

# Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of Mongolian yaks

Ivica Medugorac<sup>1</sup>, Alexander Graf<sup>2</sup>, Cécile Grohs<sup>3</sup>, Sophie Rothhammer<sup>1</sup>, Yondon Zagdsuren<sup>4</sup>, Elena Gladyr<sup>5</sup>, Natalia Zinovieva<sup>5</sup>, Johanna Barbieri<sup>6,7</sup>, Doris Seichter<sup>8</sup>, Ingolf Russ<sup>8</sup>, André Eggen<sup>9</sup>, Garrett Hellenthal<sup>10</sup>, Gottfried Brem<sup>11</sup>, Helmut Blum<sup>2</sup>, Stefan Krebs<sup>2</sup>, Aurélien Capitan<sup>3,12</sup>

1 Chair of Animal Genetics and Husbandry, Ludwig-Maximilians-University Munich, Munich, Germany

2 Laboratory for Functional Genome Analysis, Gene Center, Ludwig-Maximilians-University Munich, Munich, Germany

3 GABI, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France

4 Mongolian Yak Society, Sum, Ulan Bator, Mongolia

5 Center of Biotechnology and Molecular Diagnostics of the L.K. Ernst Institute of Animal Husbandry, Podolsk District Moscow Region, Russia

6 Institut National de la Recherche Agronomique (INRA), UMR 1388 Génétique, Physiologie et Systèmes d'Élevage, GeT-PlaGe Genomic Facility, Castanet-Tolosan, France

7 Université de Toulouse INPT ENSAT, UMR1388 Génétique, Physiologie et Systèmes d'Élevage, Castanet-Tolosan, France

8 Tierzuchtforschung e.V. München, Grub, Germany

9 AgriGenomics, Illumina, San Diego, USA

10 Genetics Institute, Faculty of Life Sciences, University College London, UK

11 Institute of Animal Breeding and Genetics Department for Biomedical Sciences University of Veterinary Medicine, Vienna, Austria

12 ALLICE, Paris, France

Correspondence:

Ivica Medugorac

E-mail: [ivica.medjugorac@gen.vetmed.uni-muenchen.de](mailto:ivica.medjugorac@gen.vetmed.uni-muenchen.de)

Aurélien Capitan

E-mail: [aurelien.capitan@inra.fr](mailto:aurelien.capitan@inra.fr)

35 **The yak is remarkable for its adaptation to high altitude and occupies a central place**  
 36 **in the economy of mountainous regions of Asia. At lower elevation, it is common to**  
 37 **hybridize yaks with cattle to combine hardiness and productivity. Hybrid males are,**  
 38 **however, sterile, preventing the establishment of stable hybrids but not a limited**  
 39 **introgression after backcrossing several generations of females. Here we inferred**  
 40 **bovine haplotypes in the genomes of 76 Mongolian yaks using high-density SNP**  
 41 **genotyping and whole genome sequencing. These inherited ~1.3% of their genome**  
 42 **from bovine ancestors after nearly continuous admixture over at least the last 1500**  
 43 **years. The introgressed regions were markedly enriched in genes involved in nervous**  
 44 **system development and function, and particularly in glutamate metabolism and**  
 45 **neurotransmission. A novel polled (i.e. hornless) mutation originating from Mongolian**  
 46 **Turano cattle was also identified. Our results support that introgressive hybridization**  
 47 **contributed to the improvement of yak management and breeding.**

48 Hybridization is not unusual in nature. Although interspecific hybrids are rare at the  
 49 population level, around 10% of animals and 25% of plants are known to occasionally  
 50 hybridize with other species<sup>1</sup>. Evaluation of the genome-wide magnitude of this phenomenon  
 51 has only recently become possible. The first results show that limited introgressions of the  
 52 genome are widespread with a potentially important role in environmental adaptation, as  
 53 suggested by incorporation of genetic material from local species into the genome of  
 54 colonizing species (e.g. Neanderthal in non-African humans<sup>2,3</sup> or *Zea mays mexicana* in  
 55 maize<sup>4</sup>). Successful analyses so far identified genes under selection but, because of some  
 56 limitations, they were rarely able to determine precisely the nature of the selective pressure,  
 57 to identify genes pathways under selection, and to pinpoint causative polymorphisms.

58 Yak and cattle diverged approximately 4.9 million years ago<sup>5</sup>. Despite anatomical and  
 59 physiological differences, both species are raised in mixed herds in Central Asia and share  
 60 similar husbandry practices. Recent studies have reported several examples of gene flow  
 61 from cattle to yaks<sup>6-8</sup> and the existence of hornless (i.e. polled) animals in both species which  
 62 do not carry the previously reported Celtic and Friesian *POLLED* mutations<sup>9-12</sup>  
 63 (**Supplementary Table 1**) raises the question of a common origin for this phenotype. For this  
 64 reason and because of the large genomic datasets available for cattle, analysis of bovine  
 65 introgression in Mongolian yaks represents an appealing model to identify exchanges of  
 66 traits of interest between domesticated species.

67 To get a first picture of bovine introgression in Mongolian yaks, we sequenced two individual  
 68 genomes (YAK13, homozygous polled and YAK40, horned) and plotted the frequency of yak  
 69 and bovine alleles for all positions of yak-specific SNPs in 70-kb sliding windows  
 70 (**Supplementary Fig. 1**; see methods). Considering as yak-specific, the variants which were  
 71 (i) homozygous for the alternate allele in the yak reference genome<sup>5</sup> (hereafter named  
 72 YAKQIU) and (ii) absent from 235 bovine genomes (**Supplementary Table 2**)<sup>13</sup>, we  
 73 estimated that at least 1.73% and 1.22% of YAK13 and YAK40 genomes were of bovine  
 74 origin (**Supplementary Fig. 1**). To identify introgression in YAKQIU itself, we used the  
 75 number of yak-specific SNPs per 70-kb interval as an indicator and estimated a bovine  
 76 proportion of 1.06% even in the yak reference sequence. To validate our WGS approach for  
 77 inference of local-ancestry, two regions suggesting homozygous cattle introgression in  
 78 YAKQIU were PCR amplified and sequenced in YAK13 and in 12 related species  
 79 (**Supplementary Table 3**). Multiple alignment and phylogenetic analyses of 2,191-bp  
 80 sequence data clustered YAK13 with gaur, banteng, bison and wisent in accordance with the  
 81 phylogeny of the tribe Bovini<sup>14</sup> whereas YAKQIU clustered with cattle thus confirming cattle  
 82 introgression in the yak reference genome too (**Fig. 1**; **Supplementary Table 4**).

83 For a systematic analysis of cattle introgression in the Mongolian yak population, we  
 84 investigated the Illumina BovineHD Beadchip genotyping data (777 k SNPs) of 76 animals  
 85 originating from different localities (**Supplementary Table 1**). Analysis of SNPs mapping to  
 86 mtDNA (N=245) identified two yaks with deviating matrilineal ancestors whereas analysis of  
 87 SNPs mapping to BTAY (N=921) revealed an absence of bovine Y chromosome in this panel

88 (**Supplementary Fig. 2, 3**). We then applied a robust forward-backward algorithm  
 89 (*RFMix*)<sup>2,15</sup> to screen for the presence of cattle haplotypes in their autosomal genomes with  
 90 the exception of the Major Histocompatibility locus (for reasons see **Supplementary Note**)  
 91 using i) WGS data from three yaks to determine alleles present in yaks, ii) a six-bovini  
 92 consensus to determine ancestral states for all SNPs, and iii) additional genotyping data from  
 93 384 cattle (**Supplementary Table 5**) as a reference panel assumed to harbour no yak  
 94 ancestry (**Supplementary Fig. 4**).

95 The proportion of the genome inferred to be of cattle ancestry ranged between 0.67% and  
 96 2.82% (mean = 1.31 ±0.36; FDR=0.05) per animal (**Supplementary Table 6**;  
 97 **Supplementary Note**), a result consistent with a severe restriction of introgression by the  
 98 culling of most of the backcross calves and the persistence of hybrid male sterility up to the  
 99 third or fourth generation of backcross<sup>16</sup>. In total, as high as 33.2% of the bovine genome  
 100 was recovered from our panel of 76 yaks with noticeable variations between chromosomes  
 101 (**Fig. 2a**; **Supplementary Tables 6 and 7**). In agreement with the “large X-effect” on hybrid  
 102 male sterility (for a review see Presgraves<sup>17</sup>), BTAX was one of the least introgressed  
 103 chromosomes and displayed the lowest medium and maximal sizes of introgressed  
 104 segments.

105 Phylogenetic analysis revealed a close genetic relationship between the admixture source  
 106 and the Turano-Mongolian cattle group (see Online Methods and **Supplementary Fig. 5**).  
 107 Simulation results of one- and multiple-date admixture followed by segment retrieval by  
 108 *RFMix* supported nearly continuous admixture along the last 1500 years with a low  
 109 proportion of cattle gametes (around 1/11000 per generation; **Fig. 2b**; **Supplementary**  
 110 **Note**). While hybridization between yak and cattle was already a common practice 1800  
 111 years ago<sup>16</sup>, we could not detect older admixture because of the limitation of the methods.  
 112 Introgression was more intense during two periods (897-1121 and 1695-1828 CE), which  
 113 coincides with the Medieval Climate Anomaly (900-1200 CE)<sup>18</sup> and the Dzungar–Qing Wars  
 114 (1687–1758 CE)<sup>19</sup>. This is most probably because of increased mortality of livestock during  
 115 these difficult times that have forced yak herders to breed all the females available to restore  
 116 their herds, including backcross animals (**Supplementary Note**).

