Supporting Appendix (SI):

Methods

Phenotype and Sample Collection

Blood samples, height at the withers, thoracic limb radiographs, and pictures (when possible), were collected from privately owned NSDTRs reported affected with skeletal dysplasia (SD), as well as phenotypically “normal” dogs. Additionally, blood samples were collected from type I intervertebral disc disease (IVDD) cases seen at the University of California, Davis School of Veterinary Medicine Teaching Hospital and privately owned NSDTRs. IVDD cases were defined by the presence of one or more mineralized thoracolumbar IVDs, as confirmed by vertebral column radiographs and/or the presence of extruded calcified degenerative disc material at surgery or necropsy. DNA was extracted from EDTA whole blood samples (Gentra Puregene DNA purification extraction kit, Qiagen, Valencia, CA).

GWAS

SD GWAS used 13 NSDTR SD cases and 15 NSDTR controls. Additionally, a GWAS was performed for 36 IVDD cases and 31 controls with no reported IVDD (number and breeds of dogs listed in Table S6)SNPs were pruned from analysis if the MAF was <5% and the call rate <90% for the SD GWAS (after pruning, there were 106,303 SNPs for analysis) and SNPs were pruned from analysis if the MAF was <10% and the call rate <90% for the across breed IVDD GWAS (after SNP pruning, there were 126,020 SNPs for Chi square analysis). Chi-square association, Bonferroni adjustments, and genomic inflation calculations were performed in Plink (1). SNPs have been pruned from analysis using recommended criteria by Kierczak et al. (2). Figures were made in R using ggplot2 and cgmisc packages (2, 3).

Whole Genome Sequencing

DNA was fragmented using the Covaris E220 sonicator (Covaris Inc.), and then 550 bp insert size fragments were selected. Illumina paired-end 150 bp libraries were prepared using PCR-free library prep kits. Reads were scanned for sequencing adaptors and low quality sequences using the Trimmomatic software package (V 0.36) (4). High quality reads were aligned to the dog reference genome canFam3 (5) using the BWA software package (v0.7.7)(6). Duplicate reads were excluded using Picard v2.2.4 tool MarkDuplicates (Picard tools, Broad Institute). Variant calling was performed using GATK HaplotypeCaller (v3.5) (7). SNPs and small indels within the associated interval were investigated for segregation with IVDD in 2 cases (1 affected NSDTR and 1 Dachshund) versus 83 controls of various normal legged breeds. BAM files for the associated interval were scanned by eye for large, segregating indels in the Integrative Genomics Viewer (IGV, Broad Institute) in the 2 cases and 2 controls (1 NSDTR and 1 Saluki). Reads were color-coded by insert size and pair orientation. BAM files for additional control dogs (1 NSDTR, 1 Weimaraner, 1 Border Collie (https://www.ncbi.nlm.nih.gov/biosample/SAMN03801652)) were used to evaluate segregation of 8 identified large indels. The remaining segregating large indels (Deletions 1-3 and Insert 5) were investigated in additional dogs using PCR (primers listed in Table S7).

Cloning

Full *FGF4* retrogene insertion sequence on CFA12 and CFA18 was obtained by cloning (TOPO TA Cloning kit with PCR2.1 TOPO, Thermo Fisher Scientific, Inc., Waltham, MA, USA). CFA12 and 18 *FGF4* insertions were amplified from genomic DNA (LongAmp Taq PCR Kit, New England Biolabs, Ipswich, MA, USA) using primers flanking the inserts (Table S8). The CFA12 *FGF4* insertion was cloned from a Beagle and the CFA18 *FGF4* insertion was cloned from a Dachshund. Plasmid DNA was extracted (QIAprep Spin Miniprep Kit, Qiagen, Valencia, CA, USA), sequenced using vector primers M13.F and M13.R (Thermo Fisher Scientific, Inc., Waltham, MA, USA), and analyzed using VectorNTI software (Thermo Fisher Scientific, Inc., Waltham, MA, USA). Internal *FGF4* primers were used to create overlapping contigs (Table S8). CFA12 and 18 *FGF4* inserts were aligned to parental *FGF4*. Polymorphisms identified were not queried in additional dogs, so could be due to sequencing or cloning error or be dog/breed specific. Retrogene sequence insertion site context and 5' UTR within each retrogene copy were compared between the two retrogenes and human genome sequence using the "Liftover" function in the USCS genome browser.

