Variants in *FtsJ RNA 2′-O-Methyltransferase 3* and *Growth Hormone 1* are associated with small body size and a dental anomaly in dogs

Sydney R. Abrams[a], Alexandra L. Hawks[a], Jacquelyn M. Evans[a,b], Thomas R. Famula[c], Mary Mahaffey[d], Gary S. Johnson[e], Jennifer M. Mason[a], and Leigh Anne Clark[a,1]

[a]Department of Genetics and Biochemistry, Clemson University, Clemson, SC 29634; [b]Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, NIH, Bethesda, MD 20892; [c]Department of Animal Science, University of California, Davis, CA 95616; [d]Department of Veterinary Biosciences and Diagnostic Imaging, College of Veterinary Medicine, University of Georgia, Athens, GA 30602; and [e]Department of Veterinary Pathobiology, College of Veterinary Medicine, University of Missouri, Columbia, MO 65211

Edited by Leif Andersson, Uppsala University, Uppsala, Sweden, and approved August 15, 2020 (received for review May 15, 2020)

Domesticated dogs show unparalleled diversity in body size across breeds, but within breeds variation is limited by selective breeding. Many heritable diseases of dogs are found among breeds of similar sizes, suggesting that as in humans, alleles governing growth have pleiotropic effects. Here, we conducted independent genome-wide association studies in the small Shetland Sheepdog breed and discovered a locus on chromosome 9 that is associated with a dental abnormality called maxillary canine-tooth mesioversion (MCM) ($P = 1.53 \times 10^{-7}$) as well as two body size traits: height ($P = 1.67 \times 10^{-5}$) and weight ($P = 1.16 \times 10^{-7}$). Using whole-genome resequencing data, we identified variants in two proximal genes: *FTSJ3*, encoding an RNA methyltransferase, and *GH1*, encoding growth hormone. A substitution in *FTSJ3* and a splice donor insertion in *GH1* are strongly associated with MCM and reduced body size in Shetland Sheepdogs. We demonstrated in vitro that the *GH1* variant leads to exon 3 skipping, predicting a mutant protein known to cause human pituitary dwarfism. Statistical modeling, however, indicates that the *FTSJ3* variant is the stronger predictor of MCM and that each derived allele reduces body size by about 1 inch and 5 pounds. In a survey of 224 breeds, both *FTSJ3* and *GH1* variants are frequent among very small "toy" breeds and absent from larger breeds. Our findings indicate that a chromosome 9 locus harboring tightly linked variants in *FTSJ3* and *GH1* reduces growth in the Shetland Sheepdog and toy breed dogs and confers risk for MCM through vertical pleiotropy.

FTSJ3 | GH1 | IGF1 | height | weight

In dogs, maxillary canine-tooth mesioversion (MCM) describes an upper canine tooth that is displaced forward toward the nose, also known as a lance canine (Fig. 1*A*, compare to normal dentition in Fig. 1*B*) (1, 2). One or both maxillary canines may be affected. MCM can cause traumatic occlusion, ulceration of the upper lip, and/or periodontal disease and may require extraction or orthodontic repositioning (1). MCM is rarely observed outside of the Shetland Sheepdog (2, 3), a breed that also has a high prevalence of hypodontia, a condition characterized by one or more congenitally missing teeth (4).

The Shetland Sheepdog is a small breed with a standard height of 13 to 16 inches at the shoulder. Derived alleles of *IGF1*, *GHR*, *HMGA2*, *IGF1R*, *SMAD2*, and *STC2* explain about half of reduced body size observed across dog breeds (5). "Small" alleles of the first four aforementioned genes have been observed in Shetland Sheepdogs in various genotypic combinations, indicating the alleles are not fixed in the breed (5). In general, smaller breeds tend to have more derived alleles at these loci than larger breeds, but even the smallest "toy" breeds are not fixed across these loci (5).

We undertook a genome-wide association study (GWAS) and whole-genome resequencing to identify genetic risk factors causing MCM and uncovered a strong correlation with body size. We identify variants in *Growth Hormone 1* (*GH1*) and *FtsJ RNA 2′-O-Methyltransferase 3* (*FTSJ3*) that are strongly associated with MCM, height, and weight and demonstrate the functional consequence of a *GH1* splice-site insertion. Our data suggest that the locus harboring *GH1* and *FTSJ3* variants is a major determinant of body size in Shetland Sheepdogs and toy dog breeds and confers risk for MCM through vertical pleiotropy.

Results

MCM Is Associated with Single-Nucleotide Polymorphisms on Chromosome 9. We obtained genetic material and phenotypic information from 86 Shetland Sheepdogs having MCM (51 bilateral, 33 unilateral, and 2 unknown laterality) and 144 Shetland Sheepdogs with correctly positioned canines (*SI Appendix*, Table S1). There was no statistical difference between males and females among cases (51 male, 35 female; $P = 0.10$). Unilateral cases were more likely to have the left canine affected (23 left, 8 right; $P = 0.01$). Bilateral cases were more likely to have hypodontia ($n = 27$) than unilateral cases ($n = 11$) ($P = 0.01$). Fourteen dogs unaffected by MCM were also missing teeth. The teeth most commonly reported as missing were the maxillary second premolars, with no side preference.

We conducted a GWAS for MCM using 39 cases, 39 controls, and 117,053 single-nucleotide polymorphisms (SNPs) after filtering and identified a single region of association exceeding Bonferroni significance on chromosome 9 (12753481, $P = 1.53 \times 10^{-7}$;

Significance

Through the study of a dental anomaly we identified a locus strongly associated with body size in the Shetland Sheepdog. Within this locus are variants in two genes: a substitution in *FtsJ RNA 2′-O-Methyltransferase 3* (*FTSJ3*) and a splice donor insertion in *Growth Hormone 1* (*GH1*). We demonstrated that the *GH1* variant causes an abnormal splicing pattern that is also observed in dominant forms of human pituitary dwarfism. Interestingly, the *FTSJ3* variant is estimated to have the greatest impact on height and weight and this gene has not been previously characterized in body size traits. Both derived alleles are found in high frequencies in very small "toy" breeds but are entirely absent from larger breeds.

Author contributions: J.M.M. and L.A.C. designed research; S.R.A., A.L.H., and J.M.E. performed research; M.M. and G.S.J. contributed samples and research materials; S.R.A., J.M.E., T.R.F., J.M.M., and L.A.C. analyzed data; and S.R.A., J.M.E., and L.A.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: lclark4@clemson.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2009500117/-/DCSupplemental.

First published September 21, 2020.