117 To identify phenotypes that have undergone positive selection, we next mined the gene  
 118 content of 365 intervals defined as the smallest exogenous segment shared for each region  
 119 showing introgression in at least 1% of the investigated haplotypes (**Supplementary Table**  
 120 **8**). Functional annotation of these 1311 transcripts using DAVID revealed a major enrichment  
 121 for genes involved in sensory perception, cognition and neurological system processes  
 122 (Benjamini corrected *P*-value <1.0E-8; **Supplementary Table 9**) which are known to be key  
 123 domestication targets<sup>20,21</sup>. Furthermore, similar results were obtained with different thresholds  
 124 on the percentage of introgression and size of intervals, indicating that selection on these  
 125 genes, which most probably contributed to taming the ferocious temper of yaks, has been a  
 126 regular and general process since the first hybridizations (**Supplementary Table 9**).

127 In total, we were able to retrieve 443 of such genes in 208 intervals after performing  
 128 complementary gene set enrichment analyses and literature review (see Online Methods).  
 129 These comprised genes related to nervous system development and function, synaptic  
 130 transmission, sensory perception and a large variety of disorders impacting learning ability,  
 131 social behavior, fear response and orientation in space in humans and animals (**Fig. 3a**;  
 132 **Supplementary Tables 8-12**). Among them we should mention *ITGA9*, a susceptibility gene  
 133 for bipolar affective disorder<sup>22</sup>, which shows the highest level of introgression with 56%  
 134 (85/152) of bovine alleles. We should also highlight the presence of nine genes from the  
 135 glutamate receptor signaling canonical pathway including each of the four subtypes of  
 136 receptors for this molecule which is the principal excitatory neurotransmitter in the brain<sup>23</sup>  
 137 (**Fig. 3b,c**). Significantly enriched canonical pathways, according to Ingenuity Pathway  
 138 Analysis, also include: (i) NAD biosynthesis from tryptophan and (ii) lysine degradation II & V  
 139 which produce L-glutamate, (iii) the visual cycle involved in the sensory transduction of light

140 in the retina, (iv) Sphingosine-1-phosphate signaling which participates to neuromodulation<sup>24</sup>,  
 141 (v) neuropathic pain signaling in dorsal horn neurons and (vi) Huntington's disease (**Fig. 3b**).

142 At the individual level, each yak carried numerous bovine genes involved in nervous system  
 143 development and function (mean = 33.03 ±10.05; **Supplementary Table 6**), although most  
 144 of them had moderate allele frequencies (median = 0.0461; **Supplementary Table 8**).  
 145 Moreover, none of the genes we investigated exhibited deleterious mutations  
 146 (**Supplementary Note**). These results are in line with previous studies which showed that  
 147 affective disorders in humans and anxiety behaviours in different species have a polygenic  
 148 basis and rely in part on the same genes (e.g.<sup>25-29</sup>). They further support our assumption that  
 149 this specific gene enrichment observed in introgressed regions in yak is due to selection on  
 150 behavioural traits.

151 Finally, with the exception of two regions encompassing *ABHD4* and *MYO6*, none of the 365  
 152 segments introgressed in our panel collocated with 182 recently reported signatures of  
 153 domestication in yak<sup>30</sup> which confirms that introgressed segments constitute a source of  
 154 favorable polymorphisms especially for genes which do not possess similar variants in yak.  
 155 This is for example the case for a *KIT* duplication causing color-sidedness in cattle<sup>7,8</sup> which  
 156 segregates in Mongolian yak (**Supplementary Fig. 6**) and presumably for a new polled  
 157 mutation. To verify this hypothesis we modeled polledness as a quantitative trait in our panel  
 158 (**Supplementary Fig. 7**) and mapped the locus to the beginning of chromosome 1 ( $P = 9.7$ -  
 159  $E9$ ; 95% CI: 1.88-2.20-Mb; **Fig. 4a**) within a bovine introgressed segment (**Fig. 4b**;  
 160 **Supplementary Table 8**). Between position 1,809,313 and 2,627,891-bp, we identified a  
 161 total of 1,024 sequence variants which were homozygous in the homozygous polled YAK13  
 162 and absent from the horned YAK40. Nearly all of them were retrieved in the genome of one  
 163 polled Turano Mongolian cattle (TM29), confirming the bovine origin of the polled mutation in  
 164 yak.

165 Genotyping of twelve indels in 604 animals originating from two yak and 21 cattle  
 166 subpopulations (**Supplementary Tables 1, 13, 14**) refined the polled locus interval to a 121-  
 167 kb segment (1,889,854-2,010,574-bp) containing 238 variants. Contrasting these with the  
 168 genomes of 234 bovines originating from Europe (**Supplementary Table 2**)<sup>13</sup>, one horned  
 169 Japanese Turano cattle<sup>20</sup>, and TM29, we excluded all but two variants originating from the  
 170 same microhomology-mediated break-induced replication event: a complex 219-bp  
 171 duplication-insertion ( $P_{219ID}$ ) beginning at 1,976,128-bp and a 7-bp deletion and 6-bp insertion  
 172 ( $P_{1ID}$ ) located 621-bp upstream of this position (**Supplementary Fig. 8,9**). This  
 173 rearrangement results in the duplication of an 11-bp motif (AAAGAAGCAAA) which is  
 174 entirely conserved among *Bovidae* (**Supplementary Fig. 10-13**) and which is also  
 175 duplicated in the 80-kb duplication responsible for Friesian polledness<sup>11</sup>. Finally, genotyping  
 176 of the  $P_{219ID}$ - $P_{1ID}$  rearrangement in yaks and cattle revealed a perfect association with  
 177 polledness of Turano-Mongolian origin, thus adding this polymorphism as third allele to the  
 178 reported allelic heterogeneity at the *POLLED* locus (**Fig. 4c,d**)<sup>9</sup>.

179 In conclusion, we present the first characterization of bovine introgression in yak at the  
 180 genomic scale. We report (i) that Mongolian yaks inherited on average 1.31% of their  
 181 genome from bovine ancestors after nearly continuous admixture over at least the last 1500  
 182 years and (ii) that these segments are significantly enriched in genes involved in nervous  
 183 system development and function which most probably have contributed to the taming of  
 184 yaks. We also show introgression of a new mutation which determines a phenotype of  
 185 primary interest in bovine and yak husbandry: the genetic absence of horns. This study  
 186 contributes to the emerging picture of the genes and pathways which have been the most  
 187 affected by domestication and highlights the beneficial role played by introgressive  
 188 hybridization in transferring favorable polymorphisms from one domestic species to another.

189

190 DATA AVAILABILITY:

191 Project accession codes (NCBI Sequence Read Archive (SRA)), PRJNA279385.

192

## 193 ACKNOWLEDGMENTS

194 A.C. acknowledges P. Boudinot and D. Boichard for interesting discussions about the Major  
 195 Histocompatibility Complex and the recombination rate in cattle respectively, M. Boussaha,  
 196 Y. Djari and C. Klopp for introducing him to the use of *SAMTOOLS* and *PINDEL* softwares and  
 197 M.-C. Deloche and C. Escoufflaire for their punctual assistance. The authors acknowledge  
 198 the Zoologischer Garten Berlin, Tierpark Cottbus and Tierpark Hellabrunn Munich  
 199 (Germany), as well as the Jardin des Plantes de Paris (France) and all the breeders in  
 200 Europe and Asia for generously providing samples and phenotypes. Apisgene is  
 201 acknowledged for funding the AKELOS research project. LAFUGA is funded by the  
 202 LMUexcellent program.

203

## 204 AUTHOR CONTRIBUTIONS

205 A.C. and I.M. conceived and coordinated the study. A.C., I.M. and S.K. designed the study.  
 206 I.M. mapped the polled locus; performed introgression analysis using SNP Chip genotyping  
 207 data; simulation analyses; and neighbor joining phylogenetic analyses. A.C. performed  
 208 variant calling, annotation and screening for candidate mutations; analysis of sequence  
 209 conservation; annotation of the gene content of the introgressed intervals; and gene set  
 210 enrichment analyses. S.K., A.G. and I.M. performed introgression analysis based on WGS  
 211 data; determination of ancestral alleles; genome and capture sequencing; and R-graphics.  
 212 C.G., S.R., and A.C. performed PCR for Sanger sequencing and for genotyping by PCR and  
 213 electrophoresis or PCR and Sanger sequencing. J.B. performed whole genome sequencing.  
 214 D.S. and I.R. performed SNP chip genotyping and whole genome sequencing. Y.Z., E.G. and  
 215 G.B. provided samples and phenotypes. H.B. provided sequencing and bioinformatics  
 216 facilities. A.E. provided Illumina BovineHD SNP chip genotyping data. G.H. provided  
 217 software and expertise in admixture analysis. A.C., I.M., S.K. and A.G. contributed to writing  
 218 the manuscript.

