Harnessing genomic information for livestock improvement

Michel Georges[1,2]*, *Carole Charlier*[1,2] *and Ben Hayes*[3]

Abstract | The world demand for animal-based food products is anticipated to increase by 70% by 2050. Meeting this demand in a way that has a minimal impact on the environment will require the implementation of advanced technologies, and methods to improve the genetic quality of livestock are expected to play a large part. Over the past 10 years, genomic selection has been introduced in several major livestock species and has more than doubled genetic progress in some. However, additional improvements are required. Genomic information of increasing complexity (including genomic, epigenomic, transcriptomic and microbiome data), combined with technological advances for its cost-effective collection and use, will make a major contribution.

Within-breed selection
A process by which sires and dams that have above average breeding values are selected as parents to produce the next generation of animals.

Genetic gains
Differences in the average breeding values of the population before and after selection. Genetic gain is a function of the amount of genetic variance, the accuracy of selection, the intensity of selection and the generation interval.

[1]*Unit of Animal Genomics, GIGA Institute, University of Liège, Liège, Belgium.*

[2]*Faculty of Veterinary Medicine, University of Liège, Liège, Belgium.*

[3]*Queensland Alliance for Agriculture and Food Innovation (QAAFI), Queensland Bioscience Precinct, The University of Queensland, Brisbane, Queensland, Australia.*

**e-mail: michel.georges@uliege.be*

https://doi.org/10.1038/s41576-018-0082-2

Since 1960, global livestock productivity (including carcass weight of meat-producing species, milk yield of dairy cows and egg production) has increased by 20–30% as a result of advances in nutrition, disease control and genetics[1]. Genetic improvement has accrued through breed substitution, cross-breeding and within-breed selection. In contrast to the one-off measures of breed substitution and cross-breeding, within-breed selection drives sustained, cumulative genetic progress. It has increasingly relied on sophisticated statistical methods, including mixed model methodology, to provide ever more accurate individual estimated breeding values (EBVs)[2]. Spectacular genetic improvements have been achieved in several species by combining within-breed selection with reproductive technologies (such as artificial insemination and embryo transfer) to more effectively disseminate elite genomes. For example, average annual milk yield per cow in the United States increased from 1,890 kg in 1924 to 9,682 kg in 2011, and more than 50% of this progress was attributed to improved genetics[1]. Between 1957 and 2001, the time for broiler chickens to reach market weight decreased threefold despite a decrease in feed consumption[3,4]. Typically, within-breed selection is expected to result in annual genetic gains of ~1–3%[1].

Currently, the most effective route to minimize the detrimental environmental impact of livestock is to increase productivity: the carbon footprint of 1 kg of milk produced in the United States in 2007 was 37% of that in 1944, and the carbon footprint of the total US dairy industry was reduced by 41% over the same period despite the 250% increase in total milk production[5]. However, exclusive emphasis on production has led to detrimental correlated responses in other traits, particularly those associated with fitness. For example, although selection for milk production in dairy cattle was extremely successful, there was a substantial undesired decline in fertility over the same period[6]. Thus, selection schemes now increasingly attempt to balance animal health, fertility, production and environmental impact.

The emergence of genomics as a discipline in the 1980s led to the concept of marker-assisted selection (MAS), in which genetic variants and genes that influence agriculturally important traits would be identified and used to further increase genetic response. A global chase for quantitative trait loci (QTL) ensued in all livestock species. QTL with large effects on economically important traits were indeed mapped, first by linkage analyses and then by genome-wide association studies (GWAS), but these did not account for a large enough proportion of the heritability to render them useful selection tools on their own[7]. MAS met with limited enthusiasm from the breeding industry until a landmark paper proposed genomic selection (GS)[8]. In its simplest form, GS makes the same assumption as standard EBV selection: the genetic variance for the traits of interest reflects the additive effects of thousands of variants with very small (and, unlike QTL, unmappable) effects that are uniformly scattered throughout the genome[2,9]. As soon as genome-wide single-nucleotide polymorphism arrays (SNP arrays) became an affordable reality, GS was tested and was soon widely adopted by the dairy cattle breeding industry as an effective and easily implemented alternative to the time-consuming and costly standard progeny testing (PT). Since 2008, more than 3 million dairy animals have been genotyped worldwide, and GS has become an essential tool for breeding companies that is expected to double genetic progress[10].

Quantitative trait loci
(QTL). Regions in the genome that encompass genetic variants with an effect on a quantitative trait of interest.

Genome-wide association studies
(GWAS). Scan of the entire genome to identify genetic variants for which variation in genotype is associated with variation for one or more phenotypes of interest.

Genomic selection
(GS). An ensemble of methods to estimate the breeding values of individual animals on the basis of genome-wide single-nucleotide polymorphism genotype information.

Single-nucleotide polymorphism arrays
(SNP arrays). Microarrays used to determine the genotype of individuals for hundreds to millions of SNPs at once.

Progeny testing
(PT). An approach by which the breeding value of an animal is estimated from phenotypic measures made on its progeny.

Genetic architecture
The description of the number, location and effects of the genetic variants that affect a phenotype of interest.

GS is increasingly being adopted by other livestock industries and in plant breeding[11,12]. Similar methods are now also used in human genetics to study the genetic architecture of common complex diseases and to predict individual disease risk[13–15]. Although GS as implemented today is expected to enable genetic progress of up to twofold in dairy cattle and layer hens, with more modest gains in other species, it is unlikely to be sufficient to meet the expected 70% increase in the world demand of animal products by 2050 (REF.[16]). Further improvements and additions to GS will be needed to meet this target.

In this Review, we examine the status of genomic resources available in the major livestock species (that is, cattle, sheep, goat, pig, poultry and salmon) and how these are being used to accelerate the discovery and management of defect-causing genes, to improve the accuracy and extend the scope of GS, to orient genome editing strategies and to develop new applications that take advantage of the genomic information that is becoming widely available.

Genomic resources for livestock species

New scaffolding methods have dramatically improved livestock reference genomes. Following the lead of the human and mouse genome projects, the animal genomics community generated draft reference genomes for the major livestock species (poultry[17], cattle[18], pig[19], goat[20], sheep[21] and salmon[22]), first using Sanger sequencing with hybrid (clone-by-clone and whole genome) shotgun approaches[23], increasingly complemented with massively parallel generation of short reads. These efforts provided initial insights into the evolution of the gene repertoire underlying adaptive features (such as plumage and beak formation, rumination and lactation, wool growth, and smell and taste specification) and into changes resulting from whole-genome duplication in salmon. They also contributed to the identification of evolutionary conserved elements[24]. However, the quality of most of these reference genomes has remained a source of concern. They were highly fragmented and littered with assembly errors, which affect positional cloning efforts and imputation accuracy, among other uses[25]. Critical mass and funding have long been missing to upgrade their status from highly fragmented drafts to high-quality finished genomes. However, the development of new scaffolding approaches, including long-read sequences (such as PacBio), optical mapping (such as Bionano Genomics) and chromatin conformation capture now provides an affordable path to high-quality reference genomes for all species[26]. The integrated use of these methods has recently enabled spectacular improvements in the quality of reference genomes for goat[27] and other livestock species (TABLE 1) (genome assemblies are available through the NCBI Genome database).

Genome-wide SNP arrays are available for the main livestock species. Draft reference genomes were typically accompanied by shallow (1–2-fold depth) sequencing of tens to hundreds of individuals representing distinct breeds and populations in order to characterize genetic variation and infer demographic history (see, for example, REFS[19,28,29]). These efforts have uncovered millions of genetic variants for all the main livestock species and profoundly changed our understanding of the domestication process (BOX 1). Databases of available SNPs (such as dbSNP) have been used to develop a large number of arrays that allow cost-effective genotyping of tens of thousands to hundreds of thousands of variants in the major livestock species (Supplementary Table 1). These arrays are extensively used to conduct GWAS, as well as GS. It is estimated that at least 3 million cattle and possibly millions of pigs and poultry have been genotyped using genome-wide SNP arrays[10,11].

Population-based resequencing for imputation-based GWAS and GS. As sequencing costs continue to decrease, livestock geneticists are resequencing the genomes of a growing number of animals. More than 2,500 cattle have had their whole genome resequenced, while the corresponding numbers are at least in the hundreds for pig, poultry, sheep and goats (M. Groenen, R. Hawken and G. Tosser-Klopp, personal communications). The best-known large-scale resequencing initiative in livestock genetics is the 1,000 Bull Genomes Project[25], but other large sequencing projects are being conducted by academic groups and breeding companies[30]. These efforts are largely inspired by the human 1,000 Genomes Project[31] and hope to achieve deep characterization of the genetic variation between and within populations. Importantly, sequencing the whole genomes of a reference population of hundreds of animals enables genotype imputation at millions of common variants in the much larger number of animals that have been genotyped with genome-wide SNP arrays. This approach can be

Table 1 | **Current status of the reference genomes for the most important livestock species**

Species	Assembly	Release date	Coverage	Number of contigs	Contig N50 (Mb)	Total (Gb)
Pig (*Sus scrofa*)	Sscrofa11.1	7 Feb 2017	65×	1,118	48.2	2.5
Goat (*Capra hircus*)	ARS1	24 Aug 2016	50×	30,399	26.2	2.9
Cattle (*Bos taurus*)	ARS-UCD1.2	11 Apr 2018	80×	2,597	25.9	2.7
Chicken (*Gallus gallus*)	GRCg6a	27 Mar 2018	82×	1,402	17.6	1
Sheep (*Ovis aries*)	Oar_rambouillet_V1.0	2 Nov 2017	126×	7,485	2.6	2.9
Zebu (*Bos indicus*)	AM293397v1	22 Feb 2018	100×	337,292	0.064	2.7
Salmon (*Salmo salar*)	ICSASG_v2	10 Jun 2015	206×	368,060	0.058	3

Box 1 | **Genetic variation provides insight into the domestication process**

Punctuated versus continuous domestication. One of the most striking insights gained from studying genetic variation in livestock species is the realization that the degree of genetic variation, measured, for instance, by the average heterozygosity per nucleotide site (π), is typically higher in livestock than in humans[19,28,29]. This observation is against expectations. It is often assumed that animal domestication occurred through rare, isolated events involving a limited number of animals, which would have caused drastic genetic bottlenecks. Further reduction in effective population size would have accompanied more recent breed creation and been accentuated by intensifying selection schemes. However, domestic animal populations remain more variable than the people who domesticated them. This realization forces us to revisit our views of the domestication process. Domestication most likely involved continuous gene flow between domestic and wild individuals from the same species, as well as from interfertile sub-species, during most of agricultural history[191]. As a result, the genomes of the majority of domestic livestock species probably have a mosaic structure that is at least as pronounced as that of the laboratory mouse[192] or human[193]. Some haplotypes segregating within pig and cattle breeds have been shown to differ approximately every 100 bp, which is a similar sequence identity to humans and chimpanzees; thus, they possibly coalesced ~5 million years ago, which is before the creation of the studied species[118,194,195].

Hard sweeps, soft sweeps and polygenic adaptation during domestication. Comparisons between the genome sequences of domestic animals and their wild extant or extinct progenitors (that is, red jungle fowl[196], rabbit[197], wild boar[191], bezoar[198] and auroch[199]) have identified chromosome regions that may have undergone hard sweeps driven by the domestication process. These regions seem to be enriched in genes that control behaviour and stature. The most convincing 40 kb hard sweep signature encompasses the G558R missense mutation in the chicken thyroid stimulating hormone receptor (TSHR), known to have a key role in metabolic regulation and photoperiod control of reproduction[196]. These genomic regions may correspond to islands of domestication that resist recurrent gene flow from wild progenitor species[191]. It is worth noting that the methods used to detect selective sweeps associated with domestication pick up only hard sweeps acting on very rare or de novo mutations. Soft sweeps acting on older and hence more common mutations (that is, standing variation in the wild progenitor) require alternative methods for their detection[200]. It is also noteworthy that evidence suggests that tame behaviour in rabbits and possibly other species evolved through shifts in allelic frequency at many loci (that is, polygenic adaptation) rather than critical changes at a few domestication loci[197].

Genotype imputation
The in silico prediction of the genotype of an individual for ungenotyped variants on the basis of known genotypes at neighbouring variants and a reference population with genotype information for all variants. Imputation exploits the nonrandom association of alleles at neighbouring variants, referred to as linkage disequilibrium.

Soft sweeps
The process by which the frequency of a favourable old variant rapidly increases in the population by positive selection until eventual fixation. Soft sweeps are not associated with the concomitant fixation of one predominant haplotype, as the variant has been distributed over multiple haplotypes by recombination before selection. Old variants that are substrates for new selection constitute the standing variation in the population.

implemented using pyramidal schemes in which a top layer of a few (possibly hundreds) highly influential animals are sequenced, an intermediate layer of 'multiplier' animals are genotyped with high-density SNP arrays and the most populated bottom layer of animals are genotyped with low-density SNP arrays. Sequence information is then projected from the upper two layers onto the animals of the bottom layer using a two-step imputation strategy[32]. Livestock represent a unique opportunity to implement this approach because samples from key ancestors of the population are often available in the form of semen straws or ampules. In the 1,000 Bull Genomes Project, bulls born in the 1960s are included in the set of sequenced animals. The availability of whole-genome sequence information for tens to hundreds of animals from specific breeds has proved extremely useful to pinpoint the causative mutations underlying monogenic defects[25,33]. Imputation of sequence information on large cohorts of phenotyped animals greatly accelerates fine-mapping and identification of causative variants for QTL detected by GWAS[25,34]. It is also anticipated that imputed sequence information could increase the accuracy of GS (see Increasing the accuracy of GS using whole-genome sequence imputation).

Epigenome maps and eQTL data sets enable functional follow-up of GWAS hits. It is increasingly recognized that regulatory (rather than coding) variants account for the majority of the genetic variation underlying complex traits, such as common complex diseases in humans or economically important traits in plants and animals[35,36]. Most of these regulatory variants are expected to affect components of gene switches, that is, proximal promoters and more distant enhancers and silencers. To aid in the identification of such regulatory variants and the genes whose expression they affect, the animal genomics community has begun to generate epigenome maps, mainly using ChIP-Seq (chromatin immunoprecipitation followed by sequencing), DNase-Seq (DNase I hypersensitive site sequencing) and ATAC-Seq (assay for transposase-accessible chromatin using sequencing), which will provide exhaustive catalogues of gene regulatory elements in livestock. Most of these efforts are coordinated through the international Functional Annotation of Animal Genomes (FAANG) project[37,38]. Liver-specific comparative enhancer maps based on histone modification data have already been generated for 20 mammals, including cow, pig and rabbit[39]. In addition, bovine DNA methylation maps have been generated for ten somatic tissues using reduced representation bisulfite sequencing[40].

Epigenome maps are complemented by multi-tissue transcriptome data sets for the analysis of expression quantitative trait loci (eQTL). In cattle, such data sets have been generated for mammary gland, liver, blood and adrenal gland and have been used to identify causative genes underlying GWAS-identified QTL[41–45]. In pigs, eQTL studies have been conducted in skeletal muscle, lung, adipose tissue and liver[46–57]. In poultry, genome-wide eQTL analyses have been reported for liver, bone, adrenal gland and hypothalamus[58–61]. The time seems right for the animal genomics community to take advantage of working with livestock species to collaboratively generate large, multi-omic, multi-tissue data sets similar to the human Genotype-Tissue Expression (GTEx) data set[62]. This approach would provide invaluable comparative information about genome function and greatly facilitate follow-up studies of GWAS and GS hits in these species.

Important Mendelian traits in livestock

The early 20th century saw a heated debate between Mendelists and Galtonists, with Galtonists claiming that Mendelian genes accounted for only a small proportion of inherited features. The debate was settled when it was realized that quantitative traits derive their continuous distribution from the combined effects of many segregating Mendelian genes (that is, they are polygenic traits). It remains true, however, that Mendelian traits — that is, phenotypes that are fully determined by one gene (monogenic) or a small number of genes (oligogenic) — are the exception rather than the rule. In humans, Mendelian traits are largely limited to blood groups and an admittedly long list of severe genetic defects that are compiled in the Online Mendelian Inheritance in Man (OMIM) database and that include the 'inborn errors of metabolism'. In addition to blood

Epigenome
The combination of chemical modifications of the DNA sequence (such as cytosine methylation) or nucleosomes (such as methylation of Lys 27 of histone H3) that mark functionally distinct segments of the genome (such as active enhancers) and are inherited mitotically and/or meiotically.

ChIP-Seq
A combination of chromatin immunoprecipitation and next-generation sequencing for genome-wide mapping of binding sites occupied by specific DNA-binding proteins or chromatin regions enriched in specific histone modifications.

DNase-Seq
A method based on next-generation sequencing for genome-wide detection of gene-switch components on the basis of their open chromatin conformation and resulting hypersensitivity to digestion by DNase I.

ATAC-Seq
An assay based on next-generation sequencing for genome-wide detection of gene-switch components on the basis of their open chromatin conformation and resulting increased accessibility to transposase Tn5.

Expression quantitative trait loci
(eQTL). Quantitative trait loci that influence the transcript levels of specific genes. *Cis*-eQTL are due to regulatory variants that control the levels of RNA molecules transcribed from gene copies located on the same DNA molecule as the variant. *Trans*-eQTL are due to regulatory variants that can also control the levels of RNA molecules transcribed from gene copies located on different DNA molecules to the variant (homologous or other chromosomes).

Pleiotropy
The ability of a genetic variant to affect more than one phenotype.

Hypomorphic
Pertaining to an allele with partial loss of function when compared with the wild-type allele.

groups and a similar list of severe genetic defects compiled in the Online Mendelian Inheritance in Animals (OMIA) database, Mendelian traits in domestic animals also include an extended list of breed-defining characteristics, such as coat colour, tegument variation, polledness, double-muscling and hyper-prolificacy.

Most breed-defining traits have been molecularly characterized. For millennia, animal breeders have performed what amounts to a mega-scale phenotype-driven mutagenesis screen. In the process, they have identified a series of mutations with large phenotypic effects that — when desirable — were selected, often becoming trademarks and breed-defining features. In many instances, mutant variants were valued because of their aesthetic effects on the animals, such as patterns of coat and plumage colour; shape of ears, horns, wattles or combs; and tonality of songs. The long-standing interest of breeders for 'fancy' animals is well illustrated by 7,000-year-old rock paintings in the Sahara[63]. In other instances, the value of the mutant variants reflects their utility. For instance, mutations with major beneficial effects on hair (such as quality of angora or cashmere) and skin texture (such as heat tolerance of slick cattle), fertility (such as twinning) and muscularity (for example, double-muscling) are all highly desired. Although some mutant phenotypes are easily recognized, others are subtler and may require human–animal proximity for their detection. An example of such a phenotype is pacing in horses, which was shown recently to result from a premature stop codon in *DMRT3*, a gene that controls spinal circuitry[64]. It is hard to imagine that such a phenotype could be detected in the systematic phenotype-driven screens that are currently being conducted in the mouse.

