotype carry the recessive lethal variant exhibiting complete penetrance for homozygous animals. We also observe a significant reduction in total number born (19.5%) and liveborn individuals (19.3%) for carrier-by-carrier matings (CxC) compared to carrier-by-non-carrier matings (CxNC). Moreover, we found an approximate fivefold increase in mummified piglets (Table 1). The difference between stillborn and mummies lies in the moment the foetus dies: The term ‘mummy’ is used for a foetus that dies mid-to-late-gestation (e.g.

**Table 1. SSC18 haplotype characteristics and phenotypic effects.** Difference is the percent difference in the average total number born (TNB), number born alive (NBA), and mummified piglets (MUM) for C x C (carrier-by-carrier) and C x NC (carrier-by-non-carrier) matings.

Position, Mb	SSC18: 39.2–40.1
Number of markers	25
Starting marker	ASGA0079708
Ending marker	ALGA0098146
Homozygotes expected (trio)	55
Homozygotes observed	0
Exact binomial test	1.12e-27
Haplotype frequency %	5.42
Carrier frequency %	10.84
C x C matings	154
Genotyped C x C progeny	218
Heterozygote C x C progeny	159 (72.9%)
Avg. TNB (difference %)	12.86 (-19.5%)
Avg. NBA (difference %)	11.69 (-19.3%)
Avg. MUM (difference %)	1.62 (476.4%)
Genes in window	BMPER, BBS9

<https://doi.org/10.1371/journal.pgen.1007661.t001>

second to third trimester) and is subsequently encapsulated and desiccated during the remainder time of the pregnancy. A foetus that dies near the end of gestation or perinatally is identified as ‘stillborn’. The reduction in total number born is slightly lower than the expected 25% based on the 1:2:1 genotype distribution expected from CxC matings. About 73% of the CxC progeny is heterozygous for the SSC18 haplotype, corresponding to the 1:2 genotype ratio expected for CxC matings that lack homozygous offspring, significantly different compared to the normal 1:2:1 Mendelian ratio ( $p = <0.00001$ ). Based on the carrier frequency, we estimate that about 1.17% of the litters within this breed are affected by the SSC18 haplotype, producing affected animals (‘mummies’), and resulting in reduced litter sizes (on average 3.08 piglets per CxC litter).

### Genotyping the offspring of carrier-by-carrier matings confirms early lethality of homozygous animals

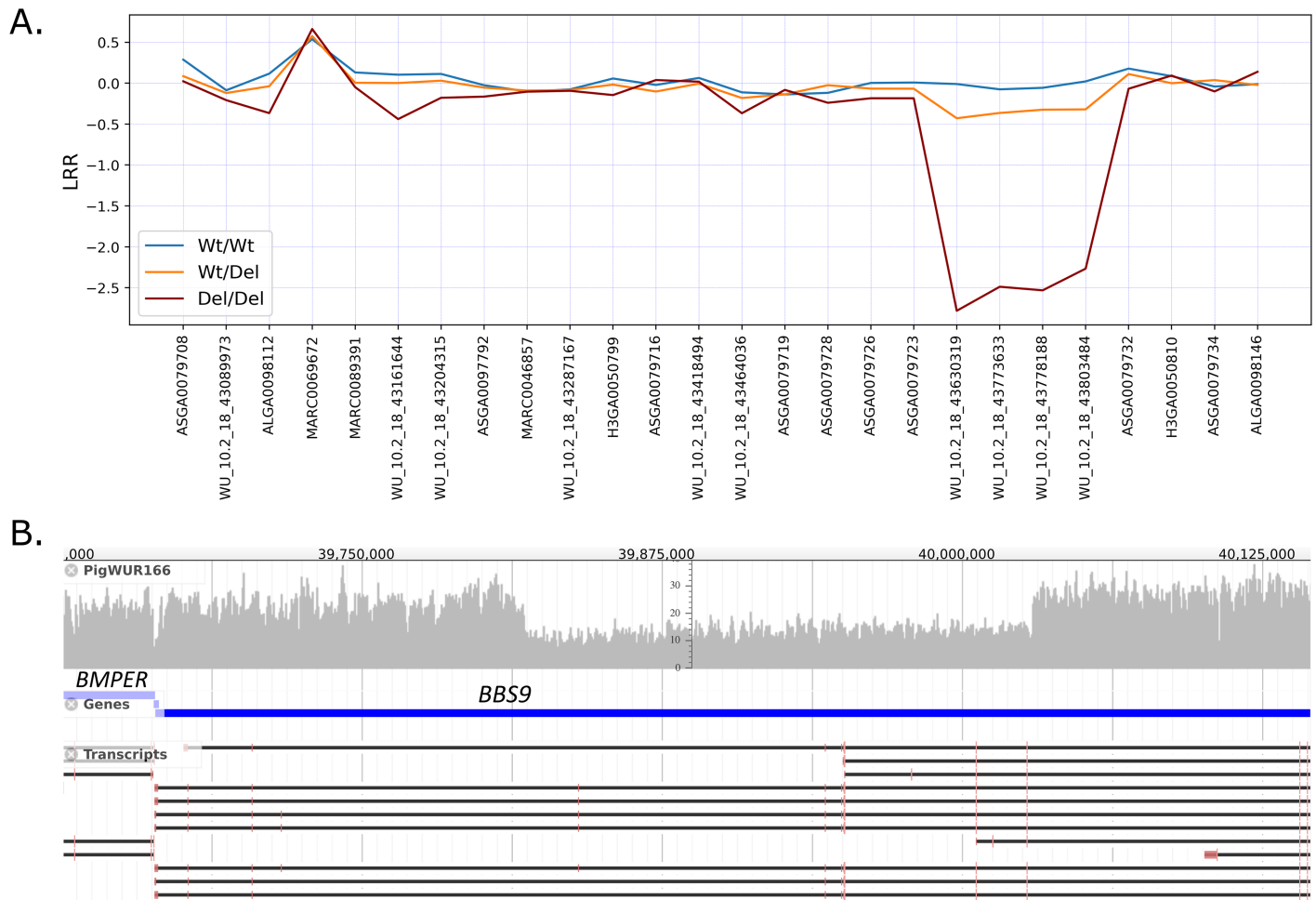
We tracked five recent CxC matings. Four pregnancies reached full term, while one resulted in spontaneous early abortion of the entire litter (Table 2). The four full-term litters produced 49 liveborn, 7 stillborn, and 14 mummified piglets. Each of these four litters produced at least 2 mummified piglets (maximum 5), significantly more than what is normally observed in this breed (on average 0.35 mummified piglets per litter,  $p = 0.0027$ ). Among the total of 48 genotyped liveborn and stillborn siblings (8 siblings were not genotyped), 16 were non-carriers, 30 were heterozygous (62.5%), and two were homozygous for the SSC18 haplotype, close to the expected 1:2 genotype ratio caused by missing homozygous offspring (S1 and S2 Tables). Among the two “fresh born” homozygous animals (i.e. piglets surviving at least until around birth), one was a stillborn piglet, the other was a liveborn but very weak piglet, that died shortly after birth.

We confirmed the homozygous status for two mummified piglets with sufficient call rate (call rate  $> 0.8$ , S1 Table), the other mummified piglets yielded insufficient DNA quality to perform genotyping and phasing (call rate  $< 0.8$ , S1 and S2 Tables). Next, we collected eight mummified piglets from one farm for phenotypic evaluation (including X-rays, S1 Fig), the other six mummified piglets were measured (length), but not stored. The approximate age when a mummified pig has died can be determined based on the length (crown to rump) and weight. The majority of the mummified piglets die approximately in the second half of the second trimester of pregnancy (50–70 days), based on the length (100–200 mm) and weight (100–190 gram) of the mummified piglets (S1 and S2 Tables, S1 Fig). Three mummified piglets from one litter (litter ID: CC4) died later in gestation as was evident from a larger size and weight (S1 Table). However, we cannot confirm the homozygous status for the SSC18 haplotype, since these animals could not be successfully genotyped due to poor DNA quality. Together these results support a broad range in the time of death between homozygous animals (supporting variation in penetrance), ranging from 50 days in gestation to 24 hours post-partum.