219

## 220 COMPETING FINANCIAL INTEREST STATEMENTS:

221 The authors declare no competing financial interests.

222

## 223 REFERENCES

- 224 1. Mallet, J. Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* **20**, 229-  
 225 237 (2005).
- 226 2. Sankararaman, S. *et al.* The genomic landscape of Neanderthal ancestry in present-day  
 227 humans. *Nature* **507**, 354-357 (2014).
- 228 3. Vernot, B. & Akey, J.M. Resurrecting Surviving Neandertal Lineages from Modern Human  
 229 Genomes. *Science* **343**, 1017-1021 (2014).
- 230 4. Hufford, M.B. *et al.* The Genomic Signature of Crop-Wild Introgression in Maize. *Plos*  
 231 *Genetics* **9**, e1003477 (2013).
- 232 5. Qiu, Q. *et al.* The yak genome and adaptation to life at high altitude. *Nature Genetics* **44**, 946-  
 233 949 (2012).
- 234 6. Qi, X.B., Jianlin, H., Wang, G., Rege, J.E.O. & Hanotte, O. Assessment of cattle genetic  
 235 introgression into domestic yak populations using mitochondrial and microsatellite DNA  
 236 markers. *Animal Genetics* **41**, 242-252 (2010).

- 237 7. Zhang, M.Q., Xu, X. & Luo, S.J. The genetics of brown coat color and white spotting in  
238 domestic yaks (*Bos grunniens*). *Animal Genetics* **45**, 652-659 (2014).
- 239 8. Durkin, K. *et al.* Serial translocation by means of circular intermediates underlies colour  
240 sidedness in cattle. *Nature* **482**, 81-84 (2012).
- 241 9. Medugorac, I. *et al.* Bovine polledness-an autosomal dominant trait with allelic  
242 heterogeneity. *PLoS ONE* **7**, e39477 (2012).
- 243 10. Allais-Bonnet, A. *et al.* Novel Insights into the Bovine Polled Phenotype and Horn  
244 Ontogenesis in Bovidae. *Plos One* **8**, e63512 (2013).
- 245 11. Rothammer, S. *et al.* The 80-kb DNA duplication on BTA1 is the only remaining candidate  
246 mutation for the polled phenotype of Friesian origin. *Genetics Selection Evolution* **46**, 44  
247 (2014).
- 248 12. Liu, W.B. *et al.* Associations of single nucleotide polymorphisms in candidate genes with the  
249 polled trait in Datong domestic yaks. *Animal Genetics* **45**, 138-141 (2014).
- 250 13. Daetwyler, H.D. *et al.* Whole-genome sequencing of 234 bulls facilitates mapping of  
251 monogenic and complex traits in cattle. *Nat Genet* **46**, 858-65 (2014).
- 252 14. Decker, J.E. *et al.* Resolving the evolution of extant and extinct ruminants with high-  
253 throughput phylogenomics. *Proceedings of the National Academy of Sciences of the United*  
254 *States of America* **106**, 18644-18649 (2009).
- 255 15. Maples, B.K., Gravel, S., Kenny, E.E. & Bustamante, C.D. RFMix: A Discriminative Modeling  
256 Approach for Rapid and Robust Local-Ancestry Inference. *American Journal of Human*  
257 *Genetics* **93**, 278-288 (2013).
- 258 16. Wiener, G., Jianlin, H. & Ruijun, L. *The yak*, (FAO Regional Office for Asia and the Pacific,  
259 2003).
- 260 17. Presgraves, D.C. Sex chromosomes and speciation in *Drosophila*. *Trends in Genetics* **24**, 336-  
261 343 (2008).
- 262 18. Adhikari, D. & Kumon, F. Climatic changes during the past 1300 years as deduced from the  
263 sediments of Lake Nakatsuna, central Japan. *Limnology* **2**, 157-168 (2001).
- 264 19. Perdue, P.C. *China marches west: the Qing conquest of central Eurasia*, (Harvard University  
265 Press, 2009).
- 266 20. Carneiro, M. *et al.* Rabbit genome analysis reveals a polygenic basis for phenotypic change  
267 during domestication. *Science* **345**, 1074-1079 (2014).
- 268 21. Schubert, M. *et al.* Prehistoric genomes reveal the genetic foundation and cost of horse  
269 domestication. *Proceedings of the National Academy of Sciences of the United States of*  
270 *America* **111**, E5661-E5669 (2014).
- 271 22. Secolin, R. *et al.* Refinement of chromosome 3p22.3 region and identification of a  
272 susceptibility gene for bipolar affective disorder. *American Journal of Medical Genetics Part*  
273 *B-Neuropsychiatric Genetics* **162B**, 163-168 (2013).
- 274 23. Snyder, S.H. & Ferris, C.D. Novel neurotransmitters and their neuropsychiatric relevance.  
275 *American Journal of Psychiatry* **157**, 1738-1751 (2000).
- 276 24. Ghasemi, R., Dargahi, L. & Ahmadiani, A. Integrated Sphingosine-1 Phosphate Signaling in the  
277 Central Nervous System: From Physiological Equilibrium to Pathological Damage.  
278 *Pharmacological research* **104**, 156-164 (2016).
- 279 25. Domschke, K. & Reif, A. Behavioral Genetics of Affective and Anxiety Disorders. in *Behavioral*  
280 *Neurogenetics* (eds. Cryan, F.J. & Reif, A.) 463-502 (Springer Berlin Heidelberg, Berlin,  
281 Heidelberg, 2012).
- 282 26. Nurnberger, J.I. *et al.* Identification of Pathways for Bipolar Disorder A Meta-analysis. *JAMA*  
283 *psychiatry* **71**, 657-664 (2014).
- 284 27. Valdar, W. *et al.* Genome-wide genetic association of complex traits in heterogeneous stock  
285 mice. *Nat Genet* **38**, 879-887 (2006).
- 286 28. Johnsson, M., Williams, M.J., Jensen, P. & Wright, D. Genetical Genomics of Behavior: A  
287 Novel Chicken Genomic Model for Anxiety Behavior. *Genetics* **202**, 327-340 (2016).
- 288 29. Zapata, I., Serpell, J.A. & Alvarez, C.E. Genetic mapping of canine fear and aggression. *BMC*  
289 *Genomics* **17**, 572 (2016).

290 30. Qiu, Q. *et al.* Yak whole-genome resequencing reveals domestication signatures and  
 291 prehistoric population expansions. *Nature Communications* **6**, 10283 (2015).  
 292

293

## 294 FIGURE LEGENDS

295 **Figure 1. Phylogenetic analyses of sequence data on chromosomes 9 and 25 confirm**  
 296 **bovine introgression in the individual sequenced for generating the yak reference**  
 297 **genome.** a) Bovine introgression plot based on WGS data. Blue and red dots show the  
 298 relative frequencies of homozygous and heterozygous genotypes for yak-specific  
 299 polymorphisms from the sequencing results used for generating the yak reference genome.  
 300 Each dot represents the number of genotypes in a 70-kb sliding window divided by expected  
 301 (genome-wide) number of polymorphisms in 70-kb windows. Background color of each  
 302 interval is shaded according to read depth, ranging from 0 (dark gray) to >40 (white).  
 303 Genotype frequencies from white shaded intervals are likely to be affected by artefacts from  
 304 repeat expansions in yaks, and those from dark shaded intervals by poor mappability or  
 305 deletions. Results with neutral grey background are regarded as more robust. Introgressed  
 306 intervals are identified by a break in the red line (circular binary segmentation of mean  
 307 homozygous genotype frequency) and a drop in the frequency of homozygous genotypes for  
 308 yak-specific alleles. Heterozygous genotypes for yak-specific alleles and the yellow line  
 309 (circular binary segmentation of mean heterozygous genotype frequency) serve as a control  
 310 to distinguish homozygous versus heterozygous cattle introgression. These statistics suggest  
 311 homozygous cattle introgression in two regions (BTA09:68.495-70.115-Mb and  
 312 BTA25:17.345-19.995-Mb) in the individual sequenced for generating the yak reference  
 313 genome. b) Details of five exons and flanking sequences from four genes (*L3MBTL3*,  
 314 *SAMD3*, *ACSM2B* and *MGC134577*) which were sequenced in 14 *Bovini* animals for  
 315 validation of our introgression analysis based on WGS data (**Supplementary Table 4**). c)  
 316 Neighbor-joining phylogeny of 14 haplotypes representing yaks, cattle and 10 *Bovidae*  
 317 species, supporting homozygous cattle introgression in the reference yak genome. This  
 318 analysis was based on sequence data from the five regions of chromosome 9 and 25  
 319 presented above and totalizing 2,191 nucleotides. The reliability of the tree branches was  
 320 tested by 1,000 bootstrap replicates.