Over the past 10 years, as genomic resources and methods improved, the causative genes and mutations underlying most of these breed-defining characteristics have been identified, and a number of dominant themes have emerged[65] (TABLE 2). Most (75%) of the corresponding mutations are at least partially dominant, that is, heterozygotes express a phenotype; such mutations would have been easier to detect and maintain in the population than recessive ones. A large proportion of mutations affect gene regulation (43%), resulting in gain-of-function phenotypes through ectopic gene expression. Regulatory mutations are often structural (64%) and involve duplications, insertions (including of retroelements), inversions or combinations thereof. The same phenotype is often determined by mutations in the same gene in different species and by allelic series within species, which indicates that mutations of only that gene can generate the corresponding phenotype without major deleterious pleiotropy. Different phenotypes are sometimes caused by allelic series that have evolved one from the other by serial accumulation of multiple mutations.

The molecular dissection of breed-defining traits has revealed some remarkable biology, including the demonstration of serial translocation by circular intermediates, which is likely to be an ancient exon shuffling mechanism[66], the identification of a hypomorphic *MSTN* mutation resulting from the acquisition of an illegitimate microRNA target site[67] and the interplay between *cis*-effects and microRNA-mediated *trans*-effects underlying polar overdominance of the callipyge phenotype in sheep[68]. However, it is noteworthy that the molecular underpinnings of the cashmere and mohair wool types in goat and the very widespread recessive piebald phenotype in cattle remain unknown.

About the number of defect-causing recessive mutations carried per individual. Diploidy has enabled an increase in genome size while ensuring that most individuals in the population have at least one functional copy of each gene. Concomitantly, most individuals are expected to be heterozygous for loss-of-function (LoF) alleles in a number of haplosufficient genes. The analysis of whole-genome sequences from large numbers of individuals indicates that this number is ~100 in humans[69]. It appears to be very similar in livestock species, including the cow[30]. This estimate is much higher than expected from epidemiological studies, which suggest that humans carry on average ~0.5–1 allele that is lethal when homozygous[70]. This apparent conundrum can be explained by the observation that for the majority of genes (~75%), homozygosity for LoF mutations is viable but confers a modest selective disadvantage that is sufficient to preclude fixation of the mutations, which explains the evolutionary conservation of the corresponding gene. Indeed, data from the International Mouse Phenotype Consortium indicate that only ~25% of mammalian genes are essential in the sense that at least one functional allele is needed for survival until reproductive age. Homozygosity (or compound heterozygosity) for LoF mutations in the corresponding genes are lethal, either before (embryonic lethal (EL)) or after birth. The number of such recessive lethal alleles that are carried, on average, by healthy individuals has been of considerable interest for a long time[71]. Indeed, this number determines, for instance, the increased morbidity endured by offspring of consanguineous marriages or matings. Simulations for mammalian genomes[30] suggest that this number increases with effective population size (N_e), from ~0.5 for $N_e = 100$ (which is the N_e for many livestock populations) to ~5 for $N_e = 10{,}000$ (which is the N_e of the human population). Approximately 1% (independent of N_e) of conceptuses succumbs from homozygosity (or compound heterozygosity) for at least one of around ten common EL mutations (frequency >0.02) in livestock compared with at least one among thousands of rare EL mutations in humans[30]. This observation suggests that managing severe genetic defects in livestock populations (including EL mutations) is a tractable problem that requires the identification and tracking of around ten such common mutations per population. It is noteworthy that in humans (and probably in other mammals), an estimated 3,000 genes are haploinsufficient and hence LoF-intolerant[72].

Identifying causative mutations for recessive defects has become trivial. Livestock populations are characterized by recurrent outbursts of genetic defects. This is particularly true for species such as cattle in which artificial insemination allows elite sires to have tens

Table 2 | **Breed-defining traits that have been characterized at the molecular level**

Gene	Species	Transmission	Phenotype	Mutations	
				Coding	Regulatory
Coat or feather colour: melanocyte development					
KIT	Bovine	dom	Colour-sided	–	DUPC6[a,b], DUPC29[a,b]
	Bovine	1/2 dom	Degree of white spotting	–	Unknown
	Pig	dom	White	(SS …	… + DUP1[a] + DUP2–4[a])[b]
	Pig	dom	Patch	–	DUP1[a,b]
	Pig	dom	Belt	–	(DUP2–4)[a,b]
	Pig	codom	Roan	SS	–
KITLG	Bovine	codom	Roan	MS	–
	Goat	codom	Roan	Unknown	–
MIFT	Bovine	dom	White, blue eyes, hearing loss	MS	3bpDEL
	Bovine	1/2 dom	Degree of white spotting	–	Unknown
SOX10	Chicken	rec	Dark brown	–	DEL[a]
TWIST2	Bovine	dom	White belt	–	QUAD[a]
CDKN2A	Chicken	Z-linked dom	Extreme dilution	–	(SNP1–2)[b]
	Chicken	Z-linked 1/2 dom	Dilution	(MS1 …	… + SNP1–2)[b]
	Chicken	Z-linked dom	Barring	(MS2 …	… + SNP1–2)[b]
EDNRA	Goat	1/2 dom	Degree of white spotting	(MS …	… + CNV[a])
EDNRB2	Chicken	rec	Mottled	MS	–
	Chicken	rec	White	MS	–
Coat or feather colour: melanin synthesis					
MC1R	Bovine	dom	Black	MS	–
	Bovine	rec	Red	FS	–
	Bovine	rec	Telstar	–	Unknown
	Pig	dom	Black	MS1, MS2	–
	Pig	rec	Red	MS3 + MS4	–
	Pig	–	Coat-colour diversity	MS1–8, FS1	–
	Pig	som	Black spotting	(FS + MS2)	–
	Sheep	dom	Black	MS1, MS2	–
	Goat	dom	Black	MS	–
	Goat	rec	Red	SG	–
	Chicken	dom	Extended black	MS1[b]	–
	Chicken	rec	Buttercup	(MS1 + MS2)[b]	–
ASIP	Bovine	1/2 dom	Brindle	–	INS (LINE)[a]
	Sheep	dom	White or tan	–	DUP[a]
	Sheep	rec	Self-colour black	FS, 9bpDEL, MS	–
TYR	Bovine	rec	Albino	FS	–
	Chicken	rec	Albino	6bpDEL	–
	Chicken	rec	White	–	INS (ERV)[a]
TYRP1	Bovine	rec	Dun	MS	–
	Pig	1/2 dom	Brown or blond	6bpDEL	–
	Sheep	rec	Light coat	MS	–
	Goat	dom	Brown	MS	–
Coat or feather colour: melanin transport					
PMEL	Bovine	rec	Dilution	MS, 3bpDEL	–
	Chicken	dom	White	9bpINS[b]	–
	Chicken	dom	Smokey	(9bpINS + 12bpDEL)[b]	–
	Chicken	dom	Dun	15bpDEL	–

Table 2 (cont.) | **Breed-defining traits that have been characterized at the molecular level**

Gene	Species	Transmission	Phenotype	Mutations	
				Coding	Regulatory
Coat or feather colour: melanin transport (cont.)					
MLPH	Bovine	rec	Cool grey	FS	–
	Chicken	rec	Lavender	MS1	–
	Chicken	rec	Grey dilution	MS2	–
COPA	Bovine	dom	Red	MS	–
SLC45A2	Bovine	rec	Oculocutaneous albinism	MS1, MS2	–
	Chicken	Z-linked rec	Silver	MS1, MS2	–
Muscle mass					
MSTN	Bovine	1/2 dom	Double-muscling	MS1–3, FS1–2, SG1–3, SS1	–
	Sheep	1/2 dom	Double-muscling	FS1, FS2	3′ UTR SNP
DLK1 + *PEG11*	Sheep	Polar overdominance	Callipyge	–	IG SNP
Dermal features: horns					
BTA1 (1.7–1.9 Mb)	Bovine	dom	Polled	–	INDEL1[a], INDEL2[a], DUP1[a]
TWIST1	Bovine	dom	Scurs type 2	FS	–
RXFP2	Sheep	dom	Polled	–	3′ UTR INS (PG)[a]
	Sheep	1/2 dom, sex-dep	Horn size	–	3′ UTR SNP
FOXL2	Goat	dom	Polled	–	DEL[a]
Dermal features: hair, feather or shell structure					
KRT27	Bovine	dom	Curly coat	MS	–
PRLR	Bovine	dom	Slick coat	SG	–
	Chicken	Z-linked dom	Slow feathering	–	DUP[a] + INS[a] (ERV)
IRFBP2	Sheep	rec	Short and woolly fleece	–	3′ UTR INS (PG)[a]
FMN1/GREM1	Goat	dom	Wattles	Unknown	–
PPARD	Pig	1/2 dom	Large and floppy ears	MS	–
SLCO1B3	Chicken	1/2 dom	Blue eggshell	–	INS1 (ERV)[a], INS2 (ERV)[a]
KRT75	Chicken	1/2 dom	Frizzled feathering	69bpDEL	–
CYP19A1	Chicken	dom, sex-dep	Henny feathering	–	INS (ERV)[a]
PDSS2	Chicken	rec	Silk feathering	–	SNP
HOXB8	Chicken	1/2 dom	Muffs and beard	–	CNV[a]
GDF7	Chicken	1/2 dom	Naked neck	–	INS[a]
FGF20	Chicken	rec	Featherless or scaleless	SG	–
Dermal features: comb					
SOX5	Chicken	dom	Pea comb	–	DUP[a]
MNR2	Chicken	dom	Rose comb	–	Complex CNV[a]
SOX5 + *MNR2*	Chicken	epistasis	Walnut comb	–	(*SOX5* DUP[a] + *MNR2* CNV[a])
EOMES	Chicken	1/2 dom	Duplex comb	–	DUP[a]
Dermal features: pigmentation					
BCO2	Sheep	rec	Yellow fat	SG	–
	Chicken	rec	Yellow skin and shank colour	SG	–
EDN3	Chicken	dom	Silkie and dermal hyperpigmentation	–	Complex CNV[a]

Table 2 (cont.) | **Breed-defining traits that have been characterized at the molecular level**

Gene	Species	Transmission	Phenotype	Mutations	
				Coding	Regulatory
Fertility					
BMP15	Sheep	X-linked 1/2 dom	Prolificity	MS1–5, SG1–2, FS1–2	–
BMPR1B	Sheep	1/2 dom	Prolificity	MS	–
GDF9	Sheep	1/2 dom	Prolificity	MS1–5	–
B4GALNT2	Sheep	1/2 dom	Prolificity	–	(SNP1 + SNP2)

Multiple mutations constituting one allele are bracketed. CNV, copy number variant; codom, co-dominant; DEL, deletion; dom, dominant; DUP, duplication; ERV, endogenous retrovirus; FS, frameshift; IG, intergenic; INDEL, insertion and/or deletion; INS, insertion; LINE, long interspersed nuclear element; MS, missense; PG, pseudogene; QUAD, quadruplication; rec, recessive; sex-dep, sex-dependent; SG, stop gain; SNP, single-nucleotide polymorphism; som, somatic; SS, splice site; UTR, untranslated region; xbpINS, *x* bp insertion; ybpDEL, *y* bp deletion. [a]Structural variants. [b]Allelic series involving the serial accumulation of multiple mutations.

Overdominance
The phenotypic superiority (for example, on a quantitative scale) of heterozygotes ('Aa') over both homozygous classes ('AA' and 'aa').

Haplosufficient
Pertaining to genes for which one functional copy is sufficient to ensure normal development and function.

Compound heterozygosity
Pertaining to the inheritance of two distinct mutations in different alleles of the same gene, one from each parent.

Autozygosity mapping
Mapping of a recessive mutation on the basis that all affected individuals will be homozygous for the same (autozygous) haplotype. Typically applied in genetically isolated populations in which the hypothesis of allelic homogeneity is reasonable.

Modifier locus
A locus with variants that may (depending on the genotype of the individual) affect the phenotypic expression conferred by specific variants at another locus. The effects of modifier loci include suppression and epistasis.

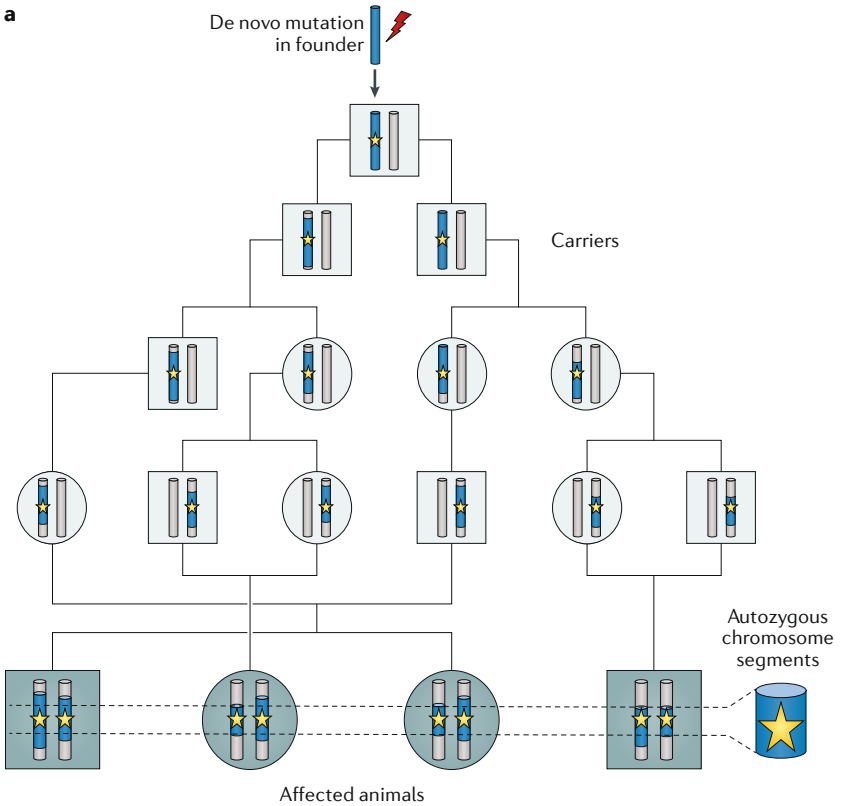
Reverse genetic screens
Process aimed at completing the phenotype–genotype map by sorting individuals according to their genotype at a variant with unknown function and searching for shared phenotypes, as opposed to forward genetics, which consists of sorting individuals according to a phenotype and searching for shared variants.

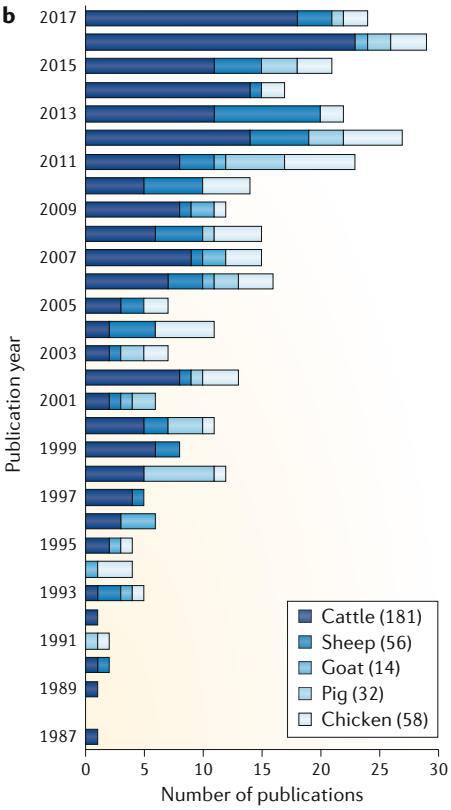
of thousands to hundreds of thousands of offspring, thereby disseminating their deleterious mutations in the general population and leading to the emergence of affected individuals in subsequent generations. Before the advent of genomics, determining whether an elite sire carried such a common defect required time-consuming and expensive PT, that is, mating the sire with affected dams when possible (such as for mule-foot)[73] or to known carriers or daughters and verifying the occurrence of affected offspring. In the 1980s, it became possible to positionally clone the responsible gene, with the aim to develop a diagnostic marker. However, this was a very tedious multi-year effort that often required the generation of informative pedigrees to prove the genetic nature of the condition (which was not always obvious otherwise) and allow linkage mapping. This situation changed dramatically when genome-wide SNP arrays became available, which enabled effective autozygosity mapping using only a few affected individuals[74]. More recently, whole-genome resequencing of small numbers of affected individuals has enabled autozygosity mapping and, in approximately half of cases, simultaneous identification of the causative mutations in a matter of days, especially when taking advantage of sequenced reference populations such as the 1,000 Bull Genomes Project[25] (FIG. 1a). An increasingly common approach for the management of recessive defects in livestock is to systematically collect samples from animals with severe defects and to sequence the whole genome of small sets of animals with similar symptoms as soon as they become available. Identifying an autozygous genomic segment — or better, a causative mutation — shared by all affected individuals confirms the genetic origin of the defect and readily provides a diagnostic test. As a result of these advances, the list of characterized defects in livestock in OMIA has markedly increased in recent years (FIG. 1b).