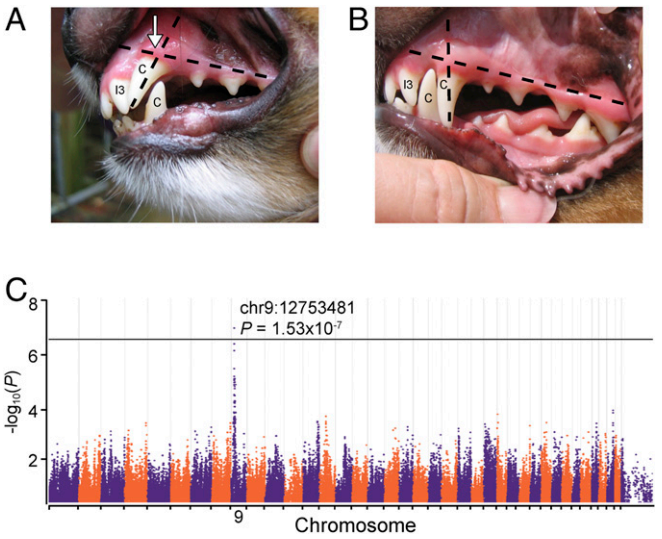


Fig. 1. MCM in the Shetland Sheepdog. (*A*) The maxillary canine (upper C) of an affected dog is closer to the third incisor (I3) and rostral to the mandibular canine (lower C). The long axis of the affected canine (indicated by the vertical dashed line) is rostral compared to *B* normal dentition in an unaffected dog. (*C*) Manhattan plot of MCM GWAS using 39 cases and 39 controls. The $-\log_{10}$ *P* values for 117,053 SNPs are plotted on the *y* axis against chromosome position. The black horizontal line represents the threshold for Bonferroni significance. The position and *P* value of the lead SNP are included.

Fig. 1*C* and *SI Appendix*, Table S2). There was no evidence of genomic inflation (lambda = 1.03). Sixty-seven percent of cases were homozygous for the risk allele of the lead SNP (chr9:12753481), while 31% were heterozygous. The risk allele was also present among the control dogs: 8% were homozygous risk and 72% were heterozygous, likely indicating a complex pattern of inheritance. Using the lead SNP, we calculated regional linkage disequilibrium (LD) ($r^2 > 0.5$) to define a conservative 1.9-Mb candidate interval from 11.6 Mb to 13.5 Mb on chromosome 9 harboring about 30 genes.

***FTSJ3* and *GH1* Harbor Candidate Causal Variants.** We initially searched for candidate causal variants using a VCF file generated from 722 dogs, including 1 Shetland Sheepdog having bilateral MCM, 2 Shetland Sheepdogs with unknown dentition, 665 dogs of various pure and mixed breeds, and 54 wild canids. The affected Shetland Sheepdog was homozygous for the risk allele at the lead SNP and for the associated haplotype across the candidate interval. The affected dog possessed no unique variants, even when excluding the other two Shetland Sheepdogs from the analysis, indicating that the causal variant is likely a polymorphism. Manual scanning of the critical interval in IGV revealed no major structural changes unique to the Shetland Sheepdog.

We then used whole-genome resequencing data from the aforementioned Shetland Sheepdog with bilateral MCM to identify variants that were 1) homozygous, 2) present within the coding regions or splice sites of genes within our critical interval, and 3) absent from the genomes of five Collies, a medium- to large-sized, genetically similar ancestor of Shetland Sheepdogs (6) in which MCM has never been described. We identified two synonymous SNPs, one nonsynonymous SNP, and one splice-site variant and considered the latter two to be potentially deleterious. A substitution, g.11775131T > C, occurs in *FTSJ3* (*FTSJ3*^{mut}, hereafter) and predicts a p.Lys797Glu missense in a highly conserved position of the canine protein (XP_005624307; *SI Appendix*, Fig. S4). This change is categorized as "probably damaging" by in silico programs PANTHER (1500) and PolyPhen2 (0.979). *GH1* harbors a single base insertion in the splice donor of exon 3: g.11833343_11833344insA (*GH1*^{mut}, hereafter). We also further investigated a g.14561174A > G substitution in *AXIN2*, located 1 Mb outside of our critical interval, because mutations in the orthologous human gene result in dental anomalies (7). The *AXIN2* variant predicts a p.Lys103Arg missense variant, a change categorized as probably damaging by PANTHER (456) and benign by PolyPhen2 (0.02). While we cannot exclude the possibility that a noncoding variant is contributing to MCM, we focused the remainder of the study on these three potentially deleterious coding variants.

We genotyped all three candidate variants in 78 MCM cases and 125 controls. *FTSJ3*^{mut} and *GH1*^{mut} were strongly associated with MCM ($P = 8.2 \times 10^{-16}$ and 1.7×10^{-13}, respectively), while the *AXIN2* variant showed a weaker association ($P = 8.5 \times 10^{-7}$). We thus eliminated *AXIN2* as a candidate causal gene.

We investigated the frequencies of *FTSJ3*^{mut} and *GH1*^{mut} using publicly available genomes from 1,049 dogs representing 224 different breeds (*SI Appendix*, Table S3). Across all breeds, the derived alleles had frequencies of 4.5% and 3.7%, respectively. In addition to the Shetland Sheepdog, these alleles were found only in the following toy breeds: Affenpinscher, Biewer Terrier, Chihuahua, Miniature Pinscher, Papillion, Pomeranian, Toy Poodle, Cavalier King Charles Spaniel, Toy Fox Terrier, and Yorkshire Terrier.

MCM Is Highly Correlated with Body Height and Weight. Because growth hormone is necessary for normal body growth (8), we looked for a correlation between body size and MCM. Affected dogs were significantly shorter and weighed less than controls ($P = 4.6 \times 10^{-6}$ and 5.6×10^{-11}, respectively) (Fig. 2 *A* and *B*). As female Shetland Sheepdogs are often smaller in height and weight compared to males ($P = 0.007$ and 0.22, respectively in our cohort), we confirmed that the correlations were significant within each sex (*SI Appendix*, Fig. S2).

Because of the strong correlation between MCM and body size, we conducted another GWAS with height, weight, and sex as covariates (34 cases vs. 37 controls). The chromosome 9 signal was diminished but persisted as the most significant association (chr9:12753481, $P = 1.04 \times 10^{-6}$) (*SI Appendix*, Fig. S3*A*). Independent GWASs for height and weight indicate the locus on chromosome 9 affects both phenotypes ($P = 1.67 \times 10^{-5}$ and 1.16×10^{-7}, respectively) (Fig. 2 *C* and *D*). The impact of this locus on either phenotype cannot be detected when we include MCM as a covariate, consistent with a shared cause (i.e., pleiotropy) (*SI Appendix*, Fig. S3 *B* and *C*) (9).