**Table 2. Tracked CxC matings for the SSC18 haplotype.** Phenotypes and genotypes of 4 litters from CxC matings from two different farms. The number of successfully genotyped individuals are indicated between parentheses for each birth type. Litter CC3 contains two fresh born homozygous individuals. An overview presenting the haplotypes and carrier status of the four litters is provided in S3 and S4 Tables.

Litter	Farm	Parity	Liveborn	Stillborn	Mummified	# Non-carriers	# Carriers	# Confirmed homozygotes
CC1	1	5	10 (6)	1 (1)	4 (0)	2	5	-
CC2	2	1	12 (11)	0	3 (1)	3	8	1
CC3	1	5	17 (15)	3 (3)	2 (0)	7	9	2
CC4	2	3	10 (10)	3 (2)	5 (1)	4	8	1

<https://doi.org/10.1371/journal.pgen.1007661.t002>

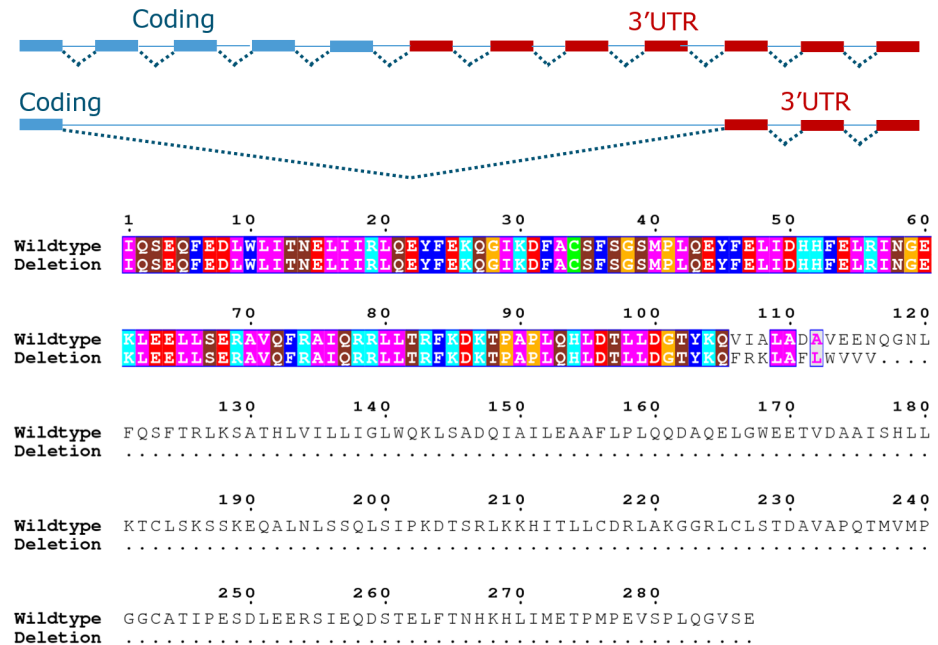


**Fig 1. A)** Log R Ratio (LRR) signal intensities on the 50K SNPchip for homozygotes (del/del) carriers (wt/del), and non-carriers (wt/wt). Four markers within the 212kb deletion show reduced LRR intensities. **B)** Screen capture of the alignment of carrier animal PigWUR166. The aligned region on SSC18 shows reduced coverage in the deletion within the *BBS9* gene.

<https://doi.org/10.1371/journal.pgen.1007661.g001>

### Carriers exhibit a 212kb deletion affecting the *BBS9* gene

To identify candidate causal mutations, we analysed whole genome sequence data from 73 individuals from the same Large White population and identified 10 carrier animals for the SSC18 haplotype (S5 Table). We first annotated loss-of-function and (deleterious) missense mutations within and surrounding the haplotype region (+/- 5 Mb) uniquely found in the SSC18 haplotype carriers. However, none of the mutations were predicted to have high impact (Variant Effect Predictor, build 90 [17]). Next, we assessed the presence of structural variation within the same region and identified a large deletion in complete LD with the SSC18 haplotype of approximately 212kb (position 39,817,373 to 40,029,300), spanning a part of the *BBS9* gene (Fig 1A and 1B). The deletion is supported by both split-reads and discordantly mapped pairs in carrier samples (S2 Fig). Moreover, carrier animals show reduced signal intensities (referred to as Log R Ratio; Fig 1A, S3 Fig), and increased homozygosity for four markers on the Porcine50K SNPchip located within deletion, caused by the absence of a second haplotype for the deletion region. In addition, several markers neighbouring the deletion show an excess of heterozygosity, caused by the absence of homozygous del/del animals.



**Fig 2. BBS9 “Wild-type” (top) and mutant (bottom) transcripts.** The deletion transcript skips four coding and four 3’UTR exons, resulting in a frameshift (indicated with an arrow in the alignment) introducing 11 AAs before a preliminary stop codon.

<https://doi.org/10.1371/journal.pgen.1007661.g002>

### SSC18 deletion produces a truncated BBS9 protein

We analysed RNA-seq data from one carrier animal in eight different tissue types (sample: Pig-Wur166, [S6 Table](#)) to investigate the impact of the deletion on the expression of *BBS9*. Moderate gene expression levels for *BBS9* were observed for the majority of the examined tissues, except for muscle, and with highest gene expression in testis ([S6 Table](#)). We evaluated the effect of the deletion on the *BBS9* mRNA and show that the deletion induces skipping of 4 coding, and 4 3’UTR exons for the *BBS9* canonical transcript ([Fig 2](#), RefSeq ID: XM\_021079336.1), resulting in direct splicing from exon 19 to exon 28 (3’UTR). The mutant transcript results in a frameshift introducing 11 novel amino acids before a premature stop codon, generating a truncated *BBS9* protein of 694 amino acids (including 11 novel amino acids) instead of the wild type 865 amino acids. This truncated *BBS9* protein will likely be non-functional ([Fig 2](#)), supported by pathogenic mutations identified in humans affecting the same C-terminal tail of the *BBS9* protein [[18](#), [19](#)]. Moreover, the affected protein coding exons exhibit a negative subRVIS score, indicating intolerance to loss-of-function mutations [[20](#)]. Finally, we evaluated the expression of *BBS9* using a RT-qPCR on 8 carrier and 10 non-carrier samples from whole blood using primers that target exons located within the deletion. The results show a 50% lower expression of the wild-type *BBS9* gene in carrier animals ([S4 Fig](#)).

### SSC18 deletion lowers *BMPER* expression by affecting cis-regulatory elements

To evaluate the impact of the deletion on the downstream *BMPER* gene we investigated possible allelic imbalance for the *BMPER* gene within the same carrier animal. The *BMPER* gene is highly expressed in lung, while moderately expressed in the other tissue types ([S6 Table](#)). One heterozygous coding synonymous mutation within the fourth exon of the *BMPER* canonical

**Table 3. Allele specific expression of the *BMPER* gene for a SSC18 carrier animal.** One heterozygous coding synonymous SNP within the fourth exon of the *BMPER* canonical transcript (XM\_013990842.2) was used to test for allelic imbalance.

Locus	Gene	Del-allele	Alt-allele	Del- Count	Wt- Count	Ratio	FDR-p
18:39594479	<i>BMPER</i>	C	T	24	73	0.753	2.35e <sup>-05</sup>

<https://doi.org/10.1371/journal.pgen.1007661.t003>

transcript (XM\_013990842.2) was used to test for allelic imbalance. Interestingly, we observed a three-fold higher expression of the *BMPER* allele for the wild-type haplotype (T allele) compared to the del haplotype (in lung tissue, Table 3). By contrast, three homozygous wild-type animals showed no allele specific differences in expression for the *BMPER* gene (S7 Table), suggesting that the region affected by the 212kb deletion contains *BMPER* cis-regulatory elements. To support the presence of *BMPER* regulatory elements within the deletion we aligned liver ChipSeq (H3K27Ac, H3K4Me3) data [21] to the Sscrofa11.1 genome build. Two strong enhancer peaks are observed within the deletion region, while only weak signals are observed outside the deletion region (S5 Fig). In addition, the sequence of the 212kb deletion was mapped to the human genome to identify the homologous sequence on the human genome (GRCh38: Chr7:33.50–33.71). This region contains several conserved regulatory elements, identified from the Regulatory Element Database [22], one non-coding RNA (LOC105375227), and several enhancer sites, of which at least two are annotated to enhance *BMPER* expression according to the human EnhancerAtlas [23].