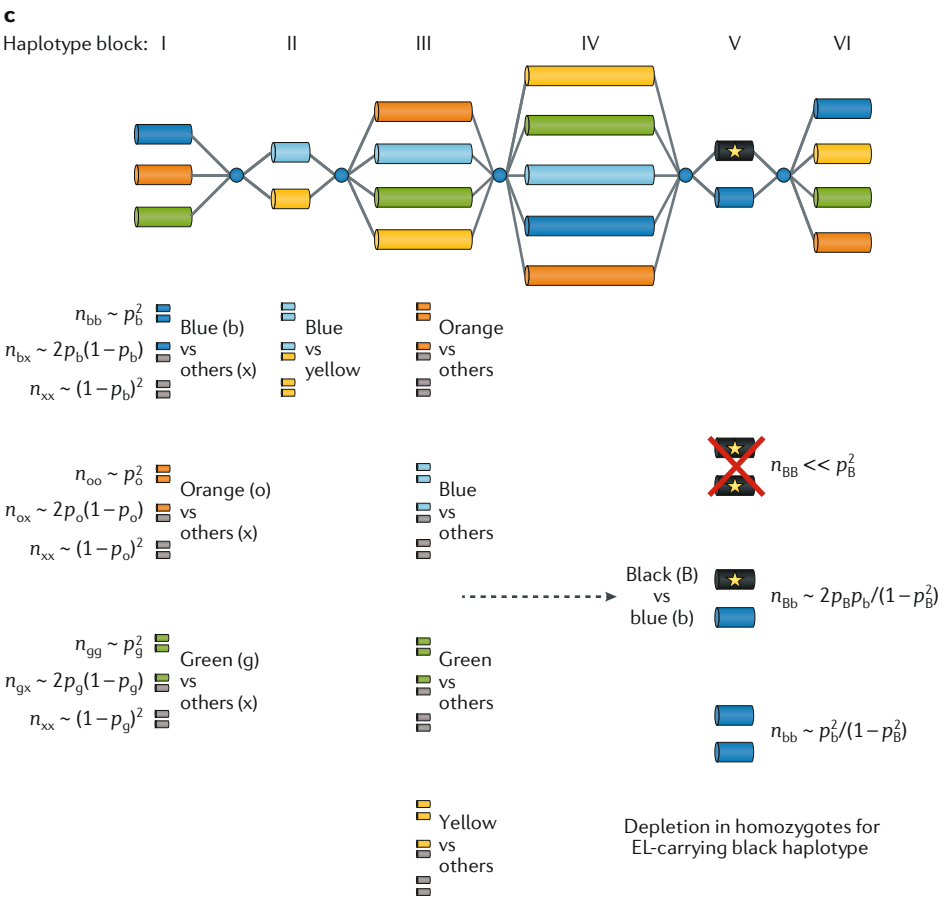
Germline mosaicism for deleterious dominant mutations is common. A number of genetic defects that were assumed to be recessive because of their rarity — only a small number of affected animals were detected among thousands of offspring from a healthy sire — could not be mapped by autozygosity mapping. In several cases, these defects were demonstrated to be caused by dominant de novo mutations (DNMs) for which the sire was germline mosaic[25,33,75,76]. Unlike in humans, who typically produce only one affected offspring, samples were available from a number of affected animals, which greatly facilitated the identification of the causative mutation. In at least one instance, neurocristopathy, it also enabled a modifier locus that affected disease severity to be mapped[33].

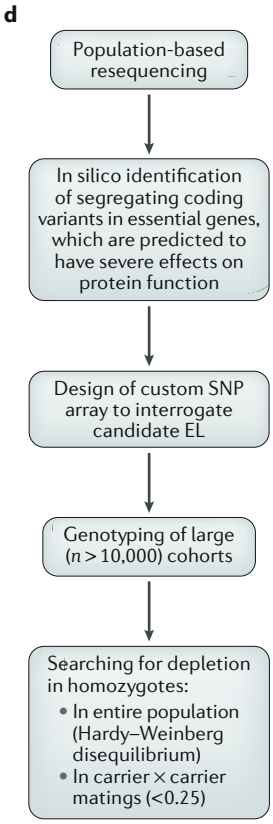
Preliminary studies of the bovine germ line have shown that sires are mosaic for ~30% of DNM present in a sperm cell and that dams are mosaic for ~50% of DNM present in an oocyte[77]. These numbers may be considerably higher than those in humans. Our preliminary data suggest that early embryonic cleavage is 20-fold more mutation-prone than cell divisions occurring later in development[77] and that this increased rate of DNM might be due, in part, to reproductive technologies routinely used in livestock, such as the combination of oocyte pick-up, in vitro oocyte maturation and in vitro oocyte fertilization (C.C., unpublished observations).

Reverse genetic screens for EL mutations that compromise carrier fertility. It is likely that for the majority of LoF variants in essential genes, homozygosity will result in embryonic or fetal lethality rather than a genetic defect manifesting at or after birth. The economic effect of such mutations when carried by influential sires, in terms of lost pregnancies, can run into hundreds of millions of dollars[78,79]. Embryonic lethality is a complex phenotype that is not observed directly and hence is difficult to study. Daughters of sires that carry EL mutations are expected to be less fertile (as measured, for instance, by the probability to return to heat after insemination), but the effect on the sire's EBV is detectable only for EL mutations that have reached high frequencies in the population[80]. Reverse genetic screens have been devised to overcome this issue. First, scientists have mined large genome-wide SNP data sets for haplotypes with considerable autozygous depletion, that is, haplotypes for which no individuals are found to be homozygous despite some being expected (FIG. 1c). At least 17 such haplotypes have been identified in cattle[78,81–83] and 4 in pig[84,85] (Supplementary Table 2). In a number of cases, mining of population-based resequencing data sets in the corresponding genomic regions has identified the causative mutations[25,79,86–88]. However, to be effective, this approach requires that the haplotype and the causative


a
De novo mutation in founder
Carriers
Autozygous chromosome segments
Affected animals


b
Publication year
2017
2015
2013
2011
2009
2007
2005
2003
2001
1999
1997
1995
1993
1991
1989
1987
0
5
10
15
20
25
30
Number of publications
Cattle (181)
Sheep (56)
Goat (14)
Pig (32)
Chicken (58)


c
Haplotype block: I
II
III
IV
V
VI
$n_{bb} \sim p_b^2$
$n_{bx} \sim 2p_b(1-p_b)$
$n_{xx} \sim (1-p_b)^2$
Blue (b) vs others (x)
Blue vs yellow
Orange vs others
$n_{oo} \sim p_o^2$
$n_{ox} \sim 2p_o(1-p_o)$
$n_{xx} \sim (1-p_o)^2$
Orange (o) vs others (x)
Blue vs others
$n_{BB} << p_B^2$
Black (B) vs blue (b)
$n_{Bb} \sim 2p_Bp_b/(1-p_B^2)$
$n_{gg} \sim p_g^2$
$n_{gx} \sim 2p_g(1-p_g)$
$n_{xx} \sim (1-p_g)^2$
Green (g) vs others (x)
Green vs others
$n_{bb} \sim p_b^2/(1-p_B^2)$
Yellow vs others
Depletion in homozygotes for EL-carrying black haplotype


d
Population-based resequencing
In silico identification of segregating coding variants in essential genes, which are predicted to have severe effects on protein function
Design of custom SNP array to interrogate candidate EL
Genotyping of large ($n > 10,000$) cohorts
Searching for depletion in homozygotes:
• In entire population (Hardy–Weinberg disequilibrium)
• In carrier × carrier matings (<0.25)

◀ Fig. 1 | **Identification of mutations and genes causing monogenic defects in livestock. a** | Autozygosity mapping: a defect-causing mutation (yellow star) in a haplosufficient gene appears by de novo mutation on a specific haplotype (labelled in blue) in a founder individual. The mutation spreads in the population as it is transmitted to descendants of the founder. In subsequent matings between healthy carriers, 25% of individuals are affected and are homozygous for the mutation and at least part of the blue haplotype of the founder. **b** | Number of monogenic traits for which the causative mutation was discovered by publication year and livestock species (data compiled from Online Mendelian Inheritance in Animals). **c** | Identification of haplotypes carrying embryonic lethal (EL) mutations: haplotypes are identified within blocks, and a Hardy–Weinberg-based test for their depletion in homozygotes in the population is performed for each haplotype. As the limits between haplotype blocks are poorly defined in livestock, window-based or hidden Markov model-based approaches are used. **d** | Flow chart of the steps involved in a sequence-based reverse genetic screen for EL mutations in livestock. SNP, single-nucleotide polymorphism.

mutation are in near perfect linkage disequilibrium (LD), and this is certainly not always the case. An alternative approach, therefore, is to mine the sequence data for candidate EL mutations (that is, LoF variants in essential genes) and to genotype these directly in large cohorts. The absence of homozygous individuals, particularly in the offspring of carrier-by-carrier matings, strongly supports the EL nature of the corresponding variants. Using this approach, nine EL were uncovered in cattle that would not have been detected using haplotype-based approaches[30] (FIG. 1d; Supplementary Table 2).

Culling carriers of deleterious recessive variants is not the right approach. Once information became available about EL mutations, the spontaneous reaction of breeding companies was to cull breeding animals, particularly sires, that carry known deleterious mutations in order to assure customers that their animals will not have a known recessive defect. However, as the number of identified mutations increases, the proportion of 'mutation-free' animals becomes vanishingly small. Furthermore, excessive use of such mutation-free sires is bound to cause the emergence of novel defects in subsequent generations and hampers genetic progress for other traits. Instead, the costs of the defect should be properly modelled and weighted against the economic value of the other traits in a selection index. Alternatively, the dam population could be genotyped to avoid matings between parents carrying the same deleterious variants (see From selecting animals to selective matings using genomic information), or cross-breeding could be used, as it is less likely that the same EL defects segregate in different breeds.

GS for complex agricultural traits

With the exception of the breed-defining characteristics, inherited defects and EL mutations discussed above, nearly all economically important traits in livestock are complex polygenic traits. They include milk yield and composition (that is, milk protein and fat yield), carcass yield, composition and quality (for example, marbling and tenderness), egg yield, growth rate, feed efficiency, fertility and disease resistance. Although some of these traits are categorical, most are continuously distributed quantitative traits. The phenotypic variation of these traits is assumed to reflect the combined effects of developmental noise, differences in environmental exposure and genotypic differences at a large number of loci scattered throughout the genome. The heritability (that is, the proportion of the trait variance that is caused by genetic factors) of these economically important traits in livestock typically ranges from ~5% to 50%. For a handful of traits, QTL mapping studies and GWAS analyses have identified loci with relatively large effects on the considered trait. However, the joint effects of these detectable loci typically explain only a limited fraction of the overall heritability. The most convincing evidence indicates that the remainder of the heritability is highly polygenic, corresponding to hundreds if not thousands of genetic variants that each has a very small effect on the trait of interest[89]. The contribution of each variant to overall heritability is assumed to be mainly additive, that is, dominance and epistasis account for only a modest fraction of the variance[9].

Before genomic tools made it possible to identify specific loci by mapping experiments, quantitative geneticists made the simplifying assumption that heritability of quantitative traits was determined by an infinitely large number of variants of infinitely small effect that were spread uniformly throughout the genome. This model is known as Fisher's infinitesimal model and is one of the pillars of quantitative genetics theory[2]. The infinitesimal model makes predictions about the expected phenotypic similarity of relatives as a function of their relatedness, underpins the methods used to estimate heritabilities and breeding values (BVs) of individuals in pedigreed populations and guides the design of selection schemes in plant and animal breeding. The proven efficacy of plant and animal breeding and the good fit of genetic progress to predictions based on the infinitesimal model[89] strongly suggest that a sizeable proportion of the genetic architecture of most quantitative traits can be approximated reasonably well by the infinitesimal model.

GS under the infinitesimal model. One of the key factors that determines the rate of genetic progress in a breeding programme is the accuracy of selection, that is, the accuracy with which animals with the best BVs are chosen as parents of the next generation. Among the methodological breakthroughs that have contributed most to improving the accuracy of selection is the development and use of mixed models to estimate BVs[90]. In these linear models, known environmental factors are fitted together with the BVs of individual animals, which are fitted as random effects. In attempting to find the best solutions for the random BVs, prior knowledge of their joint distribution is used. For instance, because full-sibs or parents and their offspring are closely related, their EBVs are expected to be more similar than those of unrelated animals. The infinitesimal model makes precise predictions about the shape of this joint distribution: BVs are normally distributed, and the covariance between the BVs of two animals is a simple function of their relatedness (that is, their kinship coefficient)[91,92] and the genetic variance of the trait. BVs for millions of animals can be simultaneously estimated from corresponding sets of linear equations, yielding so-called best linear unbiased predictors (BLUPs) of the EBVs (BOX 2).

Haplotypes
A combination of alleles at multiple variant positions transmitted by a gamete. The term is often used to describe variants that are located close to each other in the genome.

Linkage disequilibrium
(LD). The nonrandom association of alleles at two or more loci, which is manifest by the over-representation of specific haplotypes and the concomitant under-representation of others.

Selection index
A weighted sum of breeding values for several traits, each weighted by economic or perceived relevance.

Kinship coefficient
A measure of genetic relatedness between two individuals. The kinship coefficient corresponds to the probability that two alleles (one from each individual) drawn at random from the two possible alleles (maternal and paternal) for each individual for a randomly selected locus in the genome are identical by descent. The kinship coefficient between two individuals corresponds to the expected inbreeding coefficient of their putative offspring.

Box 2 | **BLUP versus GBLUP**

BLUP uses pedigree information to estimate BVs. Livestock selection has relied for decades on the use of mixed model methodology to estimate breeding values (BVs). The animals' phenotypes are assumed to reflect differential exposure to known environmental factors (such as herd–year–season effects), as well as differences in intrinsic genetic ability (that is, the animals' BVs). BVs are modelled as random effects with multivariate normal distribution. Following standard quantitative genetics theory, the covariance structure of the BVs is assumed to reflect genetic relatedness:

$$\sigma_{ij} = 2\Theta_{ij}\sigma_A^2 \quad (1)$$

where Θ_{ij} is the kinship coefficient of individuals *i* and *j* and σ_A^2 is the additive genetic variance, where the matrix of $2\Theta_{ij}$ values is known as the additive relationship matrix *A*. Note that this model implicitly assumes the infinitesimal model. Kinship coefficients are computed from pedigree information and used to constrain the BV solutions, yielding so-called best linear unbiased predictors (BLUPs)[2,90]. BLUP-type approaches are extremely efficient and are able to extract information from millions of equations. An interesting feature of BLUP is that it allows estimation of BVs for animals without phenotypic records, including BVs for number of eggs or quantity of milk for males and BVs for unborn animals. Thus, BLUP can be seen as a method of statistical learning, in which the model is trained on a reference population that has both information on phenotypes and genetic relatedness and can then be used to predict the phenotype (including individual BVs) for animals for which only relatedness information is available. For a non-inbred offspring with no record of its own but with perfect information for the parents, the maximum accuracy of BLUP is 0.707 (the square root of 0.5, as half the genetic variance is between families and half is within).

GBLUP uses genome-wide SNP information to predict Mendelian sampling. In genomic BLUP (GBLUP), the kinship coefficient, Θ_{ij}, between two individuals is typically estimated as the correlation between standardized (that is, standard deviations from the mean) allelic dosages across all single-nucleotide polymorphisms (SNPs)[201,202]. The matrix of $2\Theta_{ij}$ values computed from SNP data is known as the genomic relationship matrix *G*. The accuracy of GBLUP is a function of the size of the training population (*N*), the trait heritability (h^2) and the number of loci affecting the trait (M_e)[203,204]. A simple equation that incorporates these parameters is:

$$M_e = \sqrt{Nh^2 / \left(Nh^2 + M_e\right)} \quad (2)$$

Under the infinitesimal model, M_e corresponds to the effective number of independent chromosome segments in the population. Various estimates for M_e are available, with the simplest being:

$$M_e = 2N_eL \quad (3)$$

where N_e is the effective population size and *L* is the length of the genome in Morgans[204]. M_e is approximately 6,000 in Holstein–Friesian cattle and 45,000 in Merino sheep. Assuming the same-sized reference population, genomic estimated BV (GEBV) will be more accurate in cattle[205–207]. Note that this formula for the accuracy of GBLUP is population-based — M_e is the number of independent chromosome segments in the population. If there are large full-sib or half-sib families in the population, accuracy is increased in some cases by taking advantage of the fact that within these families, M_e is much smaller[204]. Estimates of M_e can also take into account the effect of family structure between the reference population and selection candidates, as well as the covariance among chromosome segment effects, when N_e is low[208]. The maximum accuracy of GBLUP approaches 1 for progeny with no records of their own, given an extremely large reference population. GBLUP can be seen as estimating SNP effects but also as using an estimator (based on SNPs) of realized instead of expected (based on pedigree) relationships[209]. The difference between expected and realized relationships results in the improved accuracy[204].

Until approximately 10 years ago, kinship coefficients (needed to estimate BVs) were computed using genealogical information. Once genome-wide SNP arrays became available for livestock species, it became possible to estimate kinship coefficients from SNP genotypes rather than pedigree data[93–95]. Estimating EBVs with mixed models using kinship coefficients deduced from SNP information is referred to as genomic BLUP (GBLUP) as opposed to standard pedigree-based BLUP, the corresponding estimates are known as GEBVs instead of EBVs, and the corresponding approach to breeding is called genomic selection (GS) (BOX 2).

One advantage of GBLUP over BLUP is that it is not reliant on pedigree information, which may not always be available or accurate. However, its main advantage is that SNP information better tracks and captures Mendelian sampling than does pedigree information. For example, on the basis of pedigree information only, full-sibs all have the same EBV. By contrast, SNP information enables allelic transmission to be tracked at positions where the parents are heterozygous and can therefore differentiate sibs. This information can have a major effect on the accuracy and utility of EBVs. Before GS, candidate elite dairy sires that had identical EBVs based on pedigree information (for instance, because they were full-sibs) required expensive and time-consuming PT to expose differences in the BVs: their individual EBVs were estimated from the performances of tens to hundreds (depending on the country) of daughters, and PT took at least 5 years at a cost of ~US$50,000 per bull[10]. By contrast, as long as the reference population is large enough, GBLUP provides information of a similar accuracy to PT but at birth or earlier if SNP genotyping is performed on blastocyst biopsies (FIG. 2). For traits with high heritability (such as milk yield, $h^2 \approx 0.3$), the accuracy of GBLUP is typically somewhat lower than that of a combination of PT and BLUP. However, this modest penalty is more than compensated for by the fact that the information becomes available 5 years earlier. For low heritability traits (such as fertility, $h^2 \approx 0.05$), GBLUP information is not only available 5 years earlier but can even be more accurate than results obtained with PT combined with BLUP, depending on the size of the daughter groups used for PT. It is, therefore, no surprise that GBLUP was eagerly adopted worldwide by the dairy cattle breeding industry. As large-scale SNP genotyping rapidly became an outsourced commodity, the implementation of within-breed GBLUP in place of BLUP initially required only modest methodological adjustments, which also greatly facilitated the transition[10].

The use of phenotypic records of hundreds of thousands to millions of pedigreed animals across many generations made a substantial contribution to the accuracy of BLUP. Most of these animals could obviously not be retrospectively SNP genotyped. However, statistical methods have been devised to combine valuable pedigree information with SNP data when available. These single-step approaches are now routinely used in practice[96–101].

