bioRxiv preprint doi: https://doi.org/10.1101/442012; this version posted October 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1 # Balancing selection at a premature stop mutation in the *myostatin* gene underlies a
2 recessive leg weakness syndrome in pigs

3 Oswald Matika[1], Diego Robledo[1], Ricardo Pong-Wong[1], Stephen C. Bishop[1@], Valentina
4 Riggio[1], Heather Finlayson[1], Natalie R. Lowe[1], Annabelle E. Hoste[2], Grant A. Walling[2],
5 Alan L. Archibald[1], John A. Woolliams[1‡], and Ross D. Houston[1‡]

6 [1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh,
7 Midlothian, EH25 9RG, UK
8 [2]JSR Genetics, Southburn, Driffield, East Yorkshire, YO25 9ED, UK.

9 [@]Deceased

10 ‡ These authors contributed equally to this work

11 ## ABSTRACT

12 Balancing selection provides a plausible explanation for the maintenance of deleterious
13 alleles at moderate frequency in livestock, including lethal recessives exhibiting
14 heterozygous advantage in carriers. In the current study, a leg weakness syndrome
15 causing mortality of piglets in a commercial line showed monogenic recessive
16 inheritance, and a region on chromosome 15 associated with the syndrome was identified
17 by homozygosity mapping. Whole genome resequencing of cases and controls identified
18 a mutation coding for a premature stop codon within exon 3 of the porcine *Myostatin*
19 *(MSTN*) gene, similar to those causing a double-muscling phenotype observed in several
20 mammalian species. The *MSTN* mutation was in Hardy-Weinberg equilibrium in the
21 population at birth, but significantly distorted amongst animals still in the herd at 110 kg,
22 due to an absence of homozygous mutant genotypes. In heterozygous form, the *MSTN*
23 mutation was associated with a major increase in muscle depth and decrease in fat depth,
24 suggesting that the deleterious allele was maintained at moderate frequency due to

bioRxiv preprint doi: https://doi.org/10.1101/442012; this version posted October 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

25 heterozygous advantage. Knockout of the porcine *MSTN* by gene editing has previously
26 been linked to problems of low piglet survival and lameness. This *MSTN* mutation is an
27 example of putative balancing selection in livestock, providing a plausible explanation
28 for the lack of disrupting MSTN mutations in pigs despite many generations of selection
29 for lean growth.

30 **Keywords: leg weakness, *MSTN*, balancing selection, heterozygous advantage, pigs**

31 Introduction

32 Leg weakness is heterogeneous condition causing lameness in pigs, and has negative
33 impacts on both animal welfare and productivity [1, 2]. Significant heritability estimates
34 have been reported for leg weakness traits [reviewed in 3], with moderate to high
35 estimates in certain pig breeds, e.g. $h^2 = 0.45$ in Landrace [4]. Several quantitative trait
36 loci (QTL) have been identified for these traits, albeit they are generally not consistent
37 across studies and breeds [5-8], which may be partly due to the heterogeneity of this
38 condition. Interestingly, significant genetic correlations between leg weakness and other
39 production traits (such as growth and muscle depth) have been detected [4]. Further, in a
40 divergent selection experiment in Duroc lines, selection for high leg weakness was
41 associated with a significant increase in muscle length and weight [9]. Taken together,
42 these results suggest a degree of antagonistic genetic relationship between leg weakness
43 and muscle growth traits in pigs, potentially explaining increases in the syndrome
44 observed with intense selection for lean growth in recent decades.

45 Deleterious alleles can be maintained at relatively high frequency in commercial livestock
46 populations due to heterozygous advantage for traits under selection [10]. Examples of
47 such balancing selection in cattle include a frame-shift mutation in the *mannose receptor*
48 *C typed 2* (*MRC2)* gene responsible for crooked tail syndrome and also associated with

bioRxiv preprint doi: https://doi.org/10.1101/442012; this version posted October 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

49 increased muscle mass in Belgian Blue [11], and a large deletion with antagonistic effects
50 on fertility and milk production traits in Nordic Red [12, 13] is likely to have caused an
51 increase in incidence of porcine stress syndrome (also known as malignant hyperthermia)
52 in the 1970s and 1980s, due to the association of the causative missense mutation with
53 reduced backfat – a trait under selection. More recently, a recessive embryonic lethal
54 deletion in the *bone structure and strength QTL 9* (*BSS9)* gene was associated with
55 heterozygous advantage for growth rate, explaining an unexpectedly high frequency of
56 this allele in a commercial pig line [14].

57 We investigated genetic parameters and mode of inheritance for a leg weakness syndrome
58 causing piglet mortality in a commercial line of Large White pigs. A monogenic recessive
59 inheritance was observed, and homozygosity mapping was used to identify a genomic
60 region on *sus scrofa* chromosome (SSC) 15 associated with the trait. A mutation causing
61 a premature stop codon in exon 3 of the *Myostatin* (*MSTN)* gene (similar to mutations
62 causing the 'double-muscling' phenotype in cattle [15] was the outstanding functional
63 candidate in the region. Comparison of *MSTN* genotype frequencies at birth and 110 kg
64 supported this hypothesis, and carriers were shown to have significantly higher muscle
65 mass and reduced fat depth than wild type homozygotes. Therefore, we propose that the
66 *MSTN* mutant allele is highly deleterious in this population in homozygous form, but was
67 maintained at moderate frequency due to heterozygous advantage.

68 **RESULTS**

69 **The piglet leg weakness trait shows a recessive mode of inheritance**

70 Estimates of heritability for the leg weakness syndrome (analysed as a binary trait on the
71 underlying liability scale) was high (0.57 ± 0.10 in the sire and dam model) with low
72 (0.17 ± 0.02 and 0.11 ± 0.02) but significant effects observed for permanent

bioRxiv preprint doi: https://doi.org/10.1101/442012; this version posted October 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

73 environmental effects due to the dam and litter, respectively (Table 2, Table S1). The
74 overall prevalence of leg weakness in the commercial cohort was 6.3 % (Table 1). When
75 only affected litters were considered, the mean proportion of affected piglets was 23 % ±
76 0.7. This within-litter prevalence is consistent with the expectation under the hypothesis
77 of a single recessive locus (i.e. 25 %). Complex Bayesian segregation analysis suggested
78 that almost all the variation was explained by a single locus with almost no environmental
79 variation. The estimate of the additive effect was 0.50 ± 0.001 and dominance effect was
80 -0.50 ± 0.001, which is in precise agreement with a recessive locus model.

81 **Causative variant maps to chromosome 15**

82 Homozygosity mapping was used to map the underlying recessive variant, and the longest
83 homozygous segment was a region of ~ 8.3 Mbp on SSC 15. In this region, all 55
84 informative single nucleotide polymorphisms (SNPs) on the Illumina PorcineSNP60 SNP
85 chip [16] were homozygous in affected animals, but contained SNPs that were
86 heterozygous or homozygous for the alternative allele in the unaffected animals. The
87 segment started with SNP ALGA0110636 (rs81338938) at position 86,745,668 in the
88 current Sscrofa11.1 reference genome assembly (Genbank assembly accession
89 GCA_000003025.6) and finished with SNP H3GA0044732 (rs80936849) at position
90 95,062,143 (Fig 1a and b). The *MSTN* gene was located within this homozygous segment,
91 from position 94,620,269 - 94,628,630. The 8.3 Mbp segment was assumed to represent
92 a selective sweep likely to contain the underlying causative mutation, and became the
93 focus of further analyses to discover and characterise this mutation.

