FGF4 retrogene on CFA12 is responsible for chondrodystrophy and intervertebral disc disease in dogs

Emily A. Brown[a], Peter J. Dickinson[b], Tamer Mansour[a], Beverly K. Sturges[b], Miriam Aguilar[a], Amy E. Young[c], Courtney Korff[a], Jenna Lind[a], Cassandra L. Ettinger[d], Samuel Varon[e], Rachel Pollard[b], C. Titus Brown[a,d], Terje Raudsepp[f], and Danika L. Bannasch[a,d,1]

[a]Department of Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis, CA 95616; [b]Department of Surgical and Radiological Sciences, School of Veterinary Medicine, University of California, Davis, CA 95616; [c]Department of Animal Science, University of California, Davis, CA 95616; [d]Genome Center, University of California, Davis, CA 95616; [e]Barney and Russum Animal Clinic, Fairfield, CA 94533; and [f]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843

Edited by Harris A. Lewin, University of California, Davis, CA, and approved September 18, 2017 (received for review June 2, 2017)

Chondrodystrophy in dogs is defined by dysplastic, shortened long bones and premature degeneration and calcification of intervertebral discs. Independent genome-wide association analyses for skeletal dysplasia (short limbs) within a single breed ($P_{Bonferroni}$ = 0.01) and intervertebral disc disease (IVDD) across breeds ($P_{Bonferroni}$ = 4.0 × 10^{-10}) both identified a significant association to the same region on CFA12. Whole genome sequencing identified a highly expressed *FGF4* retrogene within this shared region. The *FGF4* retrogene segregated with limb length and had an odds ratio of 51.23 (95% CI = 46.69, 56.20) for IVDD. Long bone length in dogs is a unique example of multiple disease-causing retrocopies of the same parental gene in a mammalian species. FGF signaling abnormalities have been associated with skeletal dysplasia in humans, and our findings present opportunities for both selective elimination of a medically and financially devastating disease in dogs and further understanding of the ever-growing complexity of retrogene biology.

GWAS | inherited | genetic | dysplasia | chondrodysplasia

Variation in domestic dog (*Canis familiaris*, CFA) morphology has long fascinated both scientists and pet owners. Domestication of the dog from the wolf and the subsequent variation in size and shape within purebred dog breeds is a remarkable feat of animal breeding and selection. One of the most extreme examples of dog breed differences is in limb length, as extremely short limbs define many breeds. This morphological feature is present in breeds from all over the world and from all American Kennel Club (AKC) groups, indicating that the underlying genetic causes are likely very old.

Extensive examination of growth plates has been performed on many of these short-legged dog breeds (dachshund, Pekingese, French bulldog, spaniels, beagle), as these breeds are also prone to intervertebral disc disease (IVDD) (1–3). Histopathological analysis of the bones of puppies from these breeds demonstrated that their short stature is due to defects in endochondral ossification, the process whereby cartilage is replaced with bone, in the developing limb. The long bone growth plates show disorganization of the proliferative zone and reduction in the depth of the maturation zone (1–4). In addition to the long bones, similar but more subtle changes exist in endochondral ossification of the vertebral bodies (1, 2).

The intervertebral disc (IVD), which sits between vertebral bodies, is composed of an outer fibrous basket, called the annulus fibrosis, made of 70% collagen and an inner gel-like layer that is a remnant of the embryonic notochord, called the nucleus pulposus (5). Together, these structures and the cartilaginous endplates allow for flexibility of the vertebral column. In chondrodystrophic dogs, the nucleus pulposus is gradually replaced by chondrocyte-like cells in chondroid metaplasia (or metamorphosis) that occurs between birth and 1 y of age (1, 2). Recent studies have shown that in advanced stages of degeneration in nonchondrodystrophoid dogs there is also replacement of notochordal cells by chondrocyte-like cells, similar to the changes observed in chondrodystrophoid dogs, although this happens at an older age (3, 6–10). The replacement of the nucleus pulposus with chondrocyte-like cells is seen in humans, and chondrodystrophoid breeds have been proposed as models for human degenerative disc disease (3, 7, 11, 12).

Hansen described the two different types of canine IVD prolapse as type I and type II. Type I occurs exclusively in chondrodystrophic breeds and is characterized by premature degeneration of all discs in young dogs. In contrast, type II occurs in older dogs and is usually limited to a single disc with only partial protrusion. In type I disc disease, the calcified nucleus pulposus may undergo an explosive herniation through the annulus fibrous into the vertebral canal, resulting in inflammation and hemorrhage and causing severe pain and neurological dysfunction (1, 2). In veterinary hospital population studies, breeds with a significant increased risk of IVDD include the beagle, cocker spaniel, dachshund, French bulldog, Lhasa apso, Pekingese,

Significance

Chondrodystrophy, characterized by short limbs and intervertebral disc disease (IVDD), is a common phenotype in many of the most popular dog breeds, including the dachshund, beagle, and French bulldog. Here, we report the identification of a *FGF4* retrogene insertion on chromosome 12, the second *FGF4* retrogene reported in the dog, as responsible for chondrodystrophy and IVDD. Identification of the causative mutation for IVDD will impact an incredibly large proportion of the dog population and provides a model for IVDD in humans, as FGF-associated mutations are responsible for IVDD and short stature in human achondroplasia. This is a report of a second retrogene copy of the same parental gene, each causing complementary disease phenotypes in a mammalian species.

Author contributions: E.A.B. and D.L.B. designed research; E.A.B., P.J.D., T.M., B.K.S., M.A., A.E.Y., C.K., J.L., C.L.E., and D.L.B. performed research; P.J.D., B.K.S., S.V., R.P., C.T.B., T.R., and D.L.B. contributed new reagents/analytic tools; E.A.B. and D.L.B. analyzed data; and E.A.B., P.J.D., and D.L.B. wrote the paper.

Conflict of interest statement: The University of California, Davis, has filed a provisional patent entitled: "Methods of Diagnosing Intervertebral Disc Disease and Chondrodystrophy in Canines," on May 30, 2017.

This article is a PNAS Direct Submission.

Data deposition: The sequence data reported in this paper has been deposited in the Sequence Read Archive (SRA Bioproject no. PRJNA377155) and in the GenBank database (accession nos. MF040221 and MF040222).