An evaluation of the impact of 7 years of GS in US dairy cattle[102] shows that the rates of genetic gain per year increased 50–100% for high heritability traits, such as milk yield, and 300–400% for low heritability traits, such as somatic cell counts (SCCs, a measure of udder health) and daughter pregnancy rate (a measure of female fertility). Thus, GS enabled genetic gain that

Hard sweeps
The process by which the frequency of a favourable new variant rapidly increases in the population by positive selection until eventual fixation of the variant and the haplotype upon which it occurred.

Balancing selection
A selective force on a locus that leads to a steady state whereby multiple alleles are simultaneously maintained in the population, rather than one allele becoming fixed at the expense of the others.

was not dominated by production but was instead balanced between traits. As expected, the generation interval dramatically shortened, from 7 years to 2.5 years in the 'sire of bull' path (that is, elite bulls used to breed the next generation of bulls). The rate of inbreeding did not seem to be affected. Although the gains are less than those in dairy cattle, GS has also been applied and deemed profitable in defined contexts in beef cattle[103,104], pigs[105], sheep[106–108] and layer chickens[109].

Another potential advantage of GS is that new traits can be included in selection indices to maximize genetic gain for profit. Provided a suitable reference population for these new traits can be constructed, genotyped selection candidates will have a GEBV for these traits, and selection decisions can be made accordingly. For example, including GEBV for feed efficiency in the selection index for dairy cattle is expected to improve genetic gain for profitability by at least 3%[110].

Accommodating non-normally distributed gene effects with Bayesian approaches. As with BLUP, GBLUP assumes that all segments of the genome contribute equally to the heritability of the trait, in accordance with the infinitesimal model. It can be shown that this is mathematically equivalent to a BLUP model that fits the effect of individual SNPs[111]. However, GWAS indicate that this assumption (that the effect of each SNP comes from a normal distribution, with the same variance across all SNP, such that all SNP effects are small) may not always be appropriate. Indeed, effects of a magnitude that is virtually impossible under this model have been identified and with GBLUP, their effects will be over-conservatively regressed downwards in genomic predictions. Examples of such major gene effects segregating within breeds include, among others, variants in *MSTN* in cattle[112–115] and sheep[67] and *RYR1*, *PRKAG3* and *IGF2* in pig[116–118], which all affect muscularity; *DGAT1*, *GHR* and *ABCG2*, which affect milk yield and composition in cattle[119–121]; and *PLAG1*, *HMGA2* and *LCORL*, which affect stature in cattle[122,123].

To better accommodate such large effects in GS, alternative prior distributions have been considered, mostly in a Bayesian framework. For instance, BayesA assumes a Student's *t* distribution of effects, which makes somewhat larger effects more likely[8]; BayesB assumes a mixture distribution with a large spike of zero effects and a Student's *t* distribution for the remaining effects[8]; BayesCπ assumes many effects at zero and the rest following a normal distribution[124]; and BayesR assumes SNP effects follow a multi-normal distribution[125]. The accuracy of GEBV will be highest when the prior distribution best matches the true distribution of SNP effects[111]. For example, for milk production traits (particularly fat percentage) affected by the *DGAT1* mutation in cattle, accuracies of GS are higher with BayesA and BayesB than with GBLUP[124].

The same methods can be used to study the genetic architecture of traits of interest. For example, a Bayesian method and a large number of dairy cattle with imputed whole-genome sequence data and milk production phenotypes were used to estimate the number of loci affecting these traits[32]. The study estimated that for milk yield, 4,330 SNPs had a non-zero effect, with only 7 SNPs explaining 1% or more of the genetic variation. Similar numbers were obtained for milk fat and protein yield.

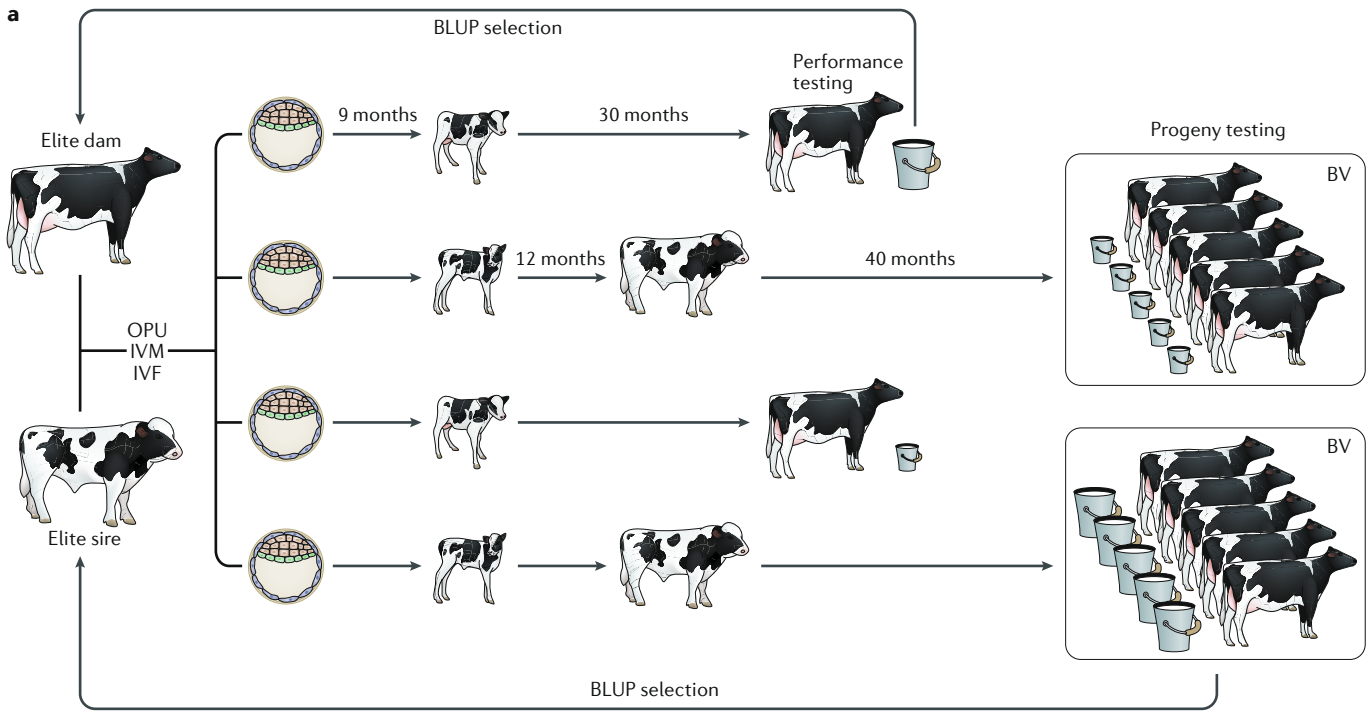
Balancing selection for variants with large effects is common in livestock. The occurrence of variants with large effects on complex quantitative traits seems to be much more common in livestock than in humans. This may be due, in part, to the strong directional selection to which nearly all livestock populations are subjected. DNMs with large effects on the selected traits sequentially undergo hard sweeps, causing large effects detectable by GWAS until the corresponding variants reach fixation[126]. Furthermore, genomic studies have provided ample evidence that several variants with large effects are maintained in livestock populations by balancing selection through a variety of mechanisms[127].

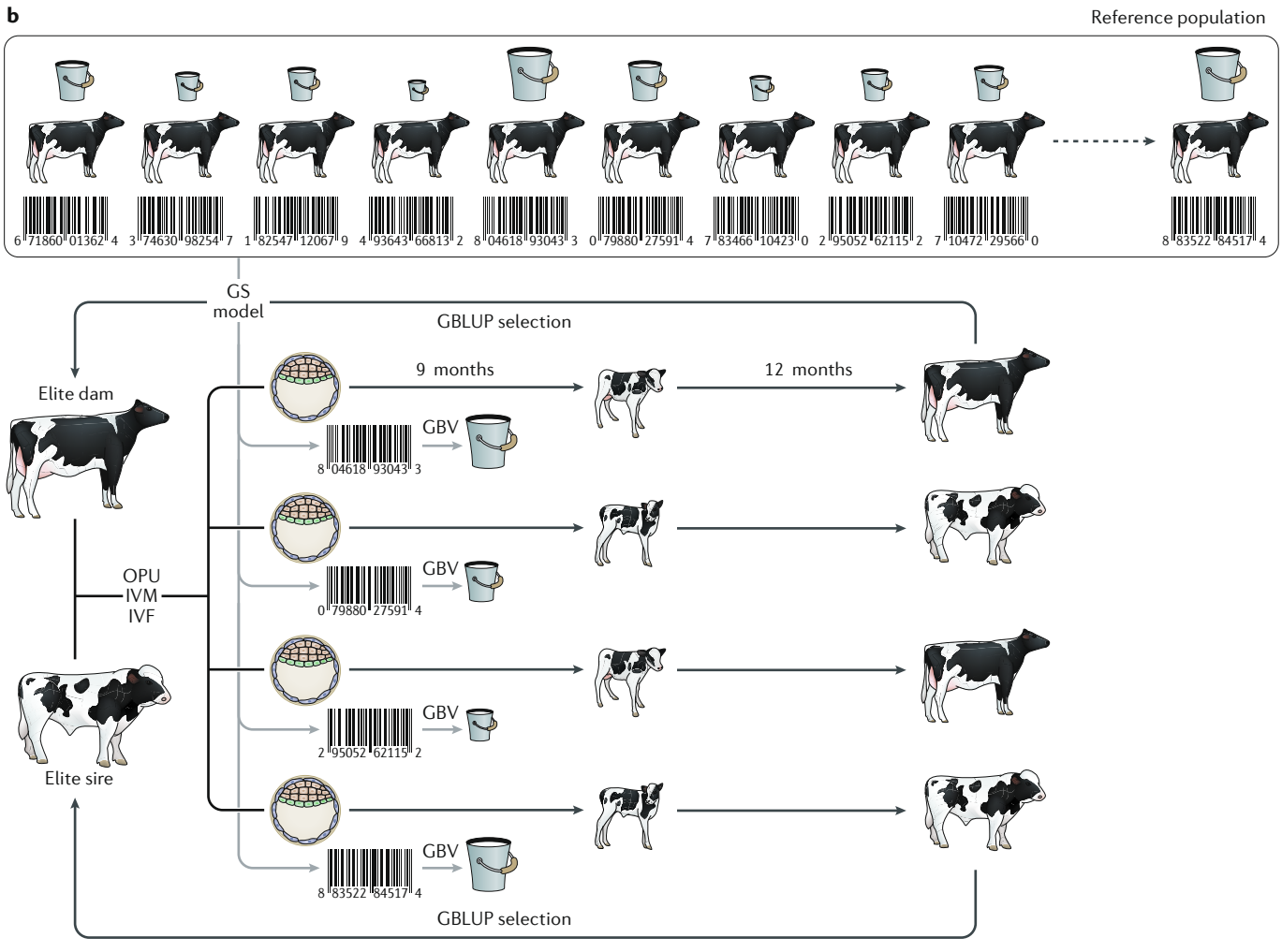
For instance, a growing list of variants is known to improve performance in heterozygotes but cause a defect in homozygotes, and these variants are hence subject to balancing selection. Classic examples include a *RYR1* variant in pigs that increases carcass yield in heterozygotes but causes porcine stress syndrome and related syndromes in homozygotes[112] and bovine *MSTN* LoF variants that increase muscle mass in heterozygotes but cause birthing difficulties for mothers of homozygous calves. Accordingly, the double-muscled phenotype, which is caused by homozygosity or compound heterozygosity for *MSTN* LoF variants, is avoided in most cattle breeds. In Belgian Blue cattle, disruptive variants in at least four genes (*MRC2*, *RNF11*, *WWP1* and *ATP2A1*) are known to increase muscularity in heterozygotes but affect viability or fitness in homozygotes[30,128–130]. Mutations in the ovine *BMP15* and *GDF9* genes increase litter size in heterozygotes but cause sterility in homozygotes[131,132]. In Scandinavian dairy cattle, a specific haplotype increases milk production in heterozygotes but is lethal in homozygous embryos[80].

Balancing selection also affects pleiotropic variants that have a positive effect on a desired trait but a negative effect on another trait. For example, a missense mutation in the SH2 domain of *SOCS2* in sheep increases stature and milk yield but also increases susceptibility to mastitis[133]. Similarly, an allele of the relaxin-like receptor 2 (*RXFP2*) locus in Soay rams increases horn size (and hence reproductive success) but reduces survival[134].

The *CLPG* mutation, which causes callipyge muscular hypertrophy in sheep, is an example of polar overdominance, by which only heterozygous animals inheriting the mutation from a particular parent (the sire in the case of *CLPG*) express the phenotype[68,135,136]. Salmon provide a remarkable example of balancing selection of *VGLL3* variants, in which antagonistic selection for age at maturation in males versus females creates sexual conflict that is partially resolved by sex-dependent dominance[137,138].

Other examples of balancing selection may reflect breeding objectives that change over time or differ between countries. For example, the K232A *DGAT1* variant in cattle results in increased fat yield but decreased protein yield. As breeding objectives evolved, this allele went from being favoured to penalized to neutral[119]. The segregation of *PLAG1* variants that have a major


a
BLUP selection
Performance testing
Elite dam
9 months
30 months
Progeny testing
BV
12 months
40 months
OPU
IVM
IVF
Elite sire
BV
BLUP selection


b
Reference population
6 71860 01362 4
3 74630 98254 7
1 82547 12067 9
4 93643 66813 2
8 04618 93043 3
0 79880 27591 4
7 83466 10423 0
2 95052 62115 2
7 10472 29566 0
8 83522 84517 4
GS model
GBLUP selection
Elite dam
9 months
12 months
GBV
8 04618 93043 3
OPU
IVM
IVF
0 79880 27591 4
2 95052 62115 2
8 83522 84517 4
Elite sire
GBLUP selection

◀ Fig. 2 | **Selection procedure of elite dairy sires and cows. a** | Before genomic selection (GS), elite sires and cows with the highest estimated breeding values (EBVs) were mated, often using reproductive technologies such as oocyte pick-up (OPU), in vitro maturation (IVM) and in vitro fertilization (IVF) to generate multiple embryos. Female calves (born after 9 months) would produce milk after ≥30 months, providing direct information about their value (in a performance test). Male calves would produce semen after ≥12 months and, ≥40 months later, milking daughters providing information about their breeding values (BVs) (in a progeny test). Best linear unbiased predictors (BLUPs) of the BVs of the new generation of cows and sires were computed using mixed models fitting environmental effects and using pedigree information to constrain the solutions of the EBVs. **b** | With GS, elite cows and sires are mated (often using OPU, IVM and IVF), and genomic DNA is extracted from offspring at birth or even as embryos before implantation and genotyped with genome-wide single-nucleotide polymorphism (SNP) arrays. Genomic BLUPs (GBLUPs) of the offspring BVs are then computed using DNA information to constrain the solutions of the genomic BVs (GBVs). The statistical models for GS are trained on a reference population of animals that have both SNP genotypes and phenotypic information.

effect on stature may occur because, after an extended period of selection for smaller cattle (in comparison to the ancestral auroch), larger cows are now preferred for some breeds in some countries but not others[112,139]. Moreover, the *PLAG1* variants that increase stature may negatively affect fertility, particularly age at puberty[140].

A more complete understanding of balancing selection operating at specific loci could be exploited to prioritize or avoid specific matings in breeding programmes (see From selecting animals to selective matings using genomic information).

Increasing the accuracy of GS using whole-genome sequence imputation. To control costs, GS is typically conducted using low-density or, at best, medium-density SNP arrays that interrogate 10,000–50,000 SNPs. With the exception of the handful of thoroughly studied major genes described above, causative variants remain unknown and have, therefore, seldom been included on the arrays. GS presumably works through indirect association, that is, because causative variants are in LD with one or several of the genotyped variants. Directly interrogating the causative variants would most certainly be better. Indeed, LD between causative and tagging variants is likely not perfect (that is, $r^2 < 1$) most of the time. LD will further decay — and the accuracy of GS will decrease — as the number of generations separating the reference population from the selection candidates increases[141]. Reproductive technologies, such as in vitro fertilization (IVF) of oocytes from prepubertal heifers, will increasingly enable selection based on SNPs only over multiple shortened generations, further exacerbating this issue.

One way to compensate for the fact that most causative SNPs are not directly interrogated on the arrays is to impute full sequence information on genotyped animals. Population-based resequencing efforts have been undertaken to generate reference populations needed for accurate imputation of at least the common single-nucleotide variants; rare variants, structural variants (including copy number variants (CNVs)) and DNMs are typically poorly imputed or ignored in this process. However, as the number of common variants to consider increases from tens of thousands to millions without a concomitant increase in phenotypic data points, so too does the curse of dimensionality; that is, it becomes more difficult to accurately estimate the effects of the growing number of SNPs. Although numerical methods have been developed to handle millions of variants[101,142–145], indiscriminate use of whole sequence information has only modestly increased (≤5%) the accuracy of genomic predictions[146–148]. Strategies to further improve the accuracy of whole-genome-sequence-based GS currently involve either selecting or assigning more weight to a subset of imputed variants that are more likely to be causative.

A first approach consists of ranking the variants based on their strength of association with the trait of interest. Variant effects should ideally be estimated by fitting them all simultaneously in the model. Increases in GS accuracy of up to 6% have been obtained by adding 1,623 top variants detected by sequence-based GWAS to the default 50,000 genome-wide SNPs in European dairy cattle[146].

A second approach aims to exploit prior biological information. Causative variants are either coding or regulatory. Coding variants can be readily identified by mining the sequence data. For example, more than 100,000 non-synonymous variants, which are expected to be enriched in causative variants, have been identified in cattle, and these can be ranked according to the predicted severity of the amino acid substitution on protein function[25,30]. Regulatory variants are more difficult to identify. Nearly 1 million evolutionarily constrained elements that overlap potential promoters, enhancers and insulators have been identified[24]. SNPs in these elements are likely to be enriched in regulatory variants. However, growing evidence suggests that enhancers, in particular, are rapidly evolving[39,149]. Thus, the establishment of exhaustive catalogues of regulatory elements in livestock will likely require the generation of species-specific and tissue-specific epigenome maps, and this is the aim of the FAANG project[37]. Much remains to be learned about how variants perturb regulatory elements, including whether they need to be within the element or can influence regulatory function from a distance. Transposable elements (TEs) are also a possible source of causative regulatory variants. TEs are still active in the genome of livestock species and generate numerous polymorphic insertion sites, of which several are likely to influence the expression levels of neighbouring genes[150].