94 **Sequence analysis reveals *MSTN* mutation as causative candidate**

95 To identify candidates for the causative variant, whole genome sequence data from ten
96 cases, six presumed heterozygous carrier dams, and 22 controls were analysed. A total of

bioRxiv preprint doi: https://doi.org/10.1101/442012; this version posted October 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

97 40 SNPs identified within the homozygous segment fitted the pattern of a potential
98 causative variant assuming a recessive mode of inheritance. Functional annotation of
99 these SNPs revealed that 19 were intergenic, 19 were intronic, 1 was in a pseudogene,
100 and 1 caused a premature stop codon. There were also 10 InDels identified, 3 of which
101 were intergenic and 7 of which were intronic (Figure S1). The outstanding functional
102 candidate was a mutation in the third exon of the *MSTN* locus that results in the
103 replacement of a codon for glutamic acid with a stop codon in exon 3 at position 274
104 (c.820G>T; p.E274*) (Figure 2A). The mutation is located in a region that is highly
105 conserved across multiple species, and is predicted to result in truncation of the protein
106 (Figure 2B). Functional annotation of all other variants detected in the selective sweep
107 region did not reveal any other obvious causative candidates. This SNP and stop gain
108 mutation was not present on the Ensembl variation database, accessed 25th July 2018.

109 **Changes in mutant allele frequency from piglets to adults**

110 The *MSTN* c.820G>T mutation showed no statistically significant deviation from Hardy
111 Weinberg equilibrium (HWE) in the 486 piglets sampled at birth ($q = 0.22$, $\alpha = 0.019$,
112 $\chi^2=0.18$, $P > 0.05$). Random mating of the dioecious population would be expected to
113 result in a value of α that is slightly negative [17], and the value observed does not differ
114 significantly from this value. However, the mutation deviated significantly from HWE at
115 40 kg ($q = 0.17$ $\alpha = - 0.180$, $\chi^2=12.2$, $P > 0.001$) and at 110 kg ($q = 0.17$ $\alpha = - 0.210$,
116 $\chi^2=11.45$, $P > 0.001$). This was due to the loss of homozygous mutant piglets, with all but
117 one dying (or being euthanized) shortly after birth, and the remaining piglet being
118 euthanized due to poor health before it reached 110 kg. The large change in α in a negative
119 direction is quantitative evidence of the selective disappearance of homozygote
120 genotypes, as opposed to disappearance as a result of selection against the allele itself.
121 There were no significant changes in the relative genotype frequencies (GG, GT) over the

bioRxiv preprint doi: https://doi.org/10.1101/442012; this version posted October 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

122 total period or any sub-period from birth to the end of the test, confirming all changes in
123 q and a are due to the selective loss of homozygotes, and that any other mortality or
124 culling was at random with respective to *MSTN* genotype.

125 **Association of the *MSTN* mutation with performance traits**

126 The association of the porcine *MSTN* c.820G>T mutation with performance traits was
127 assessed on 384 pigs which had completed a commercial performance test. Given the loss
128 of the homozygous mutant animals the effect of the *MSTN* c.820G>T mutation was only
129 estimated by the difference between the heterozygotes and the wild type pigs. The
130 genotype means and differences are shown in Table 3, with the most notable of these
131 being a major increase in muscle depth and a reduction in fat depth in the carriers ($p <$
132 0.001), with no evidence of a difference in live weight at 110 kg. Approximately 31 % of
133 the genetic variation in muscle depth and 18 % of the genetic variation in fat depth was
134 explained by this single variant (Table 3). The heterozygous animals had on average
135 5 mm increased muscle depth, and 1.7 mm decreased backfat depth when compared with
136 wild type homozygous animals.

137 **DISCUSSION**

138 A piglet leg weakness syndrome identified in a commercial line of Large White pigs
139 showed moderate to high heritability, consistent with the upper range of estimates
140 reported in the literature [18, 19, 2]. The within-litter incidence of the syndrome and the
141 complex Bayesian segregation analysis both pointed to a monogenic recessive condition.
142 Homozygosity mapping revealed a 8.3 Mbp segment on SSC15 as a putative selective
143 sweep likely to contain the causative variant. The outstanding functional candidate variant
144 identified by whole genome sequencing of cases and controls caused a premature stop
145 codon within exon 3 of the *MSTN* gene (Figure 2), and this variant was in HWE at birth

bioRxiv preprint doi: https://doi.org/10.1101/442012; this version posted October 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

146 but significantly distorted by 110 kg due to an absence of homozygous mutant animals.
147 Knockout of *MSTN* by gene editing in previous studies has been associated with poor
148 health and mortality of knockout piglets [20-23], including observations of a piglet leg
149 weakness syndrome with an inability to stand or walk [21], strikingly similar to the
150 syndrome described in the current study.

151 *MSTN* is a member of the *transforming growth factor beta* (*TGF-β*) superfamily, which
152 is highly conserved across species, and is typically expressed in developing and mature
153 skeletal muscle as a key regulator of muscle growth [24]. The *MSTN* gene has been a
154 gene of interest to animal breeders for over twenty years since the discovery of loss-of-
155 function mutations in the cattle *MSTN* gene, which cause muscle hypertrophy leading to
156 double muscling phenotypes [15, 24, 25]. Interestingly, these loss-of-function mutations
157 causing double muscling are frequently due to premature stop mutations in the highly
158 conserved exon 3 of *MSTN [25], (*Figure 3). Further, in the case of Marchigiana beef
159 cattle, the *MSTN* variant causing double muscling results is due to replacement of
160 Glutamic Acid with a stop codon (the same change observed in the current study) in the
161 equivalent position in the protein as observed in the current study [26]. This observation
162 provides additional indirect evidence that the porcine *MSTN* mutation described herein is
163 likely to cause the increase in muscle depth and decrease in fat depth associated with
164 carrier animals.

165 In livestock breeding schemes, *MSTN*-inactivating mutations have been retained due to
166 selection for increased lean growth associated with meat production, particularly for
167 double-muscled cattle. However, despite many generations of selection for lean growth
168 in pigs no such mutations have been reported previously. Associations between
169 polymorphisms within the porcine *MSTN* gene and production traits have been shown in
170 small-scale studies [27-29]. Further, a genome-wide association study in a sire line Large

bioRxiv preprint doi: https://doi.org/10.1101/442012; this version posted October 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

171 White population (related to the animals in the current study) detected SNPs on SSC 15
172 significantly associated with rib fat between 81.1 and 87.8 Mbp [30]. While these SNPs
173 are between 7 and 13 Mbp closer to the centromere than *MSTN*, two of these SNPs
174 overlap with the region on homozygosity indicative of a selective sweep in the current
175 study. Therefore, it is plausible that this association may be due to linkage disequilibrium
176 (LD) with the *MSTN* variant, or that other variants impacting performance traits exist on
177 SSC15. The selection index applied in this line explicitly benefited animals with positive
178 muscle depth and negative fat depth estimated breeding values, with analysis of the index
179 showing that heterozygotes had a slight but significant advantage ($p<0.05$). Therefore, it
180 is likely that the MSTN c.820G>T mutation was maintained at moderate frequency in this
181 line despite its deleterious impact on piglet mortality due to heterozygous advantage for
182 muscle and fat traits. Similar examples of balancing selection have observed to explain
183 the maintenance of deleterious alleles at moderate frequency in commercial cattle [10,
184 31] and pig [14] populations. The results described herein have major implications for the
185 targeted ablation of *MSTN* via gene editing to increase lean growth in pigs, and provide a
186 plausible explanation of why *MSTN* loss-of-function mutations have not previously been
187 reported in pigs despite decades of selection for lean growth.