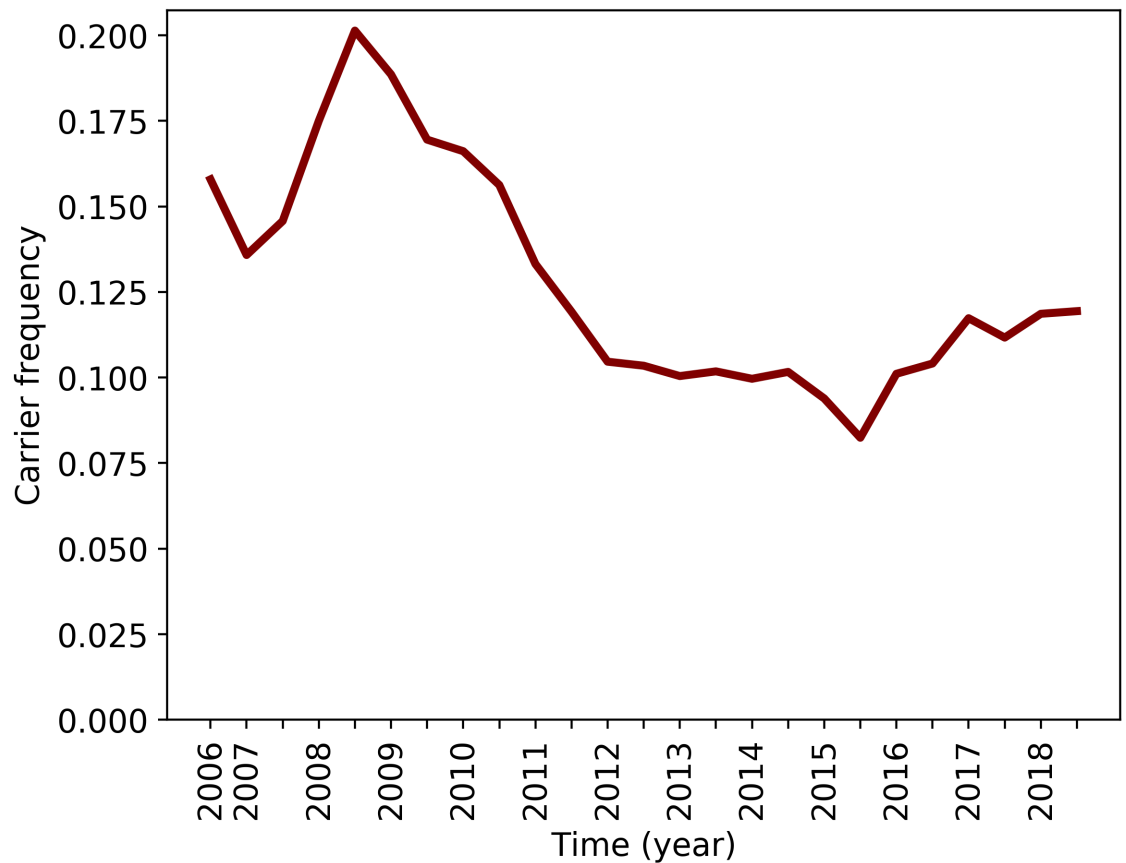
### Tracing the origin of the deletion

To investigate the origin of the deletion, we analysed the frequency of the deletion over the last decades. The first born animals within our genotyped set are from February 2006, allowing the tracking of the frequency of the deletion over the past decade. The number of genotyped animals was lower in the period 2006–2010. However, we genotyped over 320 animals in the (live) population from 2008 onwards, providing reliable frequency estimates (S8 Table). The SSC18 haplotype carrier frequency was high (>15%) over the period 2006–2010 (maximum 20% in 2008) and then decreased to a relative stable ~10% carrier frequency from 2012 onwards (Fig 3).

The Large White population under study has been created out of the consolidation of a number of Dutch breeding organizations around the turn of the last century [24]. During the consolidation phase, which resulted in merging of populations and phasing out of other populations, sperm of breeding boars was deposited at the Dutch Centre for Genetic Resources (CGN). The current Large White pure line descends from two different populations, the StamBoek-Z and the Dumeco-W line [24]. Both breeds were merged around 2003 to form the current Large White breeding line. From the 11 StamBoek-Z boars available at CGN, none were carrier of the deletion. However, from the 56 genotyped Dumeco-W boars available at CGN, five were carrier for the deletion haplotype (8.9%). These boars were born in 2000 and 2001 (S9 Table), showing that the deletion derives from this ancestral line and has been maintained in the Large White population for the past eighteen years (~ 15 generations).

### Carriers of the deletion have increased growth rate and feed intake

We examined whether the current carrier frequency is purely the result of genetic drift, or whether carriers exhibit selective advantage for important traits within the breeding program. We first simulated genetic drift for a lethal recessive allele in the current Large White population (S7 Fig). The results show that lethal alleles can reach allele frequencies up to about 10% by drift alone (although extremely rare), at which the lack of homozygotes is preventing



**Fig 3. SSC18 carrier frequency from 2006–2018.** The frequency has changed significantly over the past 12 years ( $p = 0.012$ ).

<https://doi.org/10.1371/journal.pgen.1007661.g003>

further increase. Next, we tested whether deletion carriers exhibit heterozygote advantage, by performing an association study for both carrier and non-carrier animals using deregressed estimated breeding values (DEBVs) for 16 production traits available from the Topigs-Norsvin breeding program (Table 4).

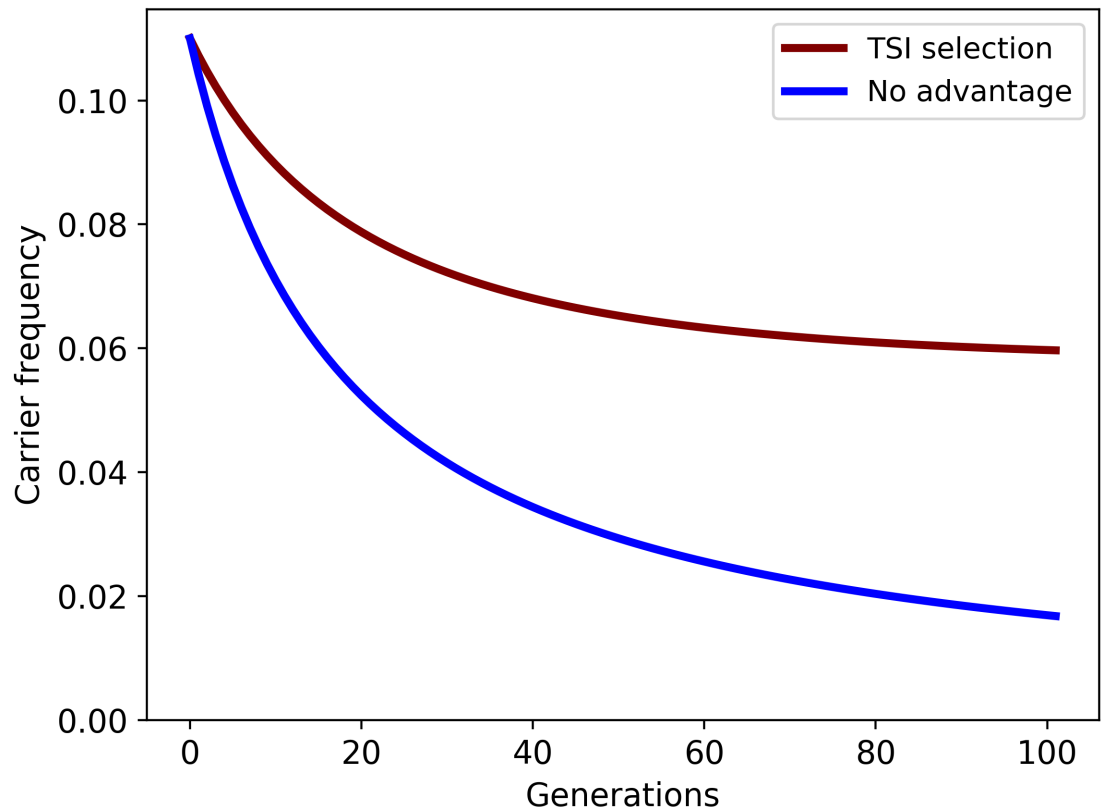
The carriers grow faster (TGR and LGR), have smaller loin depth (LDE), produce litters that are lighter (LBW), show higher mortality in their litters (LMO) and have a higher feed intake (DFI) when compared to the non-carriers. Selection on growth has not significantly