321

322 **Figure 2. Analysis of the size distribution of introgressed intervals reveals three major**  
 323 **introgressions events.** a) Minimal, maximal, average and median length of introgressed  
 324 intervals on each of the 30 chromosomes (x-axis) over 76 yak animals genotyped with the  
 325 Illumina BovineHD SNP chip. The genome-wide average and median lengths are presented  
 326 by green and red dotted lines, respectively. b) Distribution of the size of the bovine DNA  
 327 segments introgressed into the yak genomes as estimated by our *RFMix* procedure.  
 328 Absolute counts of fragments observed in (i) all 76 yak animals (green curve); (ii) 26 yaks  
 329 sampled in Mongolia (black curve); (iii) 50 yaks of Mongolian descent sampled in Europe  
 330 (blue curve); (iv) simulated three-date admixture in 76 de-introgressed yaks with a proportion  
 331 of cattle DNA of 0.0005 at 250, 0.011 at 150 and 0.0045 at 37 generations ago (orange  
 332 curve); and (v) continuous admixture with a proportion of cattle DNA of 0.00045 every 5  
 333 generations in a period between 40 to 220 generations ago (red curve) were divided by the  
 334 number of considered haploids in each of the four groups. The size of the introgressed  
 335 segments detected varied between 108-kb and 24.63 Mb with a median length of 601-kb.  
 336 Ten Mb (Chr23:22.0 to 32.0 Mb) comprising the MHC region were not considered in this  
 337 distribution. The figure presents intervals up to a maximal length of 5,000-kb. Longer  
 338 intervals had frequencies of 0 or 1% and are not all shown here for reasons of clarity.

339

340 **Figure 3. Bovine introgressed segments show a major enrichment for genes related to**  
 341 **nervous system development and function.** Word cloud illustrating the major enrichment  
 342 for genes related to nervous system development and function, behavior, neurological  
 343 diseases and psychological disorders revealed by the Ingenuity Pathway Analysis. A total of  
 344 1311 genes associated to 365 intervals showing at least 1% (i.e. 2 alleles) of bovine genome  
 345 introgression in our panel of 76 yaks (**Supplementary Table 8**) were considered for the  
 346 Ingenuity Pathway Analysis (IPA). A unique keyword was attributed to each significantly  
 347 enriched pathway in the “Diseases and Bio Function analysis” according to the “Diseases or  
 348 Functions Annotation” (see Online Methods). Keywords referring to ubiquitous cellular or  
 349 organismal processes are not represented to not overload the cloud. Font size is proportional  
 350 to the number of occurrence of the keywords. b) Venn-like diagram presenting the canonical  
 351 pathways which are significantly enriched ( $P < 0.01$ ) in introgressed segments according to  
 352 IPA and details on associated genes (**Supplementary Table 11**). These consist in five  
 353 pathways related to nervous system development, function or pathologies and to two  
 354 pathways resulting in the production of L-glutamate which is the principal excitatory  
 355 neurotransmitter in the brain. c) Localization at the neuron synapse level of the main proteins  
 356 belonging to the glutamate receptor signaling canonical pathway (adapted from IPA).  
 357 Proteins encoded by genes listed in (b) are highlighted in pink. GRIA4: Glutamate Receptor,  
 358 Ionotropic, AMPA 4; GRIK3: Glutamate Receptor, Ionotropic, Kainate 3; GRIN2A: Glutamate  
 359 Receptor, Ionotropic, NMDA 2A; GRIN3A: Glutamate Receptor, Ionotropic, NMDA 3A;  
 360 GRIP1: Glutamate Receptor Interacting Protein 1; GRM4: Glutamate Receptor, Metabotropic  
 361 4; CAMK4: Calcium/Calmodulin-Dependent Protein Kinase IV; DLG4: Discs, Large Homolog  
 362 4 (*Drosophila*). NMDA: N-methyl-D-aspartate; AMPA:  $\alpha$ -Amino-3-hydroxy-5-methyl-4-  
 363 isoxazolepropionic acid. CALM belonging to the Calcium/Calmodulin-Dependent Protein  
 364 Kinase group and PSD-95, the postsynaptic density protein 95, are respectively encoded by  
 365 CAMK4 and DLG4. EPSPs: Excitatory Postsynaptic Potentials. Note the presence of each of  
 366 the four subtypes of glutamate receptors (Ionotropic, AMPA; Ionotropic, Kainate; Ionotropic,  
 367 NMDA; and Metabotropic) among these proteins.

368

369 **Figure 4. Introgression of a novel and complex mutation at the *POLLED* locus from**  
 370 **bovines causes polledness in Mongolian yaks.** a) Mapping of the *POLLED* locus with  
 371 Illumina BovineHD Beadchip genotyping data from 36 polled and 40 horned animals and  
 372 polledness modeled as a quantitative trait. b) Bovine introgression plot based on WGS data.  
 373 Orange and turquoise dots show the mean frequency of bovine and yak alleles at positions  
 374 of yak specific variants in a 70-kb sliding window. Background color of each interval is  
 375 shaded according to read depth ranging from 0 (dark gray) to >40 (white). Introgressed  
 376 intervals are identified by a break in the red line (circular binary segmentation of mean allele  
 377 frequency) and a drop of yak-allele frequency below 0.5. Note that YAK13 is homozygous for  
 378 a bovine introgressed segment encompassing the mapping interval of the polled locus ( $P =$   
 379  $9.7 \cdot 10^{-9}$ ; 95% CI: 1.88-2.20-Mb). This result is independently supported by a reduction of the  
 380 divergence of YAK13 genome sequence from the UMD3.1 bovine reference sequence in the  
 381 polled region (0.28 %) between position 1,809,313 and 2,627,891-bp as compared with the  
 382 average divergence of 1.08 % exhibited by both yaks at the genome level. c) Scheme  
 383 presenting the nature and location of the three different mutations identified at the polled  
 384 locus in bovine as compared with the wild type allele. Red boxes represent the segments  
 385 that are duplicated in the Celtic, Friesian and Mongolian alleles whereas light and dark grey  
 386 boxes represent the original segments. Note that none of the three polled mutations affect  
 387 coding regions and that the molecular mechanism underlying polledness remains unknown  
 388 at the present time. d) Details of the complex Polled Mongolian mutation which results in the  
 389 duplication of a 11-bp motif which is entirely conserved among *Bovidae* and well conserved  
 390 among vertebrates (**Supplementary Fig. 10-13**). Boxes from different colors are used to  
 391 show segmental duplications.

392



393

394

## 395 **ONLINE METHODS**

396

### 397 **Animals**

398 In total 120 yaks, 1025 cattle from a wide diversity of breeds originating from Eurasia and  
399 Africa, as well as representatives of nine other bovid species were considered in at least one  
400 of the analyses performed in this study. Briefly, they consist of animals used for mapping of  
401 POLLED locus in yak and Mongolian turano cattle (**Supplementary Table 1** and  
402 **Supplementary Fig. 7**); sets of whole genome sequences of yak and cattle  
403 (**Supplementary Table 2**) used for introgression analyses (**Supplementary Fig. 1**) and  
404 filtering of candidate mutations; bovid species used for target sequencing and phylogenetic  
405 analyses (**Supplementary Tables 3** and **4**); and sets of Illumina BovineHD chip genotypes  
406 used for admixture and mapping analyses (**Supplementary Table 5**).

407

### 408 **Horn/polled phenotypes and derived genotypes**

409 The polled phenotype is an autosomal dominant trait in cattle<sup>31</sup> and yak<sup>12</sup>, readily measurable  
410 on any animal older than six months. Artificial dehorning of yak and cattle is not practiced in  
411 the sampling area in Central Asia. Therefore, any polled yak descending from one polled and  
412 one horned parent is necessarily heterozygous, i.e. Pp, at the underlying POLLED locus.  
413 One horned offspring with confirmed paternity is sufficient to declare a polled parent as Pp.  
414 Animals having two polled parents and consecutively ten or more polled offspring with  
415 horned mates are declared as homozygous polled PP. Similar animals having less than ten  
416 offspring (all polled) with horned mates are either PP or Pp and were declared as P•. Finally,  
417 all horned animals are pp. Derived genotypes of yak animals used for mapping are  
418 presented in **Supplementary Fig. 7**.