***GH1*^{mut} Causes Exon 3 Skipping.** To further assess *GH1* as a candidate gene, we sought to determine if the splice-site insertion alters gene splicing. We expressed mutant and wild-type gene fragments in HEK293 cells and evaluated splicing patterns through complementary DNA (cDNA) sequencing. Cells carrying the wild-type fragment produced a primary band at 475 bp. The mutant cells yielded the wild-type band and a brighter, smaller band of 358 bp (Fig. 3). Sequencing confirmed that the larger band represented normal splicing of exons 2 through 5. The 358-bp band is a spliceoform in which all 117 bp of exon 3 are skipped, predicting a protein lacking 39 of 216 amino acids. Proteins resulting from *GH1* transcripts without exon 3 have been identified in human patients with growth hormone deficiency type II and shown to act in a dominant negative manner (8). Both cell lines also yielded a small amplicon of 196 bp that sequencing confirmed to be an isoform in which exons 3 and 4 are skipped (10).

***FTSJ3*^{mut} Impacts Body Size in an Additive Fashion.** *FTSJ3* and *GH1* are separated by ~58 kb on chromosome 9 and their alleles are predictably in high LD (D′ = 0.99). We observed three haplotypes in our population (Fig. 4*A*). The absence of an *FTSJ3*^{mut}-*GH1*$^{+}$ haplotype suggests that the *GH1* mutation occurred first and that recombination between chr9:11775131 and chr9:11833343 is

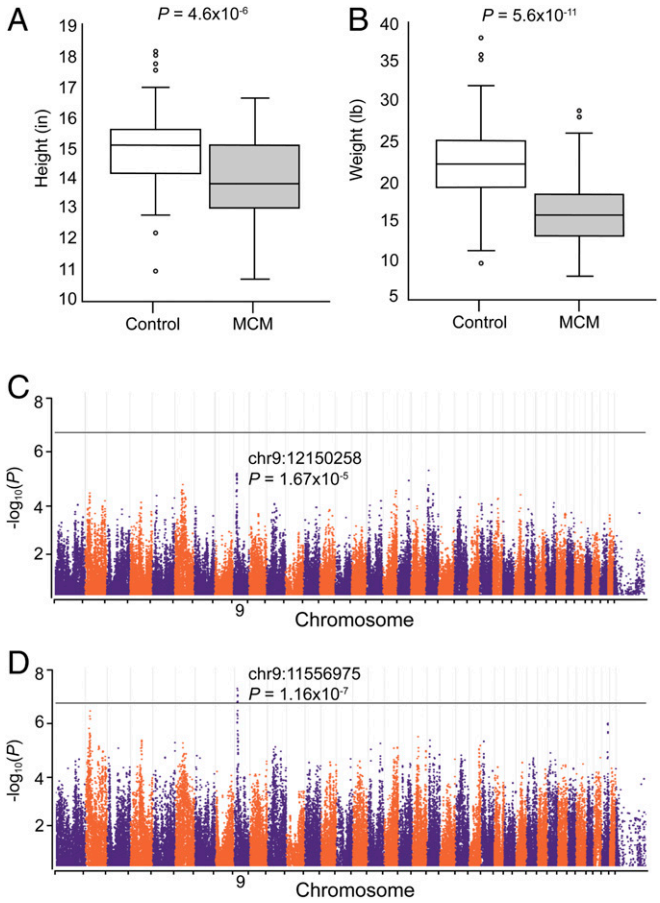


Fig. 2. Height and weight correlate with MCM. Boxplots illustrate (*A*) lower heights in affected dogs ($n = 63$, mean = 13.97 inches) compared to controls ($n = 104$, mean = 14.93 inches) and (*B*) reduced weights in affected dogs ($n = 62$, mean = 16.20 pounds) compared to controls ($n = 99$, mean = 21.97 pounds). Two-sample *t* test *P* values are reported above. Manhattan plots for (*C*) height and (*D*) weight GWASs using 34 cases and 37 controls. The $-\log_{10}$ *P* values for 117,053 SNPs are plotted on the *y* axis against chromosome position (CanFam3.1). The black horizontal line represents the threshold for Bonferroni significance. The positions and *P* values of the lead SNPs are included.

rare. The *FTSJ3*$^{+}$-*GH1*^{mut} haplotype was uncommon but found significantly more often in dogs having normal dentition ($P = 0.04$).

To determine if both alleles or just *GH1*^{mut} impact body size, we investigated height and weight within two subpopulations that were fixed at one locus but variable at the other (Fig. 4*B*). We found that the heights and weights of dogs having only one risk allele at *FTSJ3* were significantly greater than those having two copies, suggesting that *FTSJ3*^{mut} reduces body size in an additive fashion. When *FTSJ3*^{mut} is neutralized, we found no significant differences in the heights and weights of dogs having one vs. two *GH1*^{mut} alleles. This is consistent with a dominant negative effect of the GH1 mutant protein. We did not have the genotypic combinations in our population to enable us to study the impact of having at least one *GH1*^{mut} allele vs. homozygous wild type.

***FTSJ3* and *IGF1* Are Highly Predictive of MCM.** To evaluate genotypic and phenotypic contributions to MCM, we performed statistical modeling using Akaike information criterion (AIC), where smaller numbers indicate the best and most parsimonious fit to the disease data (Fig. 4*C*). Consistent with our association statistics, we found that *FTSJ3* is a better predictor of MCM than *GH1* and that weight is a better predictor of MCM than height. While *FTSJ3* alone is a better predictor of MCM than weight alone, the best predictive model includes both *FTSJ3* and weight. In homozygosity, *FTSJ3*^{mut} confers a strong risk for MCM, regardless of weight. However, in heterozygosity, weight governs risk, such that a lightweight dog has a much higher risk than a heavy dog (Fig. 4*D*).

We then considered the impact of alleles of other known body size genes segregating in the Shetland Sheepdog: *IGF1*, *GHR*, *HMGA2*, and *IGF1R*. Using associated SNPs from the BeadChip and our GWAS cohort (39 cases vs. 39 controls), we found that the derived allele of *IGF1* was significantly overrepresented among cases ($P = 0.04$), and even more so when we only consider dogs heterozygous for *FTSJ3*^{mut} ($P = 0.016$). The derived allele of *IGF1R* was significantly overrepresented among controls ($P = 0.04$) but was not significant in the heterozygous subset. We detected no significant differences in the allele frequencies of *GHR* or *HMGA2*.