[1]To whom correspondence should be addressed. Email: dlbannasch@ucdavis.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1709082114/-/DCSupplemental.

Pembroke Welsh corgi, and shih tzu (13–15). Pet insurance data suggests a conservative "lifetime prevalence" for IVDD in dogs of 3.5% in the overall population; however, in the chondrodystrophic breed with the highest risk, the dachshund, the lifetime prevalence is between 20–62% with a mortality rate of 24% (9, 10, 16–18). The effect of this disease on dogs and the financial burden to pet owners is enormous.

Skeletal dysplasia (SD), a general term to classify abnormalities of growth and development of cartilage and/or bone resulting in various forms of short stature, occurs in humans and dogs in many forms (19). With advances in molecular genetics, many of the diseases in humans are being reclassified based on the specific underlying causative mutations (20). To a lesser degree, progress has also been made in understanding the molecular nature of SD and the extreme interbreed limb-length variation observed in dogs (21–24). While the mutations causing some subtypes of SD in dogs have been determined, there are still many unexplained types of SD observed within and across dog breeds.

In 2009, the genetic basis for extreme differences in limb length in dogs was investigated by Parker et al. (25) using an across-breed genome-wide association approach. They determined that a *FGF4* retrogene insertion on CFA18 ~25 Mb from the parental copy of the *FGF4* locus was responsible for the "chondrodysplasia" phenotype in a number of breeds, such as the basset hound, Pembroke Welsh corgi, and dachshund. However, the *FGF4* retrogene insertion on CFA18 failed to explain breeds such as the American cocker spaniel, beagle, and French bulldog, that in addition to dachshunds, were the breeds originally classified as chondrodystrophoid based on histopathologic and morphologic analysis by Hansen (1) and Braund (3). The FGF gene family has similarly been implicated in SD in humans, with mutations in *FGFR3* found to be responsible for achondrodysplasia, the most common form of dwarfism, characterized by shortened limbs and abnormal vertebrae and IVDs (20, 26–30). FGF genes are involved in a number of embryological development processes, and specific levels of ligand and receptor are key for appropriate growth and development (31–33).

In this study, genome-wide association analysis in a cohort of Nova Scotia duck tolling retrievers (NSDTRs) with and without severe SD identified a significant association on CFA12 due to a 12-Mb associated haplotype, of which 1.9 Mb was found to be shared in chondrodystrophoid breeds. Subsequent genome-wide association analysis of Hansen's type I IVDD across breeds localized the same 1.9-Mb region on CFA12, suggesting that the locus responsible for SD in the NSDTR is also responsible for type I IVDD and the chondrodystrophoid phenotype across dog breeds. A previous genetic investigation of IVDD in dachshunds and limb-length morphology in Portuguese water dogs both identified the same CFA12 locus; however, neither study reported a causative mutation (34, 35). Here, using Illumina paired-end genomic sequencing, we uncover a second *FGF4* retrogene insertion (chr12: 33.7 Mb [*Canis familiaris* {canFam} 3]) in the canine genome and show that it is not only responsible for SD in the NSDTR, but also chondrodystrophy, including the predisposition to Hansen's type I IVDD, across all dog breeds.

Results

Genome-Wide Association Studies for SD and IVDD. A form of SD is common in the NSDTR and is characterized by variable decrease in limb length and associated abnormalities, including long-bone bowing, physeal widening, and joint incongruity (Fig. 1 *A* and *B*). On physical examination, in addition to shorter limbs, SD dogs may also have valgus limb deformities and larger ears (pinnae). While SD is a common phenotype in the breed, the degree of severity is highly variable.

To determine a region of the genome associated with SD in the NSDTR, genome-wide association analysis was performed using 13 NSDTRs with severe SD and 15 NSDTR controls without severe SD. There were 41 single-nucleotide polymorphisms

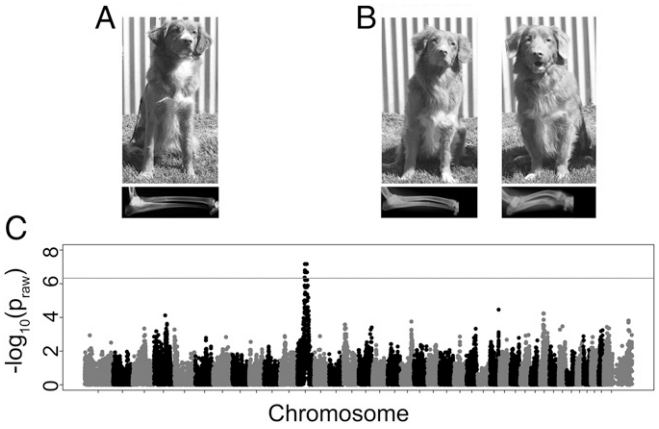


Fig. 1. Skeletal dysplasia in the NSDTR. (*A*) Picture and lateral thoracic limb radiograph of unaffected NSDTRs (ages 1 and 4 y old, respectively). (*B*, *Left*) Picture and lateral thoracic limb radiograph of mildly SD-affected NSDTRs (ages 4 and 2 y old, respectively); (*Right*) picture and lateral thoracic limb radiograph of more severely SD-affected NSDTRs (ages 3 y old and 6 mo old, respectively). Relative to the unaffected dog, the mildly SD-affected NSDTR has cranial bowing of the radius. Radiographic changes in the more severely SD-affected NSDTR include moderate cranial bowing of the radius, physeal widening, and incongruity of the elbow joint with the shape of the semilunar notch of the ulna being elongated. Pictures and radiographs are representatives of each phenotype and not paired. (*C*) SD in NSDTR GWAS. Manhattan plot showing $-\log_{10}$ of the raw *P* values for each genotyped SNP by chromosome (*x* axis). Genomic inflation was 1.02. Line denotes genome-wide significance based on Bonferroni-corrected *P* values.

(SNPs) that were genome-wide significant with a $P_{\text{Bonferroni}}$ of <0.05, all present between chr12: 35,413,695 and 46,117,273 (top SNP chr12: 36,790,324 $P_{\text{Bonferroni}} = 0.01$) (canFam2) (Fig. 1*C* and *SI Appendix*, Fig. S1).