Finally, it is possible that QTL information for intermediate phenotypes, including eQTL, will help pinpoint causative regulatory variants[122]. The contribution of regulatory variants to intermediate phenotype variation is likely to be higher than that for the agricultural trait that they influence, which could facilitate fine-mapping[151]. Intermediate phenotypes of particular interest in livestock are fatty acid profiles, as these affect the health and monetary values of animal food products. Across cattle, pigs and sheep, mutations of moderate to large effect have been mapped to or close to a small number of key genes involved in fatty acid synthesis[152–154]. eQTL information can certainly help to identify the target genes whose expression is perturbed by these regulatory variants[41,42,155].

Methods are also being developed to better take advantage of prior association or biological information for GS. Variants that are more likely to be

Non-synonymous
Variants that cause a change in the amino acid sequence of a protein. By contrast, synonymous variants are variants in the open reading frame of a protein-coding gene that do not change the amino acid sequence. Most non-synonymous variants affect the first and second codon positions, while most synonymous variants affect the third codon position.

Intermediate phenotypes
Phenotypes that mediate the link between a causative variant and the end-point disease or agricultural phenotype of interest — includes transcript, protein and metabolite levels.

functional based on biological information, including gene expression, can receive a higher prior probability in the model[156] or be assigned a distinct probability distribution of effect classes in BayesRC[32]. An alternative approach is to give additional weight to preselected variants (for example, in proportion to the amount of genetic variation they explain in GWAS) in the genetic relationship matrix (G matrix) (BOX 2) of the single-step method widely used for routine genetic evaluations[157].

Within-breed GS versus across-breed GS. To be effective, GS requires reference populations of tens of thousands of individuals. Many livestock breeds are numerically too small to make this a realistic scenario. Across-breed GS has been proposed as an alternative approach for smaller populations, in which GS models trained in a large breed are used to make predictions in a small breed. However, this method has proved to be ineffective when using data from low-density or even medium-density SNP arrays[158]. Several factors may explain this failure.

First, the linkage phase between causative variants and distant genotyped SNPs may differ between breeds. Indeed, predictions for across-breed GS improved using data from higher density arrays containing 600,000 SNPs, which are bound to include SNPs closer to the causative variants[154,159–163]. The improvement was highest (+8.7%) when using imputed sequence-based selection with Bayesian models[32]. Note that for this imputation-based strategy to be effective, whole-genome sequence data from a sufficient number (>100) of animals representing the small breed need to be available.

Second, the effectiveness of across-breed GS may be limited by the degree to which segregating causative variants are shared across breeds, which will be determined by the proportion of causative variants that existed before breed formation and how much gene flow there has been between the breeds. It is reasonable to assume that the degree of sharing will be inversely proportional to the effect size. Indeed, the allelic frequency of small-effect variants will change more slowly under selection (that is, polygenic adaptation) than will large-effect variants, which are expected to undergo selective sweeps that lead to rapid fixation. Moreover, the sharing of large-effect variants between breeds is highly suggestive of balancing selection, which may reveal associated deleterious effects. GWAS conducted in multiple dairy cattle breeds indicate that ~50% of QTL might be shared across breeds and might predate breed formation if not domestication[161,164]. Some sharing is also the result of more recent between-breed introgression. For example, European breeders reportedly imported Asian pigs to improve local stock at the end of the 19th century, and this has resulted in the near fixation of Asian haplotypes conferring desired features at, for instance, the *IGF2*, *FASN*, *ME1* and *KIT* loci[118,165].

Finally, it is important to remember that the effects of genetic variants are neither constant across populations nor across generations. Indeed, in the case of dominance, the effects of allelic substitution are a function of allelic frequency and may be affected by gene-by-gene (GxG) and gene-by-environment (GxE) interactions[2].

From selecting animals to selective matings using genomic information. The prevailing paradigm in present-day animal breeding is to identify the animals that have the highest EBVs within their generation as genitors of the next. However, at some point, genetic progress may be limited by the rate at which favourable alleles combine by recombination. It is conceivable that the fastest route towards the ideal allelic combination does not involve selecting the best animals from each generation but instead requires an approach that uses temporarily suboptimal individuals. Thus, a new paradigm may be to find the fastest mating scheme to concentrate a maximum of favourable alleles in one individual. In its simplest version, such an approach may consist of identifying matings that have the potential to produce the best offspring, that is, genomic mate selection. For example, a corresponding sire and dam may not have the highest EBVs, but their genotypes might complement each other in a way that produces superior offspring. Furthermore, the availability of genomic information makes it possible to control inbreeding at the genome level while simultaneously making rapid gains from GS[166–168]. However, it is important to estimate EBVs and inbreeding using the same source of information (that is, either based on pedigree or genomic information) when controlling inbreeding using the approach of optimum contribution selection[166].

Editing livestock genomes

Programmable nucleases have revived interest in editing livestock genomes. Early reports describing the generation of transgenic animals[169,170] led to expectations of a revolution in animal breeding. Slow-pace selection would make way for the engineering of plant and animal genomes. Genetic improvement would no longer require the variants to pre-exist in the breed of interest. Beneficial mutations could be transferred between populations, transferred between species or even designed at will. However, the anticipated revolution has yet to occur. Generating transgenic animals remained extremely arduous and expensive, demonstrably useful transgene constructs few, and western consumers apprehensive of genetically modified organisms. The initial pronuclear microinjection technique was not only inefficient but carried with it the risk of insertional inactivation of endogenous genes. Retroviral vectors provided limited cargo capacity, were subject to epigenetic inactivation and could also perturb endogenous genes upon insertion. The inability to derive embryonic stem cells prevented homologous recombination-based techniques until the development of somatic cell nuclear transfer (SCNT)[171], which enabled refined gene replacement by homologous recombination in cultured fetal fibroblasts followed by nuclear transfer to enucleated oocytes. However, the exceedingly low rate of spontaneous homologous recombination severely limited its applicability (reviewed elsewhere[172]).

The development of programmable nucleases, including Fok1-based zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), and, more recently, of the CRISPR–Cas9-based nucleases profoundly changed the picture[173]. These nucleases

Gene flow
The passage of alleles between populations as a result of migration or interbreeding.

Polygenic adaptation
The process by which a phenotype caused by many genes evolves in a population under selection, not by massive changes in the frequency of a few variants with major effects on the phenotype (hard and soft sweeps) but by very small changes in the frequency of many variants with minor effects on the phenotype.

induce double-stranded breaks in DNA with high efficiency, which facilitate the generation of LoF mutations in any gene of interest by inducing error-prone non-homologous end-joining (NHEJ) and result in a ~100-fold increase in the rate of gene replacement by homology-dependent repair (HDR) compared with spontaneous homologous recombination. The ease of development of target-specific CRISPR–Cas9 reagents and the rapidly expanding toolbox of CRISPR–Cas9-derived methods[174] make them the preferred option over ZFNs and TALENs. Since 2011, a growing number of genome-edited domestic animals have been produced through the combined use of SCNT and programmable nucleases (reviewed elsewhere[172,175]). Indeed, the rate of double-stranded break-induced NHEJ is now high enough that, despite mosaicism commonly reducing germline transmission, it has become more effective (in terms of the number of embryos required to obtain an edited offspring) to circumvent SCNT and inject the programmable nucleases directly into the zygote when aiming to generate LoF mutations[172]. Although the efficiency of HDR may still be too low for direct zygotic injection to be an attractive proposition for gene replacement, the development of effective CRISPR–Cas9-dependent knock-in methods, such as precise integration into target chromosomes (PITCh)[176] and homology-independent targeted integration (HITI)[177], expands the number of applications of direct zygotic injection.

Genomic advances are uncovering a growing list of target genes and mutations. Thus far, efforts in editing the genome of livestock have mostly concentrated on largely uncontroversial human health applications, such as generating animal models of human genetic diseases, producing biopharmaceuticals and xenotransplantation. However, genomic and other advances are uncovering a growing list of putative target genes and mutations for which editing has been or is being pursued in order to improve aspects of livestock production[175,178,179], such as growth rate, muscle mass, meat fat content, milk composition, wool growth, resistance to bacterial, viral and parasitic diseases, temperature tolerance, animal welfare and environmental impact (TABLE 3). A specific application that is worth mentioning is the generation of male-specific double-muscling by inserting a *trans*-acting inhibitor of *MSTN* on the Y chromosome, which could have a major economic impact by improving the carcass yield and thereby the value of male dairy calves[180].

Combining GS with genome editing. As mentioned above, nearly all economically important traits in livestock are highly polygenic. As the number of animals with both phenotypic records and SNP genotypes increases into the millions and as fine-mapping methods continue to improve, a growing number of causative variants (particularly those with the largest effects) is bound to be identified, as has been shown to occur for common complex diseases in humans[36]. The efficiency of CRISPR-based genome editing has increased to the point that multiple edits can be introduced simultaneously in cells with relative ease[181,182]. A strategy termed 'promotion of alleles by genome editing' (PAGE) has been proposed, which combines genome editing with GS[183]. In PAGE, sires would first be selected based on their GEBV using GS. Before dissemination, the sire's genome would be edited for a number of causative variants to render them homozygous for the favourable allele. Under optimal conditions, it is estimated that PAGE could increase genetic response 2–4-fold over GS alone. Although far from trivial, schemes for the practical implementation of PAGE have been proposed[175,184] that involve establishing, SNP genotyping and editing fetal fibroblast cell lines followed by SCNT. Although it will initially be limited by the number of known causative variants and technical hurdles, the feasibility of PAGE is likely to be explored by a number of breeding organizations in the near future.

New applications of genomic technology

Detecting cows with subclinical mastitis by bulk genotyping of tank milk. As growing numbers of animals are genotyped with genome-wide SNP arrays as part of GS programmes, new applications emerge, such as the early detection of cows with subclinical mastitis by bulk genotyping of tank milk[185] (FIG. 3). Dairy farms typically comprise tens to hundreds of cows whose milk is stored in large tanks before daily collection by milk processing factories. One of the major health issues on dairy farms is mastitis[186]. Milk from affected cows is characterized by an increase in SCCs as immune cells migrate into the udder and milk. The normal range for SCCs is <100,000 cells per ml, but it may reach millions in cows suffering clinical mastitis, and even before overt symptoms appear, cows with subclinical mastitis may already have >200,000 cells per ml. At that stage, their milk yield is typically decreased, and the milk quality of the entire tank is affected. Early detection of cows with clinical mastitis is therefore paramount. A new cost-effective way to achieve this is to SNP genotype the tank milk, which yields estimates of the allelic ratio (the B-allele frequency) in the milk for tens of thousands to hundreds of thousands of markers. This ensemble of allelic ratios reflects the combination of the cows' known SNP genotypes, and the unknown proportion of DNA contributed by each cow to the tank milk. These proportions can be accurately estimated by solving a corresponding set of linear equations and converted to SCCs for individual cows. The number of SNPs needed to achieve adequate accuracy depends on the number of cows on the farm: tens of thousands of SNPs are sufficient for farms with tens of cows, but hundreds of thousands of SNPs are needed for farms with several hundred cows. However, low-density SNP arrays, comprising 10,000 SNPs, are most commonly used to genotype cows. Nonetheless, whole-genome sequence information can be imputed for the cows, and the tank milk can be genotyped with high-density SNP arrays or by shallow sequencing (M.G., unpublished observations). Sequencing also enables simultaneous microbiome analysis to identify bacterial contaminants.

Importantly, the ability to determine the relative contribution of genotyped individuals or their genotyped ancestors to samples of bulk animal products provides novel opportunities to trace and authenticate animal food products[187].

Mosaicism
The occurrence of mutations in some but not all cells of an organism that is entirely derived from a single zygote.

Table 3 | **Target genes and mutations for editing livestock genomes**

Phenotype	Gene	Editing type	Effect	Species	Refs	
					Gene editing	Naturally occurring variation
Carcass yield and composition						
Increased muscle mass	*GH*	RKI	Ectopic expression in all tissues	Pig, sheep, salmon	210–214	–
Increased muscle mass	*GHRF*	RKI	Ectopic expression in all tissues	Pig, sheep	211,212	–
Increased muscle mass	*IGF1*	RKI	Ectopic expression in all tissues	Pig	211,215	–
Increased muscle mass	*cSKI*	RKI	Ectopic expression in all tissues	Mouse	216	–
Increased muscle mass	*MSTN*	KO or KD	Constitutive or conditional KD	Cattle, sheep, goat	217–222	67,113–116
Increased muscle mass	*IGF2*	PAGE	–	–	–	118
Increased muscle mass	*FST*	RKI	Ectopic expression in muscle	Mouse	223	–
Increased muscle mass	*FLRG*	RKI	Ectopic expression in muscle	Mouse	223	–
Increased linoleic and alpha-linoleic content	*FAD2*	RKI	Ectopic expression in adipose	Pig	224	–
Increased omega-3 fatty acids content	*FAT1*	RKI	Ectopic expression in all tissues	Pig, cattle, sheep	225–227	–
Decreased fat content	*UCP1*	HRKI	Constitutive expression	Pig	228	229
Milk yield and composition						
Increased milk volume	*GHR*	PAGE	–	–	–	120
Increased casein content	*CSN* (ABK)	RKI	Increased (mammary gland) expression	Cattle	230	231
Humanized milk	*BLG*	KO or KD	Constitutive or conditional KD	Cattle, goat	232–235	–
Increased nutritional quality	*LALBA*	RKI	Ectopic expression in mammary gland	Pig	236	–
Humanized milk	*LALBA*	RKI	Ectopic expression in mammary gland	Cattle	237	–
Low-lactose milk	*LPH*	RKI	Ectopic expression in mammary gland	Mouse	238	–
Fatty acid altered composition	*SCD*	RKI	Ectopic expression in mammary gland	Goat	239	–
Increased omega-3 fatty acids content	*FAT1*	RKI	Ectopic expression in all tissues	Cattle	226	–
Wool yield and composition						
Increased fleece growth rate	*GH*	RKI	Ectopic expression in all tissues	Sheep	214	–
Increased fleece growth rate	*IGF1*	RKI	Ectopic expression in wool follicle	Sheep	240	–
Altered fleece quality	*K2.10*	RKI	Ectopic expression in wool follicle	Sheep	241	–
Increased fleece growth rate	*CYS* (E, M, K)	RKI	Ectopic expression in all tissues	Sheep	242	–
Resistance to infectious diseases						
Mastitis resistance	*HLZ*	RKI or HRKI	Ectopic expression in mammary gland	Goat, cattle	243,244	–
Mastitis resistance	*LSS*	RKI or HRKI	Ectopic expression in mammary gland	Cattle	245,246	–
Resistance to *Flavobacterium columnare* and *Edwardsiella ictaluri*	*CecB*	RKI	Ectopic expression in all tissues	Catfish	247	–
Resistance to tuberculosis	*HBD3*	RKI	Ectopic expression in lung	Cattle	248	–
Resistance to *Actinobacillus pleuropneumoniae*	*PBD2*	RKI	Ectopic expression in all tissues	Pig	249	–
Resistance to tuberculosis	*SP110*	HRKI	Ectopic expression in all tissues	Cattle	250	251–253
Resistance to foot-and-mouth disease	Anti-FMDV shRNA	–	–	Pigs	254	–
Resistance to influenza and classical swine fever	*MX1*	RKI	Ectopic expression in all tissues	Pigs	255,256	–

Table 3 (cont.) | **Target genes and mutations for editing livestock genomes**

Phenotype	Gene	Editing type	Effect	Species	Refs	
					Gene editing	Naturally occurring variation
Resistance to infectious diseases (cont.)						
Resistance to influenza	shRNA decoy	RKI	Ectopic expression in all tissues	Chicken	257	–
Resistance to visna	Envelope protein of visna	–	–	Sheep	258	–
Resistance to avian leukosis virus	*ALV6*	RKI	Ectopic expression in all tissues	Chicken	259	–
Resistance to PRRS	*CD163*	KO/ HRKO	Constitutive KO	Pig	260,261	–
Resistance to African swine fever	*RELA*	KO or HRKI	Constitutive expression	Pig	262,263	–
Resistance to BSE	*PRP*	RKI or KO	Constitutive KO or KD	Cattle, goat, sheep	264–268	269
Resistance to trypanosomiasis	*APOL1*	RKI	Ectopic (haematocyte) expression	Cattle	270	271
Animal welfare and environmental impact						
Heat tolerance	*PRLR*	PAGE	–	Cattle	–	75
Cold tolerance	*UCP1*	HRKI	Constitutive expression	Pig	228	229
Cold tolerance	*wfiAFP6*	RKI	Ectopic expression in all tissues	Salmon	272	273
Hornless	*POLL*	HRK or PAGE	–	Cattle	274	275–277
Reduced faecal phosphorus output	*APPA*	RKI	Ectopic expression in salivary gland	Pig	278	–

BSE, bovine spongiform encephalopathy; HRKI, knock-in by homologous recombination; HRKO, knockout by homologous recombination; KD, knockdown; KO, knockout; PAGE, promotion of alleles by genome editing; PRRS, porcine reproductive and respiratory syndrome; RKI, random knock-in; shRNA, short hairpin RNA.

Combined genomic and metagenomic predictions of high-value traits. In addition to an animal's own genome, the species composition and abundance of the gut microbiome have been shown to be associated with some traits. For example, two studies have demonstrated that rumen microbiome profiles are associated with methane emission levels in cattle and sheep[188,189]. It is worth noting that BLUP models can be used to evaluate the effect of the microbiome on phenotypes of interest by replacing the genetic relationship matrix with a microbiome similarity matrix.

It is not yet clear whether predictions of future phenotypes for high-value traits, such as methane emissions and feed efficiency, can be improved by integrating genomic predictions from SNP genotypes with gut (or rumen for cattle) microbiome profiles. Although high-throughput sequencing enables rapid and cost-effective profiling of gut microflora, the number of animals with high-value traits measured, gut microflora profiled and SNPs genotyped is still quite small. However, a very preliminary study tested the concept in a small sample of 28 Holstein–Friesian dairy cattle for which 30,000 SNP genomic predictions for feed efficiency and rumen microbiome profiles were available[190]. The genomic and microbiome profile predictions were combined using a linear regression model, and although the results must be interpreted with caution because of the small data set, the prediction accuracy in cross validation was maximized when both SNP and rumen microbiome profiles were used ($r_{SNP} = 0.33$; $r_{SNP+M} = 0.57$). These results are encouraging, and larger scale studies may lead to future phenotype predictions that will enable selection of animals that will perform well over their lifetimes.