The *IGF1* association with MCM strengthened in an expanded cohort of 65 cases and 100 controls ($P = 0.0007$). We therefore added *IGF1* genotypes into the model with *FTSJ3* and weight and observed that weight was no longer a significant factor. A risk prediction plot using *FTSJ3* and *IGF1* genotypes illustrates an additive effect (Fig. 4*E*). Consistent with an additive model, individuals with the fewest derived alleles across the three genes have greater heights and weights and the lowest frequencies of MCM, while dogs having the most derived alleles are shorter, weigh less, and have higher frequencies of MCM (Table 1).

***FTSJ3* Explains Most Weight Variance in Shetland Sheepdogs.** When evaluated simultaneously (e.g., height as predicted by *FTSJ3* and *GH1*), the individual contributions of *FTSJ3* and *GH1* to height and weight cannot be disentangled because the genes are in high LD. In independent models (e.g., height as predicted by *FTSJ3*), *FTSJ3* is the most impactful, reducing body size by about 1 inch and 5 pounds with each derived allele and accounting for 37% of height variance and 73% of weight variance in our cohort. *GH1* is slightly less impactful, explaining about 32% and 67% of height and weight variances, respectively. Comparatively, *IGF1* explains only 7% of height and 28% of weight in Shetland Sheepdogs.

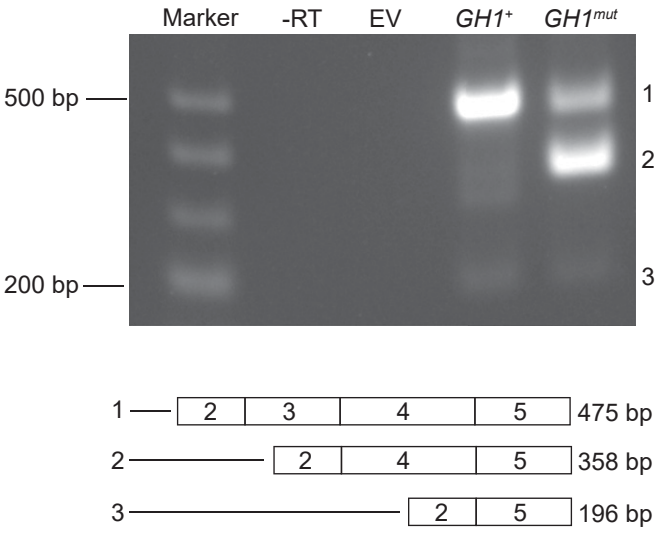
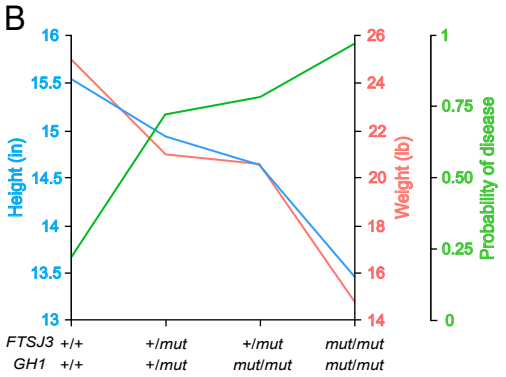


Fig. 3. Amplification of *GH1* cDNAs. Agarose gel electrophoresis of a minus reverse transcriptase control (–RT, lane 2), empty vector control (EV, lane 3), wild-type *GH1* cDNA (*GH1*$^{+}$, lane 4), and *GH1* cDNA harboring the splice-site insertion (*GH1*^{mut}, lane 5). The primary wild-type band at 475 bp (1), mutant band at 358 bp (2), and faint wild-type band at 196 bp (3) were each extracted and characterized via Sanger sequencing. The splicing pattern of each isoform for exons 2 to 5 is depicted below the gel image.

A

Haplotype	Cases (2*n* = 156)	Controls (2*n* = 250)	OR	95% CI	*P*-value
FTSJ3^{+}-*GH1*$^{+}$	26	130	**0.19**	**0.11-0.30**	**1.2x10^{-12}**
FTSJ3^{+}-*GH1*^{mut}	4	19	**0.32**	**0.11-0.96**	**0.04**
FTSJ3^{mut}-*GH1*^{mut}	126	101	**6.20**	**3.87-9.93**	**8.2x10^{-16}**



C

Predictive Model	AIC
FTSJ3	157
GH1	168
Height	193
Weight	172
FTSJ3 + Height	159
***FTSJ3* + Weight**	**155**
GH1 + Height	169
GH1 + Weight	160

