**Derived Variants at Six Genes Explain Nearly Half of Size Reduction in Dog Breeds**

Maud Rimbault[1]*, Holly C. Beale[1]*, Jeffrey J. Schoenebeck[1], Barbara C. Hoopes[2], Jeremy J. Allen[3], Paul Kilroy-Glynn[4], Robert K. Wayne[5], Nathan B. Sutter[3], Elaine A. Ostrander[1#]

[1.] Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. [2.] Department of Biology, Colgate University, Hamilton, New York 13346, USA. [3.] Department of Clinical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, New York 14853, USA. [4.] School of Biotechnology, Dublin City University, Dublin, Ireland. [5.] Department of Ecology and Evolutionary Biology, University of California at Los Angeles, Los Angeles, California 90095, USA.

* These authors contributed equally to this work.

[#] Corresponding author: eostrand@mail.nih.gov

## SUPPLEMENTAL DATA

**Supplemental Results** **2**

**Supplemental Methods** **7**

**Supplemental Figure Legends (S1-S4)** **10**

**Supplemental Tables** **13**

Supplemental Tables S1 and S7 13

Supplemental Table Legends (S2-S6, S8-S9) 15

**References** **17**

## Supplemental Results

### Fine-mapping the size loci

The locus on canine chromosome 4 at 67 Mb coincided with the growth hormone receptor (*GHR*) gene (Fig. 1A). At this locus, as with other loci, variant discovery was conducted by sequencing intergenic regions, exons and proximal intronic sequences of candidate genes in sets of 12 dogs (split between breeds with a standard breed weight < 5.9 kg (13 lbs) and > 36.3 kg (80 lbs)). Among the variants found (Supplemental Tables 2 and 3), we identified five exonic SNPs in the *GHR* gene: two nonsynonymous SNPs in the fifth exon (CFA4:67,040,898 G/A and CFA4:67,040,939 C/T) (Table 1 and Fig. 1A) and three silent mutations in the ninth (last) exon at positions 67,023,946, 67,024,057 and 67,024,675. Genotyping a larger set of dogs (n = 315) at these two nonsynonymous SNPs validated their association with body size ($p = 2.98 \times 10^{-66}$ and $1.52 \times 10^{-13}$ respectively) (Supplemental Table 3). The nonsynonymous SNP on CFA4 at 67,040,898 changes a glutamic acid codon (GAG) at amino acid 191 to lysine (AAG) (E191K). The second nonsynonymous SNP at 67,040,939 changes a proline (CCA) at amino acid 177 to leucine (CTA) (P177L). Subsequently, we refer to these SNPs as *GHR*(1) and *GHR*(2) respectively. Cross-species alignments revealed that the glutamic acid corresponding to *GHR*(1) is highly conserved in mammals and birds, and, the proline corresponding to *GHR*(2) is highly conserved in birds, fishes and in all mammals excluding primates, which have an alanine (Supplemental Fig. 1).

The association signal on CFA10 at eight Mb spanned *HMGA2* (Fig. 1B). Syntenic comparisons with human and mouse suggested that the first exon of *HMGA2* was located within a gap of the dog genome assembly. We sequenced across the gap and identified a region homologous to the first exon of human and mouse *HMGA2*. By sequencing mRNA, we

confirmed that the putative exon in the gap is expressed (Supplemental Methods). We next sequenced all the exons (including the 5' and 3' UTRs) and proximal intronic sequence of the *HMGA2* gene in sets of 12 dogs as previously described (Supplemental Table 3). Several variants were identified in the gene (Supplemental Table 2), but none were nonsynonymous. The most highly associated of these variants was a SNP in the 5' UTR of *HMGA2* (CFA10:8,348,804), ~730 bp before the start codon of the first exon (Table 1) ($p$ = 1.47 x $10^{-06}$, n = 26).

The locus on CFA4 at 39-41 Mb included two candidate genes: *STC2* and fibroblast growth factor 18 (*FGF18*) (Supplemental Fig. 2A). One hundred and fifty-two dogs were genotyped at 44 SNPs with a SNPlex assay, confirming and refining the association (Supplemental Table 4). The most highly associated SNP was located within *STC2* (Supplemental Fig. 2B) (CFA4:39,164,250, $p$ = 2.65 x $10^{-20}$, $n_{genotyped}$ = 146 of 152). Conditional analysis of the SNPlex1 and the CanMap data sets indicated two independent signals in *STC2-FGF18* locus (data not shown). Given its comparatively strong association signal in both the CanMap GWAS and SNPlex1 assay, we focused on the *STC2* locus. Marker discovery was conducted on sets of 12 dogs as described previously (Supplemental Table 3), and 39 SNPs were genotyped using DNA from 226 dogs in a second SNPlex assay (Supplemental Fig. 2C and Supplemental Table 4). The SNPlex data revealed three contiguous linkage disequilibrium (LD) blocks as defined by Haploview (Barrett et al. 2005). These blocks are formed by 14 markers that span *STC2* and a region of 28 kb downstream of its stop codon. The majority of dogs carried one of two haplotypes spanning the LD block that differed at only one position, 39,182,836 (dbSNP canine build 131: rs24111740) (Supplemental Fig. 2E). Ninety-two percent of the dogs carrying Haplotype 1 (120 of 131)