188 **Materials and Methods**

189 **Ethics statement**

190 All samples were collected on a commercial nucleus farm as part of standard husbandry
191 and management procedures in the nucleus herd, which complied with conventional UK
192 red tractor farm assurance standards (https://assurance.redtractor.org.uk/) where sick or
193 injured livestock that do not respond to treatment are promptly and humanely euthanized
194 by a trained and competent stockperson.

bioRxiv preprint doi: https://doi.org/10.1101/442012; this version posted October 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

195 **Animals**

196 A piglet leg weakness syndrome was characterised in a Large White sire line, reared in a
197 nucleus herd, under standard conditions but with additional data recording. Leg weakness
198 trait observations were collected on 19,006 piglets from 2007 to 2010, during which time
199 a high incidence of the syndrome was detected. In addition, DNA was sampled on a
200 further 119 piglets in 2011 and 486 piglets from the same population in 2012, of these
201 384 also had weight and carcass phenotypes available. The cohort born in 2007 to 2010
202 will be referred to as the commercial cohort and those in 2011 and 2012 as the survey
203 cohort. Pedigree was available for all animals and spanned 9 generations. Details of the
204 data and pedigree structure are presented in Table 1.

205 The leg weakness in the phenotyped animals was visually classified as normal or affected
206 (0 / 1 respectively). The leg defect is characterised by the piglet not being able to
207 straighten its legs to stand, this being most apparent for the front legs, and being slow to
208 suckle. Videos exemplifying the syndrome are given in Additional File 1. Detailed post
209 mortems were conducted on two affected individuals but the results were ineffective in
210 providing additional diagnostic aids (see Additional File 2). The Online Mendelian
211 Inheritance in Animals database (OMIA: http://omia.org/OMIA000585/9823/) was
212 searched for previous reports of leg weakness in pigs but the syndrome observed here did
213 not appear. These problems frequently resulted in death from either starvation or being
214 crushed by the sow. Where the outcome resulted in poor welfare the piglets were
215 euthanised in accord with the Ethics Statement. Additional farrowing data were collected
216 on the females in the commercial cohort including numbers born alive, dead or
217 mummified, parity and year of birth. Body weights, ultrasound muscle and fat depths
218 were obtained for individuals in the survey cohort that were retained in the herd until
219 commercial slaughter age (when average weight is 110kg). The ultrasonic measurements

bioRxiv preprint doi: https://doi.org/10.1101/442012; this version posted October 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

220 taken were the average depth of the *m. longissimus dorsi* and overlaying subcutaneous fat
221 layer across the last four ribs.

222 **Statistical analyses of genetic parameters**

223 Initial studies to establish the genetic basis of the leg weakness syndrome were undertaken
224 by fitting linear mixed models to the commercial cohort. The binary record of the
225 syndrome was modelled on the observed 0 / 1 scale and the full model fitted was:

226 $$\mathbf{y}=\mathbf{X}\boldsymbol{\beta} + \mathbf{Z_1u} + \mathbf{Z_2v} + \mathbf{Z_3w} + \mathbf{e} \qquad \textbf{(1)}$$

227 where **y** is the vector of leg weakness phenotypes; **β**, a vector of fixed effects for month
228 of observation (43 df), sex (1 df), parity (7 df), numbers in litter born dead (1 df) or alive
229 (1 df), with design matrix **X**; **u**, additive polygenic effects, assumed to be distributed
230 MVN(0, $\mathbf{A}\sigma^2_a$) with design matrix $\mathbf{Z_1}$; **v**, random litter effects assumed to be distributed
231 MVN(0, $\mathbf{I}\sigma^2_v$), with design matrix $\mathbf{Z}_2$; **w**, maternal environment effects across litters
232 assumed to be distributed MVN(0, $\mathbf{I}\sigma^2_w$) with design matrix $\mathbf{Z}_3$; and **e** residuals assumed
233 to be distributed MVN(0, $\mathbf{I}\sigma^2_e$). Variations in this model were fitted replacing the
234 individual polygenic effects with sire and/or dam effects, assumed to be distributed
235 MVN(0, $\mathbf{A}\sigma^2_s$) and MVN(0, $\mathbf{A}\sigma^2_d$) respectively. All models were fitted using the
236 ASREML software [32]. Likelihood ratio tests were used to assess the random effects.
237 In addition, an analogous threshold mode with an underlying continuous liability was
238 fitted with a logit link function and sire and dam effects associated with the pedigree, but
239 not with individual polygenic effects, following recommendation of Gilmour et al. [36].
240 For the full model, the phenotypic variance was calculated as $\sigma^2_p = \sigma^2_u + \sigma^2_v + \sigma^2_w + \sigma^2_e$.
241 Where sire and dam models were used, σ^2_u was replaced by $\sigma^2_s + \sigma^2_d$. Heritability (h^2)
242 was calculated as σ^2_u/σ^2_p or $2(\sigma^2_s + \sigma^2_d)/\sigma^2_p$ depending on the model. The proportion of

bioRxiv preprint doi: https://doi.org/10.1101/442012; this version posted October 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

243 variance explained by the litter and maternal environmental effects were estimated as
244 σ^2_v/σ^2_p and σ^2_w/σ^2_p, respectively. Heritabilities on the observed scale (0/1) was
245 transformed to an underlying liability scale following Dempster and Lerner [33] using the
246 observed prevalence of the syndrome in the commercial population which was 6.3%

247 Inspection of the data suggested that the syndrome may be due to a single gene with the
248 predisposing deleterious allele showing a recessive mode of inheritance, and this
249 hypothesis was tested using chi square tests and segregation analyses. An initial test of a
250 monogenic recessive mode of inheritance was carried out by pooling all affected litters,
251 estimating the probability of being affected conditional on being born in an affected litter,
252 and using chi-squared to test the null hypothesis that the probability of being affected was
253 0.25. A weakness of this approach is that some litters by chance will have no affected
254 offspring, so a more complex segregation model was fitted. This model included all
255 known phenotypes and pedigree data, and assumed a monogenic inheritance with
256 environmental variation fitted by Gibbs sampling [34].

257 **Homozygosity mapping of the recessive mutation**

258 Ten affected animals from different litters and 10 unaffected full-sib controls from the
259 commercial cohort were genotyped using the Illumina PorcineSNP60 SNP chip [16].
260 Only those SNPs that mapped to known positions on autosomal chromosomes and were
261 not fixed nor completely heterozygous were retained. This left 38,570 segregating
262 autosomal SNPs for the use in homozygosity mapping. Homozygous regions were
263 assessed by alignment with the Sscrofa11.1 reference genome assembly sequence
264 (Genbank assembly accession GCA_000003025.6).