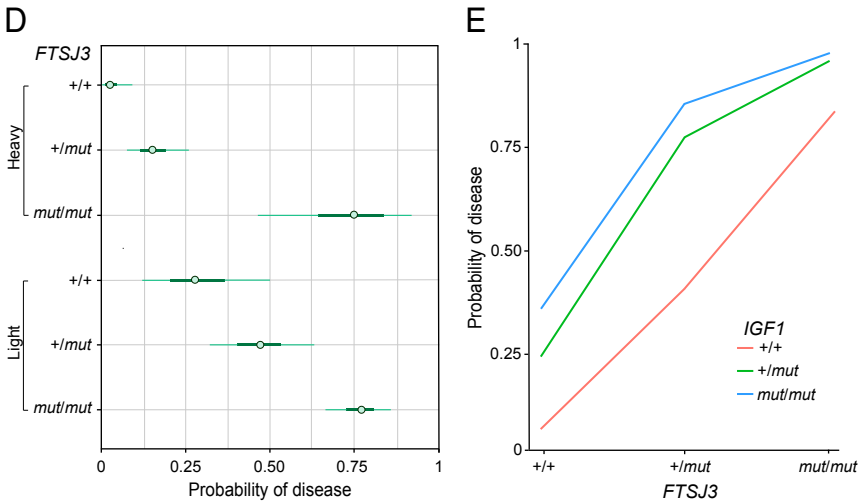


Fig. 4. Understanding the roles of *FTSJ3*, *GH1*, and *IGF1* in MCM and body size. (*A*) Odds ratios (OR), 95% CIs, and Fisher's exact two-tailed *P* values are calculated for each of the three observed haplotypes for MCM. Haplotypes with *FTSJ3*$^{+}$ are protective, while the haplotype with both *FTSJ3*^{mut} and *GH1*^{mut} confers risk. (*B*) Height and weight averages (calculated using all available data) and MCM risk are plotted for the four most frequent genotypic combinations of *FTSJ3* and *GH1*. One copy of each derived allele ($n = 74$) markedly reduces height and weight compared to wild type ($n = 40$; $P = 0.003$ and 0.0002, respectively). Compared to heterozygotes, homozygosity for the *GH1*^{mut} allele alone ($n = 23$) does not significantly alter height ($P = 0.3$) or weight ($P = 0.7$). Among *GH1*^{mut} homozygotes, adding a second copy of *FTSJ3*^{mut} ($n = 67$) significantly reduces height ($P = 0.0002$) and weight ($P = 7.8 \times 10^{-5}$), illustrating that *GH1* is not solely responsible for the variation in body size and disease risk attributed to this locus. (*C*) AIC values are shown for different genotypic and/or phenotypic predictive models of MCM. The small AIC value for *FTSJ3* + weight (bold) indicates a more parsimonious fit to the disease data. (*D*) Probability of disease plotted against *FTSJ3* genotypes at two extremes: heavy (27 pounds) and light (12 pounds) shows the importance of weight on risk for MCM. (*E*) Probability of disease for all combinations of *FTSJ3* and *IGF1* genotypes shows a greater role for *IGF1* among *FTSJ3* heterozygotes.

Discussion

Domesticated dogs vary in body size more than any other terrestrial species, but genetically isolated purebred populations must conform to standards that restrict variation within a breed. Each breed therefore possesses a set of polymorphisms that dictate growth, and these loci are under constant artificial selective pressures (5). While many inherited diseases of dogs correlate with body size [e.g., Legg–Calve–Perthes disease (11), osteosarcoma (12), and gastric dilatation-volvulus (13)], surprisingly few studies to date have directly attributed disease phenotypes to alleles underlying growth (14–16). In this study, we discovered that a dental anomaly in the Shetland Sheepdog correlates with small body size and that a large proportion of height and weight variation within the breed is attributed to a locus on chromosome 9 that harbors *FTSJ3* and *GH1*.

GH1 is an excellent candidate gene for a growth phenotype and was, in fact, recently suggested to harbor variants contributing to small size in dogs (17). *GH1* encodes pituitary growth hormone, the release of which is the first step in the GH1–IGF1 axis that regulates cell proliferation, differentiation, and apoptosis. We demonstrate here that a single base insertion in the 3′ donor splice site of *GH1* exon 3 causes incomplete alternative splicing. In humans, weak splice sites surrounding this exon increase the importance of intron 3 splicing sequences (8). Mutations that occur within the first six bases of the donor splice site or interrupt splicing enhancer elements cause exon 3 skipping

Table 1. Distribution of genotypic combinations across dental and morphometric phenotypes in 65 MCM cases and 100 controls

FTSJ3	*GH1*	*IGF1*	Unilateral	Bilateral	Controls	95% CI*	*P* value*	Mean weight, pounds	Mean height, in
+/+	+/+	+/+	0	0	12	**0–0.20**	**0.045**	26.53[†]	15.92[†]
+/+	+/+	+/*mut*	1	0	18	**0–0.21**	**0.005**	23.64[†]	15.22
+/+	+/+	*mut*/*mut*	0	0	5	0–0.37	0.083	26.40	15.91
+/+	+/*mut*	+/*mut*	0	1	0	—	—	18.70	15.80
+/*mut*	+/*mut*	+/+	2	0	18	**0.02–0.28**	**0.002**	21.46[†]	14.99
+/*mut*	+/*mut*	+/*mut*	3	3	20	**0.13–0.45**	**0.021**	21.92[†,‡]	15.06[‡]
+/*mut*	+/*mut*	*mut*/*mut*	3	6	4	0.42–0.88	0.19	17.85[†,§]	14.48[§]
+/*mut*	*mut*/*mut*	+/+	0	2	2	—	—	22.70	14.91
+/*mut*	*mut*/*mut*	+/*mut*	1	1	8	0.05–0.49	0.064	19.16[§]	14.20[§]
+/*mut*	*mut*/*mut*	*mut*/*mut*	0	0	1	—	—	21.00	15.25
mut/*mut*	*mut*/*mut*	+/+	5	7	6	0.45–0.84	0.153	16.54[‡]	13.79[§]
mut/*mut*	*mut*/*mut*	+/*mut*	5	9	4	**0.55–0.92**	**0.028**	14.67	13.53
mut/*mut*	*mut*/*mut*	*mut*/*mut*	4	11	2	**0.67–0.97**	**0.007**	13.40[§]	13.12[§]

*Calculated for genotypes observed at least five times in all 65 MCM cases and 100 controls. Significant statistics are bolded. Laterality status was unknown for one case with the genotype +/*mut* +/*mut* +/*mut*.
[†]One control had measurements that were not available.
[‡]Two cases had measurements that were not available.
[§]One case had measurements that were not available.

(18). The mutant protein misfolds and exerts a dominant negative effect through disruption of the trafficking and stability of wild-type GH1 (19, 20), resulting in pituitary dwarfism. Orofacial anomalies are common in growth hormone deficiencies and include underdevelopment of the maxilla and mandible, delays in tooth eruption, and tooth agenesis or malposition (21, 22).

FTSJ3 has not been previously described as having a role in canine body size, although genome-wide studies have associated the locus with various height- (23) and weight-related phenotypes in other species, including humans (24–27). *FTSJ3* encodes an RNA 2′-O-methyltransferase, located in the nucleolus, and is ubiquitously expressed and evolutionarily conserved (28). Since 1974 it has been known that 2′-O methyl groups are added to pre-rRNA during its synthesis (29), and it is now clear that these modifications are made by FTSJ3 (30). FTSJ3 has also been shown to add internal 2′-O methyl groups to HIV RNA (31), raising the possibility that it internally modifies cellular pre-messenger RNA (mRNA) and/or mRNA. Knockdown of FTSJ3 activity decreases the rate of cell proliferation (30), which may be a consequence of decreased ribosome production and/or mRNA supply. In any case, a decrease in the net rate of cell proliferation in developing puppies of toy breeds might account for the markedly slower rates of growth in these breeds as compared to larger breeds. In an alternative scenario, mutation of *FTSJ3*, or a linked noncoding variant, could disrupt the complex regulatory pattern of the *GH1* locus and impact growth hormone signaling (32).