Underlying this strong association for SD in NSDTRs was an ~12-Mb critical interval from chr12: 36–48 Mb (canFam2). Since the NSDTR SD phenotype is not uncommon in different dog breeds, we investigated haplotype sharing across breeds and observed that a portion of this associated haplotype was shared with two breeds of dog considered classically chondrodystrophic: the American cocker spaniel and beagle (1, 3). By plotting the minor allele frequency (MAF) across this interval for 7 American cocker spaniels, 14 beagles, and 13 SD-affected NSDTRs, the critical interval identified via GWAS for SD was shortened to a shared haplotype from chr12: 36.4–38.3 Mb (canFam2) (Fig. 2*A*).

To test the hypothesis that the same locus was responsible for SD and chondrodystrophy, a second genome-wide association study (GWAS) was performed using IVDD-affected cases (n = 36) and unaffected controls (n = 31) across 26 dog breeds (listed in *SI Appendix*, Table S6). The most highly associated SNP was located on CFA12 [chr12: 36,909,311 (canFam2)] with a $p_{\text{raw}} = 3.2 \times 10^{-15}$, $P_{\text{Bonferroni}} = 4.0 \times 10^{-10}$, and odds ratio of 32.67 (Fig. 2*B*). Observing linkage disequilibrium with the highest associated SNP using r^2 values >0.06, the critical interval identified via GWAS for IVDD overlaps with that seen when mapping MAF across breeds and SD in the NSDTR (Fig. 2*C*).

Identification of FGF4 Insertion on CFA12. To identify a causative variant for SD and IVDD, paired-end whole-genome sequences of two cases, one SD-affected NSDTR and one IVDD-affected dachshund, and 83 unaffected controls were investigated in the associated interval. The average coverage for these samples was 8.71×. There were 9,156 SNP variants and 7,877 insertion/deletion (indel) variants identified from chr12: 33.1–35.5 Mb (canFam3) [chr12: 36.1–38.5 Mb (canFam2)]; however, none segregated with the IVDD phenotype. The same interval was also investigated by visual inspection of BAM files to flag mate

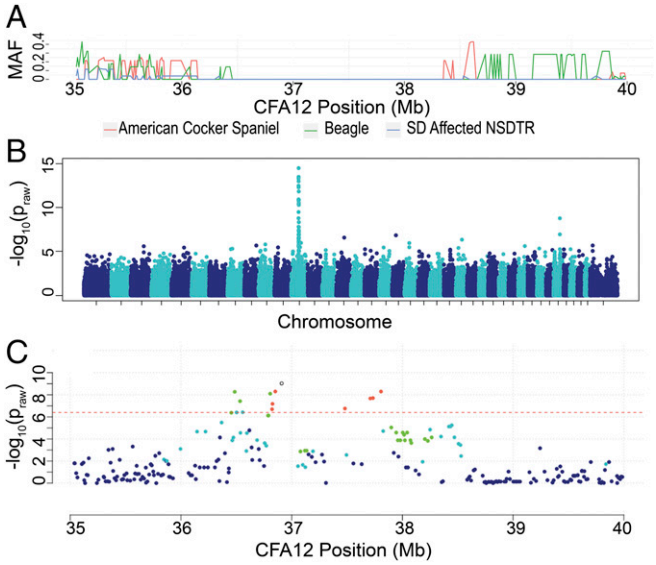


Fig. 2. Across breed investigation of SD-IVDD locus. (*A*) MAF on the *y* axis and base pair on CFA12 on the *x* axis plotted by breed. SD-affected NSDTR ($n = 13$), American cocker spaniel ($n = 7$), and beagle ($n = 14$). (*B*) Manhattan plot for the IVDD GWAS showing $-\log_{10}$ of the raw *P* values (*y* axis) for each genotyped SNP by chromosome (*x* axis). Genomic inflation was 1.63. (*C*) SNPs in 5-Mb region surrounding most highly associated SNP [chr12: 36,909,311 (canFam2)] plotted by base pair on the *x* axis and *P* value on the *y* axis. SNPs are color coded by r^2 value to show the extent of linkage disequilibrium (orange, r^2 0.6–0.8; green, r^2 0.4–0.6; light blue, r^2 0.2–0.4; and dark blue, r^2 0–0.2).

pairs with unusual insert sizes in an effort to identify any large indels. Using the two cases and two controls (one NSDTR and one Saluki) eight large indels (>200 bp) were identified within the interval (*SI Appendix*, Table S1). Four large indels did not segregate when investigated in additional control genomes, while the remaining four were eliminated after PCR showed lack of segregation between cases and controls.

Visual inspection of the BAM files for read pairs mapping to a different chromosome location identified a region, located at approximately chr12: 33,710,200 (canFam3) that segregated with the two cases and two controls (*SI Appendix*, Fig. S2). At this location, read mates mapped to chr18: 48.4 Mb (canFam3) and chr7: 68.3Mb (canFam3) in the NSDTR and dachshund cases, but none of the controls. The reads that mapped to CFA18 aligned to parental *FGF4*, which was highly suggestive of a *FGF4* retrogene insertion at this location. The reads that mapped to CFA7 were investigated by PCR and appear to mark a genome assembly error or a mutation within the dog used for the genome assembly (canFam3).

To investigate the potential *FGF4* insert on CFA12, the region was PCR amplified using primers flanking the insertion site from genomic DNA of an IVDD-affected beagle. Wild-type dogs without the insert had a single 615-bp band, while dogs homozygous for the CFA12 *FGF4* insertion had an ∼4-kb product. Sanger sequencing showed the insertion on CFA12 is 3,209 bp long (GenBank accession no. MF040221) and includes parental *FGF4* cDNA (i.e., *FGF4* exons spliced without introns), as shown in the insert schematic comparing parental *FGF4* to the CFA12 insert (Fig. 3). The insert also contains a majority of the predicted 5′-untranslated region (UTR), which includes the transcription start site (TSS) as only PCR primers *FGF4*_TSS.F1 and *FGF4*.R1 yielded a product in RT-PCR using cDNA from neonatal beagle IVD (*SI Appendix*, Table S8). The insertion location is intergenic between the 3′-UTR of OGFRL1 ∼9.5 kb on the proximal side and ∼350 kb to the RIMS1 gene on the distal side. It is also intronic to an unvalidated antisense transcript (NASEQ_As_0027291).