Conclusions and future perspectives

The field of animal breeding just completed a prototypical, once-in-a-lifetime Gartner hype cycle. The innovation trigger was the emergence of genomics as a new scientific discipline at the end of the 1980s. The peak of inflated expectations coincided with the initial investment of the public and private sectors in efforts to map QTL that would enable transformative MAS in the 1990s. A trough of disillusionment followed the realization that mapped QTLs explained insufficient variation to be of practical use. A slope of enlightenment followed the landmark paper that introduced GS[8] and the development of cost-effective SNP genotyping arrays. It turned remarkably rapidly into widespread adoption of the technology, with millions of animals being enlisted in GS programmes over the past 10 years. GS has revolutionized animal breeding and allowed an estimated doubling of genetic gains in some species. The success of the underlying methodology and the ensuing support for the infinitesimal model have influenced other disciplines, including plant breeding and medical genomics.

Understandably, attention to and investments in research and development faded somewhat as breeding

Gartner hype cycle
A model first proposed by the Gartner firm to explain the phases of maturation, adoption and social application of new technologies.

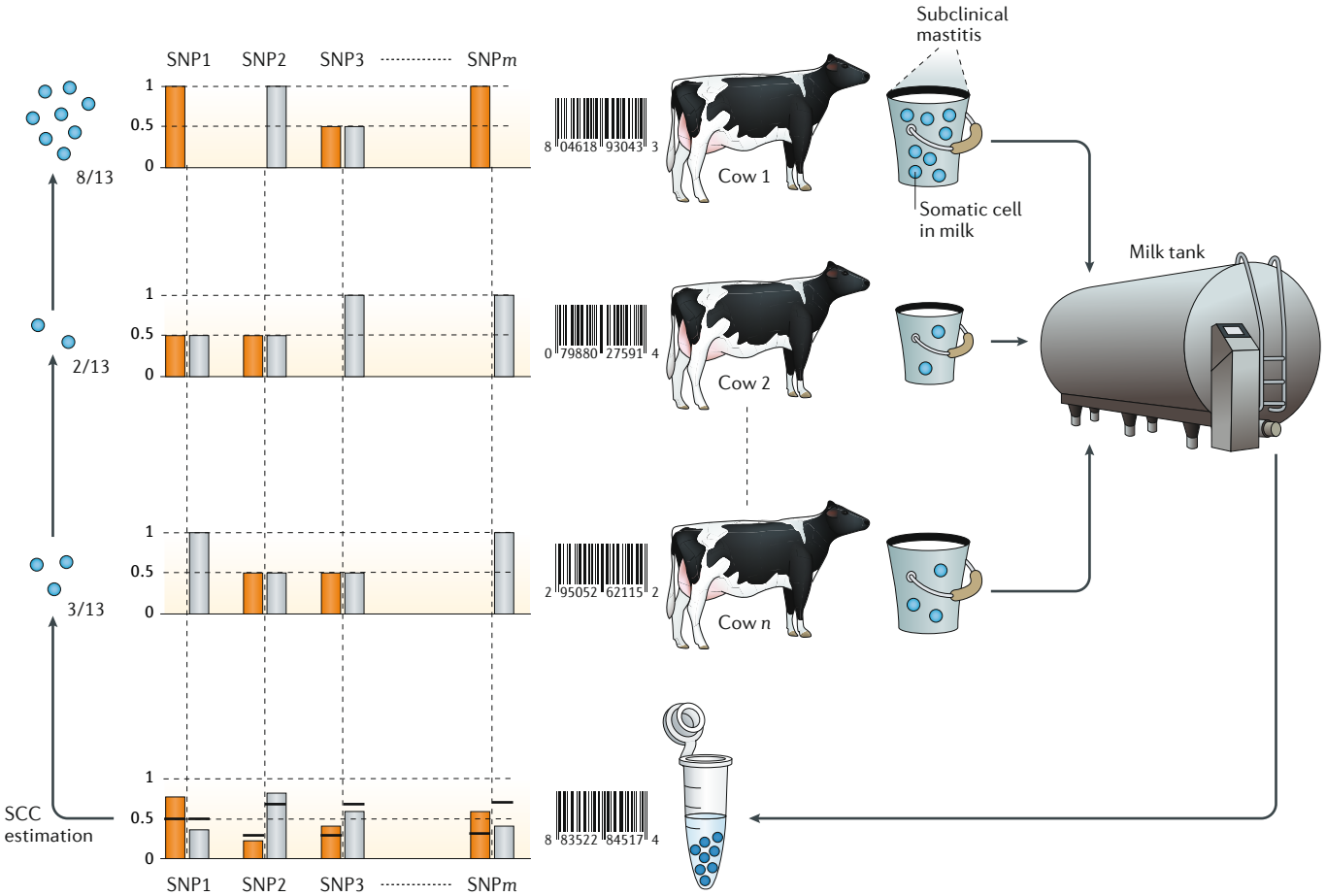


Fig. 3 | **Identifying cows with subclinical mastitis by bulk genotyping of tank milk.** Assuming that all cows on the farm that contributed milk to the tank have been genotyped with whole-genome single-nucleotide polymorphism (SNP) arrays, it is possible to determine what proportion of DNA each cow contributed to the tank by SNP genotyping the tank milk and measuring the allelic ratio (B-allele frequency) for each SNP. If the milk volume contributed by each cow is known and if the somatic cell counts (SCCs) in the tank are known, the SCC for each individual cow can be determined. The bar graphs represent the allelic dosages for allele A (orange) and allele B (grey) for *m* SNPs in individual cows (upper three graphs) and tank milk (bottom graph). The horizontal lines in the bottom graph correspond to the expected allelic dosage, assuming that all cows contributed equal amounts of DNA to the tank milk. The deviation from expectation allows the contribution of each cow to be estimated.

organizations focused on the implementation and consolidation of GS. Now that these goals have been largely accomplished and GS has been widely adopted, its limitations as implemented today are recognized and better understood. Logically, there is, therefore, renewed interest and investment in the next round of innovation. Identified objectives include more effective forward and reverse genetic screens to identify genes with major effects, efforts to systematically identify and incorporate causative variants into improved models of GS, renewed attention to the CRISPR–Cas9-based editing of livestock genomes combined with GS and the development of novel uses of the growing body of genomic data. Thus, genomic information will be a critical component of the global response to the world's pressing nutritional requirements.

Published online: 04 December 2018

1. Thornton, P. K. Livestock production: recent trends, future prospects. *Phil. Trans. R. Soc. B* **365**, 2853–2867 (2010).
2. Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer Associates, Sunderland, MA, 1998).
3. Havenstein, G. B., Ferket, P. R. & Qureshi, M. A. Growth liability and feed conversion of 1957 versus 2001 broilers when fed representative 1957 and 2001 broiler diets. *Poult. Sci.* **82**, 1500–1508 (2003).
4. Sigel, P. B. Evolution of the modern broiler and feed efficiency. *Annu. Rev. Anim. Biosci.* **2**, 375–385 (2014).
5. Capper, J. L. & Bauman, D. E. The role of productivity in improving the environmental sustainability of ruminant production systems. *Annu. Rev. Anim. Biosci.* **1**, 469–489 (2013).
6. Pryce, J. E., Royal, M. D., Garnsworthy, P. C. & Mao, I. L. Fertility in the high-producing dairy cow. *Livestock Prod. Sci.* **86**, 125–135 (2004).
7. Goddard, M. E. & Hayes, B. J. Mapping genes for complex traits in domestic animals and their use in breeding programs. *Nat. Rev. Genet.* **10**, 381–391 (2009).
8. Meuwissen, T. H., Hayes, B. J. & Goddard, M. E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829 (2001).
This is a landmark paper that triggered the adoption of GS by the breeding industry.
9. Hill, W. G., Goddard, M. E. & Visscher, P. Data and theory point to mainly additive genetic variance for complex traits. *PLOS Genet.* **4**, e1000008 (2008).
10. Wiggans, G. R., Cole, J. B., Hubbard, S. M. & Sonstegard, T. S. Genomic selection in dairy cattle: the USDA experience. *Annu. Rev. Anim. Biosci.* **5**, 309–327 (2017).
11. Van Eenennaam, A. L., Weigel, K. A., Young, A. E., Cleveland, M. A. & Dekkers, J. C. Applied animal genomics: results from the field. *Annu. Rev. Anim. Biosci.* **2**, 105–139 (2014).
12. Crossa, J. et al. Genomics selection in plant breeding: methods, models and perspectives. *Trends Plant Sci.* **22**, 9661–9975 (2017).
13. Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
14. Gusev, A. et al. Quantifying missing heritability at known GWAS loci. *PLOS Genet.* **9**, e1003993 (2013).

15. Wray, N. R. et al. Pitfalls of predicting complex traits from SNPs. *Nat. Rev. Genet.* **14**, 507–515 (2013).
16. Herrero, M. et al. Livestock and the environment: what have we learned in the past decade? *Annu. Rev. Environ. Resour.* **40**, 177–202 (2015).
17. International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004).
18. Bovine Genome Sequencing and Analysis Consortium. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* **324**, 522–528 (2009).
19. Swine Genome Sequencing Consortium. Analysis of pig genomes provide insight into porcine demography and evolution. *Nature* **491**, 393–398 (2012).
20. Dong, Y. et al. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nat. Biotechnol.* **31**, 135–141 (2013).
21. Jiang, Y. et al. The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* **344**, 1168–1173 (2014).
22. Lien, S. et al. The Atlantic salmon genome provides insights into rediploidization. *Nature* **533**, 200–205 (2016).
23. Green, E. D. Strategies for the systematic sequencing of complex genomes. *Nat. Rev. Genet.* **2**, 573–583 (2001).
24. Lindblad-Toh, K. et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).
25. Daetwyler, H. D. et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.* **8**, 858–865 (2014).
This is a report by the 1,000 Bull Genomes Project consortium that introduces a community resource to accelerate the detection and use of causative variants in cattle.
26. Worley, K. C. A golden goat genome. *Nat. Genet.* **49**, 485–486 (2017).
27. Bickhart, D. M. et al. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat. Genet.* **49**, 643–650 (2017).
This paper brilliantly illustrates the utility of novel scaffolding techniques in dramatically improving the quality of reference genome sequences in an affordable way.
28. International Chicken Polymorphism Map Consortium. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature* **432**, 717–722 (2004).
29. Bovine HapMap Consortium. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* **324**, 528–532 (2009).
30. Charlier, C. et al. NGS-based reverse genetic screen for common embryonic lethal mutations compromising fertility in livestock. *Genome Res.* **26**, 1–9 (2016).
31. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
32. MacLeod, I. M. et al. Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genomics* **17**, 144 (2016).
33. Bourneuf, E. et al. Rapid discovery of de novo deleterious mutations in cattle enhances the value of livestock as model species. *Sci. Rep.* **7**, 11466–11485 (2017).
34. Kadri, N. K. et al. Coding and non-coding variants in HFM1, MLH3, MSH4, MSH5, RNF212 and RNF212B affect recombination rate in cattle. *Genome Res.* **26**, 1323–1332 (2016).
35. Schaub, M. A., Boyle, A. P., Kundaje, A., Batzoglou, S. & Snyder, M. Linking disease associations with regulatory information in the human genome. *Genome Res.* **22**, 1748–1759 (2012).
36. Huang, H. et al. Association mapping of inflammatory bowel disease loci to single variant resolution. *Nature* **547**, 173–178 (2017).
37. Andersson, L. et al. Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome Biol.* **16**, 57–63 (2015).
38. Tuggle, C. K. et al. GO-FAANG meeting: a gathering on functional annotation of animal genomes. *Anim. Genet.* **47**, 528–533 (2016).
39. Villar, D. et al. Enhancer evolution across 20 mammalian species. *Cell* **160**, 554–566 (2015).
40. Zhou, Y. et al. Reduced representation bisulphite sequencing of ten bovine somatic tissues reveals DNA methylation patterns and their impacts on gene expression. *BMC Genomics* **17**, 779 (2016).
41. Littlejohn, M. D. et al. Expression variants of the lipogenic AGPAT6 gene affect diverse milk composition phenotypes in *Bos taurus*. *PLOS ONE* **9**, e85757 (2014).
42. Littlejohn, M. D. et al. Sequence-based association analysis reveals an MGST1 eQTL with pleiotropic effects on bovine milk composition. *Sci. Rep.* **6**, 25376–25390 (2016).
43. Kemper, K. E. et al. Leveraging genetically simple traits to identify small-effect variants for complex phenotypes. *BMC Genomics* **17**, 858–867 (2016).
44. Brand, B. et al. Adrenal cortex expression quantitative trait loci in a German Holstein x Charolais cross. *BMC Genetics* **17**, 135–146 (2016).
45. Lopdell, T. J. et al. DNA and RNA-sequence based GWAS highlights membrane-transport genes as key modulators of milk lactose content. *BMC Genomics* **18**, 968 (2017).
46. Ponsuksili, S., Murani, E., Brand, B., Schwerin, M. & Wimmers, K. Integrating expression profiling and whole-genome association for dissection of fat traits in a porcine model. *J. Lipid Res.* **52**, 6668–6678 (2011).
47. Liaubet, L. et al. Genetic variability of transcript abundance in pig peri-mortem skeletal muscle: eQTL localized genes involved in stress response, cell death, muscle disorders and metabolism. *BMC Genomics* **12**, 548–565 (2011).
48. Ernst, C. W. & Steibel, J. P. Molecular advances in QTL discovery and application in pig breeding. *Trends Genet.* **29**, 215–224 (2013).
49. Heidt, H. et al. A genetical genomics approach reveals new candidates and confirms known candidate genes for drip loss in a porcine resource population. *Mamm. Genome* **24**, 416–426 (2013).
50. Chen, C. et al. Genetic dissection of blood lipid traits by integrating genome-wide association study and gene expression profiling in a porcine model. *BMC Genomics* **14**, 848–859 (2013).
51. Ponsuksili, S., Murani, E., Trakooljul, N., Schwerin, M. & Wimmers, K. Discovery of candidate genes for muscle traits based on GWAS supported by eQTL-analysis. *Int. J. Biol. Sci.* **10**, 327–337 (2014).
52. Reiner, G. et al. Pathway deregulation and expression QTLs in response to *Actinobacillus pleuropneumoniae* infection in swine. *Mamm. Genome* **25**, 600–617 (2014).
53. Ma, J. et al. A splice mutation in the PHKG1 gene causes high glycogen content and low meat quality in pig skeletal muscle. *PLOS Genet.* **10**, e1004710 (2014).
54. Kogelman, L. J. et al. An integrative systems genetics approach reveals potential causal genes and pathways related to obesity. *Genome Med.* **7**, 105–120 (2015).
55. Martinez-Montes, A. M. et al. Deciphering the regulation of porcine genes influencing growth, fatness and yield-related traits through genetical genomics. *Mamm. Genome* **28**, 130–142 (2017).
56. Gonzalez-Prendes, R. et al. Joint QTL mapping and gene expression analysis identify positional candidate genes influencing pork quality traits. *Sci. Rep.* **7**, 39830–39839 (2017).
57. Maroilley, T. et al. Deciphering the genetic regulation of peripheral blood transcriptome in pigs through expression genome-wide association study and allele-specific expression analysis. *BMC Genomics* **18**, 967 (2017).
58. Blum, Y. et al. Complex trait subtypes identification using transcriptome profiling reveals an interaction between two QTL affecting adiposity in chicken. *BMC Genomics* **12**, 567–575 (2011).
59. Johnsson, M., Jonsson, K. B., Andersson, L., Jensen, P. & Wright, D. Quantitative trait locus and genetical genomics analysis identifies putatively causal genes for fecundity and brooding in the chicken. *G3* **6**, 311–319 (2015).
60. Johnsson, M., Williams, M. J., Jensen, P. & Wright, D. Genetical genomics of behavior: a novel chicken genomic model for anxiety behavior. *Genetics* **202**, 327–340 (2016).
61. Fallahsharoudi, A. et al. QTL mapping of stress related gene expression in cross between domesticated chickens and ancestral red junglefowl. *Mol. Cell. Endocrinol.* **446**, 52–58 (2017).
62. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
63. Dunne, J. et al. First dairying in green Saharan Africa in the fifth millennium BC. *Nature* **486**, 390–394 (2012).
64. Andersson, L. S. et al. Mutations in *DMRT3* affect locomotion in horses and spinal circuit function in mice. *Nature* **488**, 642–646 (2012).
This paper demonstrates the value of domestic animals in uncovering functions of mammalian genes by studying unique selected phenotypes, in this case 'ambling'.
65. Andersson, L. Molecular consequences of animal breeding. *Curr. Opin. Genet. Dev.* **23**, 295–301 (2013).
66. Durkin, K. et al. Serial translocations by means of circular intermediates underlies color sidedness in cattle. *Nature* **482**, 81–84 (2012).
This study identified a novel CNV-generating mechanism that may underlie exon shuffling by studying coat-colour variation in cattle.
67. Clop, A. et al. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat. Genet.* **38**, 813–818 (2006).
This paper describes one of the most convincing examples in mammals of a phenotype resulting from perturbed microRNA-mediated gene regulation.
68. Georges, M. et al. in *Epigenetics and Complex Traits* (eds Naoumova, A. K. & Greenwood, C. M. T.) 89–106 (Springer, New York, NY, 2013).
This paper reviews the current molecular understanding of the unique phenomenon of polar overdominance at the ovine callipyge locus.
69. MacArthur, D. G. et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–828 (2012).
70. Bittles, A. H. & Neel, J. V. The costs of human inbreeding and their implications for variations at the DNA level. *Nat. Genet.* **8**, 117–121 (1994).
71. Simmons, M. J. & Crow, J. F. Mutations affecting fitness in *Drosophila* populations. *Annu. Rev. Genet.* **11**, 49–78 (1977).
72. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
73. Baker, R. D., Snider, G. W., Leipold, H. E. & Johnson, J. L. Embryo transfer tests for bovine syndactyly. *Theriogenology* **13**, 87 (1980).
74. Charlier, C. et al. Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nat. Genet.* **40**, 449–454 (2008).
This study is one of the first illustrations of how genome-wide SNP arrays accelerated the identification of mutations causing genetic defects in domestic animals.
75. Littlejohn, M. D. et al. Functionally reciprocal mutations of the prolactin signalling pathway define hairy and slick cattle. *Nat. Commun.* **5**, 5861–5869 (2014).
76. Agerholm, J. S. et al. A de novo missense mutation of FGFR2 causes facial dysplasia syndrome in Holstein cattle. *BMC Genetics* **18**, 74–83 (2017).
77. Harland, C. et al. Frequency of mosaicism points towards mutation-prone early cleavage cell divisions. Preprint at bioRxiv http://biorxiv.org/content/early/2016/10/09/079863 (2016).
78. 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).
This study is one of the first demonstrations of how the availability of genome-wide SNP genotypes for large populations can be used to effectively identify haplotypes harbouring EL mutations affecting fertility.
79. 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).
80. Kadri, N. K. 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.* **10**, e1004049 (2014).
81. Sahana, G., Nielsen, U. S., Aamand, G. P., Lund, M. P. & Guldbrandtsen, B. Novel harmful recessive haplotypes identified for fertility traits in Nordic Holstein cattle. *PLOS ONE* **8**, e82909 (2013).
82. 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).
83. Pausch, H. et al. Homozygous haplotype deficiency reveals deleterious mutations compromising