265 **Whole genome resequencing**

bioRxiv preprint doi: https://doi.org/10.1101/442012; this version posted October 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

266 The genomes of the ten cases used for homozygosity mapping and six separate dams with
267 affected offspring, assumed heterozygotes, were whole genome shotgun sequenced on an
268 Illumina HiSeq 2500 platform. The dams were individually sequenced with a 10x genome
269 coverage. The piglets were barcoded and individually sequenced at 3x coverage to
270 achieve 30x coverage for the pool. The full sequencing output resulted in ~1.3 billion
271 paired-end reads with an average of 48 million paired-end reads/sample for the piglets
272 and 157 million paired-end reads/sample for the dams. Quality filtering and removal of
273 residual adaptor sequences was conducted on read pairs using Trimmomatic v.0.32 [35].
274 Only reads where both pairs had a length greater than 32 bp post-filtering were retained,
275 leaving a total of ~1.2 bn paired-end reads.

276 Whole genome resequencing was followed by alignment to the Sscrofa11.1 assembly;
277 using the Burrows-Wheeler Aligner with default parameters [36]. The average alignment
278 rate of properly paired reads was of 92 %. PCR duplicates were marked using Picard
279 Tools (http://broadinstitute.github.io/picard). Variant calling was performed using the
280 Genome Analysis Toolkit (GATK) HaplotypeCaller after read recalibration [37]. The
281 parameter setting for the hard filters that were applied to the raw genotypes were:
282 QualByDepth < 2.0, FisherStrand > 60.0, RMSMappingQuality < 40.0,
283 MappingQualitySumTest < -12.5, ReadPosRankSumTest < -8.0.

284 Candidate loci were identified from the sequences of the 16 animals from this study plus
285 22 additional *Sus scrofa* control sequences obtained from a public database [38],
286 comprising 7 domesticated breeds (Duroc, Hampshire, Jiangquhai, Landrace, Large
287 White, Meishan and Pietrain) and wild boar. The following criteria were used to identify
288 candidate SNPs:

289 i. Homozygous for the same allele in all the affected piglets

bioRxiv preprint doi: https://doi.org/10.1101/442012; this version posted October 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

290 ii. Heterozygous in parents of affected piglets (i.e. putative carriers)
291 iii. Heterozygous or homozygous for the alternative allele (i.e. the one not observed
292 in the affected offspring) in the control (unaffected) animals

293 It is worth noting that the limited sequencing depth means that both alleles will not be
294 detected for all bases in all individuals. This limitation is particularly relevant to the
295 reliable detection of heterozygous SNPs.

296

297 **Genotyping**

298 A 'kompetitive allele specific PCR' (KASP) assay was designed by LGC Genomics
299 (Teddington, UK) to enable genotyping of the mutation in the *MSTN* stop codon in large
300 numbers of animals. The survey cohort of 486 piglets sampled at birth were genotyped
301 by LGC. Of these, 265 remained as candidates for the final selection at 110 kg with a
302 complete record of their performance test, together with another 119 pigs phenotyped at
303 slaughter age (total n = 374). In both the 486 piglets and the subsequent subsets surviving
304 to 40 and 110 kg, the frequency of the *MSTN* mutation (q) was calculated by counting,
305 and the departure from Hardy Weinberg equilibrium the genotypes was estimated as α
306 $=1-H_{obs}/H_{exp}$ [17] where H_{obs} is the observed heterozygosity and H_{exp} is the expected
307 heterozygosity calculated as 2q(1-q). The significance of departure from true random
308 mating genotype frequencies ($\alpha = 0$) was tested using a chi-squared test.

309 **Association analysis**

310 Associations between the *MSTN* c.820G>T locus and variation in performance traits in
311 commercial testing conditions were examined in the survey cohort (n = 384). The
312 performance test was started at a target weights of 40 kg, at an average age of 85 days,

bioRxiv preprint doi: https://doi.org/10.1101/442012; this version posted October 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

313 and continued for 54 (s.d. 12) days. The performance tests were performed over two
314 distinct periods. The traits available were live weights and ages at the start and the end of
315 the test, ultrasonic muscle and fat depths measured at the end of the test, days from birth
316 to 40 kg and days from 40 to 110 kg. Univariate mixed models were fitted to these data
317 in ASReml-R4 using the following model:

318 $$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}_1\boldsymbol{\beta} + \mathbf{X}_2 b + \mathbf{Z}\mathbf{u} + e \qquad \textbf{(2)}$$

319 where **y** is the vector of phenotypes; μ, a fitted mean, and **1**, a vector of 1's; **β**, a vector
320 of fixed nuisance effects with design matrix $\mathbf{X}_1$; *b*, a scalar fixed effect for the effect of
321 SNP genotype with design matrix $\mathbf{X}_2$, this has only 1 df due to the absence of homozygotes
322 completing the test; **u** , additive polygenic effects assumed to be distributed MVN(0,
323 $\mathbf{A}\sigma^2_a$), with design matrix **Z**; and **e**, residuals assumed to be distributed MVN(0, $\mathbf{I}\sigma^2_e$).
324 For all traits, the sex of the piglet (1 df), parity of dam (4 df) and period of testing (1 df)
325 were fitted as nuisance factors, together with cubic smoothing splines for the start date of
326 the test fitted separately within each period [39]. The age at the time of measurement was
327 fitted as covariate (1 df) for all traits other than days to 40 kg and days from 40 to 110 kg.
328 The significance of fixed effects was assessed using Wald tests [ASReml Manual].

329 **ACKNOWLEDGEMENTS**

330 Institute Strategic Funding Grants to The Roslin Institute (BBS/E/D/20211553
331 BBS/E/D/30002275). Edinburgh Genomics is partly supported through core grants from
332 NERC (R8/H10/56), MRC (MR/K001744/1) and BBSRC (BB/J004243/1). Diego
333 Robledo is supported by a Newton International Fellowship of the Royal Society
334 (NF160037)

335 **REFERENCES:**

bioRxiv preprint doi: https://doi.org/10.1101/442012; this version posted October 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