To compare the CFA12 *FGF4* retrogene to the previously identified CFA18 *FGF4* retrogene, it was necessary to obtain the full-length sequence of the CFA18 insertion (25). The cloned product was sequenced using the flanking and common internal primers (*SI Appendix*, Table S8), yielding a 2,665 bp insert (GenBank accession no. MF040222). While it contained the same length 5′-UTR and *FGF4* cDNA as that seen in the CFA12 *FGF4* insert, the 3′-UTR was shortened in comparison. The 3′-UTR of the CFA18 *FGF4* insert was followed by a sequence containing 30 adenine and one guanine residues and a different target site duplication (TSD) sequence (AAG TCA GAC AGA G). The 5′-UTR shared between the two canine FGF4 retrogenes was compared with the human FGF4 gene and was 295 bp long and 91% identical at the nucleotide level. Within the highly conserved 5′-UTR that was transposed, there are 79 conserved transcription factor binding sites between dog and human (https://ecrbrowser.dcode.org). There are also multiple DNase I hypersensitivity sites as well as H3K4me3 and H3K9ac marks (www.genome.ucsc.edu) within the human sequence.

Association of FGF4 Retrogene with SD, Height, and IVDD. To assay the insertions in additional dogs, insertion- and allele-specific PCR-based genotyping assays were developed for both the CFA12 *FGF4* insertion and the previously identified CFA18 *FGF4* insertion (Fig. 4*A*). Twelve SD NSDTR cases from the GWAS were genotyped and were homozygous for the CFA12 *FGF4* insertion, while all controls were heterozygous or wild type. Additionally, IVDD cases ($n = 7$) from the NSDTR breed were collected and were either homozygous mutant or heterozygous for the CFA12 *FGF4* insertion (*SI Appendix*, Table S3). All NSDTRs tested for the CFA18 *FGF4* insertion ($n = 31$) were wild type, including SD and IVDD cases. NSDTRs with known height ($n = 20$ males) at the withers were genotyped for the CFA12 *FGF4* insertion to investigate the association of height with genotype status. Height and genotype were significantly associated in a dose-dependent manner (all $P < 0.04$) when comparing wild-type, heterozygous, and homozygous dogs (Fig. 4*B*).

To assess the significance of association of the CFA12 *FGF4* insertion with IVDD across breeds, dogs used in the IVDD GWAS were genotyped for both insertions. All dogs' genotypes were concordant with phenotype except for one case, a rottweiler (*SI Appendix*, Table S2). When associated with IVDD, the CFA12 *FGF4* insertion was more highly associated than both the most highly associated SNPs from the GWAS, as well as the CFA18 *FGF4* insertion (Fig. 4*C*). To further investigate the association of the CFA12 *FGF4* insertion with IVDD, 33 additional cases were genotyped for

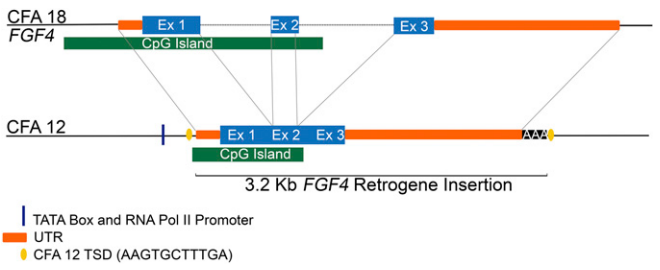
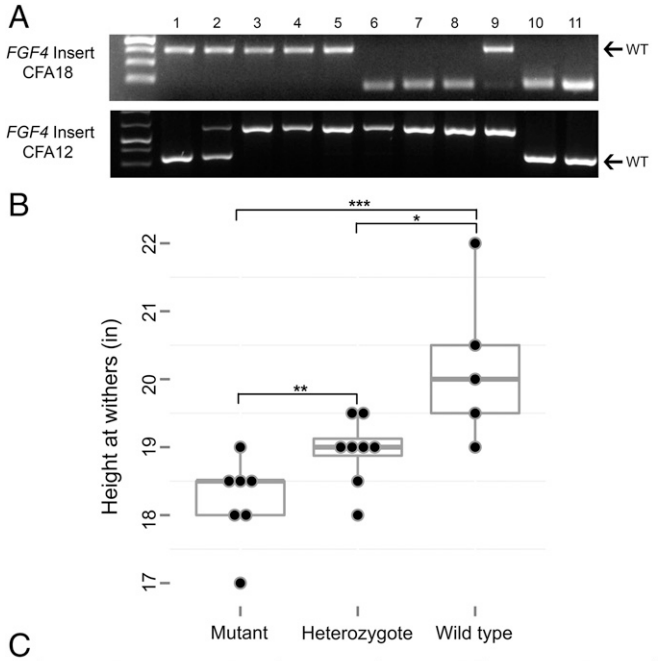


Fig. 3. Schematic of parental *FGF4* (CFA18) retrotransposition to CFA12: 33,710,178 (canFam3): Predicted TATA box at chr12: 33,709,940–947 (canFam3) and predicted RNA Pol II promoter at chr12: 33,709,964–976 (canFam3) (54). Parental *FGF4* UTRs are unknown in the dog; however, they were approximated based on human and mouse TransMap data (www.genome.ucsc.edu). The predicted 5′-UTR spans from chr18: 48,413,185 to 48,413,480 (canFam3), but RT-PCR in beagle IVD suggests that the TSS is between chr18: 48,413,315 and 48,413,402 (canFam3) (*SI Appendix, SI Methods*). The insert includes all predicted 3′-UTR, followed by 42 adenine residues and the duplicated TSD sequence [chr12: 33,710,168–33,710,178 (canFam3)].