reproductive and rearing success in cattle. *BMC Genomics* **16**, 312 (2015).
84. Häggman, J. & Uimari, P. Novel harmful recessive haplotypes for reproductive traits in pigs. *J. Anim. Breed. Genet.* **134**, 129–135 (2017).
85. Derks, M. F. L. et al. A systematic survey to identify lethal recessive variation in highly managed pig populations. *BMC Genomics* **18**, 858–870 (2017).
86. 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**, e92769 (2014).
87. Sonstegard, T. S. et al. Identification of a nonsense mutation in *CWC15* associated with decreased reproductive efficiency in Jersey cattle. *PLOS ONE* **8**, e54872 (2013).
88. Fritz, S. et al. An initiator codon mutation in *SDE2* causes recessive embryonic lethality in Holstein cattle. *J. Dairy Sci.* **101**, 6220–6231 (2018).
89. Martinez, V., Bünger, L. & Hill, W. G. Analysis of response to 20 generations of selection for body composition in mice: fit to infinitesimal model assumptions. *Genet. Sel. Evol.* **32**, 3 (2000).
90. Henderson, C. R. Sire evaluation and genetic trends. *J. Anim. Sci.* **1973**, 10–41 (1973).
91. Wright, S. Coefficients of inbreeding and relationship. *Am. Nat.* **56**, 330–338 (1922).
92. Malecot, G. *Les mathématiques de l'hérédité* (Masson, Paris, 1948).
93. Ritland, K. Estimators of pairwise relatedness and individual inbreeding coefficients. *Genet. Res.* **67**, 175–185 (1996).
94. Toro, M. et al. Estimation of coancestry in Iberian pigs using molecular markers. *Conserv. Genet.* **3**, 309–320 (2002).
95. Garant, D. & Kruuk, L. E. B. How to use molecular marker data to measure evolutionary parameters in wild populations. *Mol. Ecol.* **14**, 1843–1859 (2005).
96. Aguilar, I. et al. A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* **93**, 743–752 (2010).
This study is one of the first demonstrations that pedigree, genomic and phenotypic information from millions of animals could be combined in a single-step approach to predict GEBVs.
97. Christensen, O. F. & Lund, M. S. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* **42**, 2 (2010).
98. Liu, Z., Goddard, M. E., Reinhardt, F. & Reents, R. A single-step genomic model with direct estimation of marker effects. *J. Dairy Sci.* **97**, 5833–5850 (2014).
99. Fernando, R. L., Dekkers, J. C. & Garrick, D. J. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. *Genet. Sel. Evol.* **46**, 50 (2014).
100. Fernando, R. L., Cheng, H., Golden, B. L. & Garrick, D. J. Computational strategies for alternative single-step Bayesian regression models with large numbers of genotyped and non-genotyped animals. *Genet. Sel. Evol.* **48**, 96 (2016).
101. Misztal, I. & Legarra, A. Invited review: efficient computation strategies in genomic selection. *Animal* **11**, 731–736 (2017).
102. García-Ruiz, A. et al. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proc. Natl Acad. Sci. USA* **113**, 3995–4004 (2016).
This study demonstrates the impact of GS on genetic gain in the US dairy industry.
103. Van Eenennaam, A. L., van der Werf, J. H. & Goddard, M. E. The value of using DNA markers for beef bull selection in the seedstock sector. *J. Anim. Sci.* **89**, 307–320 (2011).
104. Lourenco, D. A. et al. Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus. *J. Anim. Sci.* **93**, 2653–26662 (2015).
105. Abell, C. E., Dekkers, J. C. M., Rothschild, M. F., Mabry, J. W. & Stalder, K. J. Total cost estimation for implementing genome-enabled selection in a multi-level swine production system. *Genet. Sel. Evol.* **46**, 32 (2014).
106. Shumbusho, F. et al. Economic evaluation of genomic selection in small ruminants: a sheep meat breeding program. *Animal* **10**, 1033–1041 (2016).
107. Brito, L. F. et al. Prediction of genomic breeding values for growth, carcass and meat quality traits in a multi-breed sheep population using a HD SNP chip. *BMC Genetics* **18**, 7 (2017).
108. Santos, B. F. S., van der Werf, J. H. J., Gibson, J. P., Byrne, T. J. & Amer, P. R. Assessment of the genetic and economic impact of performance recording and genotyping in Australian commercial sheep operations. *J. Anim. Breed. Genet.* **135**, 221–237 (2018).
109. Wolc, A. et al. Response and inbreeding from a genomic selection experiment in layer chicken. *Genet. Sel. Evol.* **47**, 59 (2015).
110. Gonzalez-Recio, O., Pryce, J. E., Haile-Mariam, M. & Hayes, B. J. Incorporating heifer feed efficiency in the Australian selection index using genomic selection. *J. Dairy Sci.* **97**, 3883 (2014).
111. Goddard, M. E. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* **136**, 245–257 (2009).
112. Grobet, L. et al. A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. *Nat. Genet.* **17**, 71–74 (1997).
113. Kambadur, R., Sharma, M., Smith, T. P. & Bass, J. J. Mutations in myostatin (*GDF8*) in double-muscled Belgian Blue and Piedmontese cattle. *Genome Res.* **7**, 910–916 (1997).
114. McPherron, A. C. & Lee, S. J. Double muscling in cattle due to mutations in the myostatin gene. *Proc. Natl Acad. Sci. USA* **94**, 12457–12461 (1997).
115. Grobet, L. et al. Molecular definition of an allelic series of mutations disrupting the myostatin function and causing double-muscling in cattle. *Mamm. Genome* **9**, 210–213 (1998).
116. Fujii, J. et al. Identification of a mutation in the porcine ryanodine receptor associated with malignant hyperthermia. *Science* **253**, 448–451 (1991).
117. Milan, D. et al. A mutation in *PRKAG3* associated with excess glycogen content in pig skeletal muscle. *Science* **288**, 1248–1251 (2000).
118. Van Laere, A. S. et al. A regulatory mutation in *IGF2* causes a major QTL effect on muscle growth in the pig. *Nature* **425**, 832–836 (2003).
119. Grisart, B. et al. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine *DGAT1* gene with major effect on milk yield and composition. *Genome Res.* **12**, 222–231 (2002).
120. Blott, S. et al. Molecular dissection of a QTL: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics* **1663**, 253–2666 (2003).
121. Cohen-Zinder, M. et al. Identification of a missense mutation in the bovine *ABCG2* gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Res.* **15**, 936–944 (2005).
122. Karim, L. et al. Variants modulating the expression of a chromosome domain encompassing *PLAG1* influence bovine stature. *Nat. Genet.* **43**, 405–413 (2011).
123. Bouwman, A. C. et al. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nat. Genet.* **50**, 362–367 (2018).
124. Habier, D., Fernando, R. L., Kizilkaya, K. & Garrick, D. J. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics* **12**, 186 (2011).
125. Erbe, M. et al. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J. Dairy Sci.* **95**, 4114–4129 (2012).
126. Druet, T. et al. Selection in action: dissecting the molecular underpinnings of the increasing muscle mass in Belgian Blue Cattle. *BMC Genomics* **15**, 796 (2014).
127. Hedrick, P. W. Heterozygote advantage: the effect of artificial selection in livestock and pets. *J. Hered.* **106**, 141–154 (2015).
128. Fasquelle, 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.* **5**, e1000666 (2009).
129. Sartelet, A. et al. A splice site variant in the bovine *RNF11* gene compromises growth and regulation of the inflammatory response. *PLOS Genet.* **8**, e1002581 (2012).
130. Sartelet, A. et al. Allelic heterogeneity of Crooked Tail Syndrome fits the balancing selection hypothesis. *Anim. Genet.* **43**, 591–594 (2012).
131. Galloway, S. M. et al. Mutations in an oocyte-derived growth factor gene (*BMP15*) cause increased ovulation rate and infertility in a dosage-sensitive manner. *Nat. Genet.* **25**, 279–283 (2000).
132. Hanrahan, J. P. et al. Mutations in the genes for oocyte-derived growth factors GDF9 and BMP15 are associated with both increased ovulation rate and sterility in Cambridge and Belcaler sheep (*Ovis aries*). *Biol. Reprod.* **70**, 900–909 (2004).
133. Rupp, R. et al. A point mutation in suppressor of cytokine signalling 2 (*SOCS2*) increases the susceptibility to inflammation of the mammary gland while associated with higher body weight and size and higher milk production in a sheep model. *PLOS Genet.* **11**, e1005629 (2015).
134. Johnston, S. E. et al. Life history trade-offs at a single locus maintains sexually selected genetic variation. *Nature* **502**, 93–95 (2013).
135. Cockett, N. E. et al. Polar overdominance at the ovine callipyge locus. *Science* **273**, 236–238 (1996).
136. Freking, B. A. et al. Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. *Genome Res.* **12**, 14966–11506 (2002).
137. Barson, N. J. et al. Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature* **528**, 405–498 (2015).
This paper describes a remarkable example of balancing selection occurring in a natural population as a result of antagonistic selection in males and females.
138. Ayllon, F. et al. The *vgll3* locus controls age at maturity in wild and domesticated atlantic salmon (*Salmo salar* L.) males. *PLOS Genet.* **11**, e1005628 (2015).
139. Utsunomiya, Y. T. et al. A *PLAG1* mutation contributed to stature recovery of modern cattle. *Sci. Rep.* **7**, 17140 (2017).
140. Fortes, M. R. S. et al. Evidence for pleiotropism and recent selection in the PLAG1 region in Australian Beef cattle. *Anim. Genet.* **44**, 6636–6647 (2013).
141. Pszczola, M. & Calus, M. P. Updating the reference population to achieve constant genomic prediction reliability across generations. *Animal* **10**, 1018–1024 (2016).
142. Calus, M. P. Right-hand-side updating for fast computing of genomic breeding values. *Genet. Sel. Evol.* **46**, 24 (2014).
143. Calus, M. P., Bouwman, A. C., Schrooten, C. & Veerkamp, R. F. Efficient genomic prediction based on whole-genome sequence data using split-and-merge Bayesian variable selection. *Genet. Sel. Evol.* **48**, 49 (2016).
144. Wang, T. et al. Application of a Bayesian non-linear model hybrid scheme to sequence data for genomic prediction and QTL mapping. *BMC Genomics* **18**, 618 (2017).
145. van den Berg, I. et al. Multi-breed genomic prediction using Bayes R with sequence data and dropping variants with a small effect. *Genet. Sel. Evol.* **49**, 70 (2017).
146. Brøndum, R. F. et al. Quantitative trait loci markers derived from whole genome sequence data increases the reliability of genomic prediction. *J. Dairy Sci.* **98**, 4107–4116 (2015).
147. Veerkamp, R. F., Bouwman, A. C., Schrooten, C. & Calus, M. P. Genomic prediction using preselected DNA variants from a GWAS with whole-genome sequence data in Holstein-Friesian cattle. *Genet. Sel. Evol.* **48**, 95 (2016).
148. VanRaden, P. M., Tooker, M. E., O'Connell, J. R., Cole, J. B. & Bickhart, D. M. Selecting sequence variants to improve genomic predictions for dairy cattle. *Genet. Sel. Evol.* **49**, 32 (2017).
149. Long, H. K., Prescott, S. L. & Wysocka, J. Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* **167**, 1170–1187 (2016).
150. Wragg, D. et al. Genome-wide analysis reveals the extent of EAV-HP integration in domestic chicken. *BMC Genomics* **16**, 784–794 (2015).
151. Kemper, K. E., Jayes, B. J., Daetwyler, H. D. & Goddard, M. E. How old are quantitative trait loci and how widely do they segregate? *J. Anim. Breed. Genet.* **132**, 121–134 (2015).
152. Yokota, S. et al. Contributions of *FASN* and *SCD* gene polymorphisms on fatty acid composition in muscle from Japanese Black cattle. *Anim. Genet.* **43**, 790–792 (2012).
153. Zhang, W. et al. Genome-wide association studies for fatty acid metabolic traits in five divergent pig populations. *Sci. Rep.* **6**, 24718 (2016).
154. Bolormaa, S. et al. Detailed phenotyping identifies genes with pleiotropic effects on body composition. *BMC Genomics* **17**, 224 (2016).