336 1. Hill MA. Economic relevance, diagnosis, and countermeasures for degenerative joint disease
337 (osteoarthrosis) and dyschondroplasia (osteochondrosis) in pigs. Journal of the American
338 Veterinary Medical Association. 1990;197(2):254-9.
339 2. Rothschild MF, Christian LL. Genetic-control of front-leg weakness in duroc swine .1. Direct
340 response to 5 generations of divergent selection. Livestock Production Science. 1988;19(3-
341 4):459-71.
342 3. Fukawa K, Kusuhara S. The genetic and non-genetic aspects of leg weakness and
343 osteochondrosis in pigs - Review. Asian Australas J Anim Sci. 2001;14(1):114-22.
344 4. Jorgensen B, Vestergaard T. Genetics of leg weakness in boars at the Danish pig breeding
345 stations. Acta Agriculturae Scandinavica. 1990;40(1):59-69.
346 5. Guo YM, Ai HS, Ren J, Wang GJ, Wen Y, Mao HR et al. A whole genome scan for quantitative
347 trait loci for leg weakness and its related traits in a large F-2 intercross population between
348 White Duroc and Erhualian. J Anim Sci. 2009;87(5):1569-75.
349 6. Uemoto Y, Sato S, Ohnishi C, Hirose K, Kameyama K, Fukawa K et al. Quantitative trait loci for
350 leg weakness traits in a Landrace purebred population. Anim Sci J. 2010;81(1):28-33.
351 7. Laenoi W, Uddin MJ, Cinar MU, Phatsara C, Tesfaye D, Scholz AM et al. Molecular
352 characterization and methylation study of matrix gla protein in articular cartilage from pig with
353 osteochondrosis. Gene. 2010;459(1-2):24-31.
354 8. Laenoi W, Uddin MJ, Cinar MU, Grosse-Brinkhaus C, Tesfaye D, Jonas E et al. Quantitative trait
355 loci analysis for leg weakness-related traits in a Duroc x Pietrain crossbred population. Genet Sel
356 Evol. 2011;43:7.
357 9. Draper DD, Rothschild MF, Christian LL. Effects of divergent selection for leg weakness on
358 muscle and bone characteristics in Duroc swine. Genet Sel Evol. 1992;24(4):363-74.
359 10. Hedrick PW. Heterozygote Advantage:The Effect of Artificial Selection in Livestock and Pets.
360 J Hered. 2015;106(2):141-54.
361 11. Fasquelle C, Sartelet A, Li WB, Dive M, Tamma N, Michaux C et al. Balancing Selection of a
362 Frame-Shift Mutation in the MRC2 Gene Accounts for the Outbreak of the Crooked Tail
363 Syndrome in Belgian Blue Cattle. PLoS Genet. 2009;5(9):9.
364 12. Otsu K, Khanna VK, Archibald AL, Maclennan DH. Cosegregation of porcine malignant
365 hyperthermia and a probable causal mutation in the skeletal-muscle ryanodine receptor gene in
366 backcross families. Genomics. 1991;11(3):744-50.
367 13. Fujii J, Otsu K, Zorzato F, Deleon S, Khanna VK, Weiler JE et al. Identification of a mutation in
368 porcine ryanodine receptor associated with malignant hyperthermia. Science.
369 1991;253(5018):448-51.
370 14. Derks MFL, Lopes MS, Bosse M, Madsen O, Dibbits B, Harlizius B et al. Balancing selection
371 on a recessive lethal deletion with pleiotropic effects on two neighboring genes in the porcine
372 genome. PLoS Genet. 2018.
373 15. Grobet L, Martin LJR, Poncelet D, Pirottin D, Brouwers B, Riquet J et al. A deletion in the
374 bovine myostatin gene causes the double-muscled phenotype in cattle. Nature Genetics.
375 1997;17(1):71-4.
376 16. Ramos AM, Crooijmans R, Affara NA, Amaral AJ, Archibald AL, Beever JE et al. Design of a
377 High Density SNP Genotyping Assay in the Pig Using SNPs Identified and Characterized by Next
378 Generation Sequencing Technology. PLoS One. 2009;4(8):13.
379 17. Robertson A. Inbreeding in artificial selection programmes. Genetical Research.
380 1961;2(02):189-94.
381 18. Bereskin B. Genetic-aspects of feet and legs soundness in swine. J Anim Sci. 1979;48(6):1322-
382 8.
383 19. Jorgensen B, Andersen S. Genetic parameters for osteochondrosis in Danish Landrace and
384 Yorkshire boars and correlations with leg weakness and production traits. Anim Sci.
385 2000;71:427-34.
386 20. Kang JD, Kim S, Zhu HY, Jin L, Guo Q, Li XC et al. Generation of cloned adult muscular pigs
387 with myostatin gene mutation by genetic engineering. RSC Adv. 2017;7(21):12541-9.

bioRxiv preprint doi: https://doi.org/10.1101/442012; this version posted October 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

388 21. Kang Q, Hu Y, Zou Y, HU W, Li L, Chang F et al., editors. Improving pig genetic resistance and
389 muscle production through molecular biology. Proceedings, 10th World Congress of Genetics
390 Applied to Livestock Production; 2014 August 17-22; Vancouver, BC, Canada.
391 22. Wang KK, Ouyang HS, Xie ZC, Yao CG, Guo NN, Li MJ et al. Efficient Generation of Myostatin
392 Mutations in Pigs Using the CRISPR/Cas9 System. Sci Rep. 2015;5.
393 23. Cyranoski D. Super-muscly pigs created by small genetic tweak. Nature. 2015;523(7558):13-
394 4.
395 24. McPherron AC, Lee SJ. Double muscling in cattle due to mutations in the myostatin gene.
396 Proceedings of the National Academy of Sciences of the United States of America.
397 1997;94(23):12457-61.
398 25. Bellinge RHS, Liberles DA, Iaschi SPA, O'Brien PA, Tay GK. Myostatin and its implications on
399 animal breeding: a review. Animal Genetics. 2005;36(1):1-6.
400 26. Marchitelli C, Savarese MC, Crisa A, Nardone A, Marsan PA, Valentini A. Double muscling in
401 Marchigiana beef breed is caused by a stop codon in the third exon of myostatin gene. Mamm
402 Genome. 2003;14(6):392-5.
403 27. Stinckens A, Luyten T, Bijttebier J, Van den Maagdenberg K, Dieltiens D, Janssens S et al.
404 Characterization of the complete porcine MSTN gene and expression levels in pig breeds
405 differing in muscularity. Animal Genetics. 2008;39(6):586-96.
406 28. Tu PA, Shiau JW, Ding ST, Lin EC, Wu MC, Wang PH. The association of genetic varitations in
407 the promoter region of myostatin gene with growth traits in Duroc pigs. Animal Biotechnology.
408 2012;23(4):291-8.
409 29. Tu PA, Lo LL, Chen YC, Hsu CC, Shiau JW, Lin EC et al. Polymorphisms in the promoter region
410 of myostatin gene are associated with carcass traits in pigs. J Anim Breed Genet.
411 2014;131(2):116-22.
412 30. Fowler KE, Pong-Wong R, Bauer J, Clemente EJ, Reitter CP, Affara NA et al. Genome wide
413 analysis reveals single nucleotide polymorphisms associated with fatness and putative novel
414 copy number variants in three pig breeds. BMC Genomics. 2013;14:15.
415 31. Kadri NK, Sahana G, Charlier C, Iso-Touru T, Guldbrandtsen B, Karim L et al. A 660-Kb Deletion
416 with Antagonistic Effects on Fertility and Milk Production Segregates at High Frequency in Nordic
417 Red Cattle: Additional Evidence for the Common Occurrence of Balancing Selection in Livestock.
418 PLoS Genet. 2014;10(1):11.
419 32. Gilmour AR, Gogel BJ, Cullis BR, Thompson R. ASReml User Guide Release 3.0. 3.0 ed.
420 Hempstead,HP1 1ES, UK.: VSN Int. Ltd Hempstead; 2009.
421 33. Dempster ER, Lerner IM. Heritability of threshold characters. Genetics. 1950;35:212-36.
422 34. Walling GA, Bishop SC, Pong-Wong R, Gittus G, Russel AJF, Rhind SM. Detection of a major
423 gene for litter size in Thoka Cheviot sheep using Bayesian segregation analyses. Anim Sci J.
424 2002;75:339-47.
425 35. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
426 Bioinformatics. 2014;30(15):2114-20.
427 36. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform.
428 Bioinformatics. 2010;26(5):589-95.
429 37. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C et al. A framework for
430 variation discovery and genotyping using next-generation DNA sequencing data. Nature
431 Genetics. 2011;43(5):491-+.
432 38. Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF et al. Analyses
433 of pig genomes provide insight into porcine demography and evolution. Nature.
434 2012;491(7424):393-8.
435 39. White IMS, Thompson R, Brotherstone S. Genetic and environmental smoothing of lactation
436 curves with cubic splines. J Dairy Sci. 1999;82(3):632-8.