Investigation of Potential Insert from CFA7

The integrity of the region on CFA7 (chr7:68,371,500-68,374,000) was tested using PCR (LongAmp Taq PCR Kit, New England Biolabs, Ipswich, MA, USA) with primers spanning the potentially inserted segment (Table S7) followed by sequencing (Applied Biosystems 3500 Genetic Analyzer and Big Dye Terminator Sequencing Kit, Life Technologies, Burlington, ON, Canada) for 6 cases (1 NSDTR, 1 Beagle, 1 Basset Hound, 1 French Bulldog, and 1 Maltese) and 1 control (Boston Terrier). If the genome assembly is correct, PCR product size should be 2,077 bp, if not, the product would only be 842 bp. Products were aligned to the UCSC genome browser (genome.ucsc.edu/).

Genotyping

Three primer PCR was performed using a forward and reverse primer flanking the respective insert, as well as an additional forward primer located within the insert (Table S9). For the CFA18 *FGF4* insertion: wild type dogs had a 388 bp product, homozygous mutant dogs had a 168 bp product, and heterozygous dogs had both. For the CFA12 *FGF4* insertion, a 333bp product was present in wild type dogs, a 654 bp band was present in the homozygous mutant dogs, and both were present in heterozygous dogs. Height at the withers was collected for 20 male NSDTRs to associate dog height with the CFA12 *FGF4* insertion. Significance was determined using a one-tailed T-test and a threshold cutoff of $p<0.05$. To compare the significance of association of the CFA12 *FGF4* insertion to the most highly associated SNP and the CFA18 *FGF4* insertion, 34 of 36 IVDD cases and 31 controls were genotyped for both the CFA12 and CFA18 *FGF4* insertions. Chi square and odds ratio analysis were performed in Plink (1).

TSS Investigation

To ensure the CFA12 *FGF4* insertion included the TSS. PCR was performed using cDNA with primers at varying positions 5' to the *FGF4* start codon (Table S8). cDNA was synthesized from RNA extracted from neonatal intervertebral disc (IVD) and vertebral body (VB), as described below, from a Beagle.

qRT-PCR

cDNA was prepared from RNA extracted from IVD or VB dissected from neonatal canine tail samples, skeletal muscle, and testis, using the QuantiTect Reverse Transcription Kit and the RNeasy Fibrous Tissue Mini Kit (Qiagen, Valencia, CA, USA), respectively. Additionally, cDNA was made from commercially available Beagle skeletal muscle and testis RNA (Zyagen, San Diego, CA, USA). Quantitative RT-PCR was performed for *FGF4* using cDNA synthesized from 500ng of RNA extracted from IVD and VB. Samples were run in triplicate with 20ng of cDNA. A technical replicate was removed from analysis if the standard deviation of the 3 technical replicates was greater than 1, and a sample was removed from analysis if there were less than 2 technical replicates that met this criteria.

Semi-quantitative RT-PCR was performed for genes within and near the critical interval, including *COL9A1*, *SMAP1*, *B3GAT2*, *OGFRL1*, *LINC00472*, *RIMS1*, *KCNQ5*, and *COL12A1*, as well as *FGF4,* for VB, IVD, skeletal muscle, and testis cDNA from a case and control (SI Methods). *RPS5* was a housekeeping gene control. All case samples were collected from a Beagle, while control VB and IVD were from a Cane Corso and skeletal muscle and testis from a Labrador Retriever. Primers were designed using Primer3, except for *RPS5* in which the primers were as recommended by Brinkhof et al. (8, 9). Quantitative RT-PCR was performed for *FGF4* using cDNA synthesized from IVD and VB RNA of 4 cases (4 Beagle) and 5 controls (1 Rottweiler and 4 Cane Corso) using the Rotor-Gene SYBR Green PCR Kit (Qiagen, Valencia, CA, USA) on the Rotor Gene Q real-time PCR system. *FGF4* transcript levels were normalized to *RPS5* and fold change in expression was calculated by taking $2^{-(\Delta\Delta CT)}$ with statistical significance assessed using a Mann-Whitney-Wilcoxin test (additional criteria in SI Methods).