C

Locus	Chi Square	p-value	Odds Ratio
Chr12:36,909,311 (CanFam2)	62.14	3.2×10^{-15}	32.67 (30.12, 35.44)
CFA 12 *FGF4* Insertion	67.32	2.3×10^{-16}	51.23 (46.69, 56.20)
CFA 18 *FGF4* Insertion	23.51	1.2×10^{-6}	6.38 (5.99, 6.79)

Fig. 4. Association of *FGF4* insertion genotypes with height and IVDD. (*A*) Genotyping results for CFA18 and CFA12 *FGF4* insertions. Arrows indicate wild-type (WT) band. Numbered lanes: ladder and 1–3, NSDTR; 4, beagle; 5, American cocker spaniel; 6, dachshund; 7, basset hound; 8, Pembroke Welsh corgi; 9, coton de Tulear; 10, cairn terrier; and 11, West Highland white terrier. (*B*) Height at the withers in inches (in) (*y* axis) by genotype (*x* axis) for 20 NSDTRs. Seven SD NSDTR cases were homozygous mutant for the CFA12 *FGF4* insertion and their mean height was 18.22 in. Thirteen NSDTRs were unaffected with SD: 5 were wild type and had a mean height of 20.2 in; 8 were heterozygous for the CFA12 *FGF4* insertion and had a mean height of 18.94 in. ***$P = 0.01$, **$P = 0.02$, *$P = 0.03$. (*C*) Association of loci with IVDD. The 95% confidence interval for odds ratio appears in parentheses.

the CFA12 *FGF4* insertion: 10 were heterozygous and 23 were homozygous for the CFA 12 *FGF4* insertion (*SI Appendix*, Table S3).

To investigate the breed distribution of the CFA12 *FGF4* retrogene insertion, 568 dogs from 50 breeds were genotyped [*SI Appendix*, Table S4, data for breeds genotyped for the CFA18 *FGF4* retrogene insertion by Parker et al. (25) are included for comparison]. The CFA12 *FGF4* insertion segregates in the majority of breeds where it occurs and is present in small- and medium-sized dog breeds with high frequency (Fig. 5). Interestingly, all of the dogs with the CFA12 *FGF4* insertion also have large external ears (pinnae), which is consistent with the phenotype seen in the NSDTR (photographs available through AKC.org).

Based on occurrence of IVDD at the Pritchard Veterinary Medicine Teaching Hospital at University of California, Davis (UC Davis), the breeds with a statistically higher frequency of IVDD are also those with a higher frequency of the CFA12 *FGF4* insert allele, while the breeds with a statistically lower frequency of IVDD are those with a lower frequency of the CFA12 *FGF4* insert allele (*SI Appendix*, Table S5).

CFA12 FGF4 Retrogene Expression. To investigate the gene expression environment in which *FGF4* inserted on CFA12, semi-quantitative RT-PCR (semi-qPCR) was performed for genes across the IVDD-associated interval. Using cDNA derived from neonatal vertebral body (VB) and IVD, skeletal muscle, and testis, expression levels of genes across the CFA12-associated interval were assayed in a beagle case and cane corso or Labrador retriever control, including: *COL9A1*, *SMAP1*, *B3GAT2*, *OGFRL1*, *LINC00472*, *RIMS1*, *KCNQ5*, and *COL12A1*. Expression differences between case and control were not apparent in these genes; however, we confirmed that all except *RIMS1* are expressed in both neonatal VB and IVD, supporting that *FGF4* inserted itself in a gene milieu conducive to expression in IVD. Semi-qPCR for total *FGF4* (parental and retrogene products) in the same tissues showed increased expression across all tested tissue types in the case versus the control (*SI Appendix*, Fig. S3).

To evaluate the effect of the CFA12 *FGF4* retrogene insertion on overall *FGF4* transcript levels, quantitative RT-PCR was performed. A comparison between samples homozygous for the CFA12 *FGF4* insertion and samples with only the parental copy of *FGF4* (i.e., wild type for both the CFA12 and CFA18 *FGF4* insertions) showed a 19.47× higher ($P = 0.03$) and 2.16× higher ($P = 0.03$) expression of *FGF4* in neonatal IVD and VB, respectively (*SI Appendix*, Fig. S4).

Discussion

In this study, we report the identification of a *FGF4* retrogene insertion in the dog genome responsible for chondrodystrophy across dog breeds, characterized by both short limbs and susceptibility to Hansen's type I intervertebral disc disease. A region was identified on CFA12 due to association with a segregating form of skeletal dysplasia observed in the NSDTR. While NSDTRs can be variably affected, the use of severely affected dogs enabled identification of the locus through GWAS. Haplotype sharing with chondrodystrophoid breeds and genome-wide association analysis for type I IVDD identified the same region on CFA12. Evaluation of mismapped mate pairs allowed the identification of an *FGF4* retrogene, which leads to a ∼20-fold increase in expression of *FGF4* in neonatal intervertebral disc. Due to the embryonic expression pattern of *FGF4*, it is probable that these expression changes are also impacting endochondral ossification. This is the second *FGF4* retrogene identified in dogs that affects limb length. While the *FGF4* retrogene on CFA18 impacts limb length, the *FGF4* retrogene on CFA12 explains the chondrodystrophoid phenotype, which includes limb length and IVDD (significant odds ratio > 50).

Fibroblast growth factor 4 (*FGF4*) is a growth factor gene expressed in specific tissues and at specific times throughout embryonic development in the mouse (36). *FGF4* is highly expressed in the apical ectodermal ridge of the developing limb bud, as well as somites and the notochord that will form the vertebral column and IVDs (36–38). FGF signaling is required

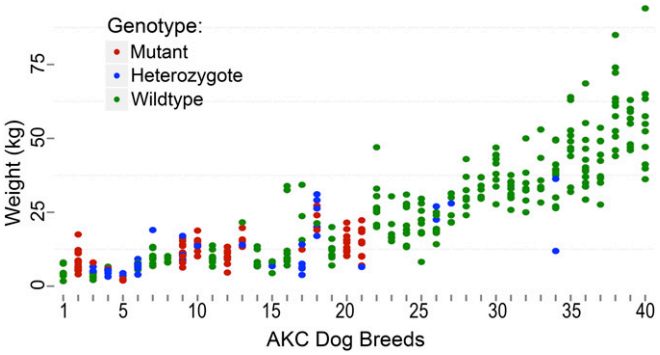


Fig. 5. CFA12 *FGF4* genotypes across breeds: Genotypes for the CFA12 *FGF4* insertion across dog breeds ordered by breed standard height from shortest to tallest (*x* axis), plotted by dog weight in kilograms (kg) (*y* axis). Dog breeds can be found in *SI Appendix*, Table S4. Only genotyped dogs with weights available ($n = 376$) are shown. Dogs are color coded by genotype status.