437

bioRxiv preprint doi: https://doi.org/10.1101/442012; this version posted October 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

439 Table 1 Number of animals in pedigree and records used for variance components
440 analyses

Description	Data
Number of records	19,006
Number of pedigree records	27,501
Number of generations	9
Number of litter	1,903
Number of Sires	346
Number of Sires of Sire	175
Number of Dams of Sire	239
Number of Dams	1,929
Number of Sires of Dam	240
Number of Dams of Dam	882
Prevalence %	6.3

441

bioRxiv preprint doi: https://doi.org/10.1101/442012; this version posted October 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

442 Table 2 Genetic parameter estimates and standard errors for the trait of leg weakness on
443 the liability scale using the logit transformation showing the outcomes of fitting sire or
444 dam models with or without maternal environment (σ^2_w). All models have litter variance
445 (σ^2_v) fitted

Model[¥]	Sire	Sire and Dam Models	
	σ^2_w >0	σ^2_w >0	σ^2_w =0
σ^2_s	1.069 (0.283)	1.100 (0.290)	1.087 (0.286)
σ^2_d		0.664 (0.231)	1.110 (0.186)
σ^2_v	0.681 (0.123)	0.681 (0.123)	0.769 (0.123)
σ^2_w	0.999 (0.148)	0.437 (0.203)	
σ^2_p	6.049 (0.307)	6.182 (0.322)	6.267 (0.325)
h^2_u	0.71 (0.16)	0.57 (0.10)	0.70 (0.08)
σ^2_v/σ^2_p	0.11 (0.02)	0.11 (0.02)	0.12 (0.02)
σ^2_w/σ^2_p	0.17 (0.2)	0.07 (0.03)	

446 ¥ - σ^2_s, σ^2_d, σ^2_v and σ^2_w are variances due to the sire and dam genetic effects, litter and
447 maternal environment effects. σ^2_p is phenotypic variance. h^2_u is estimated heritability on
448 observed scale, and σ^2_v/σ^2_p and σ^2_w/σ^2_p are the proportion of phenotypic variance
449 explained by the litter and maternal environment effects respectively

450

451

452

453

bioRxiv preprint doi: https://doi.org/10.1101/442012; this version posted October 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

454 Table 3. Estimates and statistical significance of the effect of the *MSTN* c.820G>T locus
455 on the growth and carcass traits of pigs obtained from a commercial performance test.
456 Estimates are shown in absolute units and standardised by phenotypic standard
457 deviations (σ_P). Standard errors are in parentheses. The traits are categorised into: live
458 weights and live weight gain; muscle and fat depths measured by ultrasound at the end
459 of the test either conditional on age or on live weight; and periods to achieve growth
460 targets.

Trait	*MSTN* c.820G>T genotypes			σ_P	TG-GG	Wald
	TG	GG	TG-GG		σ_P units	test
Live weights:						
At start (kg)	36.83 (0.84)	38.18 (0.85)	-1.36 (0.44)	3.81 (0.18)	-0.36	P<0.01
At end (kg)	84.89 (1.59)	87.13 (1.64)	-2.24 (0.94)	7.96 (0.33)	-0.28	P<0.05
Daily gain (kg/d)	1.02 (0.02)	1.03 (0.02)	-0.01 (0.01)	0.11 (0.01)	-0.09	NS
Depths for age:						
Fat depth (mm)	8.42 (0.62)	10.19 (0.62)	-1.76 (0.31)	2.72 (0.13)	-0.65	P<0.001
Muscle depth (mm)	53.21 (1.11)	48.38 (1.10)	4.83 (0.68)	5.95 (0.25)	0.81	P<0.001
Depths for weight:						
Fat depth (mm)	9.25 (0.57)	10.74(0.56)	-1.49 (0.28)	2.49 (0.12)	-0.60	P<0.001
Muscle depth (mm)	54.89 (1.05)	49.49 (1.02)	5.40 (0.63)	5.54 (0.24)	0.97	P<0.001
Periods:						
To 40kg (d)	91.66 (2.12)	88.62 (2.11)	3.04 (1.01)	9.01 (0.45)	0.36	P<0.01
From 40 to 100kg (d)	66.33 (1.67)	66.62 (1.67)	-0.29 (1.81)	6.75 (0.36)	-0.04	NS

461

462

463

bioRxiv preprint doi: https://doi.org/10.1101/442012; this version posted October 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

464 **FIGURES**

465 **Fig 1**

466 **A:** Homozygosity mapping of the leg weakness syndrome assuming a single underlying
467 recessive mutation on Chromosome 15 (SSC15). Vertical blocks in red and yellow
468 represent homozygous genotypes, and blue the heterozygote genotype. The ten cases
469 (above) and ten controls (below) are shown (one per line). A summary of homozygosity
470 mapping is provided cases vs control at the bottom of Fig1A in which if all the animals
471 within a group (i.e. case or control) are homozygous for the same allele, then the relevant
472 colour (red or yellow) is shown and if any animal within a group (i.e. case or control) is
473 heterozygous then the SNP is coloured blue

474 **B**: This is an extract from Fig1a, showing the longest shared haplotype segment (55 SNPs)
475 in the cases on SSC15 ranging from ALGA0110636 (rs81338938) to H3GA0044732
476 (rs80936849) and corresponds to position 86,745,668– 95,062,143 in the new pig
477 reference genome assembly Sscrofa11.1 GCA_000003025.6). The first ten lines are the
478 cases; the second ten lines are the controls with genotypes shown only for the region of
479 homozygosity shared across the controls. Genotypes shown for controls are shown only
480 if they are different to cases (blank genotype in controls within the targeted segment
481 means that they share the same genotype as cases). Finally, the bottom two lines are
482 summary lines shown on the same basis as the summary lines in Fig1A.

bioRxiv preprint doi: https://doi.org/10.1101/442012; this version posted October 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