**Table 4. Traits significantly associated with heterozygous carriers of the deletion.** Effect shows the direction of the association, se shows the standard error. Table shows increased DEBVs for growth rate (TGR: growth rate in test period, ~25–120Kg, LGR: lifetime growth rate), daily feed intake (DFI), and litter mortality (LMO), while decreased DEBV for litter birth weight (LBW, grams), loin depth (LDE), and longevity (LGY) are observed for carriers. The symbols "+" and "-" indicate positive and negative effects. The effect on DFI can be considered both positive and negative. If TGR is increasing, DFI tends to increase a bit. However, it should not increase too much because it will affect feed conversion. An overview of all traits tested is provided in S10 Table.

Trait (unit)	Non-carriers	Carriers	P	-log10(P)	effect	se
TGR (gr/day) <sup>+</sup>	15013	1605	0.000046	4.34	11.46	2.81
LDE (mm) <sup>-</sup>	15011	1598	0.000198	3.70	-0.45	0.12
LGR (gr/day) <sup>+</sup>	15116	1616	0.000315	3.50	6.40	1.77
LBW (gram) <sup>-</sup>	6945	824	0.001232	2.91	-16.67	5.16
LMO (%) <sup>-</sup>	7345	871	0.001248	2.90	0.67	0.21
DFI (gr/day) <sup>+/-</sup>	14671	1567	0.006764	2.17	30.61	11.30
LGY (parity) <sup>-</sup>	7250	856	0.024828	1.61	-0.08	0.04

<https://doi.org/10.1371/journal.pgen.1007661.t004>





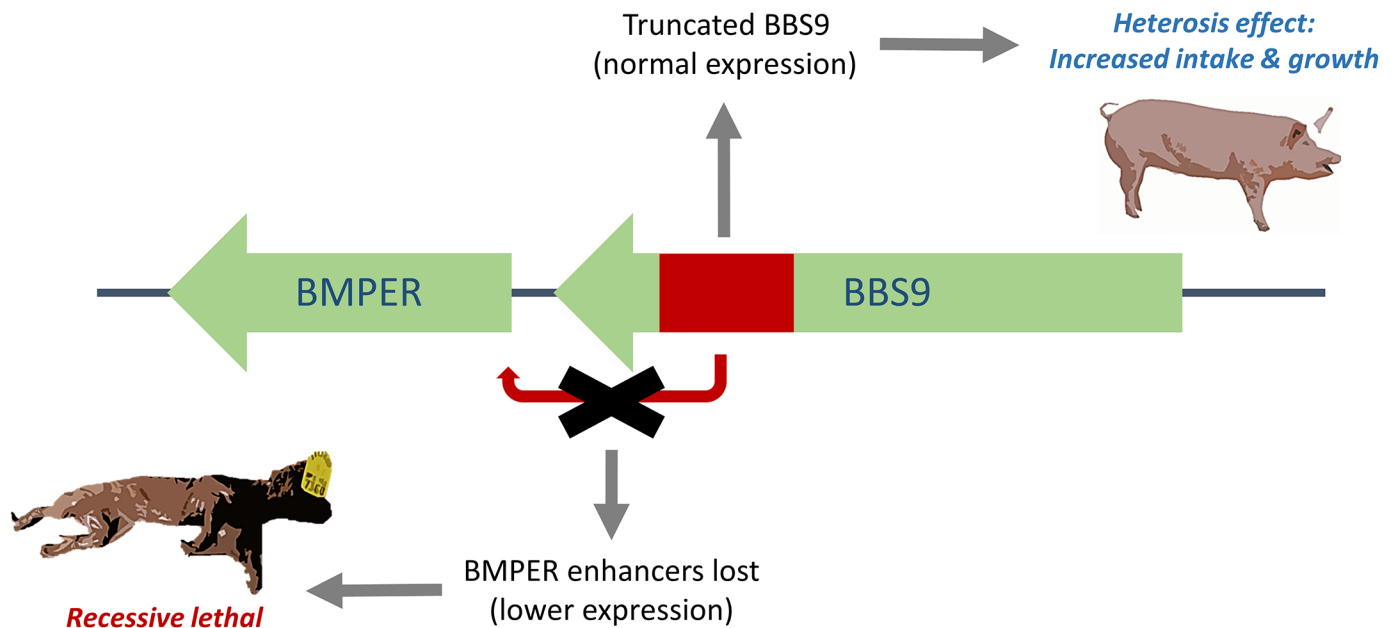
**Fig 4. Simulation of the SSC18 carrier frequency with current selective advantage over 100 generations starting with current carrier frequency (11%).** Figure shows a decrease in carrier frequency in the first generations due to the loss of homozygotes which outweighs the heterozygous advantage (~3%) perceived in the selection index (TSI). Figure shows a trade-off at approximately 6% carrier frequency at which the heterozygous advantage is compensated by the loss of homozygous offspring.

<https://doi.org/10.1371/journal.pgen.1007661.g004>

changed in the last decade, and there is consistent increase in genetic progress for growth in this time period (S8 Fig). To further support a balancing scenario, we evaluated the difference in the total selection index (TSI) between the carrier and non-carrier group for all animals born in 2017. Animals are ranked based on this selection index to select the top animals to produce the next generation. We observe a 2.7% higher TSI (on average) for carriers compared to non-carriers (S11 Table), caused by the positive effect on growth, that outweighs the negative effect on other traits. Next, we simulated the long term effect on the SSC18 carrier frequency based on the current heterozygous advantage and frequency (Fig 4, S9 Fig). We observe a decrease in carrier frequency in the first generations due to the loss of homozygotes, which outweighs the heterozygous advantage perceived in the selection index. However, at approximately 6% carrier frequency, the heterozygous advantage compensated for the loss of homozygous offspring, reaching a trade-off at this point. Moreover, carriers show 12.4% higher breeding values for growth compared to non-carriers (S11 Table), and we show that the carrier frequency can rise up to 22% if selection would be exclusively on growth (S10 Fig). Together these results support a balancing selection scenario showing heterozygote advantage for growth rate (Fig 5), an important selection trait in the pig breeding industry.

## Discussion

Livestock populations with small effective population size can lead to the spread of recessive lethal alleles, in which the effects of drift tend to dominate over the effect on selection [25]. On



**Fig 5. Schematic representation of the SSC18 deletion affecting *BMPER* gene expression and *BBS9* protein structure.** A heterozygous loss of function of the *BBS9* gene results in increased growth rates, while reduced expression of the *BMPER* gene results in foetal mortality in homozygous del/del animals.

<https://doi.org/10.1371/journal.pgen.1007661.g005>

the other hand, purging lethal alleles is more efficient in populations with small effective population size, which seems to contradict the relative lack of purging for the current allele under study over such long time period (almost two decades). One possible—partial—explanation is that fertility, a complex trait, is influenced by many genetic and environmental factors. As a consequence, fertility traits generally have low heritability's [26]. One of the important determinants of fertility is prenatal mortality [27]. However, prenatal mortality caused by recessive genetic defects is often difficult to capture within the pig breeding values (smaller litters will only be generated if two carriers mate), especially if the lethal variant is segregating at relative low frequency in the population.