$FTSJ3^{mut}$ predicts a nonconservative substitution of glutamic acid for lysine in the conserved C-terminal domain of FTSJ3, wherein binding to preribosome complexes is mediated (33). This variation was consistently better associated with MCM and body size, despite a complete lack of recombination between the *FTSJ3* and *GH1* derived alleles. We attribute this to an ancestral haplotype ($FTSJ3^{+}$-$GH1^{mut}$) that was more frequent in larger-bodied unaffected Shetland Sheepdogs, illuminating a significantly lower risk of MCM in individuals having a wild-type copy of *FTSJ3* and reflecting the additive effect of the mutant allele. The nonrisk allele of the lead SNP also occurs on this haplotype, suggesting that the strongest associations in the GWAS were detecting the protective factor of the $FTSJ3^{+}$ allele.

Shetland Sheepdogs are variable at two other loci harboring genes encoding proteins downstream in the GH1–IGF1 axis. We observed that among individuals homozygous for $FTSJ3^{mut}$-$GH1^{mut}$, *IGF1* had a mild effect, modestly decreasing weight and increasing risk for MCM with the addition of each derived allele. In $FTSJ3^{mut}$-$GH1^{mut}$ heterozygotes, however, *IGF1* was much more impactful, taking individuals from low risk in heterozygosity to high risk in homozygosity and decreasing weight by nearly 20%. An absence of growth hormone would impact the secretion of IGF1, potentially masking the effect of an *IGF1* mutation. Our findings suggest that homozygosity for $GH1^{mut}$ results in the release of at least some growth hormone, possibly because of incomplete alternative splicing. Heterozygosity for $GH1^{mut}$ appears to permit the production of much more growth hormone, thereby allowing us to clearly see the effect of mutant *IGF1*. It is important to consider that "small" alleles of other body size genes not studied here may be impacting the trends we observed.

The chromosome 9 locus harboring *FTSJ3* and *GH1* has been associated with head and mandible lengths across breeds (34). We noted that the Shetland Sheepdog and other breeds in which MCM has been reported [e.g., Italian Greyhound and Fox Terrier (1)] are dolichocephalic, whereas the toy breeds that also possess the derived alleles are mesocephalic or brachycephalic (35). Because the frequencies of dental anomalies in dogs correlate with skull morphology (36), it is possible that the presence of MCM in the Shetland Sheepdog, and not the toy breed dogs, could be attributed to the impact of the chromosome 9 locus on their elongated skull.

Based on our data, we propose that the chromosome 9 locus demonstrates vertical pleiotropy (37), whereby its effect on dentition is mediated by its role in reducing size, specifically of the skull. Given the strong selection for height within breeds, it is plausible that multiple variants within this locus contribute to body size variation in dogs. Future studies will be necessary to determine exactly which variants, or combination of variants, in this locus are retarding growth.

Materials and Methods

Study Population. All samples were obtained with informed consent according to protocols approved by the Clemson University Institutional Review Board (IBC2018-13). Buccal cells or whole blood samples were collected from 230 Shetland Sheepdogs with unilateral MCM, bilateral MCM, or proper canine alignment. Pedigrees were obtained for all dogs. Dental status was verified by physical examination, dental photographs, radiographs, and/or veterinary records. During the course of the study, it became necessary to obtain height (measured at the withers) and weight data for each dog. Height and weight measurements were reported by owners for 167 and 161 dogs, respectively. Dogs were at least 1 y of age at the time of measurement (median age = 6 y; average age = 6.8 y). DNA was isolated following the Gentra Puregene DNA Isolation protocol (Qiagen) and concentration was quantified using a NanoDrop 1000 spectrophotometer (Thermo Scientific).

Genome-Wide Association Analyses. Inclusion in the GWAS for MCM was determined based on relatedness: selected dogs were unrelated within at least two generations. We also sought to balance geographic origin, sex, and coat color/pattern. Genotyping was performed for 78 Shetland Sheepdogs at GeneSeek, Inc. using the Illumina CanineHD BeadChip containing 220,859 markers. All samples had call rates >95%. SNPs with call rates <95%, minor allele frequencies <5%, and/or significant deviation from Hardy–Weinberg equilibrium ($P < 0.0004$) were excluded from further investigation. All filtering and statistical analyses were conducted with SNP & Variation Suite v8 (SVS; Golden Helix, Inc.). All chromosome positions are reported in CanFam3.1.

For MCM GWASs, Fisher's exact *P* values for each SNP were calculated under an additive model. Because hypodontia was underrepresented in the control GWAS population ($n = 2$) compared to the cases ($n = 13$), we confirmed our results for MCM through a second GWAS in which we excluded all but two randomly selected cases having hypodontia (28 cases vs. 39 controls) (*SI Appendix*, Fig. S1). Height (inches) and weight (pounds) were analyzed as quantitative values under an additive model. Covariates were considered using a linear regression under a full vs. reduced model.

LD pairwise analysis was performed to calculate r^2 values for the MCM lead SNP (chr9:12753481). The r^2 values were calculated in PLINK (38) using all controls.

Variant Analyses. Variant filtering within a VCF file of 722 dogs (39) was performed using Golden Helix SVS and manual scanning of the critical interval was performed in IGV (40). Variants in the affected Shetland Sheepdog (accession no. SRX4036142) were filtered for allele state and against five Collie genomes (accession nos. SRX2506416, SRX2506417, SRX2506418, SRX2506419, and SRX2506420) in Golden Helix. The impact of nonsynonymous SNPs in *FTSJ3* and *AXIN2* were characterized using two in silico programs (41, 42): PolyPhen2 scores ranging from 0.85 to 1 and PANTHER preservation times (in millions of years) >450 were considered probably damaging. Variants of *FTSJ3*, *GH1*, *AXIN2*, and *IGF1* were genotyped using Sanger sequencing. Primers are reported (*SI Appendix*, Table S4). Across breed allele frequencies for $FTSJ3^{mut}$ and $GH1^{mut}$ were calculated using whole genome data from 1,049 dogs (*SI Appendix*, Table S3).

Statistical Calculations. Associations of alleles or haplotypes with phenotypes were assessed by Fisher's exact tests using VassarStats (http://vassarstats.net/). Two-sample *t* tests in Microsoft Excel were used to evaluate differences in morphometric means between MCM cases and controls and between genotypes.