were from breeds with an SBW < 13.6 kg (30 lbs), while Haplotype 2 was evenly distributed between larger and smaller dogs. The defining SNP was located 20 kb downstream of the 3' end of the protein coding region of the *STC2* gene (Fig. 1C and Supplemental Fig. 2C-D) and was the most highly size-associated marker in this region ($p = 1.26 \times 10^{-33}$, $n_{genotyped} = 224$ of 226) (Table 1). No size-associated nonsynonymous SNPs were found in *STC2*.

The locus on CFA7 at 43 Mb was confirmed by genotyping 152 dogs at 15 SNPs with a SNPlex assay (CFA7: 43,857,216, $p = 3.54 \times 10^{-26}$, $n_{genotyped} = 151$ of 152) (Supplemental Table 4). In this region, the best candidate gene is *SMAD2* (Fig. 1D). The association interval was narrowed with the second SNPlex assay, including 38 SNPs genotyped on DNA from 211 dogs (Supplemental Table 4), and marker discovery on sets of 12 and 16 dogs, as previously described (Supplemental Table 3). The most highly associated marker was a 9.9 kb deletion located 24 kb downstream of the protein coding region of the gene (CFA7:43,794,129) ($p = 1.50 \times 10^{-40}$, $n_{genotyped} = 196$ of 211) (Fig. 1D and Table 1). A second 5.7 kb deletion located closer to *SMAD2*, 17 kb downstream of the protein coding region, was in complete LD with the 9.9 kb deletion, based on assessment of 80 dogs from 59 breeds. Unlike the 5.7 kb deletion, the 9.9 kb deletion includes a region conserved in human, mouse and rat. Within the 9.9 kb deletion, a 280 base sequence between CFA7:43,801,100-43,801,379 is highly conserved, with an average phastCons score of 0.91 (lod = 162). Based on its association, conservation, and the *in silico* prediction of a transcription factor binding site cluster (Supplemental Fig. 3), we selected the 9.9 kb deletion as our marker for further analysis.

**Model Terms and Interactions**

We tested the terms in the cSBW equation (Equation 1) for significance. The terms *STC2* and *GHR*(2) were excluded by forward and reverse AIC analysis (Akaike's Information Criterion; Venables and Ripley 2002) and were not significant by analysis of variance (ANOVA). When a dominance part of a marker's allele frequency was selected (as described in the Methods section of the main document), both the dominance part and the main term for that marker were considered selected. The loss of significance for *STC2* in the corrected data set is not a robust result, because we determined that *STC2* loses significance even in the uncorrected data set when the breeds are limited to the 65 that are in the cSBW data set. The loss of significance for *STC2* could be due to the reduced sample number or to a different distribution of the SBWs of the subset, although the means of the log SBW values of the two groups do not differ significantly (two sided t-test).

Equation 1: cSBW ~ *IGF1* + *GHR*(1) + *SMAD2* + *STC2* + *HMGA2* + *IGF1R* + *GHR*(2) + *HMGA2*_domPart + *GHR*(2)_domPart

In order to determine whether the data support the existence of inter-marker interactions, stepAIC analysis of Equation 1 was performed and all interactions were considered (AIC; Venables and Ripley 2002). The best-scoring model included four interactions and dropped the *STC2* term (Equation 2). The interaction between the main and partial dominance terms for *GHR*(2) was omitted from the subsequent analysis. The model with remaining interactions accounted for 54.1% of cSBW variance, 1.6% more than the model without interactions (52.5%). None of the remaining interactions were significant by ANOVA.

Equation 2: cSBW ~ *IGF1* + *GHR*(1) + *SMAD2* + *IGF1R* + *HMGA2* + *HMGA2*_domPart + *GHR*(2) + *GHR*(2)_domPart + *GHR*(2):*GHR*(2)_domPart + *IGF1*:*GHR*(1) + *SMAD2*:*GHR*(2)_domPart + *IGF1*:*HMGA2*_domPart

Cross-validation, which used a randomly selected 45 of the 65 breeds (69%) with PCs as a training set, was repeated 10,000 times. On average, the coefficients calculated by the training set could be used to account for 42.1%±0.2% (mean ± SEM) of the variance of the cSBW in the remaining 20 of the 65 breeds, which is 80.2% of the variance explained by the full 65 breed cSBW data set (52.5%).