419

### 420 **Whole genome sequencing of two Mongolian yaks and one Mongolian Turano cow.**

421 The genomes of one heterozygous polled Turano-Mongolian cow (TM29), one homozygous  
422 polled yak (YAK13) and one horned yak (YAK40) were sequenced with Illumina technology.  
423 Paired-end libraries were generated according to the manufacturer's instructions using the  
424 Rapid DNA library system (NuGen, San Carlos, USA) for animal TM29 and the NEXTflex  
425 PCR-Free DNA Sequencing Kit (Biooscientific) for YAK13 and YAK40. Libraries were  
426 quantified using the KAPA Library Quantification Kit (Cliniscience), controlled on a High  
427 Sensitivity DNA Chip (Agilent) and sequenced on an Illumina HiSeq1500 with 2\*110-bp read  
428 length (TM29) or on a HiSeq 2000 with 2\*101-bp read length (YAK13 and YAK40). The  
429 average sequence coverage was 8.7, 13.4 and 14.9 x, respectively. Reads were mapped on  
430 the UMD3.1 bovine sequence assembly using BWA<sup>32</sup>. Reads with multiple alignments were  
431 removed. SNPs and small indels were called using *SAMTOOLS* pileup option<sup>33</sup>. Only variants  
432 with a quality score (QUAL) of  $\geq 30$  and a mapping quality (MQ) score of  $\geq 30$  were kept.  
433 Discovery of larger indels was achieved with *PINDEL*<sup>34</sup>. Variants supported by only one read  
434 or found in the homozygous state in the three animals were considered as possible artifacts  
435 and eliminated. Detection of Copy Number Variation was performed according to Medugorac  
436 *et al.*<sup>9</sup> by calculating coverage ratios between pairs of individuals in dynamical bin sizes of  
437 5000 reads. YAK13 and YAK40 were each compared to three cattle WGS sequences in

438 order to call possible CNVs in introgressed regions. Significant CNV results were kept if they  
439 overlapped in all three comparisons of one YAK and three cattle WGS. Signals that were  
440 caused by mapping of apparently repetitive sequences getting very high coverage (>1000  
441 fold) were manually removed. The reliability and borders of the retained CNVs were verified  
442 using the Integrative Genomics Viewer (IGV)<sup>35</sup> and paired-end information. In the end, only  
443 one polymorphism was considered as a true introgressed CNV: the mutation responsible for  
444 color-sidedness in bovines that is presented in **Supplementary Fig. 6**. For the latter, the  
445 log<sub>2</sub> ratios of sequence coverages per 5000-bp window between the solid colored YAK40  
446 and the color-sided YAK13 were plotted using R and the average ratios were segmented  
447 using the circular binary segmentation implementation in the DNACopy package (v1.14.0)  
448 from Bioconductor.

449

#### 450 **Introgression analysis in WGS data of three yaks and one Mongolian Turano cow.**

451 The detection of bovine genome segments in WGS data of YAK13, YAK40 and the reference  
452 genome<sup>5</sup> (YAKQIU) was conducted as follows. First, variants which were homozygous for the  
453 alternate allele in YAKQIU and absent from TM29 and 234 additional bovine genomes<sup>13</sup> were  
454 identified and considered as yak-specific. Then, the mean frequency of yak and bovine  
455 alleles for each of these variants was estimated for sliding windows of 70-kb along the  
456 genomes of YAK13, YAK40, and TM29. This window size corresponds to half of the  
457 expected mean size of segments which would have been introgressed from the earliest  
458 possible hybridization between cattle and yak (**Supplementary Note**). To detect  
459 introgression in the yak reference sequence itself, the number of homozygous and  
460 heterozygous genotypes for yak-specific polymorphisms were estimated in 70-kb windows  
461 and compared to the expected numbers based on genome-wide observations (see Figure 1).  
462 Introgressed intervals were identified by circular binary segmentation (CBS) of mean allele  
463 frequency and a drop of yak-allele frequency below 0.5. CBS is implemented in the R-  
464 package DNACopy (v1.14.0) from Bioconductor. Frequencies of yak-specific alleles in both,  
465 TM29 and YAKQIU, served as a control.

466

#### 467 **Conventional Sanger sequencing of target genomic regions with cattle ancestry in** 468 **reference yak genome.**

469 Two regions suggesting homozygous cattle introgression in the yak reference genome<sup>5</sup>  
470 (Chr9:68,495,000-70,115,000 and Chr25:17,345,000-19,995,000; **Fig.1** and **Supplementary**  
471 **Fig. 1**) were selected to test the reliability of our approach. For each region, two PCR  
472 products were amplified in 13 animals representing 12 bovid species (**Supplementary Table**  
473 **3 and 4**). PCR was performed using the Go-Taq Flexi system (Promega) according to the  
474 manufacturer's instructions on a Mastercycler pro thermocycler (Eppendorf). Amplicons were  
475 purified and bidirectionally sequenced by Eurofins MWG using conventional Sanger  
476 sequencing. The resulting sequences were aligned with the corresponding sequences from  
477 the yak reference genome using the *CLUSTALW* (<http://www.clustal.org>) software<sup>36</sup> which is  
478 part of the *MEGA* software<sup>37</sup> package version 6.06 (<http://www.megasoftware.net>). Then they  
479 were trimmed to get equal lengths for most animals and fragments. Finally, phylogeny was  
480 inferred based on a total of 2191 nucleotides of sequence using the Neighbor-Joining  
481 methods<sup>38</sup> implemented in *MEGA* software<sup>37</sup>. The percentage of replicate trees in which the  
482 associated taxa clustered together was determined by the bootstrap test<sup>39</sup> (1,000 replicates).

483 A similar approach was used to study the MHC locus and to estimate the false discovery rate  
484 of the *RFMix* analysis, as described in **Supplementary Note**.

485

#### 486 **Analysis of Illumina HD genotypes**

487 *General information.* Illumina BovineHD BeadChip genotypes from 467 *Bovidae* animals  
488 were considered. These consisted of 76 yaks (36 polled and 40 horned; **Supplementary**  
489 **Fig. 7**), a panel of 384 individuals representative of the world wide diversity of cattle and  
490 assumed to harbor no yak ancestry (**Supplementary Note**, **Supplementary Table 5**), and  
491 representatives of six *bovini* species (two gaur, one wood bison, one European bison, one  
492 banteng, one water buffalo and one nilgai; **Supplementary Table 3**). Of note, the panel of  
493 384 cattle comprised 11 polled and 14 horned turano cattle from Mongolian and Yakutian  
494 breeds (**Supplementary Tables 1** and **5**). A total of 697,172 SNP markers were successfully  
495 genotyped in three to six *bovini* species. Only 42,230 SNP (5.43%) were informative in yaks.  
496 Haplotypes were inferred and missing genotypes imputed using hidden Markov models  
497 (software package *BEAGLE*)<sup>40</sup> and three cohort types; namely trios (two parents, one  
498 offspring), pairs (one parent, one offspring) and unrelated animals. Marker order was based  
499 on release UMD3.1 of the *Bos taurus* genome  
500 ([http://www.cbcb.umd.edu/research/bos\\_taurus\\_assembly.shtml](http://www.cbcb.umd.edu/research/bos_taurus_assembly.shtml)).

501 *Inferring maternal and paternal phylogenies.* To avoid artifacts, only SNP from the  
502 mitochondrial genome and Y chromosome showing high call rates (>99%) and complete  
503 homozygosity within each single animal (n= 245/343 and 921/1224 markers respectively)  
504 were used. Moreover, only animals with less than 5% of missing genotypes for mitochondrial  
505 or Y chromosome markers were considered. Maternal and paternal phylogenies were  
506 constructed with the Neighbor-Joining methods<sup>38</sup> implemented in *MEGA* software<sup>37</sup> version  
507 6.06. The percentage of replicate trees in which the associated taxa clustered together was  
508 determined by the bootstrap test<sup>39</sup> (1,000 replicates).

509 *Introgression analysis in 76 Mongolian yaks.* Yak-specific alleles were inferred from  
510 homozygous SNPs located in genomic regions of YAK13, YAK40 or YAKQIU that are free of  
511 cattle ancestry, based on previous analyses of WGS data (**Supplementary Fig. 1**). Since  
512 WGS data didn't provide clear introgression status in two specific region (Chr22:31,682,450-  
513 31,842,000 and Chr23:24,661,105-29,153,851; **Supplementary Fig. 1**), we used genotypes  
514 of 76 yaks to define the major allele (frequency $\geq$ 0.90) as yak-specific. For all remaining SNPs  
515 (<1.00%) the major allele (frequency  $\geq$ 0.75) in six *bovini* species was considered as  
516 ancestral and yak-specific. Then, a rapid and robust forward-backward algorithm  
517 implemented in the software package *RFMix*<sup>2,15</sup> was used to screen for the presence of cattle  
518 haplotypes in 76 Mongolian yaks. This algorithm uses designated reference haplotypes to  
519 infer local ancestry in designated admixed haplotypes which supposes to include pure yak  
520 and pure cattle in the analysis. Since there is no genetic and historical support for  
521 introgression of yak into cattle, we considered the 384 cattle (**Supplementary Table 5**) as a  
522 reference panel assumed to harbor no yak ancestry.

523 On the other hand, we were not able to find a complete yak genome without cattle ancestry  
524 but we detected complete chromosomes or large chromosomal fragments with pure yak  
525 ancestry. These chromosomal regions as well as yak-specific alleles were used to create a

526 synthetic pure yak genome (YAKYAK) which served as reference yak in initial *RFMix*  
527 analyses.<sup>2</sup>

528 For each chromosome we started two rounds of *RFMix* analyses. The first round used 384  
529 cattle genomes as reference cattle population and only YAKYAK as reference yak  
530 population. The admixed sample consisted of all 76 yak genomes. For each chromosome,  
531 initial *RFMix* analyses detected different subsets of yak haploids as pure. These pure yak  
532 chromosomes supplemented the YAKYAK chromosome in the second round of *RFMix*  
533 analyses to produce final results for a specific chromosome.