GENETICS

for appropriate embryonic axial growth and segmentation, and *FGF4*/*FGF8* murine hypomorphs are characterized by altered vertebral morphology and smaller limb buds (39, 40). Additionally, *FGF8* hypomorphs are observed to have either hypoplastic or nonexistent external ear structures (41). In mice, creation of a gain-of-function *FGF4* copy to replace an inactive *FGF8* gene was able to rescue limb development; however, it also caused abnormal tissue deposition and postaxial polydactyly, highlighting that levels of *FGF* throughout embryonic development must be properly controlled for normal limb formation (31). While the specific embryonic expression pattern of *FGF4* in dogs with four to six copies of the gene is unknown, we hypothesize that the insertion site milieu on CFA12 versus CFA18 is contributing to differences in expression between the retrogenes, leading to the differences in phenotype.

A survey of retrogenes in the canine reference genome reported ∼70 functional retrogenes in the dog; however, only the previous CFA18 *FGF4* retrogene insertion has been reported to be associated with a disease-causing phenotype (25, 42). Similarly in humans, the formation of processed pseudogenes in general, as well as those that retain their intended function and cause disease, is rare (43–46).

Both copies of the canine *FGF4* retrogenes have signatures of having arisen from RNA retrotransposed by LINE-1 integrase and reverse transcriptase, including flanking TSDs and polyA tracts (class 1 templated sequence insertion polymorphism) (47). The insertion sites of the two *FGF4* retrogenes are very different. The CFA18 site is within a LINE element and the CFA12 insertion site is intergenic between *OGFRL1* and *RIMS*. The CFA18 *FGF4* retrogene insertion was predicted to be expressed due to insertion near sequences with promoter properties (25). While the CFA12 *FGF4* insertion is placed near a potential TATA box and RNA Pol II promoter, it is more likely that the CpG island included in the retrogene is driving expression (48–50). This hypothesis is supported by the finding that a majority of retrogene expression is actually due to genomic context and contribution of CpG islands, not through the use of nearby promoters (51). Human *FGF4* shares the large CpG island observed in dogs and other species. Within the highly conserved 5′-UTR that was transposed, there are many conserved transcription factor binding sites between dog and human as well as multiple methylation marks further supporting that both CFA12 and CFA18 *FGF4* retrogene insertions contain the necessary components for transcription. To our knowledge, this is a unique documentation of a second retrogene insertion of the same parental gene resulting in a disease phenotype in a mammalian species. Due to the lack of resources available to identify these types of mutations, it is likely that there are other phenotype inducing retrocopies present in the canine genome that have yet to be discovered.

Chondrodystrophy-associated mutation events occurred a very long time ago, as there are descriptions of short-legged dogs dating back over 4,000 y (52). In addition, both mutations occur concurrently in very unrelated dog breeds from diverse breed groupings and geographical locations. The fact that *FGF4* has been retrotransposed twice in dogs in the last 3–4,000 y makes it likely that this has happened at other times. The large CpG island in the 5′-end of the parental *FGF4* gene may enable phenotypic consequences more readily than for other retrogenes. Once the *FGF4* retrogene appeared and produced an obvious phenotype, strong selection was likely applied to retain it, aided by the semidominant nature of the mutation.

The NSDTR is the smallest of the retriever dog breeds, and based on the association of the CFA12 *FGF4* insertion with height, we hypothesize that the heterozygous phenotype is aesthetically desirable and that selection is maintaining the insertion at a relatively high allele frequency. Investigation of the CFA12 *FGF4* insertion in additional breeds also showed high allele frequency in multiple small- and medium-sized dog breeds. In breeds also containing the CFA18 *FGF4* insertion, there is an even more dramatic decrease in height (e.g., basset hound, Cardigan Welsh corgi, dachshund, etc.), supporting that both *FGF4* retrogenes affect long-bone length.

In addition to segregating with height, the CFA12 *FGF4* insertion also segregates with Hansen's type I IVDD susceptibility. Of the IVDD cases genotyped for the CFA12 *FGF4* insertion, all were homozygous mutant or heterozygous, except for one, suggesting that one additional copy of *FGF4* on CFA12 is sufficient to cause type I IVDD. The single discordant case was a rottweiler, a breed that does not fit the chondrodystrophic phenotype. It is possible that there is another cause of IVDD in non-chondrodystrophoid dog breeds occurring without endochondral ossification defects (10). IVDD-affected NSDTRs were also all either homozygous or heterozygous for the CFA12 *FGF4* insertion. This supports the idea that while the CFA12 *FGF4* insertion is semidominant with respect to height, it is dominant for altered IVDs. Given that the CFA18 *FGF4* insertion is not found in the NSDTR and was inconsistently present in the IVDD cases tested, this further supports the idea that the identified insertion on CFA12 is causing both short limbs and Hansen's type I IVDD in both the NSDTR and across dog breeds.

The breeds with a higher frequency of the CFA12 *FGF4* insertion are the same breeds identified in the last 50 y as being predisposed to IVDD. Presence of the CFA18 *FGF4* insertion is common in many breeds with IVDD, and it is possible that it may contribute to the disease; however, previous mapping within dachshunds, which are reported "fixed" for the CFA18 *FGF4* insertion, show segregation of the associated haplotype on chromosome 12 with IVDD, supporting the idea that the CFA12 *FGF4* insertion is the critical factor determining disease status (25, 34). Of particular interest is the lack of reports of IVDD cases in breeds such as the cairn terrier and West Highland white terrier, both of which have the CFA18 *FGF4* insertion, but not the CFA12 *FGF4* insertion. Similarly, the high incidence of IVDD in breeds such as the American cocker spaniel, beagle, and French bulldog that do not have the CFA18 *FGF4* insertion but a high frequency of the CFA12 *FGF4* insertion supports the idea that *FGF4* specifically from CFA12 is contributing to the IVDD phenotype.