We report a 212kb deletion that causes death of homozygous fetuses, and which also shows a positive effect on growth rate and feed intake in heterozygous pigs (Fig 5). We show that the allele has been segregating in the population for at least 18 years, despite its detrimental lethal effects in its homozygous state. The frequency of the deleterious allele was higher a decade ago, likely caused by genetic drift, because no significant changes in selection pressure for production traits have been applied during this time period. The balancing nature of the allele is clear from the higher TSI index of carrier animals. The 212kbp deletion clearly shows a net positive effect on the chances of pigs becoming selected in the breeding program, since the TSI is used to rank selection candidates. Although lower fertility is not captured in the TSI or even breeding values related to fertility directly, it is clear that the lower number of piglets born from CxC matings will result in a decreased fitness of these crosses. As a consequence, a balance is expected to arise between positive and negative selection, and indeed, the heterozygous advantage captured in the selection index compensates for the loss of homozygotes at approximately 6% carrier frequency. This frequency is somewhat lower than the current and past observed frequency. One partial explanation for this discrepancy is that we overestimate the number of CxC matings in our simulations (assuming random matings), because in practice matings between related individuals are avoided. Furthermore, during the 20<sup>th</sup> century, selection was

mainly applied on growth and carcass traits [28], indicating that the heterozygote advantage for carriers was likely stronger in the past. Interestingly, the balance between positive and negative selection would explain a rapid increase in the population under strong selection for growth traits, as has been the case for pig breeding lines at least since the introduction of modern breeding techniques.

This increase in growth rate (with the most significant effect in the test period from 25–120 Kg) and feed intake, likely results from a heterozygous loss-of-function of the *BBS9* gene. The *BBS9* protein is part of the BBSome complex, and is required for ciliogenesis [29]. *BBS9* is the central organizational component of the BBSome, having direct interactions with *BBS1*, 2, 5 and 8 [30]. Loss-of-function mutations in human *BBS9* and other members of the BBSome cause Bardet-Biedl syndrome, associated with a series of clinical features including obesity, renal anomalies, and retinopathy, with the obese phenotype as one of the key features of Bardet-Biedl syndrome patients. Studies have hypothesized that cilia defects are likely to affect feeding and satiety, causing an increased appetite and lack of satiation [29]. In addition, heterozygous carriers of a mutant BBS allele in humans show increased levels of obesity, without showing any of the other Bardet-Biedl syndrome features [31], analogous to the observed phenotype in carriers of the *BBS9* deletion. Mouse null-mutants in genes that form the BBSome complex have been associated with similar phenotypic features including obesity, lower birth weights, and partial embryonic lethality [32, 33], again supporting the *BBS9* role in increased growth rate, and the lower birth weight. We cannot completely exclude, however, that other genomic factors, in high LD with the 212kb deletion, contribute to the observed phenotype as well.

The question remains which gene or regulatory element is causal for the early death of homozygous individuals. We expect that the deletion (in homozygous state) leads to a complete loss-of-function of the *BBS9* gene, and decreased expression of the *BMPER* gene by affecting *BMPER* enhancer elements. Enhancers are important drivers of transcription and loss of enhancer elements can lead to decreased expression of the associated gene. Naturally occurring knock-outs of the *BBS9* gene does not result in fetal lethality in human [18] and is therefore not likely to be causal for the lethal phenotype. Instead, the downstream *BMPER* gene is a much stronger candidate since *BMPER* null mutants result in prenatal lethality with skeletal malformations in both mice and human [34, 35], marking the *BMPER* gene as the likely candidate underlying this phenomenon. We hypothesize that the deletion affects *BMPER* enhancer elements, resulting in insufficient expression of the *BMPER* gene in homozygous state. The lower expression of the *BMPER* gene is supported by allele specific expression for the non-deletion haplotype in carriers, not observed in individuals only carrying wild-type haplotypes. However, other BBS proteins, part of the BBSome, do cause (partial) embryonic lethality (i.e. *BBS4* [33] and *BBS7* [36]), we can therefore not exclude the possibility that a complete loss of a functional *BBS9* protein contributes to the early lethality as well. Together these results support that the deletion affects *BMPER* regulatory elements resulting in allelic imbalance for the *BMPER* gene in carrier animals. Therefore, severe downregulation of the *BMPER* gene is expected for del/del animals causing fetal death.

This work describes a striking example of balancing selection in pigs, maintaining a recessive lethal allele that shows pleiotropic effects on fertility and growth traits at moderate frequency in the population. Other examples in pigs include the Porcine Stress and Pale Soft Exudative Meat Syndrome, caused by a homozygous missense mutation in the *RYR1* gene, while heterozygotes show increased muscle mass [37]. A second example is a LINE insertion in the *SPEF2* gene causing increased litter size in sows but decreased fertility in boars [16]. Moreover, several instances of balancing selection have been described in domestic cattle breeds among which a 660kb deletion causing embryonic lethality in homozygotes, while

having increased milk yield in heterozygotes [5, 15]. Identifying balancing selection on lethal alleles can be challenging, as the only consequence observed is a (somewhat) lower fertility in the parental animals, lacking affected (liveborn) individuals. We expect that this type of balancing selection might be more prevalent within pig populations than previously thought, especially for the somewhat higher frequency lethal alleles, which are less likely to be purely the result of drift effects. Moreover, the relatively subtle effects found in this study could only be made apparent because phenotypic data derived from a very large number of pigs was available.

## Conclusion

In this study we report a 212 kb deletion with antagonistic effects on fertility and growth. We show that homozygotes for the deletion die around mid- to late-gestation, becoming mummified. Compared to other lethal alleles identified in this population, the deletion seems to be maintained at moderate frequency (10.8%) in the population. This moderate carrier frequency is likely not a result of random drift effects, as heterozygotes for the deletion-haplotype show, despite a lower birth weight, increased growth rate, and feed intake, important traits in the breeding goal. The balancing scenario observed, most likely, is a consequence of pleiotropic effects of the deletion on two different genes affecting fertility (*BMPER*) and growth (*BBS9*). The large amount of genotype data accumulating in modern breeding schemes applying genomic selection in combination with the large amount of phenotypic data deliver a powerful tool to monitor and control deleterious alleles much more efficiently.

## Material and methods

### Ethics statement

Samples collected for DNA extraction were only used for routine diagnostic purpose of the breeding programs, and not specifically for the purpose of this project. Therefore, approval of an ethics committee was not mandatory. Sample collection and data recording were conducted strictly according to the Dutch law on animal protection and welfare (Gezondheids- en welzijnswet voor dieren).

### Animals, genotypes and pre-processing

The dataset consists of 23,722 purebred Large White animals. The animals were genotyped on the Illumina GeneSeek custom 50K SNP chip (Lincoln, NE, USA). Animals with a frequency of missing genotypes  $> 0.20$  were removed. We discard markers that did not meet following filtering criteria: A minimum call rate of 0.85, a minor allele frequency  $> 0.01$ , and a Hardy-Weinberg proportions exact test p-value below  $P < 10^{-6}$ . Moreover, markers with unknown location on the Sscrofa11.1 genome build [38] were discarded, leaving 42,288 markers after filtering. All steps were performed in Plink v1.90b3.30 [39].

### Haplotype phasing

We performed haplotype phasing and imputation of missing sites in Beagle4.1 with parameter for effective population size set to 195, other settings were default [40]. Reference and test phased VCF files were merged using bcftools 1.3–27-gf31e888 [41].

### Identification of missing homozygote haplotypes

We tested the SS18 haplotype for the expected number of homozygotes using both parents haplotype information (sire, and dam) with the formula described in Fritz et al., 2013 [42]. An

exact binomial test was applied to test the number of observed homozygotes with the number of expected homozygotes. The haplotype was considered significantly depleted if  $P < 5 \times 10^{-3}$ . The difference in Mendelian ratios for CxC compared to CxNC matings was tested using a Chi-Square test.

### Pseudo genotyping for SSC18 deletion

To genotype animals directly for the SSC18 deletion, we first calculated LRR normalized signal intensities using PennCNV analysis software [43]. We built a classifier with 5 features: the LRR signal intensities for the four overlapping markers within the deletion (WU\_10.2\_18\_43630319, WU\_10.2\_18\_43773633, WU\_10.2\_18\_43778188, WU\_10.2\_18\_43803484), and the average LRR signal intensity over these four markers. Next, we applied logistic regression to distinguish carrier from non-carrier animals using the sci-kit learn Python library [44] (S6 Fig).