155. Momozawa, Y. et al. IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nat. Commun.* **9**, 2427 (2018).
156. Fang, L. et al. Use of biological priors enhances understanding of genetic architecture and genomic prediction of complex traits within and between dairy cattle breeds. *BMC Genomics* **18**, 604 (2017).
157. Fragomeni, B. O., Lourenco, D. A. L., Masuda, Y., Legarra, A. & Misztal, I. Incorporation of causative quantitative trait nucleotides in single-step GBLUP. *Genet. Sel. Evol.* **49**, 59 (2017).
158. Hayes, B. J., Bowman, P. J., Chamberlain, A. C., Verbyla, K. & Goddard, M. E. Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genet. Sel. Evol.* **41**, 51 (2009).
159. Bolormaa, S. et al. Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality traits in *Bos taurus*, *Bos indicus*, and composite beef cattle. *J. Anim. Sci.* **91**, 3088–3104 (2013).
160. Rolf, M. M. et al. Comparison of Bayesian models to estimate direct genomic values in multi-breed commercial beef cattle. *Genet. Sel. Evol.* **47**, 23 (2015).
161. Kemper, K. E. et al. Improved precision of QTL mapping using a nonlinear Bayesian method in multibreed population leads to greater accuracy of across-breed genomic predictions. *Genet. Sel. Evol.* **47**, 29 (2015).
162. Lu, D. et al. Accuracy of genomic predictions for feed efficiency traits of beef cattle using 50K and imputed HD genotypes. *J. Anim. Sci.* **94**, 1342–1353 (2016).
163. Hamidi Hay, E. & Roberts, A. Genomic prediction and genome-wide association analysis of female longevity in a composite beef cattle breed. *J. Anim. Sci.* **95**, 1467–1471 (2017).
164. Pausch, H. et al. Meta-analysis of sequence-based association studies across three cattle breeds reveals 25 QTL for fat and protein percentages in milk at nucleotide resolution. *BMC Genomics* **18**, 853 (2017).
165. Bosse, M. et al. Artificial selection on introduced Asian haplotypes shaped the genetic architecture in European commercial pigs. *Proc. Biol. Sci.* **282**, 20152019 (2015).
166. Sonesson, A. K., Woolliams, J. A. & Meuwissen, T. H. Genomic selection requires genomic control of inbreeding. *Genet. Sel. Evol.* **44**, 27 (2012).
167. Sun, C., VanRaden, P. M., O'Connell, J. R., Weigel, K. A. & Gianola, D. Mating programs including genomic relationships and dominance effects. *J. Dairy Sci.* **96**, 8014–8023 (2013).
168. Pryce, J. E., Hayes, B. J. & Goddard, M. E. Novel strategies to minimize progeny inbreeding while maximizing genetic gain using genomic information. *J. Dairy Sci.* **95**, 377–388 (2012).
169. Palmiter, R. D. et al. Dramatic growth of mice that develop from eggs microinjected with metallothionein-growth hormone fusion genes. *Nature* **300**, 611–615 (1982).
170. Hammer, R. E. et al. Production of transgenic rabbits, sheep and pigs by microinjection. *Nature* **315**, 680–683 (1985).
171. Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. & Campbell, K. H. Viable offspring derived from fetal and adult mammalian cells. *Nature* **385**, 810–813 (1997).
172. Tan, W., Proudfoot, C., Lillico, S. G. & Whitelaw, C. B. Gene targeting genome editing: from Dolly to editors. *Transgenic Res.* **25**, 273–287 (2016).
173. Kim, H. & Kim, J.-S. A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* **15**, 321–334 (2014).
174. Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* **168**, 1–17 (2017).
175. Van Eenennaam, A. L. Genetic modification of food animals. *Curr. Opin. Biotechnol.* **44**, 27–34 (2017).
176. Sakuma, T., Nakade, S., Sakane, Y., Suzuki, K. T. & Yamamoto, T. MMEJ-assisted gene knock-in using TALENs and CRISP-Cas9 with the PITCh systems. *Nat. Protoc.* **11**, 118–133 (2016).
177. Suzuki, K. et al. In vivo genome editing via CRISP/Cas9 mediated homology-independent targeted integration. *Nature* **540**, 144–149 (2016).
178. Laible, G., Wei, J. & Wagner, S. Improving livestock for agriculture — technological progress from random transgenesis to precision genome editing heralds a new era. *Biotechnol. J.* **10**, 109–120 (2015).
179. Lievens, A., Petrillo, M. & Querci Patak, M. A. Genetically modified animals: options and issues for traceability and enforcement. *Trends Food Sci. Technol.* **44**, 159–176 (2015).
180. Pirottin, D. et al. Transgenic engineering of male-specific muscular hypertrophy. *Proc. Natl Acad. Sci. USA* **102**, 6413–66418 (2005).
181. Wang, H. et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910–918 (2013).
182. Cong, L. et al. Multiplex genome engineering using CRISP/Cas systems. *Science* **339**, 819–823 (2013).
183. Jenko, J. et al. Potential of promotion of alleles by genome editing to improve quantitative traits in livestock breeding programs. *Genet. Sel. Evol.* **47**, 55 (2015).
This paper describes a strategy to combine genome editing and GS to accelerate genetic gains in livestock.
184. Kasinathan, P. et al. Acceleration of genetic gain in cattle by reduction of generation interval. *Sci. Rep.* **5**, 8674–86766 (2015).
185. Blard, G., Zhang, Z., Coppieters, W. & Georges, M. Identifying cows with subclinical mastitis by bulk SNP genotyping of tank milk. *J. Dairy Sci.* **95**, 4109–4113 (2012). **This paper describes a method to identify cows with subclinical mastitis by SNP genotyping tank milk**.
186. Hogeveen, H., Huijps, K. & Lam, T. J. Economic aspects of mastitis: new developments. *NZ Vet. J.* **59**, 16–23 (2011).
187. Nicoloso, L., Crepaldi, P., Mazza, R., Ajmone-Marsan, P. & Negrini, R. Recent advance in DNA-based traceability and authentication of livestock meat PDO and PGI products. *Recent Pat. Food Nutr. Agric.* **5**, 9–18 (2013).
188. Ross, E. M., Moate, P. J., Marett, L. C., Cocks, B. G. & Hayes, B. J. Metagenomic predictions: from microbiome to complex health and environmental phenotypes in human and cattle. *PLOS ONE* **8**, e73056 (2013).
189. Kittelmann, S. et al. Two different bacterial community types are linked with the low-methane emission trait in sheep. *PLOS ONE* **9**, e103171 (2014).
190. Wang, M., Pryce, J. E., Savin, K. & Hayes, B. J. Prediction of residual feed intake from genome & metagenome profiles in first lactation Holstein-Friesian dairy cattle. *Proc. Assoc. Adv. Breed. Genet.* **21**, 89–92 (2015).
191. Frantz, L. A. F. et al. Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nat. Genet.* **47**, 1141–1148 (2015).
This paper describes a thorough analysis of more than 100 whole-genome sequences of pigs, which strongly suggests that long-term gene flow between wild and domestic pigs counteracted by recurrent selection for domestic traits creates 'islands of domestication' in the genome.
192. Wade, C. M. et al. The mosaic structure of variation in the laboratory mouse genome. *Nature* **420**, 574–578 (2002).
193. Sankararaman, S. et al. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354–357 (2014).
194. Berry, S. et al. A triad of highly divergent polymeric immunoglobulin receptor (PIGR) haplotytpes with major effect on IgA concentration in cow's milk. *PLOS ONE* **8**, e57219 (2013).
195. Ai, H. et al. Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat. Genet.* **47**, 217–225 (2015).
196. Rubin, C.-J. et al. Whole genome resequencing reveals loci under selection during chicken domestication. *Nature* **464**, 587–591 (2010).
This paper reports the identification of convincing signatures of selective sweeps following poultry domestication.
197. Carneiro, M. et al. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* **345**, 1074–1079 (2014).
This study provides strong evidence that polygenic adaptation plays a major role in shaping the phenotype of domestic animals in response to human needs.
198. Dong, Y. et al. Reference genome of wild goat (*Capra aegagrus*) and sequencing of goat breeds provide insight into genic basis of goat domestication. *BMC Genomics* **16**, 431–442 (2015).
199. Park, S. D. E. et al. Genome sequencing of the extinct Eurasian wild aurochs, *Bos primigenius*, illuminates the phylogeography and evolution of cattle. *Genome Biol.* **16**, 234–249 (2015).
200. Field, Y. et al. Detection of human adaptation during the past 2,000 years. *Science* **354**, 760–764 (2016).
201. Nejati-Javaremi, A., Smith, C. & Gibson, J. Effect of total allelic relationship on accuracy of evaluation and response to selection. *J. Anim. Sci.* **75**, 1738–1745 (1997).
202. VanRaden, P. M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **91**, 4414–4423 (2008).
203. Daetwyler, H. D., Villanueva, B. & Wooliams, J. A. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLOS ONE* **3**, e3395 (2008).
204. Hayes, B. J., Visscher, P. M. & Goddard, M. E. Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res.* **91**, 47–60 (2009).
205. Pryce, J. E. et al. Genomic selection using a multi-breed, across-country reference population. *J. Dairy Sci.* **94**, 2625–26630 (2011).
206. Kijas, J. W. et al. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLOS Biol.* **10**, e1001258 (2012).
207. Moghaddar, N., Gore, K. P., Daetwyler, H. D., Hayes, B. J. & van der Werf, J. H. J. Accuracy of genotype imputation based on random and selected reference sets in purebred and crossbred sheep populations and its effect on accuracy of genomic prediction. *Genet. Sel. Evol.* **47**, 97 (2015).
208. Lee, S. H., Weerasinghe, W. M., Wray, N. R., Goddard, M. E. & van der Werf, J. H. Using information of relatives in genomic prediction to apply effective stratified medicine. *Sci. Rep.* **7**, 42091 (2017).
209. Hill, W. G. & Weir, B. S. Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genet. Res.* **93**, 47–64 (2011).
210. Pursel, V. G. et al. Genetic engineering of livestock. *Science* **24**, 1281–1288 (1989).
211. Pursel, V. G., Hammer, R. E., Bolt, D. J., Palmiter, R. D. & Brinster, R. L. Integration, expression and germ-line transmission of growth-related genes in pigs. *J. Reprod. Fertil. Suppl.* **41**, 77–87 (1990).
212. Rexroad, C. E. et al. Production of transgenic sheep with growth-regulating genes. *Mol. Reprod. Dev.* **1**, 164–169 (1989).
213. Nottle, M. B. et al. in *Transgenic Animals in Agriculture* (eds Murray, J. D., Anderson, G. B., Oberbauer, A. M. & McGloughin, M. M.) 145–156 (CABI Publishing, Oxon, UK, 1999).
214. Adams, N. R., Briegel, J. R. & Ward, K. A. The impact of a transgene for ovine growth hormone on the performance of two breeds of sheep. *J. Anim. Sci.* **80**, 2325–2333 (2002).
215. Pursel, V. G. et al. in *Transgenic Animals in Agriculture* (eds Murray, J. D., Anderson, G. B., Oberbauer, A. M. & McGloughin, M. M.) 131–144 (CABI Publishing, Oxon, UK, 1999).
216. Sutrave, P., Kelly, A. M. & Hughes, S. H. Ski can cause selective growth of skeletal muscle in transgenic mice. *Genes Dev.* **4**, 1462–1472 (1990).
217. Grobet, L. et al. Modulating skeletal muscle mass by postnatal, muscle-specific inactivation of the myostatin gene. *Genesis* **35**, 227–238 (2003).
218. Luo, J. et al. Efficient generation of myostatin (*MSTN*) biallelic mutations in cattle using zinc finger nucleases. *PLOS ONE* **9**, e95225 (2014).
219. Proudfoot, C. et al. Genome edited sheep and cattle. *Transgenic Res.* **24**, 147–153 (2015).
220. Ni, W. et al. Efficient gene knockout in goats using CRISPR/Cas9 system. *PLOS ONE* **9**, e106718 (2014).
221. Han, H. et al. One-step generation of myostatin gene knockout sheep via the CRISPR/Cas9 system. *Front. Agric. Sci. Eng.* **1**, 2–5 (2014).
222. Tessanne, K. et al. Production of transgenic calves expressing an shRNA targeting myostatin. *Mol. Reprod. Dev.* **79**, 176–185 (2012).
223. Lee, S. J. Quadrupling muscle mass in mice by targeting TGF-beta signaling pathways. *PLOS ONE* **2**, e789 (2007).
224. Saeki, K. et al. Functional expression of a delta 12 fatty acid desaturase transgene from spinach in transgenic pigs. *Proc. Natl Acad. Sci. USA* **101**, 6361–6366 (2004).
225. Lai, L. et al. Generation of cloned transgenic pigs rich in omega-3 fatty acids. *Nat. Biotechnol.* **24**, 345–436 (2006).
226. Wu, X. et al. Production of cloned transgenic cow expressing omega-3 fatty acids. *Transgenic Res.* **21**, 537–543 (2012).
227. Zhang, P. et al. Handmade cloned transgenic sheep rich in omega-3 fatty acids. *PLOS ONE* **8**, e55941 (2013).

228. Zheng, Q. et al. Reconstitution of UCP1 using CRISP/Cas9 in the white adipose tissue of pigs decreases fat deposition and improves thermogenic capacity. *Proc. Natl Acad. Sci. USA* **114**, E9474–E9482 (2017).
229. Berg, F., Gustafson, U. & Andersson, L. The uncoupling protein 1 gene (UCP1) is disrupted in the pig lineage: a genetic explanation for poor thermoregulation in piglets. *PLOS Genet.* **2**, e129 (2006).
230. Brophy, B. et al. Cloned transgenic cattle produce milk with higher levels of beta-casein and kappa-casein. *Nat. Biotechnol.* **21**, 157–162 (2003).
231. Martin, P., Szymanowska, M., Zwierzchowski, L. & Leroux, C. The impact of genetic polymorphisms on the protein composition of ruminant milks. *Reprod. Nutr. Dev.* **42**, 433–459 (2002).
232. Yu, S. et al. Highly efficient modification of beta-lactoglobulin (*BLG*) gene via zinc-finger nucleases in cattle. *Cell Res.* **21**, 1638–1640 (2011).
233. Jabed, A., Wagner, S., McCracken, J., Wells, D. N. & Labile, G. Targeted microRNA expression in dairy cattle directs production of beta-lactoglobulin-free, high-casein milk. *Proc. Natl Acad. Sci. USA* **109**, 16811–16816 (2012).
234. Cui, C. et al. Gene targeting by TALEN-induced homologous recombination in goats directs production of beta-lactoglobulin-free, high human lactoferrin milk. *Sci. Rep.* **5**, 10482 (2015).
235. Zhu, H. et al. Generation of beta-lactoglobulin-modified transgenic goats by homologous recombination. *FEBS J.* **282**, 4600–4613 (2016).
236. Wheeler, M. B., Bleck, G. T. & Donovan, S. M. Transgenic alteration of sow milk to improve piglet growth and health. *Reprod. Suppl.* **58**, 313–324 (2001).
237. Wang, J. et al. Expression and characterization of bioactive recombinant human alpha-lactalbumin in the milk of transgenic cloned cows. *J. Dairy Sci.* **91**, 4466–4476 (2008).
238. Jost, B., Vilotte, J. L., Duluc, I., Rodeau, J. L. & Freund, J. N. Production of low-lactose milk by ectopic expression of intestinal lactase in the mouse mammary gland. *Nat. Biotechnol.* **17**, 160–164 (1999).
239. Reh, W. A. et al. Using a stearoyl-CoA desaturase transgene to alter milk fatty acid composition. *J. Dairy Sci.* **87**, 3510–3514 (2004).
240. Damak, S., Su, H., Jay, N. P. & Bullock, D. W. Improved wool production in transgenic sheep expressing insulin-like growth factor 1. *Biotechnology* **14**, 185–188 (1996).
241. Bawden, C. S., Powell, B. C., Walker, S. K. & Rogers, G. E. Expression of a wool intermediate filament keratin transgene in sheep fibre alters structure. *Transgenic Res.* **7**, 273–287 (1998).
242. Bawden, C. S. et al. Expression of bacterial cysteine biosynthesis genes in transgenic mice and sheep: toward a new in vivo amino acid biosynthesis pathway and improved wool growth. *Transgenic Res.* **4**, 87–104 (1995).
243. Maga, E. A. et al. Production and processing of milk from transgenic goats expressing human lysozyme in the mammary gland. *J. Dairy Sci.* **89**, 518–524 (2006).
244. Liu, X. et al. Generation of mastitis resistance in cows by targeting human lysozyme gene to beta-casein locus using zinc-finger nucleases. *Proc. R. Soc. B* **281**, 20133368 (2014).
245. Wall, R. J. et al. Genetically enhanced cows resist intramammary *Staphylococcus aureus* infection. *Nat. Biotechnol.* **23**, 445–451 (2005).
246. Liu, X. et al. Zinc-finger nickase-mediated insertion of the lysostaphin gene into the beta-casein locus in cloned cows. *Nat. Commun.* **4**, 2565 (2013).
247. Dunham, R. A. et al. Enhanced bacterial disease resistance of transgenic channel catfish *Ictalurus punctatus* possessing cecropin genes. *Mar. Biotechnol.* **4**, 338–344 (2002).
248. Su, F. et al. Generation of transgenic cattle expressing human beta-defensin 3 as an approach to reducing susceptibility to *Mycobacterium bovis* infection. *FEBS J.* **283**, 776–790 (2016).
249. Yang, X. et al. Overexpression of porcine beta-defensin 2 enhances resistance to *Actinobacillus pleuropneumoniae* infection in pigs. *Infect. Immun.* **83**, 2836–2843 (2015).
250. Wu, H. et al. TALE nickase-mediated SP110 knockin endows cattle with increased resistance to tuberculosis. *Proc. Natl Acad. Sci. USA* **112**, 1530–1539 (2015).
251. Pan, H. et al. Ipr1 gene mediates innate immunity to tuberculosis. *Nature* **434**, 767–772 (2005).
252. Tosh, K. et al. Variants in the *SP110* gene are associated with genetic susceptibility to tuberculosis in West Africa. *Proc. Natl Acad. Sci. USA* **103**, 10364–10368 (2006).
253. Fox, G. J. et al. Polymorphisms of *SP110* are associated with bothpulmonary and extra-pulmonary tuberculosis among the Vietnamese. *PLOS ONE* **9**, e99496 (2014).
254. Hu, W. et al. Significant resistance to the infection of foot-and-mouth disease virus in shRNA transgenic pig. *Transgenic Res.* **21**, 901–925 (2012).
255. Muller, M., Brenig, B., Winnacker, E. L. & Brem, G. Transgenic pigs carrying cDNA copies encoding the murine Mx1 protein which confers resistance to influenza virus infection. *Gene* **121**, 263–270 (1992).
256. Yan, Q. et al. Production of transgenic pigs overexpressing the antiviral gene *Mx1*. *Cell Regen.* **3**, 11–22 (2014).
257. Lyall, J. et al. Suppression of avian influenza transmission in genetically modified chickens. *Science* **331**, 223–226 (2011).
258. Clements, J. E. et al. Development of transgenic sheep that express the visan virus envelope gene. *Virology* **200**, 370–380 (1994).
259. Crittenden, L. B. & Salter, D. W. A transgene, *alv6*, that expresses the envelope of subgroup A avian leucosis virus reduces the rate of congenital transmission of a field strain of avian leucosis virus. *Poult. Sci.* **71**, 799–806 (1992).
260. Whitworth, K. M. et al. Use of CRISPR/Cas9 system to produce genetically engineered pigs from in vitro derived oocytes and embryos. *Biol. Reprod.* **91**, 1–13 (2014).
261. Burkard, C. et al. Precision engineering for PRRSV resistance in pigs: macrophages from genome edited pigs lacking CD163 SRCR5 domain are fully resistant to both PRRSV genotypes while maintaining biological function. *PLOS Pathog.* **13**, e1006206 (2017).
262. Lillico, S. G. et al. Live pigs produced from genome edited zygotes. *Sci. Rep.* **3**, 2847–2851 (2013).
263. Lillico, S. G. et al. Mammalian interspecies substitution of immune modulatory alleles by genome editing. *Sci. Rep.* **6**, 21645–21650 (2016).
264. Richt, J. A. et al. Production of cattle lacking prion protein. *Nat. Biotechnol.* **25**, 132–138 (2007).
265. Yu, G. et al. Functional disruption of the prion protein gene in cloned goats. *J. Gen. Virol.* **87**, 1019–1027 (2006).
266. Denning, C. et al. Deletion of the alpha(1,3)galactosyl transferase (GGTA1) gene and the prion protein (PrP) gene in sheep. *Nat. Biotechnol.* **19**, 559–562 (2001).
267. Golding, M. C., Long, C. R., Carmell, M. A., Hannon, G. J. & Westhusin, M. E. Suppression of prion protein in livestock by RNA interference. *Proc. Natl Acad. Sci. USA* **103**, 5285–5290 (2006).
268. Wongsrikeao, P. et al. Combination of the somatic cell nuclear transfer method and RNAi technology for the production of a prion gene-knockdown calf using plasmid vectors harbouring the U6 or tRNA promoter. *Prion* **5**, 39–46 (2011).
269. Benestad, S. L., Anstbø, L., Tranulis, M. A., Espenes, A. & Olsaker, I. Healthy goats naturally devoid of prion protein. *Vet. Res.* **43**, 87–91 (2012).
270. Willyard, C. Putting sleeping sickness to bed. *Nat. Med.* **17**, 14–17 (2011).
271. Genovese, G. et al. Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science* **329**, 841–845 (2010).
272. Hew, C. L., Davies, P. L. & Fletcher, G. Antifreeze protein gene transfer in Atlantic salmon. *Mol. Mar. Biol. Biotechnol.* **1**, 309–317 (1992).
273. Davies, P. L. & Hew, C. L. Biochemistry of fish antifreeze proteins. *FASEB J.* **4**, 2460–2468 (1990).
274. Carlson, D. F. et al. Production of hornless dairy cattle from genome-edited cell lines. *Nat. Biotechnol.* **34**, 479–481 (2016).
This paper illustrates the engineering of a desirable phenotype in livestock by TALEN-mediated allele swapping and SCNT.
275. Medugorac, I. et al. Bovine polledness — an autosomal dominant trait with allelic heterogeneity. *PLOS ONE* **7**, e39477 (2012).
276. Allais-Bonnet, A. et al. Novel insights into the bovine polled phenotype and horn ontogenesis in Bovidae. *PLOS ONE* **8**, e63512 (2013).
277. Rothammer, S. et al. The 80-Kb DNA duplication on BTA1 is the only remaining candidate mutation for the polled phenotype of Friesian origin. *Genet. Sel. Evol.* **46**, 44 (2014).
278. Golovan, S. P. et al. Pigs expressing salivary phytase produce low-phosphorus manure. *Nat. Biotechnol.* **19**, 741–745 (2001).

Acknowledgements
C.C. is Senior Research Associate of the Fonds de la Recherche Scientifique (FRS-FNRS). Research in animal genomics conducted by the authors is funded by the European Research Council (ERC) Advanced DAMONA, H2020 GpE and WALInnov CAUSEL grants to M.G. and the DGARNE Rilouke and ULiège RetroBlue grants to C.C. The authors are grateful to T. Druet for fruitful discussions and comments on the manuscript and to M. Goddard for excellent discussions over the years.

Author contributions
All authors researched data for the article, made substantial contributions to discussions of the content and reviewed and/or edited the manuscript before submission. M.G. and B.H. wrote the article.

Competing interests
The authors declare no competing interests.

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

Reviewer information
Nature Reviews Genetics thanks D. Garrick and A. Legarra for their contribution to the peer review of this work.

Supplementary information
Supplementary information is available for this paper at https://doi.org/10.1038/s41576-018-0082-2.

RELATED LINKS
dbSNP: https://www.ncbi.nlm.nih.gov/projects/SNP
International Mouse Phenotype Consortium: http://www.mousephenotype.org
NCBI Genome database: https://www.ncbi.nlm.nih.gov/genome
Online Mendelian Inheritance in Animals: http://www.omia.org/home
Online Mendelian Inheritance in Man: https://www.omim.org