483 **Fig 2**

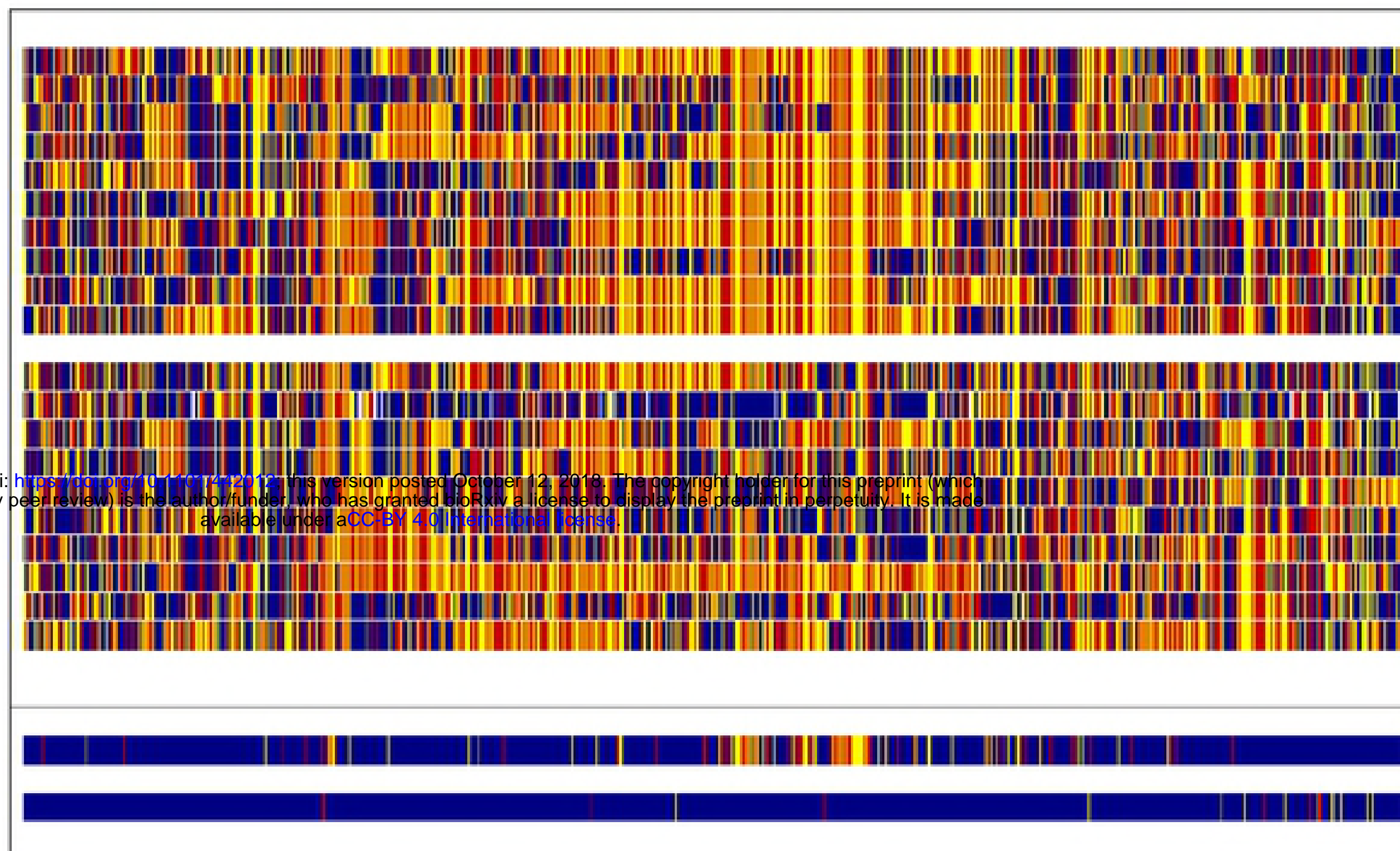
484 (A) Position of the premature stop causing mutation within the porcine myostatin locus;

485 (B) Conservation of the amino acid sequence surrounding the mutation, with the

486 consequences of the premature stop mutation highlighted in red.

1 **FIGURES**

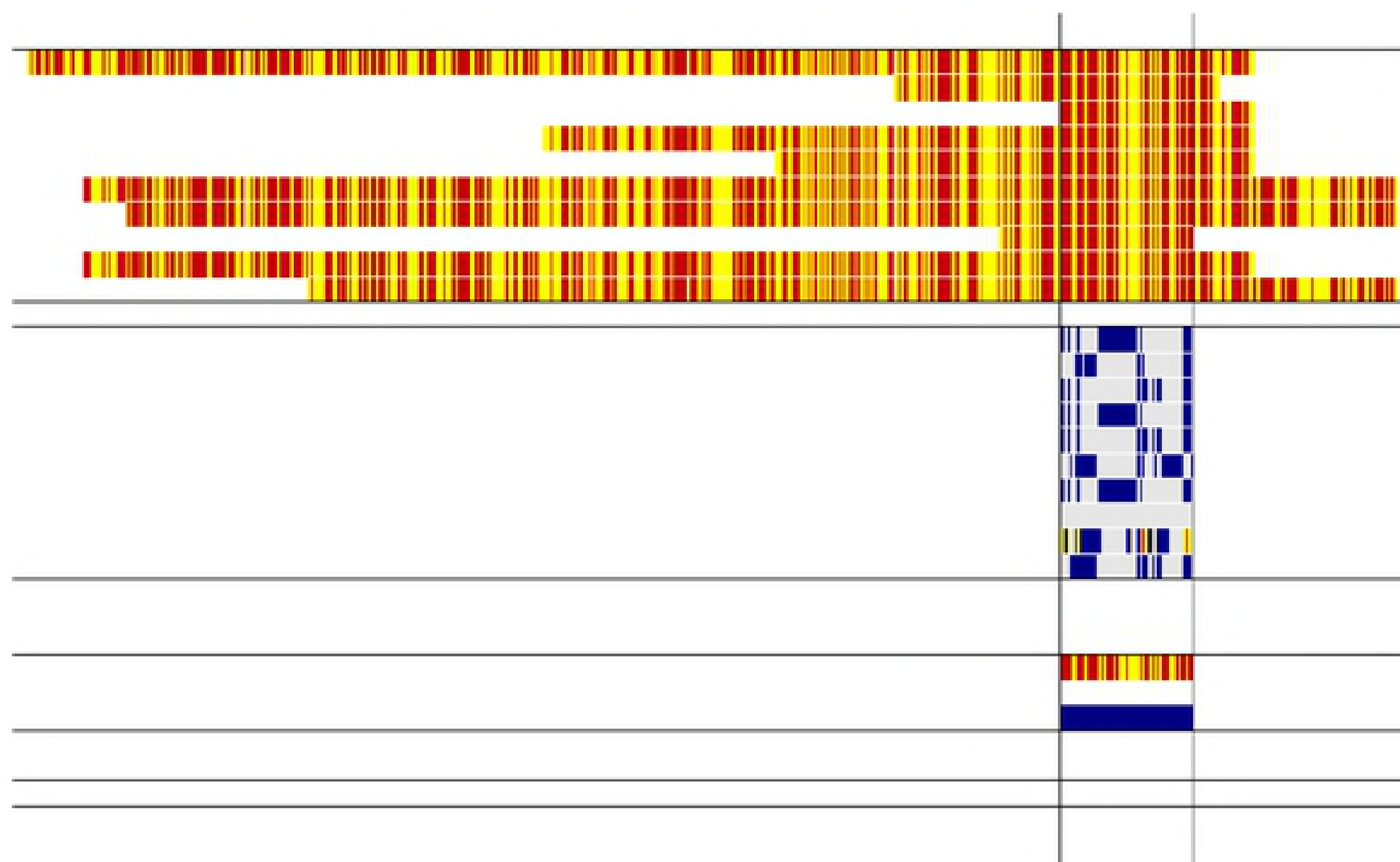
2 **1 A:**



bioRxiv preprint doi: https://doi.org/10.1101/442012; this version posted October 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