534 The *RFMix* program performs forward-backward analyses in non-overlapping windows of  
535 predefined size. In some situations, like for short segments in an unfavorable location  
536 (window transition), *RFMix* occasionally detected signatures only in the more informative half  
537 or even in none of the two windows. To deal with these problems, we set the window size at  
538 0.2 cM and performed four overlapping *RFMix* analyses (**Supplementary Fig. 4**).

539 *Source, date and number of admixture events.* *CHROMOPAINTER*<sup>41</sup> was used to decompose  
540 the chromosomes of each of the 76 Mongolian yaks as a series of haplotypic chunks inferred  
541 to be shared with at least one of the 384 cattle representing 24 breeds. In theory, given a  
542 single admixture event, ancestry chunks inherited from each source have an exponential size  
543 distribution, resulting in an exponential decay of these coancestry curves<sup>41,42</sup>. The shape of  
544 decay curve in different groups enable the estimation of admixture dates<sup>42</sup> and the  
545 determination of recipient and donor groups involved in asymmetric admixture events.  
546 Multiple admixture times result in a mixture of exponentials<sup>42</sup>; which can be tested by  
547 comparing the fit of a single exponential decay rate versus a mixture of rates. Inferences of  
548 the haplotypic makeup of admixing source groups as well as of the admixture date(s) were  
549 carried out using the *GLOBETROTTER*<sup>42,43</sup> method and complemented by simulation studies  
550 described in **Supplementary Note**.

551 The inference of the source of admixture was complemented by phylogenetic analyses of  
552 pure and admixed haploids. For 139 chromosomal segments introgressed in ten or more  
553 yaks (**Supplementary Table 8**), ten pure and ten introgressed haploids were randomly  
554 selected to constitute two yak groups. Similarly, 20 cattle groups representing 20 breeds with  
555 four or more animals genotyped with the BovineHD chip were constructed (**Supplementary**  
556 **Table 5**). These segments were divided into a total of 3076 non-overlapping blocks  
557 comprising four SNPs for which the inter-marker distance was less than 25 kb with  
558 neighbouring SNPs. Each block was considered as a multi-allelic marker in phylogenetic  
559 analysis to reduce ascertainment bias<sup>44</sup>. The proportion of shared alleles between  
560 individuals,  $PS$ <sup>45</sup>, was converted to genetic distances ( $D_{PS} = \ln(PS)$ ). A neighbor-joining tree  
561 (**Supplementary Figure 5**) reflects the averaged individual distances between groups and  
562 was constructed with the *SPLITSTREE4* program<sup>46</sup>.

563

#### 564 **Annotation of the gene content of the introgressed segments**

565 For 365 regions showing a minimum of 2 introgressed segments among the 76 animals  
566 studied, we defined the smallest portion shared by these segments. To assess the gene  
567 content of the resulting intervals, we used the “Refseq Genes” track from the UCSC Genome  
568 Browser (<http://genome.ucsc.edu>) as a primary resource. We also used the “Non-Cow  
569 Refseq Genes”, “Cow mRNAs from Genbank”, and “Cow ESTs that have been spliced

570 tracks” to recover protein genes that may have been missed during annotation of the  
 571 UMD3.1 bovine sequence assembly. These consisted in genes annotated in at least human  
 572 and mouse with no bovine RNA alignments in the orthologous region or genes with bovine  
 573 RNA alignments corresponding to at least one gene annotated in human or mouse in the  
 574 orthologous region. Intervals which did not contain genes were attributed the name of the  
 575 closest gene in 5'3' orientation and located at a maximum of 500kb downstream of its  
 576 borders according to the same orientation.

577

## 578 **Gene set enrichment analysis**

579 Gene set enrichment analyses were carried out with five software using different methods  
 580 and source of information, i.e. Gene Ontology classes for DAVID 6.7  
 581 (<http://david.abcc.ncifcrf.gov/>) and PANTHER (<http://pantherdb.org/>), bibliographic and  
 582 experimental data for Genetrail2 (<http://genetrail2.bioinf.uni-sb.de/>) and Ingenuity Pathway  
 583 Analysis (<http://www.ingenuity.com/products/ipa/>), and Mammalian Phenotype ontology  
 584 (level 3) from Mouse Genome Informatics for the specific analysis we performed with Enrichr  
 585 (<http://amp.pharm.mssm.edu/Enrichr/>)<sup>47,48</sup>. Since these analyses produced comparable  
 586 results, and for the sake of simplicity, only two of them were selected to be presented in this  
 587 study. To provide a first overview of the overrepresented groups of genes, and to test their  
 588 reliability, we performed different Gene Ontology (GO) term enrichment analyses were  
 589 performed with DAVID using different lists of genes located in chromosomal regions detected  
 590 as introgressed from cattle to yaks by *RFMix* analyses (results are presented in  
 591 **Supplementary Table 8**). Then we used Ingenuity Pathway Analysis for the precision of its  
 592 annotations. We focused on the “Top Canonical Pathways” and on the “Diseases and bio  
 593 Functions”. Only canonical pathways or annotations with a  $P$ -value  $<10^{-2}$  were retained.  
 594 Annotations related to cancer and drug metabolism which were not relevant for this study  
 595 were not considered. In addition to the IPA annotations we attributed a unique keyword to  
 596 each significantly enriched pathway according to the “Diseases or Functions Annotation” in  
 597 order to draw a word cloud. A particular attention was paid to attribute keywords related to  
 598 subcellular portions, cell types and organs rather than to general processes. Keywords which  
 599 appeared only once were finally regrouped with higher order items (e.g. celltype changed for  
 600 organ or process changed for the category defined by IPA) or with the predefined IPA  
 601 “categories” (results are presented in **Supplementary Table 10**). Finally, whereas they are  
 602 not presented in detail, results from the three other analyses were used to complete the list  
 603 of genes involved in nervous system development and function presented in **Supplementary**  
 604 **Tables 8 and 12**.

605

## 606 **Mapping of the polled locus in yaks sampled in Europe and Mongolia.**

607 Mapping of the polled locus was performed using a combined linkage disequilibrium and  
 608 linkage analysis (*cLDLA*) with horn status modelled as a quantitative trait ( $pp=0$ ,  $Pp=1$ ,  $PP=2$   
 609 and  $P=1.5$ ). Genomic relationship matrix ( $\mathbf{G}$ )<sup>49</sup> was estimated and its inverse ( $\mathbf{G}^{-1}$ ) used to  
 610 correct for population structure and possible polygenic effects in the model of the later QTL  
 611 mapping. Identical-by descent (IBD) probabilities between pairs of haplotypes<sup>50</sup> were  
 612 estimated for sliding windows of 40 SNP and summarized into a diplotype relationship matrix  
 613 ( $\mathbf{D}_{RM}$ ) which is computed in a similar way to the additive genotype relationship matrix ( $\mathbf{G}_{RM}$ )<sup>51</sup>.  
 614 *cLDLA* mapping of polledness was carried out with a procedure similar to that reported in

615 Meuwissen *et al.*<sup>52</sup>, which considers random QTL and polygenic effects. Variance component  
 616 analysis in the middle of each of the 40-SNP sliding windows was performed by the *ASREML*  
 617 package (<https://www.vsnr.co.uk/downloads/asreml/release3/UserGuide.pdf>) and a mixed  
 618 linear model:

$$619 \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{u} + \mathbf{Z}_2\mathbf{q} + \mathbf{e}$$

620 where  $\mathbf{y}$  is a vector of horn status,  $\boldsymbol{\beta}$  a vector of fixed effects (including overall mean  $\mu$ ),  $\mathbf{u}$  is a  
 621 vector of  $n$  random polygenic effects for each animal with  $\mathbf{u} \sim N(0, \mathbf{G}\sigma_u^2)$ ,  $\mathbf{q}$  a vector of  
 622 random additive genetic effects due to *POLLED* locus with  $\mathbf{q} \sim N(0, \mathbf{D}_{\text{RMP}}\sigma_q^2)$ , where  $\mathbf{D}_{\text{RMP}}$  is  
 623 the diplotype relationship matrix at position  $p$  of the putative *POLLED* locus, and  $\mathbf{e}$  a vector of  
 624 random residual effects with  $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$ , where  $\mathbf{I}$  is an identity matrix. The random effects  
 625  $\mathbf{u}$ ,  $\mathbf{q}$  and  $\mathbf{e}$  were assumed to be uncorrelated and normally distributed and their variances  
 626  $(\sigma_u^2, \sigma_q^2, \sigma_e^2)$  were simultaneously estimated using *ASREML*.

627 Using the logarithm of the likelihood estimated by *ASREML* for the model with  $(\log L_p)$  and  
 628 without *POLLED* locus effects  $(\log L_0)$ ; corresponding to the null hypothesis), we calculated  
 629 the likelihood ratio test statistic  $(LRT = -2(\log L_0 - \log L_p))$ , which is known to be  $\chi^2$ -distributed  
 630 with one degree of freedom<sup>53</sup>. Appropriately, an *LRT*-value higher than 10.8 was considered  
 631 statistically significant (equivalent to  $P < 0.001$ ).