References

1. Purcell s*, et al.* (2007) PLINK, a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* 81.
2. Kierczak M*, et al.* (2015) cgmisc: Enhanced Genome-wide Association Analyses and Visualisation. *Bioinformatics*:btv426.
3. Wickham H (2009) ggplot2: elegant graphics for data analysis. *Springer New York* 1(2):3.
4. Bolger AM, Lohse M, & Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114-2120.
5. Hoeppner MP*, et al.* (2014) An improved canine genome and a comprehensive catalogue of coding genes and non-coding transcripts. *PloS one* 9(3):e91172.
6. Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754-1760.
7. McKenna A*, et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297-1303.
8. Rozen S & Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 132:365-386.
9. Brinkhof B, Spee B, Rothuizen J, & Penning LC (2006) Development and evaluation of canine reference genes for accurate quantification of gene expression. *Anal Biochem* 356(1):36-43.

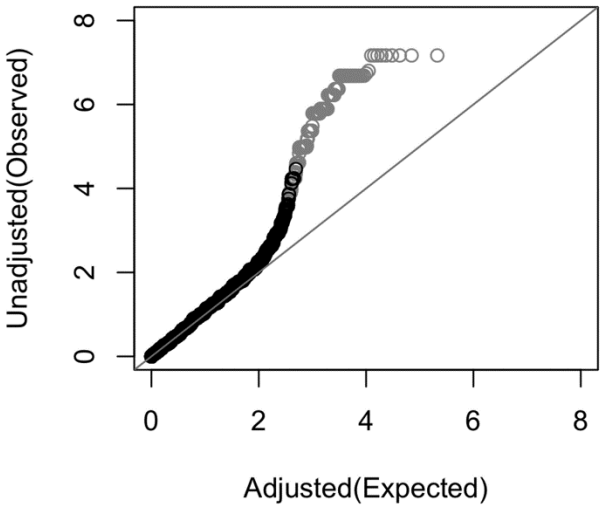

Fig. S1. Quantile-quantile (QQ) plot shows -$\log_{10}$ of the expected versus observed p-values plotted for each SNP, with the SNPs on CFA12 colored in grey.