### Phenotypic effects associated with lethal haplotypes

We examined the SSC18 haplotype for records on TNB, NSB, and MUM listed for all C x C, and C x NC matings identified in the phenotypic records, the order of C x NC matings does not reflect the sex of the parent animal and is both carrier boar and carrier sow combined. We used a Welch's t-test to assess whether the phenotypes from the C x C matings differ significantly from C x NC matings. A p-value  $< 0.05$  was considered significant.

### WGS analysis and candidate variant identification

The dataset consists of 73 whole genome sequenced Large White individuals with a total volume of 1.77 Tbp (tera base pairs) from 15.539 billion paired-end reads, ranging from 100–150 bp in length (S5 Table). The data was sequenced on Illumina HiSeq 2000. We used sickle software for quality trimming of raw reads. Next we aligned the sequences to the Sscrofa11.1 genome build [38] using BWA-MEM version 0.7.15 [45] with an average mappability of 96.11% and a sample coverage ranging from 6.6–22.7X (10X average). Samtools dedup function was used to remove PCR duplicates [41]. GATK IndelRealigner was used to perform local realignments around indels [46]. Variant calling was performed with FreeBayes v1.1.0 with following settings:—min-base-quality 10—min-alternate-fraction 0.2—haplotype-length 0—min-alternate-count 2 [47]. Variants with phred quality score  $< 20$ , and within 3 bp of an indel were discarded [41]. Variants were annotated using the Ensembl variant effect predictor (VEP, release 90) [17]. The impact of missense variants was predicted using SIFT [48]. The sequenced population was phased using Beagle4.1 [40].

### Structural variation analysis

Analysis on structural variation (SV) was performed using Lumpy with default settings [49], taking the aligned BAM files as input. Coverage information was calculated for predicted SV events using samtools depth [41], and added to the VCF format tag using PyVCF. Alignments and SV events were visualized using the JBrowse genome viewer version 1.12.1 [50].

### RNA-seq analysis and allele specific expression

We analyzed RNA-seq data on eight different tissues in one SSC18 carrier animal (sample: Pig-WUR166). In addition, we analyzed two other pigs from Duroc, and Pietrain genetic background on five different tissues. RNA-seq reads were aligned to the Sscrofa11.1 genome build using STAR 2.5.3a [51], generation of transcripts and gene expression levels were achieved with Cufflinks v2.2.1 [52]. We applied the following steps to examine allele specific expression:

First, samtools [41] was used to extract uniquely mapped reads from the BAM alignment files. Next, WASP [53] was used to reduce the mapping (reference sequence) bias. Then, GATK-SReadcounter [46] was used to obtain read counts for reference and alternative alleles at each SNP position. Lastly, a two-sided binomial test with  $p = 0.5$  (assuming no bias) and Benjamini-Hochberg false discovery rate (FDR) correction were performed in R v.3.4 at each variant position using the Stats package. The variants with FDR adjusted  $p$ -value  $< 0.05$  were considered as allele specific expression variants. Visual examination of the alignments and transcripts was performed in JBrowse [50].

### RNA-isolation and RT-qPCR

RNA was extracted from frozen whole blood using the Nucleospin RNA blood kit from Machery Nagel. cDNA was synthesized using Superscript II Reverse Transcriptase (Invitrogen) with RNA input  $\sim 100$ ng. RT-qPCR was started with: 3.75ul cDNA (1:1), 1.25ul primer forward (2uM), 1.25ul primer reverse (2uM), and 6.25ul MESA blue mix (Eurogentec). RT-qPCR was then performed with a QuantStudio 5 system using the comparative Ct (delta delta Ct) method with GAPDH as housekeeping gene for normalization. Reaction was performed as follows:

1. 50°C 2 min, 1 cycle
2. 95°C 10 min, 1 cycle
3. 95°C 15 sec, 1 cycle
4. 60°C 1 min, 40 cycles
5. 95°C 15 sec, 1 cycle
6. 60°C 1 min, 1 cycle
7. 95°C 15 sec, 1 cycle

Data was analysed with the Quantstudio Design & Analysis Software v.1.4.3. All primers and results are listed in [S12](#) and [S13](#) Tables.

### ChipSeq alignment

We downloaded three H3K27Ac, and three H3K4me3 libraries (ArrayExpress accession number: E-MTAB-2633) from liver tissue from three male pig samples described by Villar et al. 2015 [21]. Data was aligned using BWA-mem [45] and visualized in JBrowse [50].

### Frequency over time

We analyzed the frequency of the SSC18 haplotype per half-year starting from 01-jul-2006. We assessed the frequency based on total population (live animals) on each time point by looking at the proportion of carrier and non-carrier animals in the population. The number of animals per time point are provided in [S8 Table](#). We used a One-Way Repeated Measures ANOVA to test whether the frequency differs over time.

### Breeding values and association analysis

In this study, we evaluated 16 traits used in the Large White breeding program. Deregressed estimated breeding values (DEBV) were used as a response variable for each trait under study. The estimated breeding value (EBV) was separately deregressed for each trait using the methodology described by Garrick et al [54]. The EBV of each animal was obtained from the routine genetic evaluation by Topigs Norsvin using an animal model. The reliabilities per animal

for the purpose of deregression were extracted from the genetic evaluation based on the methodology of Tier & Meyer [55]. The heritabilities used for the deregression were also extracted from the routine genetic evaluation. Parent average effects were also removed as part of the deregression process to obtain more accurate estimates of the genetic merit of each individual. Finally, weighting factors based on the estimated reliability of the DEBV were also estimated according to Garrick et al [54] using a value of 0.5 for the scalar  $c$ . To ensure the quality of the DEBV, only animals with a  $w$  higher than not equal to zero and a reliability of the DEBV greater than 0.20 were used in the association analyses. The reliability of the DEBV was obtained according to Garrick et al [54].

Association analyses were performed using the software ASREML [56] applying the following model:

$$DEBV_{ij}w = \mu + R_i + a_j + e_{ij}$$

where  $DEBV_{ij}$  is the observed DEBV for the animal  $j$ ,  $w$  is weighting factor for the residual,  $\mu$  is the overall DEBV mean of the population,  $R_i$  is the carrier status of the lethal allele  $i$ ,  $a_j$  is the additive genetic effect estimated using a pedigree-based relationship matrix, and  $e_{ij}$  the residual error.

### Simulating genetic drift

We simulated changes in allele frequency across multiple populations under the model of Wright [57]. Each allele is associated with a fitness, and we set the fitness to zero for homozygotes (for lethal recessive allele) and fitness to 1 (no negative fitness effect) for carriers and non-carriers. We assume constant population size through time, and matings are simulated randomly at each generation. Changes in allele frequencies are plotted using the R package driftR (<https://github.com/cjbattey/driftR>).

### Balancing selection

Within each generation the top 5% of boars, and top 25% of gilts (based on the TSI selection index value) are used to produce the next generation. We first calculated the average TSI, and estimated breeding values for six important traits in the breeding line (S11 Table). Next, we used the ratio of carrier TSI over non-carrier TSI to estimate the selective advantage in the breeding program. Next, we simulated the long-term allele frequency change (assuming random matings) based on the selective advantage, and the loss of homozygous animals using the Hardy-Weinberg principle. Similar analysis was performed using the selective advantage on growth exclusively.

### Supporting information

**S1 Fig. Pictures and X-rays from mummified piglets.** A: Picture of a homozygous  $del/del$  mummified piglet (7360). B: X-ray of homozygous  $del/del$  mummified piglet (7360). C: X-ray of mummified piglet 7659 (status unknown). D: X-ray of homozygous  $del/del$  mummified piglet (8609).