632

### 633 **Fine mapping and identification of the Mongolian *POLLED*** 634 **mutation.**

635 The first step consisted of selecting sequence variants which were homozygous in the  
 636 homozygous polled YAK13, absent from the horned YAK40, and located between positions  
 637 1,809,313 and 2,627,891-bp on chromosome 1. This region comprises the 95% confidence  
 638 interval (1.88-2.20-Mb) obtained with the QTL mapping approach and corresponds to a  
 639 bovine chromosomal segment introgressed in yak (**Fig. 4b; Supplementary Table 8**). Then,  
 640 to narrow down the candidate region, 120 yaks and 484 Eurasian taurine cattle  
 641 (**Supplementary Table 1**) were genotyped for twelve indels using standard PCR, and  
 642 agarose gel or capillary (ABI PRISM® 377 and 3100 Genetic Analyzer, Applied Biosystems)  
 643 electrophoresis (**Supplementary Tables 13 and 14**). Of note, the same animals were also  
 644 genotyped for the Celtic ( $P_{202ID}$ ) and Friesian ( $P_{80kbID}$ ) polled mutations<sup>9-11</sup> and clearly  
 645 excluded these as possible candidates for polledness of Mongolian origin. Genotyping for  
 646 twelve indels (**Supplementary Table 13 and 14**) excluded all but two indels (LMP04 and  
 647 LMP12) as candidate mutation, and haplotype analyses reduced the polled locus interval to a  
 648 121-kb segment (1,889,854-2,010,574bp) containing 238 variants. Considering that the  
 649 Mongolian polled mutation occurred in Turano cattle and is absent even in European polled  
 650 cattle (**Supplementary Note**), these variants were subsequently filtered to retain only those  
 651 which were heterozygous in the heterozygous polled Mongolian Turano cattle TM29, and  
 652 absent in the genomes of one horned Japanese Turano cattle<sup>20</sup> and 234 bovines originating  
 653 from Europe (**Supplementary Table 2**)<sup>13</sup>.

654 Finally, to ensure that we did not miss any candidate variants for polledness, we performed  
 655 two independent verifications. We performed a new detection of structural variants in the  
 656 refined 121-kb polled interval using DELLY<sup>54</sup> and a visual examination of the whole genome  
 657 sequences of YAK13, YAK40 and TM29 in the same interval using IGV<sup>35</sup>. We did not detect

658 new candidate polymorphisms. Considering that there is no gap in the UMD3.1 bovine  
 659 genome sequence assembly and in the WGS of the homozygous polled yak (YAK13) in this  
 660 interval we can claim that we did not miss any candidate variant with our approach.

661

662 **Analysis of sequence conservation around the Mongolian *POLLED* mutation in**  
 663 **Mammals.**

664 Regions orthologous to the segments duplicated in the Mongolian *POLLED* mutation were  
 665 retrieved for 34 eutherian mammals using the EPO multiple alignment from ENSEMBL. A  
 666 consensus sequence and a sequence logo were generated using *MULTALIN*  
 667 (<http://multalin.toulouse.inra.fr/multalin/>)<sup>55</sup> and *WEBLOGO* (<http://weblogo.berkeley.edu/>)<sup>56</sup>,  
 668 respectively. After the identification of a well-conserved 11-bp motif, a novel consensus  
 669 sequence and a novel sequence logo were generated. Details on the 11-bp orthologous  
 670 sequences are presented in **Supplementary Fig. 11 and 12.**

671

672 **Analysis of sequence conservation around the Mongolian *POLLED* mutation in**  
 673 ***Bovidae*.**

674 The region encompassing the Mongolian *POLLED* mutation was PCR amplified from  
 675 genomic DNA samples of nine bovid species (**Supplementary Fig. 3** and **Supplementary**  
 676 **Table 3**). Two individuals were used for each species. PCR primers were manually designed  
 677 in regions which were conserved between bovine UMD3.1 and sheep Oar\_v3.1 genome  
 678 assemblies (**Supplementary Table 15**). PCR reactions and Sanger sequencing were  
 679 performed as previously described. The corresponding regions in cattle and yak were  
 680 obtained from the bovine UMD3.1 genome assembly and from YAK40 whole genome  
 681 sequencing data, respectively. Multispecies alignment was generated with *CLUSTALW*  
 682 software, version 2.0.1<sup>36</sup>.

683

684 REFERENCES

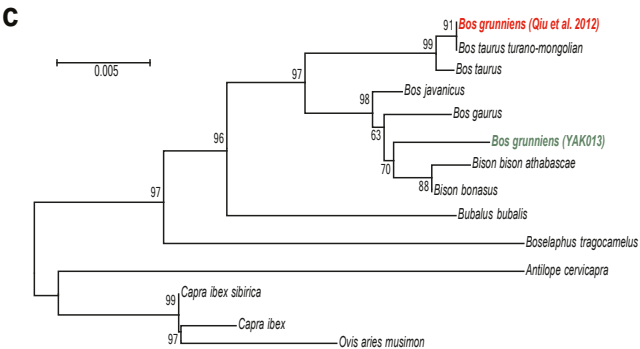
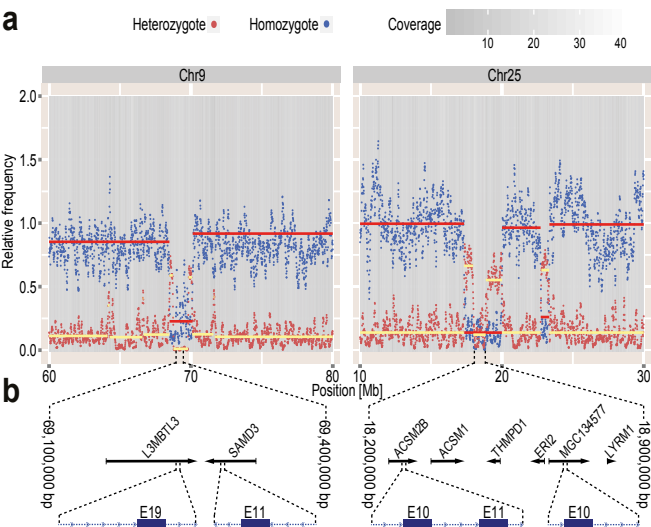
685

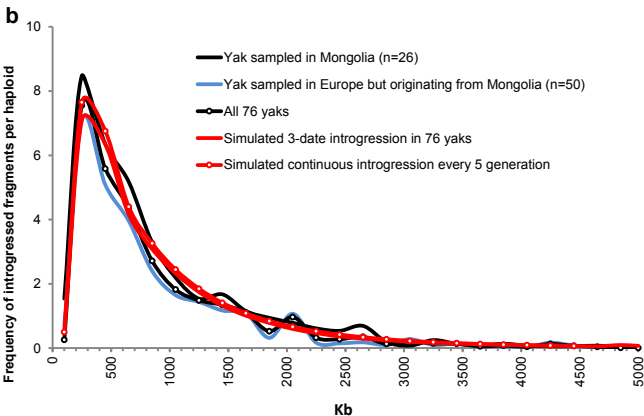
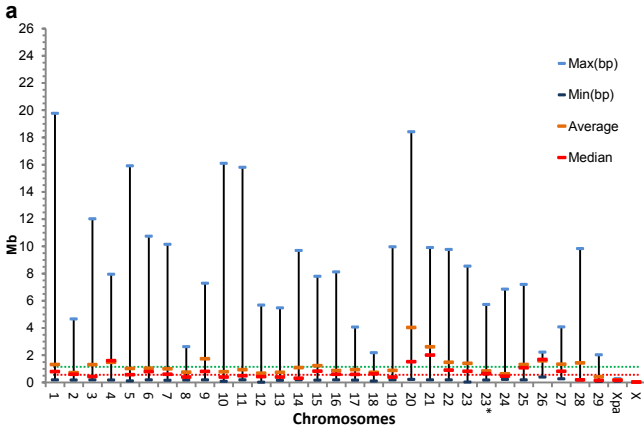
- 686 31. Dove, W.F. The physiology of horn growth: A study of the morphogenesis, the interaction of  
 687 tissues, and the evolutionary processes of a mendelian recessive character by means of  
 688 transplantation of tissues. *Journal of Experimental Zoology* **69**, 347-405 (1935).
- 689 32. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform.  
 690 *Bioinformatics* **25**, 1754-1760 (2009).
- 691 33. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-  
 692 2079 (2009).
- 693 34. Ye, K., Schulz, M.H., Long, Q., Apweiler, R. & Ning, Z.M. Pindel: a pattern growth approach to  
 694 detect break points of large deletions and medium sized insertions from paired-end short  
 695 reads. *Bioinformatics* **25**, 2865-2871 (2009).
- 696 35. Thorvaldsdottir, H., Robinson, J.T. & Mesirov, J.P. Integrative Genomics Viewer (IGV): high-  
 697 performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**, 178-  
 698 192 (2013).
- 699 36. Thompson, J.D., Higgins, D.G. & Gibson, T.J. CLUSTAL-W: improving the sensitivity of  
 700 progressive multiple sequence alignment through sequence weighting, position-specific gap  
 701 penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673-4680 (1994).