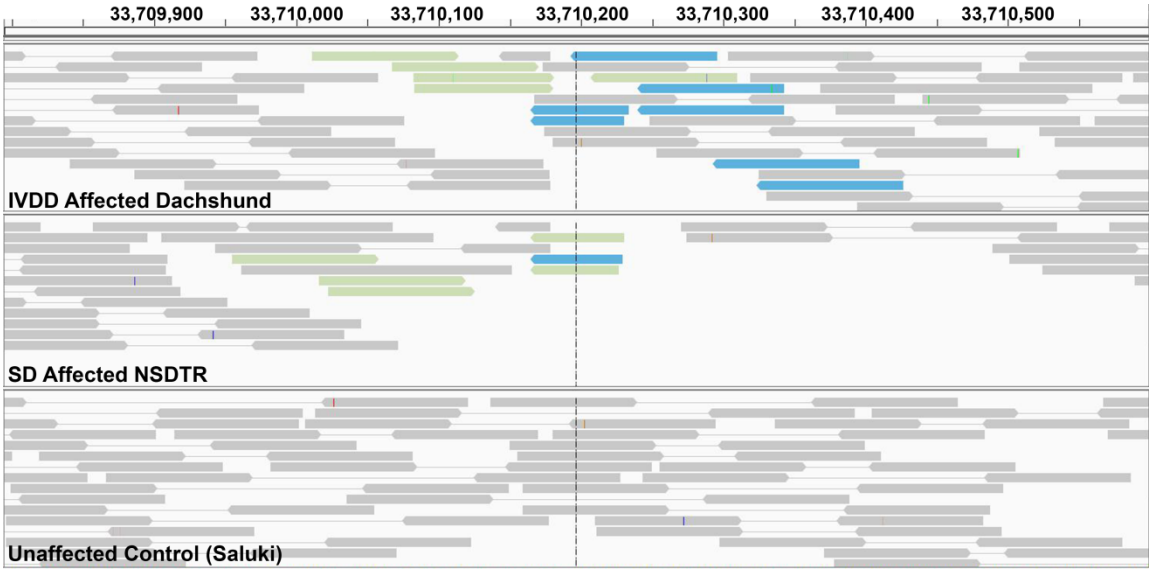

Fig. S2. Large insertion identified on CFA12: Screenshot of the Integrative Genomics Viewer (IGV-Broad Institute) showing an insertion at approximately 12:33,710,200 (canFam3) in an IVDD affected Dachshund case and a SD affected NSDTR that is not present in the Saluki unaffected control. Read mates in green map to chr18:48.4 Mb (canFam3) and the read mates in blue map to chr7:68.3 Mb (canFam3).

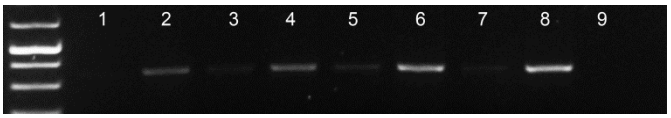

Fig. S3. Semi-quantitative RT-PCR for FGF4 across tissues in a case and control. Lane order: Ladder; 1) Control VB (Cane Corso); 2) Case VB (Beagle); 3) Control IVD (Cane Corso); 4) Case IVD (Beagle); 5) Control skeletal muscle (Labrador retriever); 6) Case skeletal muscle (Beagle); 7) Control testis (Labrador retriever); 8) Case testis (Beagle); 9) Negative control.

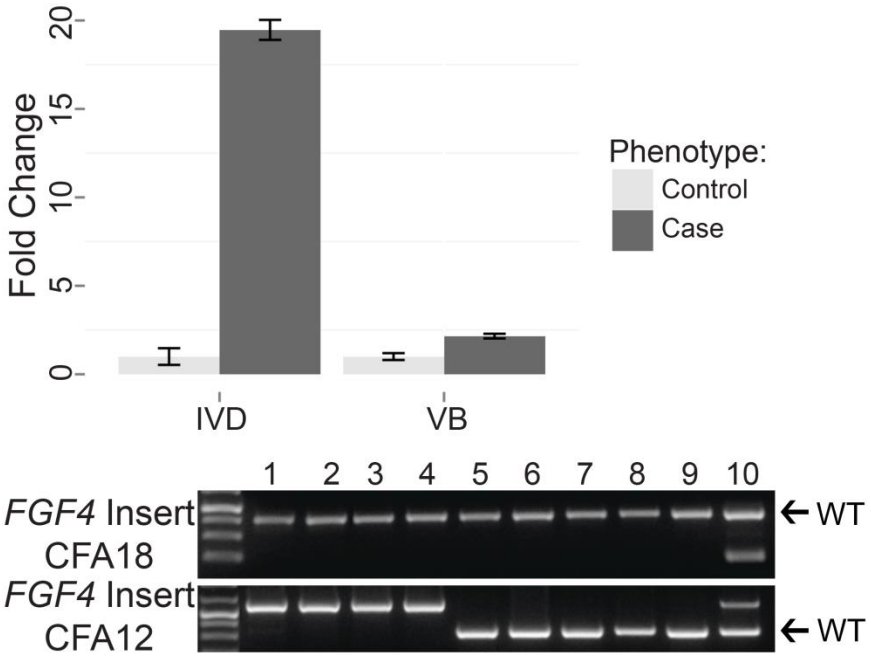

Fig. S4. *FGF4* expression: Bar graph depicting fold change differences in *FGF4* expression between controls and IVDD cases in neonatal IVD and VB. *FGF4* expression was 19.47x higher (p=0.03) in IVD and 2.16x higher (p=0.03) in VB of cases compared to controls. Error bars representative of standard error of measurement for each group. Gels depict genotypes of 4 cases (Beagles) and 5 controls (1 Rottweiler and 4 Cane Corso) used in qRT-PCR analysis. The five controls were wild type, meaning they lacked the *FGF4* insert at both the CFA12 and CFA18 locations; however, the cases, while wild type for the CFA18 *FGF4* insert, were homozygous mutant at the CFA12 locus. Lanes: 1-4: Beagle cases; 5-9: Cane Corso controls; 10: heterozygous control; 11: negative control.