The segregation of the CFA12 *FGF4* insertion within dog breeds presents an opportunity for improvement of animal health, as implementation of genetic testing over time could lead to the elimination of type I IVDD. Based on the ever-growing popularity of some breeds, the number of animals with this intervertebral disc disease mutation across the globe is in the millions. Myelopathy secondary to IVD herniation is the most commonly presenting neurological disorder of the spinal cord in dogs (53). The overall health and financial consequences across the spectrum of presentations in companion dogs is immense. Prevention of disease through breeding and eradication has the potential for far-reaching benefits beyond those achievable through advances in surgical or medical therapy. Additionally, the dog may serve as a valuable human–animal model for IVDD. Administration of a tyrosine kinase inhibitor in a mouse model with a gain-of-function mutation in *FGFR3* has been shown to overcome growth defects associated with altered FGF signaling (32). Based on the phenotype and molecular etiology of chondrodystrophy and IVDD in dogs, it has the potential to serve as a bridge between mouse and human studies evaluating the efficacy of targeted pharmacological treatment of FGF-based genetic disorders.

Given the high mortality rate of IVDD and the high cost of surgery, identification of this susceptibility locus could provide a valuable tool for owners, breeders, and veterinarians for mitigating risk of intervertebral disc herniation and resulting myelopathy (9). This could be especially useful in breeds that have both the CFA12 and CFA18 *FGF4* retrogene, as they could breed away from the CFA12 *FGF4* retrogene, while still maintaining the aesthetically desirable shortness in stature contributed

by the CFA18 *FGF4* retrogene. In breeds with only the CFA12 *FGF4* retrogene, breeders will ultimately decide if prevention of Hansen's type I IVDD outweighs any potential loss of shortness (or gain in height).

Methods

Full detailed methods are available in *SI Appendix, SI Methods*. Collection of canine samples was approved by the University of California, Davis, Animal Care and Use Committee (protocol 18561). Genome-wide SNP genotyping was performed using the Illumina Canine HD 174,000 SNP array (Illumina). Sequencing was performed on the Illumina HiSEq. 2500 with 150-bp paired-end reads. Both CFA12 and CFA18 retrogene insertions were cloned and sequenced to determine full-length insertions. The *FGF4* insertions on CFA12 and 18 were assayed using a PCR-based genotyping test. qPCR and semi-qPCR were used to assay expression of genes within the interval and *FGF4*.

ACKNOWLEDGMENTS. Funding for this work was provided by the Maxine Adler Endowed Chair Funds, Center for Companion Animal Health, School of Veterinary Medicine, University of California, Davis (UC Davis); NIH, National Institute of Dental and Craniofacial Research R01DE22532; UC Davis Signature Research in Genomics: ACTG NIH, T35 OD010956 and NIH T32 OD010931.