- 702 37. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary  
703 Genetics Analysis Version 6.0. *Molecular Biology and Evolution* **30**, 2725-2729 (2013).
- 704 38. Saitou, N. & Nei, M. The neighbor-joining method - a new method for reconstructing  
705 phylogenetic trees. *Molecular Biology and Evolution* **4**, 406-425 (1987).
- 706 39. Felsenstein, J. Confidence-limits on phylogenies - an approach using the bootstrap. *Evolution*  
707 **39**, 783-791 (1985).
- 708 40. Browning, B.L. & Browning, S.R. Efficient multilocus association testing for whole genome  
709 association studies using localized haplotype clustering. *Genetic Epidemiology* **31**, 365-375  
710 (2007).
- 711 41. Lawson, D.J., Hellenthal, G., Myers, S. & Falush, D. Inference of Population Structure using  
712 Dense Haplotype Data. *Plos Genetics* **8**, e1002453 (2012).
- 713 42. Hellenthal, G. *et al.* A genetic atlas of human admixture history. *Science (New York, N.Y.)* **343**,  
714 747-751 (2014).
- 715 43. Leslie, S. *et al.* The fine scale genetic structure of the British population. *Nature* **519**, 309-314  
716 (2015).
- 717 44. Simcic, M. *et al.* Recovery of Native Genetic Background in Admixed Populations Using  
718 Haplotypes, Phenotypes, and Pedigree Information - Using Cika Cattle as a Case Breed. *Plos*  
719 *One* **10**, e0123253 (2015).
- 720 45. Bowcock, A.M. *et al.* High-resolution of human evolutionary trees with polymorphic  
721 microsatellites. *Nature* **368**, 455-457 (1994).
- 722 46. Huson, D.H. & Bryant, D. Application of phylogenetic networks in evolutionary studies.  
723 *Molecular Biology and Evolution* **23**, 254-267 (2006).
- 724 47. Kuleshov, M.V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016  
725 update. *Nucleic Acids Research* **44**, W90-W97 (2016).
- 726 48. Chen, E.Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis  
727 tool. *BMC Bioinformatics* **14**, 128-128 (2013).
- 728 49. Powell, J.E., Visscher, P.M. & Goddard, M.E. Reconciling the analysis of IBD and IBS in  
729 complex trait studies. *Nature Reviews Genetics* **11**, 800-805 (2010).
- 730 50. Meuwissen, T.H.E. & Goddard, M.E. Prediction of identity by descent probabilities from  
731 marker-haplotypes. *Genetics Selection Evolution* **33**, 605-634 (2001).
- 732 51. Lee, S.H. & Van der Werf, J.H.J. Using dominance relationship coefficients based on linkage  
733 disequilibrium and linkage with a general complex pedigree to increase mapping resolution.  
734 *Genetics* **174**, 1009-1016 (2006).
- 735 52. Meuwissen, T.H., Karlsen, A., Lien, S., Olsaker, I. & Goddard, M.E. Fine mapping of a  
736 quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium  
737 mapping. *Genetics* **161**, 373-9 (2002).
- 738 53. Olsen, H.G. *et al.* Fine mapping of milk production QTL on BTA6 by combined linkage and  
739 linkage disequilibrium analysis. *J Dairy Sci* **87**, 690-8 (2004).
- 740 54. Rausch, T. *et al.* DELLY: structural variant discovery by integrated paired-end and split-read  
741 analysis. *Bioinformatics* **28**, i333-i339 (2012).
- 742 55. Corpet, F. Multiple sequence alignment with hierarchical-clustering. *Nucleic Acids Research*  
743 **16**, 10881-10890 (1988).
- 744 56. Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. WebLogo: A sequence logo generator.  
745 *Genome Research* **14**, 1188-1190 (2004).

746





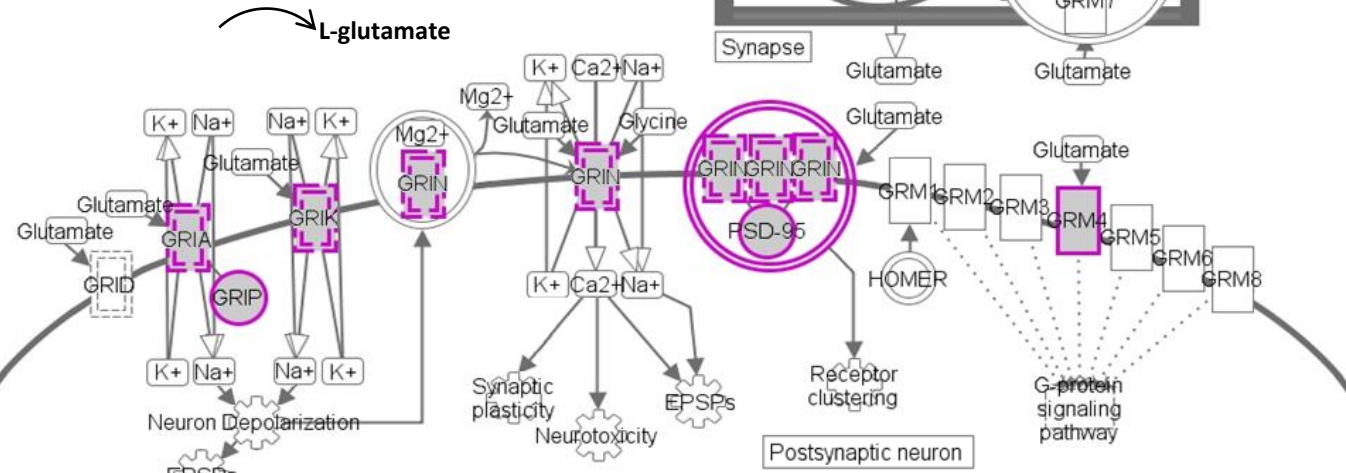
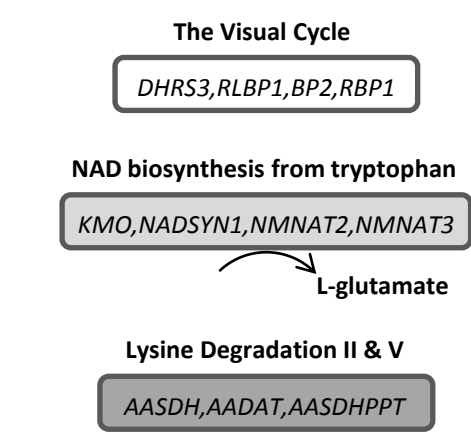
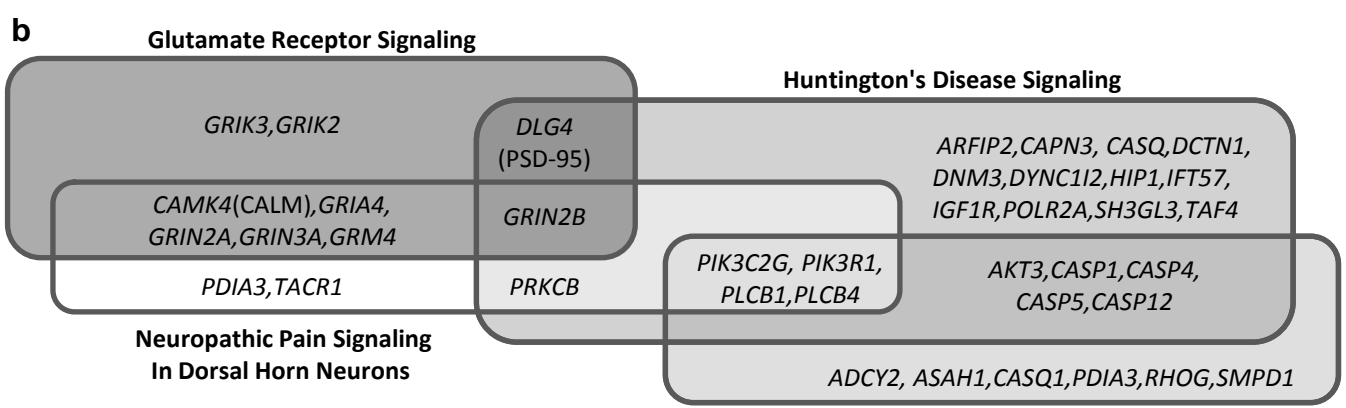
# Brain Nervous System Synapses Arteries Heart

# Neurons

# Neurological Diseases

Psychological Disorders Hematopoiesis  
 Axons Astrocytes Gornea Ear Head Blood Pressure  
 Neurites Behavior Blood Vessels  
 Heart Beats Gonads  
 Kidney GermCells  
 GenitalOrgans  
 Leukocytes Skeleton  
 Muscles  
 Fibroblasts Myoblasts  
 Lipid Metabolism Carbohydrate Metabolism

Dermatological Diseases Breast Lymphocytes Bones  
 Gastrointestinal Diseases Cilia Infectious Diseases Lymphatic System  
 Metabolic Disease Embryo Body Cavity  
 Vitamin and Mineral Metabolism Liver Adipocytes



**Glutamate Receptor Signaling**