Indel Coordinates (canFam3)	**Indel**	**Method of Elimination**
Chr12:33,927,660-33,928,003	Deletion 1	PCR
Chr12:34,256,430-34,256,530	Deletion 2	PCR
Chr12:34,467,000	Insertion 1	Additional Genomes
Chr12:34,734,000	Insertion 2	Additional Genomes
Chr12:34,758,000	Insertion 3	Additional Genomes
Chr12:34,947,000	Insertion 4	Additional Genomes
Chr12:35,228,600-35,228,800	Deletion 3	PCR
Chr12:35:498,000	Insertion 5	PCR

Table S1. Large indels identified via BAM file investigation: Coordinates and method of elimination for each of the 8 segregating large indels identified in 2 cases (1 Dachshund and 1 SD NSDTR) and 2 controls (1 NSDTR and 1 Saluki). Insertions 1-4 were eliminated based on lack of segregation with investigation of additional control genome BAM files. Deletions 1-3 and Insert 5 were eliminated based on lack of segregation demonstrated via PCR of additional cases and controls.

Breed of Dog Genotyped	Case or Control	CFA18 *FGF4* Insert Genotype			CFA12 *FGF4* Insert Genotype		
		WT	Heterozygous	Mutant	WT	Heterozygous	Mutant
Basset Hound	Case (n=1)	0	0	1	0	0	1
Beagle	Case (n=1)	1	0	0	0	0	1
Cardigan Welsh Corgi	Case (n=1)	0	1	0	0	1	0
Chihuahua	Case (n=3)	1	2	0	0	3	0
Coton de Tulear	Case (n=6)	0	2	4	0	1	5
Dachshund	Case (n=3)	0	0	3	0	0	3
French Bulldog	Case (n=3)	3	0	0	0	0	3
Maltese	Case (n=1)	0	0	1	0	0	1
Mix	Case (n=10)	1	4	5	0	6	4
NSDTR	Case (n=1)	1	0	0	0	1	0
Pembroke Welsh Corgi	Case (n=1)	0	0	1	0	0	1
Miniature Poodle	Case (n=1)	0	1	0	0	0	1
Rottweiler	Case (n=1)	1	0	0	1	0	0
Shih Tzu	Case (n=1)	0	0	1	0	1	0
Boston Terrier	Control (n=3)	3	0	0	3	0	0
Brittany	Control (n=3)	3	0	0	3	0	0
Bulldog	Control (n=1)	1	0	0	1	0	0
Collie	Control (n=2)	2	0	0	2	0	0
Coton de Tulear	Control (n=6)	0	0	6	2	4	0
Dalmatian	Control (n=1)	1	0	0	1	0	0
Ibizan Hound	Control (n=2)	2	0	0	2	0	0
Jack Russel	Control	1	0	0	1	0	0

Table S2. Dogs used for the IVDD GWAS were genotyped for both *FGF4* retrogene insertions

Breed	**Homozygous wild type**	**Heterozygous**	**Homozygous mutant**
Bichon Frise	0	1	2
Chihuahua	0	2	0
Dachshund	0	0	17
Dandie Dinmont Terrier	0	1	0
Mix	0	6	4
Nova Scotia Duck Tolling Retriever	0	5	2

Table S3. CFA12 *FGF4* insert genotyping results for an additional 40 IVDD cases.