Having combined the bilateral and unilateral MCM disease phenotypes, the observations evaluated are now classified into one of two classes (i.e., affected and control). Accordingly, for the evaluation of the binary MCM phenotype, we turn to logistic regression to model the risk of disease as a function of the observed genotypes and other potential explanatory variables. Define the probability of disease as p_i for the *i*-th dog and thus the logit of this probability as $\theta_i = log[p_i/(1-p_i)]$. Modeling the logit as a function of any explanatory variables considers this simple linear model:

$$\theta_i = b_0 + b_w Weight_i + b_h Height_i + b_s Sex_i + \sum_{j=1}^{k}(\alpha_j + \delta_j),$$

where b_0 is an unknown constant common to all dogs, b_w is the regression coefficient of body weight on MCM, b_h is the regression coefficient for height, b_s is the regression coefficient on sex (for males = 1 and 0 = females), and α_j and δ_j are the additive and dominance contributions, respectively, of the *j*-th SNP of the *k* SNPs being considered in the analysis. Estimation of the unknown model parameters is accomplished with the glm command in the public domain language R (43). Of course, various reduced models can also be considered. In the comparison of these models, that with the lowest AIC value (44) represents the best and most parsimonious fit to the disease data.

The potential impact of the observed SNPs on the measurements of height and weight was also evaluated. Height and weight were assumed to be normally distributed variables conforming to a straightforward linear model for the *i*-th dog:

$$y_i = b_0 + b_s Sex_i + \sum_{j=1}^{k}(\alpha_j + \delta_j),$$

where b_0 is an unknown constant common to all dogs, b_s is the regression coefficient on sex (for males = 1 and 0 = females), and α_j and δ_j are the additive and dominance contributions, respectively, of the *j*-th SNP of the *k* SNPs being considered in the analysis. Estimation of the unknown model parameters is accomplished with the lm command in the public-domain language R (43). As before, comparison of models was facilitated through the AIC (44).

Cloning of *GH1* Fragments into pcDNA Vector. *GH1* fragments were amplified from the genomic DNAs of two Shetland Sheepdogs, an affected dog homozygous for $FTSJ3^{mut}$-$GH1^{mut}$, and a control dog homozygous for $FTSJ3^{+}$-$GH1^{+}$, using Herculase II Fusion DNA Polymerase (Agilent) and primers with sequences that overlapped with the pcDNA3.1 vector (Invitrogen) (*SI Appendix*, Table S4). The pcDNA3.1 vector was amplified with Phusion polymerase (Thermo Scientific). $GH1^{+}$ and $GH1^{mut}$ fragments were cloned into the pcDNA3.1 vector using Gibson assembly (New England Biolabs) following the manufacturer's instructions. Insertion of *GH1* fragments into the vector was verified by Sanger sequencing.

Transfection of *GH1* into HEK293 Cells. HEK293 cells were grown at 37 °C with 5% CO_2 in Dulbecco's modified Eagle's medium (Gibco) with 10% fetal bovine serum. HEK293 cells were transfected with 4.0 μg of $GH1^{+}$ or $GH1^{mut}$ plasmid DNA using 12.0 μL of Lipofectamine LTX (Thermo Fisher) following the manufacturer's instructions. Cells were harvested 48 h posttransfection. RNA was isolated from cells using GeneJET RNA Purification Kit (Thermo Scientific) following the manufacturer's protocol. Five hundred nanograms of RNA was reverse-transcribed using MLV-RT (Optizyme) with a dT(20) primer at 42 °C for 1 h.

Sequencing of *GH1* cDNA. The cDNAs were amplified using Phire Green Hot Start II DNA Polymerase (Thermo Scientific) following manufacturer guidelines. Primers were designed to amplify the last four exons of *GH1*, with the forward primer located in exon 2 and the reverse primer located in exon 5 (*SI Appendix*, Table S4). Minus reverse transcriptase and empty vector controls were also run. PCR products were visualized via gel electrophoresis. Products were purified using the E.Z.N.A. Gel Extraction Kit (Omega Bio-Tek) and verified by Sanger sequencing.

Genotyping of Other Body Size Loci. Previously associated Illumina SNPs in the following genes influencing small body size in dogs were genotyped in the SNP chip data from our GWAS population: *IGF1R* (chr3:41758863 and 41849479), *GHR* (chr4:67040898), *IGF1* (chr15:41221438), and *HMGA2* (chr10:8183593) (5, 45).

Data Availability. Whole-genome resequencing data from affected Shetland Sheepdog, five control Collies, and all other dog genomes used herein have been deposited in the Sequence Read Archive (all accession numbers are given in *SI Appendix*, Table S3). Nucleotide sequences for Growth Hormone 1 spliceoforms are deposited in GenBank under accession numbers MT499773 to MT499775.

ACKNOWLEDGMENTS. We thank Dr. Andrei Alexandrov for assistance in cloning *GH1* fragments and Drs. Alison Starr-Moss and Mike Vaughan for many helpful discussions. L.A.C. and S.R.A. are supported by the Collie Health Foundation. Parts of this work were funded by the Clemson University Honors College and donations from the American Shetland Sheepdog Association Foundation, The Shetland Sheepdog Club of Georgia, the Shetland Sheepdog Club of Spartanburg, the Colonial Shetland Sheepdog Club, and the Edmonton Shetland Sheepdog Fanciers Club. Finally, we thank the dog owners who contributed the samples that made this research possible.

1. L. Legendre, K. Stepaniuk, Correction of maxillary canine tooth mesioversion in dogs. *J. Vet. Dent.* **25**, 216–221 (2008).
2. H. B. Lobprise, J. R. Dodd, *Wigg's Veterinary Dentistry: Principles and Practice*, (Wiley, 2019).
3. L. J. Ackerman, *The Genetic Connection: A Guide to Health Problems in Purebred Dogs*, (American Animal Hospital Association Press, 2011).
4. L. Milella, Missing teeth part 2: Congenitally missing teeth. *Companion Anim.* **14**, 55–59 (2009).
5. M. Rimbault *et al.*, Derived variants at six genes explain nearly half of size reduction in dog breeds. *Genome Res.* **23**, 1985–1995 (2013).
6. H. G. Parker *et al.*, Genomic analyses reveal the influence of geographic origin, migration and hybridization on modern dog breed development. *Cell Rep.* **19**, 697–708 (2017).
7. H. Liu, T. Ding, Y. Zhan, H. Feng, A. Novel, A novel AXIN2 missense mutation is associated with non-syndromic oligodontia. *PLoS One* **10**, e0138221 (2015).
8. A. Potter, J. Phillips, D. Rimoin, "Genetic disorders of the pituitary gland" in *Emery and Rimoin's Principles and Practice of Medical Genetics*, D. Rimoin, R. Pyeritz, B. Korf, Eds. (Elsevier, 2013), pp. 1–37.
9. H. Aschard, B. J. Vilhjálmsson, A. D. Joshi, A. L. Price, P. Kraft, Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *Am. J. Hum. Genet.* **96**, 329–339 (2015).
10. A. Palmetshofer, D. Zechner, T. A. Luger, A. Barta, Splicing variants of the human growth hormone mRNA: Detection in pituitary, mononuclear cells and dermal fibroblasts. *Mol. Cell. Endocrinol.* **113**, 225–234 (1995).