1. Hansen H-J (1952) A pathologic-anatomical study on disc degeneration in dog: With special reference to the so-called enchondrosis intervertebralis. *Acta Orthop Scand* 23 (Suppl 11):1–130.
2. Hansen H-J (1951) A pathologic-anatomical interpretation of disc degeneration in dogs. *Acta Orthop Scand* 20:280–293.
3. Braund KG, Ghosh P, Taylor TK, Larsen LH (1975) Morphological studies of the canine intervertebral disc. The assignment of the beagle to the achondroplastic classification. *Res Vet Sci* 19:167–172.
4. Martinez S, Fajardo R, Valdes J, Ulloa-Arvizu R, Alonso R (2007) Histopathologic study of long-bone growth plates confirms the basset hound as an osteochondrodysplastic breed. *Can J Vet Res* 71:66–69.
5. Bray JP, Burbidge HM (1998) The canine intervertebral disk. Part two: Degenerative changes—Nonchondrodystrophoid versus chondrodystrophoid disks. *J Am Anim Hosp Assoc* 34:135–144.
6. Hunter CJ, Matyas JR, Duncan NA (2003) The three-dimensional architecture of the notochordal nucleus pulposus: Novel observations on cell structures in the canine intervertebral disc. *J Anat* 202:279–291.
7. Cappello R, Bird JL, Pfeiffer D, Bayliss MT, Dudhia J (2006) Notochordal cell produce and assemble extracellular matrix in a distinct manner, which may be responsible for the maintenance of healthy nucleus pulposus. *Spine* 31:873–882, discussion 883.
8. Kranenburg HJ, et al. (2013) Intervertebral disc disease in dogs–Part 2: Comparison of clinical, magnetic resonance imaging, and histological findings in 74 surgically treated dogs. *Vet J* 195:164–171.
9. Bergknut N, et al. (2012) Incidence and mortality of diseases related to intervertebral disc degeneration in a population of over 600,000 dogs. *J Vet Intern Med* 26:847.
10. Bergknut N, et al. (2012) The dog as an animal model for intervertebral disc degeneration? *Spine* 37:351–358.
11. Sakai D, Nakai T, Mochida J, Alini M, Grad S (2009) Differential phenotype of intervertebral disc cells: Microarray and immunohistochemical analysis of canine nucleus pulposus and anulus fibrosus. *Spine* 34:1448–1456.
12. Oegema TR, Jr (2002) The role of disc cell heterogeneity in determining disc biochemistry: A speculation. *Biochem Soc Trans* 30:839–844.
13. Priester WA (1976) Canine intervertebral disc disease—Occurrence by age, breed, and sex among 8,117 cases. *Theriogenology* 6:293–303.
14. Fluehmann G, Doherr MG, Jaggy A (2006) Canine neurological diseases in a referral hospital population between 1989 and 2000 in Switzerland. *J Small Anim Pract* 47: 582–587.
15. Brisson BA, Moffatt SL, Swayne SL, Parent JM (2004) Recurrence of thoracolumbar intervertebral disk extrusion in chondrodystrophic dogs after surgical decompression with or without prophylactic fenestration: 265 cases (1995-1999). *J Am Vet Med Assoc* 224:1808–1814.
16. Ball MU, McGuire JA, Swaim SF, Hoerlein BF (1982) Patterns of occurrence of disk disease among registered dachshunds. *J Am Vet Med Assoc* 180:519–522.
17. Bellumori TP, Famula TR, Bannasch DL, Belanger JM, Oberbauer AM (2013) Prevalence of inherited disorders among mixed-breed and purebred dogs: 27,254 cases (1995-2010). *J Am Vet Med Assoc* 242:1549–1555.
18. Lappalainen AK, Vaittinen E, Junnila J, Laitinen-Vapaavuori O (2014) Intervertebral disc disease in Dachshunds radiographically screened for intervertebral disc calcifications. *Acta Vet Scand* 56:89.
19. Anonymous (1971) Nomenclature for constitutional (intrinsic) diseases of bones. *Pediatrics* 47:431–434.
20. Warman ML, et al. (2011) Nosology and classification of genetic skeletal disorders: 2010 revision. *Am J Med Genet A* 155A:943–968.
21. Frischknecht M, et al. (2013) A *COL11A2* mutation in Labrador retrievers with mild disproportionate dwarfism. *PLoS One* 8:e60149.
22. Goldstein O, et al. (2010) COL9A2 and COL9A3 mutations in canine autosomal recessive oculoskeletal dysplasia. *Mamm Genome* 21:398–408.
23. Neff MW, et al. (2012) Partial deletion of the sulfate transporter SLC13A1 is associated with an osteochondrodysplasia in the miniature poodle breed. *PLoS One* 7: e51917.
24. Kyöstilä K, Lappalainen AK, Lohi H (2013) Canine chondrodysplasia caused by a truncating mutation in collagen-binding integrin alpha subunit 10. *PLoS One* 8: e75621.
25. Parker HG, et al. (2009) An expressed fgf4 retrogene is associated with breed-defining chondrodysplasia in domestic dogs. *Science* 325:995–998.
26. Shiang R, et al. (1994) Mutations in the transmembrane domain of FGFR3 cause the most common genetic form of dwarfism, achondroplasia. *Cell* 78:335–342.
27. Langer LO, Jr, Baumann PA, Gorlin RJ (1968) Achondroplasia: Clinical radiologic features with comment on genetic implications. *Clin Pediatr (Phila)* 7:474–485.
28. Rousseau F, et al. (1994) Mutations in the gene encoding fibroblast growth factor receptor-3 in achondroplasia. *Nature* 371:252–254.
29. Naski MC, Wang Q, Xu J, Ornitz DM (1996) Graded activation of fibroblast growth factor receptor 3 by mutations causing achondroplasia and thanatophoric dysplasia. *Nat Genet* 13:233–237.
30. Gibson BG, Briggs MD (2016) The aggrecanopathies; an evolving phenotypic spectrum of human genetic skeletal diseases. *Orphanet J Rare Dis* 11:86.
31. Lu P, Minowada G, Martin GR (2006) Increasing Fgf4 expression in the mouse limb bud causes polysyndactyly and rescues the skeletal defects that result from loss of Fgf8 function. *Development* 133:33–42.
32. Komla-Ebri D, et al. (2016) Tyrosine kinase inhibitor NVP-BGJ398 functionally improves FGFR3-related dwarfism in mouse model. *J Clin Invest* 126:1871–1884.
33. Dailey L, Ambrosetti D, Mansukhani A, Basilico C (2005) Mechanisms underlying differential responses to FGF signaling. *Cytokine Growth Factor Rev* 16:233–247.
34. Mogensen MS, et al. (2011) Genome-wide association study in Dachshund: Identification of a major locus affecting intervertebral disc calcification. *J Hered* 102(Suppl 1): S81–S86.
35. Quignon P, et al. (2009) Fine mapping a locus controlling leg morphology in the domestic dog. *Cold Spring Harb Symp Quant Biol* 74:327–333.
36. Niswander L, Martin GR (1992) Fgf-4 expression during gastrulation, myogenesis, limb and tooth development in the mouse. *Development* 114:755–768.
37. Bagnall KM, Higgins SJ, Sanders EJ (1988) The contribution made by a single somite to the vertebral column: Experimental evidence in support of resegmentation using the chick-quail chimaera model. *Development* 103:69–85.
38. Shamim H, et al. (1999) Sequential roles for Fgf4, En1 and Fgf8 in specification and regionalisation of the midbrain. *Development* 126:945–959.
39. Boulet AM, Capecchi MR (2012) Signaling by FGF4 and FGF8 is required for axial elongation of the mouse embryo. *Dev Biol* 371:235–245.
40. Sun X, Mariani FV, Martin GR (2002) Functions of FGF signalling from the apical ectodermal ridge in limb development. *Nature* 418:501–508.
41. Abu-Issa R, Smyth G, Smoak I, Yamamura K, Meyers EN (2002) Fgf8 is required for pharyngeal arch and cardiovascular development in the mouse. *Development* 129: 4613–4625.
42. Pan D, Zhang L (2009) Burst of young retrogenes and independent retrogene formation in mammals. *PLoS One* 4:e5040.
43. Kubiak MR, Makałowska I (2017) Protein-coding genes' retrocopies and their functions. *Viruses* 9:80.
44. Hancks DC, Kazazian HH, Jr (2012) Active human retrotransposons: Variation and disease. *Curr Opin Genet Dev* 22:191–203.
45. de Boer M, et al. (2014) Primary immunodeficiency caused by an exonized retroposed gene copy inserted in the CYBB gene. *Hum Mutat* 35:486–496.
46. Breyer JP, et al. (2014) An expressed retrogene of the master embryonic stem cell gene POU5F1 is associated with prostate cancer susceptibility. *Am J Hum Genet* 94: 395–404.
47. Onozawa M, Goldberg L, Aplan PD (2015) Landscape of insertion polymorphisms in the human genome. *Genome Biol Evol* 7:960–968.
48. Antequera F (2003) Structure, function and evolution of CpG island promoters. *Cell Mol Life Sci* 60:1647–1658.
49. Hannenhalli S, Levy S (2001) Promoter prediction in the human genome. *Bioinformatics* 17(Suppl 1):S90–S96.
50. Ioshikhes IP, Zhang MQ (2000) Large-scale human promoter mapping using CpG islands. *Nat Genet* 26:61–63.
51. Carelli FN, et al. (2016) The life history of retrocopies illuminates the evolution of new mammalian genes. *Genome Res* 26:301–314.
52. Morris D (2002) *Dogs: A Dictionary of Dog Breeds* (Trafalgar Square Publishing, North Pomfret, VT).
53. Packer RMA, Hendricks A, Volk HA, Shihab NK, Burn CC (2013) How long and low can you go? Effect of conformation on the risk of thoracolumbar intervertebral disc extrusion in domestic dogs. *PLoS One* 8:e69650.
54. Solovyev VV, Shahmuradov IA, Salamov AA (2010) Identification of promoter regions and regulatory sites. *Methods Mol Biol* 674:57–83.