Breed	**Fig 5**	**CFA18 *FGF4* Insert Genotype**			**CFA12 *FGF4* Insert Genotype**		
		WT	**Heterozygous**	**Mutant**	**WT**	**Heterozygous**	**Mutant**
American Cocker Spaniel	20	8	0	0	0	1	9
Australian Cattle Dog	24	1	0	0	10	0	0
Australian Shepherd	25				10	0	0
Basset Hound	17	0	0	15	1	5	5
Beagle	19	8	0	0	0	1	17
Bernese Mountain Dog	36				10	0	0
Boston Terrier					3	0	0
Brittany	22				14	0	0
Bulldog	21	3	0	0	11	0	0
Cairn Terrier	11	0	0	7	9	0	0
Cane Corso					5	0	0
Cardigan Welsh Corgi	13	0	0	13	1	2	5
Cavalier King Charles Spaniel	12	4	0	0	0	0	9
Chesapeake Bay Retriever	33				29	7	0
Chihuahua	3	0	2	1	5	6	2
Collie		4	0	0	2	0	0
Coton de Tulear					2	5	5
Dachshund	2	0	0	14	0	1	27
Dalmatian					1	0	0
Doberman Pinscher	35	7	0	0	15	0	0
English Springer Spaniel	24				7	2	0
Fox Terrier					10	0	0
French Bulldog	9	4	0	0	0	4	28
German Shepherd Dog	32	10	0	0	10	0	0
Golden Retriever	28	4	0	0	10	0	0
Great Dane	38				10	0	0
Ibizan Hound		8	0	0	3	0	0

Table S4. Genotypes of CFA12 FGF4 retrogene from this manuscript and data for breeds genotyped for the CFA18 *FGF4* retrogene insertion by Parker et al. are included for comparison.

Breed	**VMTH Pop. %**	**IVDD %**	**Chi square**	**Significance**	**Increase or Decrease**	**CFA12 *FGF4* Insert Allele Frequency**
American Cocker Spaniel	2.3	2.99	9.15	0.003	Increase	0.95
Basset Hound	0.51	1.24	45.21	$1.77x10^{-11}$	Increase	0.68
Beagle	0.86	2.42	120.69	$4.47x10^{-28}$	Increase	0.97
Corgi	0.67	1.82	85.41	2.43x10-^{20}	Increase	0.82
Dachshund	2.68	25.95	8874.25	<0.000001	Increase	0.98
French Bulldog	0.39	1.89	248.09	$6.77x10^{-56}$	Increase	0.94
Pekingese	0.35	1.6	189.72	$3.66x10^{-43}$	Increase	0.44
Brittany	0.49	0.24	5.34	0.021	Decrease	0.00
Bulldog	0.97	0.07	35.83	$2.16x10^{-9}$	Decrease	0.00
Cairn Terrier	0.23	0.07	4.46	0.035	Decrease	0.00
Scottish Terrier	0.31	0.1	6.18	0.013	Decrease	0.05
Shetland Sheepdog	0.93	0.19	25.29	$4.95x10^{-7}$	Decrease	0.00
Springer Spaniel	0.96	0.62	5.11	0.024	Decrease	0.11
West Highland White Terrier	0.53	0.12	13.72	0.0002	Decrease	0.00
Yorkshire Terrier	1.56	0.96	10.13	0.002	Decrease	0.00

Table S5. Investigation of IVDD in breeds seen at the UC Davis School of Veterinary Medicine Teaching Hospital: Canine cases seen at the UC Davis School of Veterinary Medicine Teaching Hospital between 1980 and 2016 were queried for a clinical diagnosis of “disc/k disease” or “IVDD.” 203,958 cases were seen, of which 4,177 were diagnosed with “disc/k disease” or “IVDD.” The breeds shown have a p-value (based on Chi square test and significance threshold of $p<0.05$) associated with an increase or decrease in incidence of IVDD. Allele frequencies calculated from on dogs genotyped in Table S4.

Breed	**# of Dogs**	**Phenotype**
Basset Hound	1	Case
Beagle	1	Case
Boston Terrier	3	Control
Brittany	3	Control
Bulldog	1	Control
Cardigan Welsh Corgi	1	Case
Chihuahua	3	Case
Collie	2	Control
Coton de Tulear	12	6 Cases, 6 Controls
Dachshund	4	Case
Dalmatian	1	Control
French Bulldog	3	Case
Ibizan Hound	2	Control
Jack Russell Terrier	1	Control
Lacy Dog	1	Control
Maltese	1	Case
Miniature Poodle	1	Case
Mix	15	11 Cases, 4 Controls
NSDTR	1	Case
Pembroke Welsh Corgi	1	Case
Poodle	1	Control
Rottweiler	1	Case
Shetland Sheepdog	2	Control
Shih Tzu	1	Case
Whippet	3	Control
Yorkshire Terrier	1	Control

Table S6. Number of dogs per breed used in across breed IVDD GWAS.