11. K. Fujikawa, Comparative vascular anatomy of the hip of the miniature dog and of the normal-size mongrel. *Kurume Med. J.* **38**, 159–165 (1991).
12. G. Ru, B. Terracini, L. T. Glickman, Host related risk factors for canine osteosarcoma. *Vet. J.* **156**, 31–39 (1998).
13. E. Monnet, Gastric dilatation-volvulus syndrome in dogs. *Vet. Clin. North Am. Small Anim. Pract.* **33**, 987–1005, vi (2003).
14. E. A. Brown *et al*., *FGF4* retrogene on CFA12 is responsible for chondrodystrophy and intervertebral disc disease in dogs. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11476–11481 (2017).
15. E. Healey *et al*., Genetic mapping of distal femoral, stifle, and tibial radiographic morphology in dogs with cranial cruciate ligament disease. *PLoS One* **14**, e0223094 (2019).
16. E. C. Jeanes, J. A. C. Oliver, S. L. Ricketts, D. J. Gould, C. S. Mellersh, Glaucoma-causing *ADAMTS17* mutations are also reproducibly associated with height in two domestic dog breeds: Selection for short stature may have contributed to increased prevalence of glaucoma. *Canine Genet. Epidemiol.* **6**, 5 (2019).
17. J. J. Hayward *et al*., Imputation of canine genotype array data using 365 whole-genome sequences improves power of genome-wide association studies. *PLoS Genet.* **15**, e1008003 (2019).
18. D. Vivenza *et al*., A novel deletion in the GH1 gene including the IVS3 branch site responsible for autosomal dominant isolated growth hormone deficiency. *J. Clin. Endocrinol. Metab.* **91**, 980–986 (2006).
19. M. S. Lee *et al*., Autosomal dominant growth hormone (GH) deficiency type II: The Del32-71-GH deletion mutant suppresses secretion of wild-type GH. *Endocrinology* **141**, 883–890 (2000).
20. T. K. Graves, S. Patel, P. S. Dannies, P. M. Hinkle, Misfolded growth hormone causes fragmentation of the Golgi apparatus and disrupts endoplasmic reticulum-to-Golgi traffic. *J. Cell Sci.* **114**, 3685–3694 (2001).
21. S. Davidopoulou, A. Chatzigianni, Craniofacial morphology and dental maturity in children with reduced somatic growth of different aetiology and the effect of growth hormone treatment. *Prog. Orthod.* **18**, 10 (2017).
22. F. Ferrante, S. Blasi, R. Crippa, F. Angiero, Dental abnormalities in pituitary dwarfism: A case report and review of the literature. *Case Rep. Dent.* **2017**, 5849173 (2017).
23. X. Zhang *et al*., Linking the genetic architecture of cytosine modifications with human complex traits. *Hum. Mol. Genet.* **23**, 5893–5905 (2014).
24. R. Liu *et al*., Genome-wide association study identifies loci and candidate genes for body composition and meat quality traits in Beijing-You chickens. *PLoS One* **8**, e61172 (2013).
25. A. Gusev *et al*., Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
26. C. Zhang *et al*., "GWAS in production nucleus sows using a 650K SNP Chip to explore component traits underlying a repeatable low litter birth weight phenotype" in *Proceedings of the World Congress on Genetics Applied to Livestock Production*, (2018), Vol. vol. 11, p. 567.
27. G. Kichaev *et al*., Leveraging polygenic functional enrichment to improve GWAS power. *Am. J. Hum. Genet.* **104**, 65–75 (2019).
28. S. Gerstberger, M. Hafner, T. Tuschl, A census of human RNA-binding proteins. *Nat. Rev. Genet.* **15**, 829–845 (2014).
29. R. P. Perry, D. E. Kelly, Existence of methylated messenger RNA in mouse L cells. *Cell* **1**, 37–42 (1974).
30. L. G. Morello *et al*., The human nucleolar protein FTSJ3 associates with NIP7 and functions in pre-rRNA processing. *PLoS One* **6**, e29174 (2011).
31. M. Ringeard, V. Marchand, E. Decroly, Y. Motorin, Y. Bennasser, FTSJ3 is an RNA 2′-O-methyltransferase recruited by HIV to avoid innate immune sensing. *Nature* **565**, 500–504 (2019).
32. L. Jain *et al*., 3D interactions with the growth hormone locus in cellular signalling and cancer-related pathways. *J. Mol. Endocrinol.* **64**, 209–222 (2020).
33. F. M. Simabuco, L. G. Morello, A. Z. Aragão, A. F. Paes Leme, N. I. Zanchin, Proteomic characterization of the human FTSJ3 preribosomal complexes. *J. Proteome Res.* **11**, 3112–3126 (2012).
34. A. R. Boyko *et al*., A simple genetic architecture underlies morphological variation in dogs. *PLoS Biol.* **8**, e1000451 (2010).
35. J. J. Schoenebeck *et al*., Variation of BMP3 contributes to dog breed skull diversity. *PLoS Genet.* **8**, e1002849 (2012).
36. Z. Pavlica, V. Erjavec, M. Petelin, Teeth abnormalities in the dog. *Acta Vet. Brno* **70**, 65–72 (2001).
37. A. L. Tyler, F. W. Asselbergs, S. M. Williams, J. H. Moore, Shadows of complexity: What biological networks reveal about epistasis and pleiotropy. *BioEssays* **31**, 220–227 (2009).
38. S. Purcell *et al*., PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
39. J. Plassais *et al*., Whole genome sequencing of canids reveals genomic regions under selection and variants influencing morphology. *Nat. Commun.* **10**, 1489 (2019).
40. H. Thorvaldsdóttir, J. T. Robinson, J. P. Mesirov, Integrative genomics viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
41. I. A. Adzhubei *et al*., A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
42. H. Tang, P. D. Thomas, PANTHER-PSEP: Predicting disease-causing genetic variants using position-specific evolutionary preservation. *Bioinformatics* **32**, 2230–2232 (2016).
43. R Core Team, R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna, 2017).
44. J. J. Faraway, *Extending the Linear Model with R: Generalized Linear, Mixed Effects, and Nonparametric Regression Models*, (CRC Press, Boca Raton, FL, 2006).
45. J. J. Hayward *et al*., Complex disease and phenotype mapping in the domestic dog. *Nat. Commun.* **7**, 10460 (2016).