3

4 **1 B:**

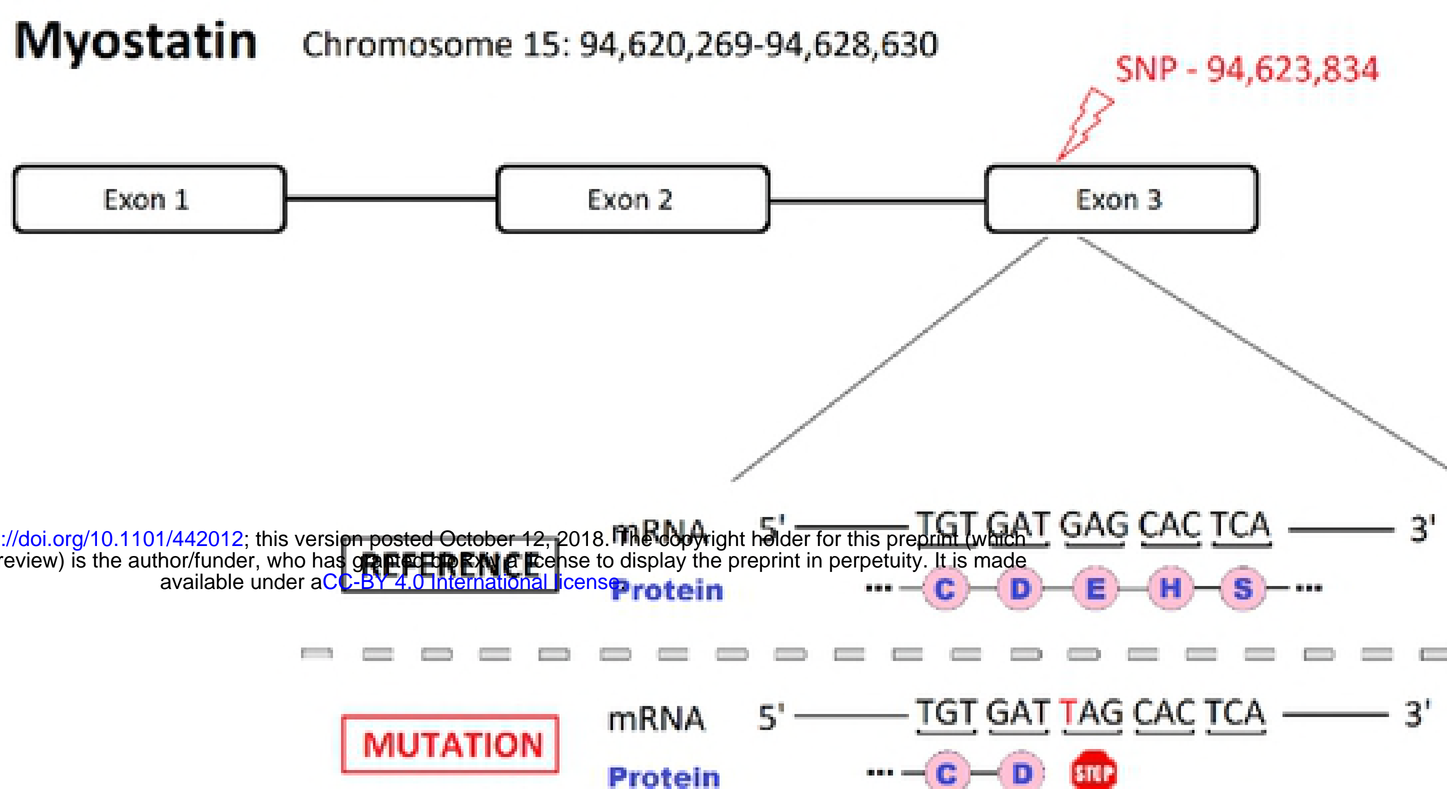


5

Figure 1

FIGURES

2 A:



bioRxiv preprint doi: https://doi.org/10.1101/442012; this version posted October 12, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2 B:

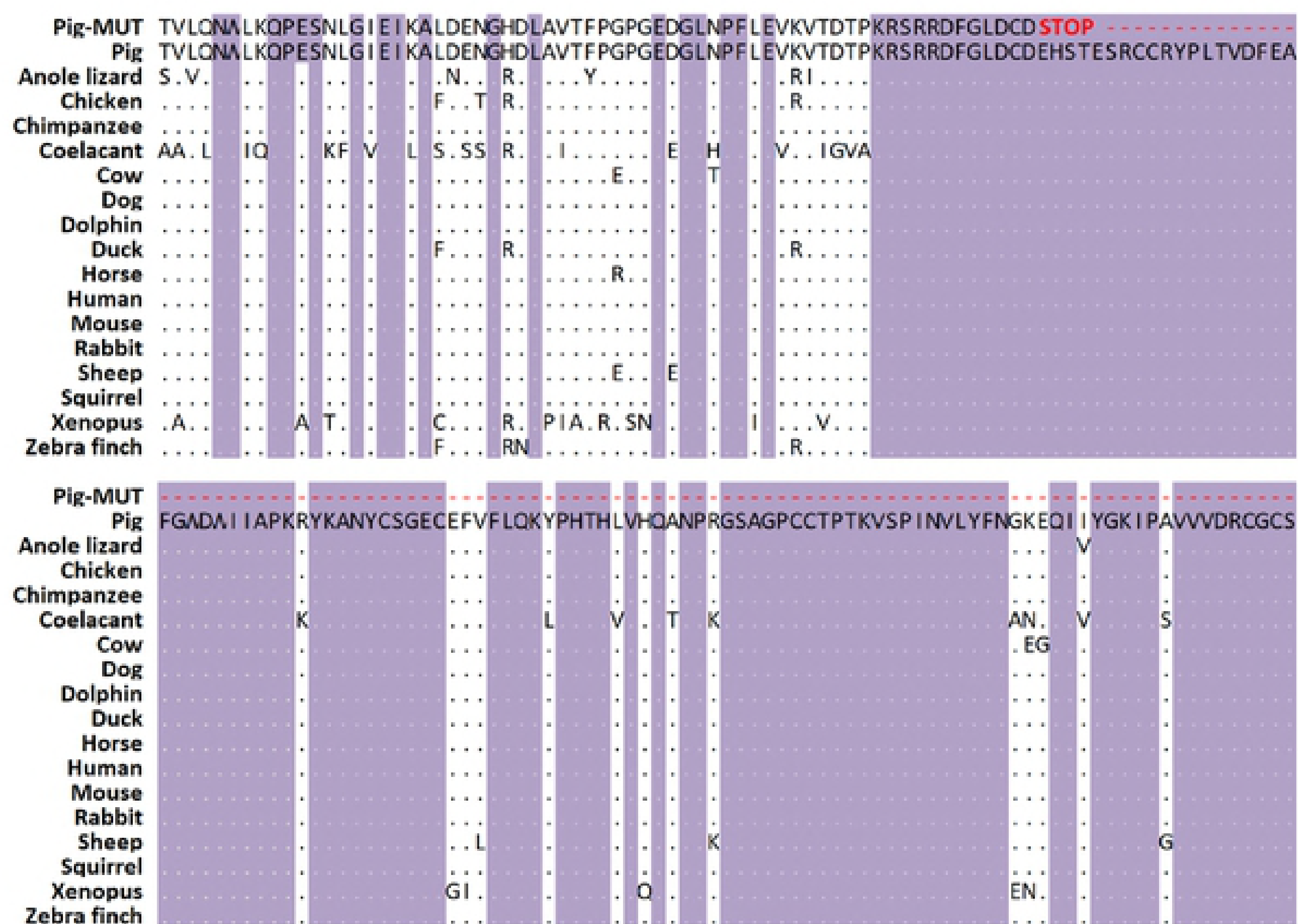


Figure 2