	Forward Primer (5'→3')	**Reverse Primer (5'→3')**
CFA7 Insert	CTCTGTGGACCTCTTTCAACG	TGACACCAGTGAGAATTGCAT
Deletion 1	TGCTTGCTCCAGCTCTGTTA	TTGGCCATAATTTTCCTTGG
Deletion 2	AAATGGCATATGGGCTGAGT	TCTGCAAAACAGCTTGCATT
Deletion 3	CACTGTTGGCAGTCCTCAAA	AAAGCCGGTTGTTGATGAAG
Insert 5	ATGCTACACCACTCCCTGCT	ATCCTTGCCAAAACTGATGG

Table S7. Primers used to large indels identified CFA12:33.1-35.5Mb (canFam3).

	Forward Primer (5'→3')	**Reverse Primer (5'→3')**
CFA12 *FGF4* Insertion	ACAGCTGGCATGGTCAGTTA	CCTGATTTTGAGACAGCCAAA
CFA18 *FGF4* Insertion	TTGGGAATGTCAAACCACTG	AGGGCCAAGTGTCCAATACA
*FGF4.*F1	GTGTTTGCATGGAGGAAGGT	
*FGF4.*F2	CTGAGCAAGAACGGGAAGAC	
*FGF4.*F3	AGCCTGATGGCTGGACTGTA	
*FGF4.*F4	GTCCGTGCGGTGAAATAAAA	
*FGF4.*R1		TTGATGCCCAGGAGGTAGTC
*FGF4.*R2		TGAGTGGGTTAAGGGTTTCG
*FGF4*_TSS.F1	GCGCTCCGCACCGAGTCC	
*FGF4*_TSS.F2	CTCCATGCAGCCCGGGTA	

Table S8. Primers used to sequence and assay the *FGF4* insert on CFA12 and 18.

	External Forward Primer (5'→3')	**Internal Forward Primer (5'→3')**	**External Reverse Primer (5'→3')**
CFA12 *FGF4* Insert Genotyping Assay	ACAGCTGGCATGGTCAGTTA	GTCCGTGCGGTGAAATAAAA	TGCTGTAGATTTTGAGGTGTCTT
CFA18 *FGF4* Insert Genotyping Assay	TTGGGAATGTCAAACCACTG	GTCCGTGCGGTGAAATAAAA	GTTCCCTCCATTTCGGTTT

Table S9. Primers for *FGF4* insert genotyping assays.

	Forward Primer (5'→3')	**Reverse Primer (5'→3')**
COL9A1	TTCTGTCGGACCAAGAGGAC	CCATCATGAAAGCCAATGGT
SMAP1	AGGCTCAGAAGCTGAACGAG	TCTGGTGTCCATTGGTCTAGG
B3GAT2	CCTACAGCCTGGAGCTGTTC	GGCTTTTGGATTGGACAAGA
OGFRL1	AAGCAACTGCCAAACCAAAG	GTTCTCTCAGGGGGAAAAGC
LINC00472	GGGCTGTACTGGCTCATTGT	AGAGCAGCACACCCAAGTCT
RIMS1	TGGCCATCTCTGCTCCTACT	ACCTCAGAACCAGCACCTGT
KCNQ5	TTGTGGAAAAGGATGCCAAT	GGCGGTGCTGTTCTTGTACT
COL12A1	CTACAGGGGACGACAGAAGG	CTGCTTCTGCTCTGGTGAGA
FGF4	GACTACCTCCTGGGCATCAA	GTCTTCCCGTTCTTGCTCAG
RPS5	TCACTGGTGAGAACCCCCT	CCTGATTCACACGGCGTAG

Table S10. Primers used for semi-qPCR and qRT-PCR experiments.