(PDF)

**S2 Fig. Screen capture of the deletion boundary in the JBrowse genome browser.**

(PDF)

**S3 Fig. LRR signal intensities within the haplotype region for two "fresh born" homozygotes of the 212kb deletion.**

(PDF)

**S4 Fig. Expression fold change of the *BBS9* gene (RT-qPCR) in 8 carriers, and 10 non-carriers.**

(PDF)

**S5 Fig. JBrowse screen capture showing aligned liver ChipSeq data (H3K27Ac, H3K4Me3) on Sscrofa11.1.**

(PDF)

**S6 Fig. Logistic regression to distinguish carrier from non-carrier animals (farm 2 litters).**

(PDF)

**S7 Fig. Genetic drift simulation for SSC18 lethal recessive with allele frequency 5.4% (10.8% carrier frequency).**

(PDF)

**S8 Fig. Genetic progress for growth (daily gain) in the Large White breed.**

(PDF)

**S9 Fig. Simulation of the SSC18 carrier frequency with current selective advantage over 200 generations starting with lower carrier frequency (2%).**

(PDF)

**S10 Fig. Simulation of the SSC18 carrier frequency if selection would be exclusively applied on growth.**

(PDF)

**S1 Table. Results from two CxC (CC1: 11-may-2017, CC3: 19-may-2017) matings on farm 1.**

(PDF)

**S2 Table. Results from two CxC (CC2: 18-may-2017, CC4: 01-jul-2017) matings on farm 2.**

(PDF)

**S3 Table. Markers and genomic positions.**

(XLSX)

**S4 Table. Haplotypes from the four tracked carrier-by-carrier matings, including parent animals.**

(XLSX)

**S5 Table. WGS sequenced individuals in the Large White breed.**

(PDF)

**S6 Table. Gene expression measured in fragments per kilobase per million (FPKM) for *BBS9* and *BMPER* gene in one SSC18 deletion carrier animal (LW) and two non-carrier pigs (DU/PI).**

(PDF)

**S7 Table. Allele specific expression test of the *BMPER* gene for three non-carriers pigs.**

Heterozygous coding variants of the *BMPER* canonical transcript (XM\_013990842.2) are used to test for allelic imbalance.

(PDF)



**S8 Table. Deletion carrier frequency and the number of genotyped animals per time point from 2006–2018.**

(PDF)

**S9 Table. Dumeco-W carriers for SSC18 deletion including birth dates.**

(PDF)

**S10 Table. Association analysis using deregressed breeding values (DEBV) for 16 traits in the Large White breed.**

(PDF)

**S11 Table. Estimated breeding values (EBV) for seven traits including the overall selection index (TSI).**

(PDF)

**S12 Table. Primer information for the genes used for RT-qPCR.**

(PDF)

**S13 Table. RT-qPCR results for *BBS9* expression in 8 carriers and 10 non-carriers.**

(PDF)

## Acknowledgments

The authors thank Meenu Bhiati for useful input on the allele specific expression analysis. We thank Rita Hoving and Ina Hulsegge for providing the CGN data and information on the consolidation of breeding lines. We thank the FAANG consortium for providing the RNA-seq data. We thank Michel Georges and Carole Charlier for useful comments on this work, Remco Pieters for generating the X-rays, and Richard Crooijmans for library preparation and lab work. We thank Marcel van Kuppenveld, Marc Martens, and Rene Sol for help with sample collection. We thank Egiel Hanenberg for providing EBVs and genetic progress information.

## Author Contributions

**Conceptualization:** Martijn F. L. Derks, Marcos S. Lopes, Martien A. M. Groenen, Hendrik-Jan Megens.

**Formal analysis:** Martijn F. L. Derks, Marcos S. Lopes, Ole Madsen, Bert Dibbits.

**Funding acquisition:** Martien A. M. Groenen, Hendrik-Jan Megens.

**Methodology:** Martijn F. L. Derks, Marcos S. Lopes, Mirte Bosse, Ole Madsen, Bert Dibbits, Barbara Harlizius.

**Visualization:** Martijn F. L. Derks, Mirte Bosse.

**Writing – original draft:** Martijn F. L. Derks, Hendrik-Jan Megens.

**Writing – review & editing:** Martijn F. L. Derks, Marcos S. Lopes, Mirte Bosse, Ole Madsen, Barbara Harlizius, Martien A. M. Groenen, Hendrik-Jan Megens.

## References

1. Leroy G. Inbreeding depression in livestock species: review and meta-analysis. *Anim Genet.* 2014; 45(5):618–28. <https://doi.org/10.1111/age.12178> PMID: 24975026
2. Yin T, Wensch-Dorendorf M, Simianer H, Swalve HH, König S. Assessing the impact of natural service bulls and genotype by environment interactions on genetic gain and inbreeding in organic dairy cattle

- genomic breeding programs. *Animal*. 2014; 8(6):877–86. <https://doi.org/10.1017/S1751731114000718> PMID: 24703184
3. Hoff JL, Decker JE, Schnabel RD, Taylor JF. Candidate lethal haplotypes and causal mutations in Angus cattle. *BMC Genomics*. 2017; 18(1):799. <https://doi.org/10.1186/s12864-017-4196-2> PMID: 29047335
  4. Sahana G, Iso-Touru T, Wu X, Nielsen US, de Koning DJ, Lund MS, et al. A 0.5-Mbp deletion on bovine chromosome 23 is a strong candidate for stillbirth in Nordic Red cattle. *Genet Sel Evol*. 2016; 48:35. <https://doi.org/10.1186/s12711-016-0215-z> PMID: 27091210
  5. Kadri NK, Sahana G, Charlier C, Iso-Touru T, Guldbrandtsen B, Karim L, et al. A 660-Kb deletion with antagonistic effects on fertility and milk production segregates at high frequency in Nordic Red cattle: additional evidence for the common occurrence of balancing selection in livestock. *PLoS Genet*. 2014; 10(1):e1004049. <https://doi.org/10.1371/journal.pgen.1004049> PMID: 24391517
  6. Knol EF, Nielsen B, Knap PW. Genomic selection in commercial pig breeding. *Animal Frontiers*. 2016; 6(1):15–22.
  7. Bosse M, Megens HJ, Madsen O, Crooijmans RP, Ryder OA, Austerlitz F, et al. Using genome-wide measures of coancestry to maintain diversity and fitness in endangered and domestic pig populations. *Genome research*. 2015; 25(7):970–81. <https://doi.org/10.1101/gr.187039.114> PMID: 26063737
  8. Derks MFL, Megens HJ, Bosse M, Lopes MS, Harlizius B, Groenen MAM. A systematic survey to identify lethal recessive variation in highly managed pig populations. *BMC Genomics*. 2017; 18(1):858. <https://doi.org/10.1186/s12864-017-4278-1> PMID: 29121877
  9. Haggman J, Uimari P. Novel harmful recessive haplotypes for reproductive traits in pigs. *Journal of animal breeding and genetics = Zeitschrift fur Tierzucht und Zuchtungsbiologie*. 2017; 134(2):129–35. <https://doi.org/10.1111/jbg.12240> PMID: 27625156
  10. Howard DM, Pong-Wong R, Knap PW, Woolliams JA. Use of haplotypes to identify regions harbouring lethal recessive variants in pigs. *Genet Sel Evol*. 2017;49.
  11. Bouquet A, Juga J. Integrating genomic selection into dairy cattle breeding programmes: a review. *Animal*. 2013; 7(5):705–13. <https://doi.org/10.1017/S1751731112002248> PMID: 23200196
  12. Hidalgo AM, Bastiaansen JWM, Lopes MS, Calus MPL, de Koning DJ. Accuracy of genomic prediction of purebreds for cross bred performance in pigs. *Journal of Animal Breeding and Genetics*. 2016; 133(6):443–51. <https://doi.org/10.1111/jbg.12214> PMID: 27087113
  13. Kelleher MM, Berry DP, Kearney JF, McParland S, Buckley F, Purfield DC. Inference of population structure of purebred dairy and beef cattle using high-density genotype data. *Animal*. 2017; 11(1):15–23. <https://doi.org/10.1017/S1751731116001099> PMID: 27330040
  14. Hedrick PW. Heterozygote advantage: the effect of artificial selection in livestock and pets. *J Hered*. 2015; 106(2):141–54. <https://doi.org/10.1093/jhered/esu070> PMID: 25524994
  15. Fasquelle C, Sartelet A, Li W, Dive M, Tamma N, Michaux C, et al. Balancing selection of a frame-shift mutation in the MRC2 gene accounts for the outbreak of the Crooked Tail Syndrome in Belgian Blue Cattle. *PLoS Genet*. 2009; 5(9):e1000666. <https://doi.org/10.1371/journal.pgen.1000666> PMID: 19779552
  16. Sironen A, Uimari P, Iso-Touru T, Vilkki J. L1 insertion within SPEF2 gene is associated with increased litter size in the Finnish Yorkshire population. *Journal of animal breeding and genetics = Zeitschrift fur Tierzucht und Zuchtungsbiologie*. 2012; 129(2):92–7. <https://doi.org/10.1111/j.1439-0388.2011.00977.x> PMID: 22394230
  17. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016; 17(1):122. <https://doi.org/10.1186/s13059-016-0974-4> PMID: 27268795
  18. Nishimura DY, Swiderski RE, Searby CC, Berg EM, Ferguson AL, Hennekam R, et al. Comparative genomics and gene expression analysis identifies BBS9, a new Bardet-Biedl syndrome gene. *Am J Hum Genet*. 2005; 77(6):1021–33. <https://doi.org/10.1086/498323> PMID: 16380913
  19. Shaheen R, Szymanska K, Basu B, Patel N, Ewida N, Faqeh E, et al. Characterizing the morbid genome of ciliopathies. *Genome Biol*. 2016;17.
  20. Gussow AB, Petrovski S, Wang QL, Allen AS, Goldstein DB. The intolerance to functional genetic variation of protein domains predicts the localization of pathogenic mutations within genes. *Genome Biol*. 2016; 17.
  21. Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, et al. Enhancer evolution across 20 mammalian species. *Cell*. 2015; 160(3):554–66. <https://doi.org/10.1016/j.cell.2015.01.006> PMID: 25635462
  22. Sheffield NC, Thurman RE, Song L, Safi A, Stamatoyannopoulos JA, Lenhard B, et al. Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and

- long-range interactions. *Genome research*. 2013; 23(5):777–88. <https://doi.org/10.1101/gr.152140.112> PMID: 23482648
23. Gao T, He B, Liu S, Zhu H, Tan K, Qian J. EnhancerAtlas: a resource for enhancer annotation and analysis in 105 human cell/tissue types. *Bioinformatics*. 2016; 32(23):3543–51. <https://doi.org/10.1093/bioinformatics/btw495> PMID: 27515742
  24. Hoving R, Hulsegge I, Hiemstra S. Varkenrassen in de genenbank. Centrum voor Genetische Bronnen, Nederland (CGN) van Wageningen University & Research. 2017;CGN rapport 37.
  25. Glemin S. How are deleterious mutations purged? Drift versus nonrandom mating. *Evolution*. 2003; 57(12):2678–87. PMID: 14761049
  26. Zak LJ, Gaustad AH, Bolarin A, Broekhuijse M, Walling GA, Knol EF. Genetic control of complex traits, with a focus on reproduction in pigs. *Mol Reprod Dev*. 2017; 84(9):1004–11. <https://doi.org/10.1002/mrd.22875> PMID: 28792084
  27. van der Lende T, Knol EF, Leenhouwers JI. Prenatal development as a predisposing factor for perinatal losses in pigs. *Reprod Suppl*. 2001; 58:247–61. PMID: 11980194
  28. Merks JWM. One century of genetic changes in pigs and the future needs. The challenge of genetic change in animal production. 2000; 27:8–19.
  29. Novas R, Cardenas-Rodriguez M, Irigoien F, Badano JL. Bardet-Biedl syndrome: Is it only cilia dysfunction? *FEBS Lett*. 2015; 589(22):3479–91. <https://doi.org/10.1016/j.febslet.2015.07.031> PMID: 26231314
  30. Klink BU, Zent E, Juneja P, Kuhlee A, Raunser S, Wittinghofer A. A recombinant BBSome core complex and how it interacts with ciliary cargo. *Elife*. 2017; 6.
  31. Croft JB, Morrell D, Chase CL, Swift M. Obesity in heterozygous carriers of the gene for the Bardet-Biedl syndrome. *Am J Med Genet*. 1995; 55(1):12–5. <https://doi.org/10.1002/ajmg.1320550105> PMID: 7702084
  32. Jiang J, Promchan K, Jiang H, Awasthi P, Marshall H, Harned A, et al. Depletion of BBS Protein LZTFL1 Affects Growth and Causes Retinal Degeneration in Mice. *J Genet Genomics*. 2016; 43(6):381–91. <https://doi.org/10.1016/j.jgg.2015.11.006> PMID: 27312011
  33. Kulaga HM, Leitch CC, Eichers ER, Badano JL, Lesemann A, Hoskins BE, et al. Loss of BBS proteins causes anosmia in humans and defects in olfactory cilia structure and function in the mouse. *Nat Genet*. 2004; 36(9):994–8. <https://doi.org/10.1038/ng1418> PMID: 15322545
  34. Funari VA, Krakow D, Nevarez L, Chen Z, Funari TL, Vatanavicharn N, et al. BMPER mutation in diaphanospondylodysostosis identified by ancestral autozygosity mapping and targeted high-throughput sequencing. *Am J Hum Genet*. 2010; 87(4):532–7. <https://doi.org/10.1016/j.ajhg.2010.08.015> PMID: 20869035
  35. Ikeya M, Kawada M, Kiyonari H, Sasai N, Nakao K, Furuta Y, et al. Essential pro-Bmp roles of cross-veinless 2 in mouse organogenesis. *Development*. 2006; 133(22):4463–73. <https://doi.org/10.1242/dev.02647> PMID: 17035289
  36. Zhang Q, Seo S, Bugge K, Stone EM, Sheffield VC. BBS proteins interact genetically with the IFT pathway to influence SHH-related phenotypes. *Hum Mol Genet*. 2012; 21(9):1945–53. <https://doi.org/10.1093/hmg/dds004> PMID: 22228099
  37. Fujii J, Otsu K, Zorzato F, de Leon S, Khanna VK, Weiler JE, et al. Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science*. 1991; 253(5018):448–51. PMID: 1862346
  38. Groenen MA, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, et al. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*. 2012; 491(7424):393–8. <https://doi.org/10.1038/nature11622> PMID: 23151582
  39. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81(3):559–75. <https://doi.org/10.1086/519795> PMID: 17701901
  40. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet*. 2007; 81(5):1084–97. <https://doi.org/10.1086/521987> PMID: 17924348
  41. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943
  42. Fritz S, Capitan A, Djari A, Rodriguez SC, Barbat A, Baur A, et al. Detection of haplotypes associated with prenatal death in dairy cattle and identification of deleterious mutations in GART, SHBG and SLC37A2. *PLoS One*. 2013; 8(6):e65550. <https://doi.org/10.1371/journal.pone.0065550> PMID: 23762392

43. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome research*. 2007; 17(11):1665–74. <https://doi.org/10.1101/gr.6861907> PMID: 17921354
44. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011; 12:2825–30.
45. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168
46. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*. 2010; 20(9):1297–303. <https://doi.org/10.1101/gr.107524.110> PMID: 20644199
47. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *arXiv*. 2012.
48. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009; 4(7):1073–81. <https://doi.org/10.1038/nprot.2009.86> PMID: 19561590
49. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol*. 2014; 15(6):R84. <https://doi.org/10.1186/gb-2014-15-6-r84> PMID: 24970577
50. Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH, JBrowse: a next-generation genome browser. *Genome research*. 2009; 19(9):1630–8. <https://doi.org/10.1101/gr.094607.109> PMID: 19570905
51. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013; 29(1):15–21. <https://doi.org/10.1093/bioinformatics/bts635> PMID: 23104886
52. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012; 7(3):562–78. <https://doi.org/10.1038/nprot.2012.016> PMID: 22383036
53. van de Geijn B, McVicker G, Gilad Y, Pritchard JK. WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat Methods*. 2015; 12(11):1061–3. <https://doi.org/10.1038/nmeth.3582> PMID: 26366987
54. Garrick DJ, Taylor JF, Fernando RL. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet Sel Evol*. 2009; 41:55. <https://doi.org/10.1186/1297-9686-41-55> PMID: 20043827
55. Tier B, Meyer K. Approximating prediction error covariances among additive genetic effects within animals in multiple-trait and random regression models *Journal of Animal Breeding and Genetics*. 2004; 121(2):77–89.
56. Gilmour A, Gogel B, Cullis B, Thompson R. ASReml User Guide. 2009.
57. Wright S. Evolution in Mendelian Populations. *Genetics*. 1931; 16(2):97–159. PMID: 17246615