## Supplemental Methods

*SNPlex assays*

Two custom SNPlex genotyping assays (Applied Biosystems) were designed to target SNPs surrounding the strongest association peaks at two loci, CFA4: 39-41 Mb and CFA7: 43 Mb. The first SNPlex assay included 59 SNPs and was performed on 75 dogs from 15 breeds with an SBW < 9.1 kg (20 lbs) and 77 dogs from 14 breeds with an SBW > 24.9 kg (55 lbs), following the manufacturer's protocol (Applied Biosystems) (Supplemental Table 4). Additional SNP genotyping was then performed with a second SNPlex assay for these two loci: for CFA4: 39-41 Mb, 226 dogs (112 dogs from 22 breeds with an SBW < 12.2 kg (27 lbs) and 114 dogs from 26 breeds with an SBW ≥ 18.1 kg (40 lbs)) were genotyped at 39 SNPs and for CFA7: 43 Mb, 211 dogs (108 dogs from 22 breeds with an SBW < 12.2 kg and 103 dogs from 25 breeds with an SBW ≥ 18.1 kg) were genotyped at 38 SNPs (Supplemental Table 4). Genotypes were called with GeneMapper software v4.0 (Applied Biosystems). SNPs with high failure rates (> 40% missing data) and non-polymorphic SNPs in our data set were removed. Individuals with > 45% missing genotypes were also excluded from further analyses. SNP positions are listed in Supplemental Table 2.

*Sanger Sequencing*

Marker discovery was performed on sets of 12 dogs (six dogs from breeds with an SBW < 5.9 kg (13 lbs) and six dogs from breeds with an SBW ≥ 36.3 kg (80 lbs)) or on a set of 16 dogs (six dogs from breeds with an SBW < 5.9 kg and ten from breeds with an SBW ≥ 36.3 kg). At least 50% of the small and 50% of the large dogs were successfully genotyped at all but seven markers. 125, 90, 506 and 145 amplicons were designed to fine-map the loci at CFA4: 39-41

Mb, CFA4: 67 Mb, CFA7: 43 Mb and CFA10: 8 Mb respectively. Primers were designed with Primer3 software (Rozen and Skaletsky 2000) using standard parameters. Primers were designed to amplify all exons with surrounding intronic sequence of all candidate genes. Additional primers were designed at regular intervals across the whole regions of interest. Primers are listed in Supplemental Table 5.

PCR amplification was performed in a final volume of 10 µl containing 10 ng of dog DNA, 0.25 U AmpliTaq Gold DNA polymerase (Applied Biosystems), 1x PCR Gold Buffer, 0.2 µM of each specific primer, 0.1 mM of each dNTP and 1.5 mM $MgCl_2$. Touchdown PCR was carried out as follows: initial denaturation at 94°C for 10 min, followed by 20 cycles of 94°C for 30 s, 65°C for 30 s with a touchdown process (-0.5°C per cycle) and 72°C for 1 min, 20 cycles of 94°C for 30 s, 55°C for 30 s, 72°C for 1 min, and a final extension phase at 72°C for 10 min. For regions with high GC content, amplification was performed using the KOD DNA polymerase (Novagen) with the following PCR conditions: initial denaturation at 94°C for 2 min, followed by 35 cycles of 98°C for 10 s and 68°C for 2 min, and a final extension phase at 72°C for 5 min. The resulting PCR products were cleaned using an exonuclease/shrimp alkaline phosphatase reaction (37°C for 30 min and then 80°C for 15 min) and sequenced using the BigDye V3.1 Cycle Sequencing Kit (Applied Biosystems), according to the manufacturer's instructions. Sequence data were collected on an ABI 3730xl capillary sequencer (Applied Biosystems). Traces were aligned and analyzed using PhredPhrap/Polyphred and Consed (Nickerson et al. 1997; Ewing et al. 1998; Ewing and Green 1998; Gordon et al. 1998). For the fine-mapping of 16 dogs at chr7: 43 Mb, variants were called using Polyphred 6.11 and the indel detector DIPDetector (N. Hansen,

unpublished). Only biallelic markers (SNPs, insertions and deletions) were considered for further analyses. Markers positions are listed in Supplemental Table 2.

*Filling gap in reference genome*

To close the gap and determine the integrity of *HMGA2*, we generated a PCR product from primers on either side of the gap. This ~2 kb product was then further amplified and sequenced with internal primers (Internal primers in Supplemental Table 5). The spanning PCR product included ~950 bases of unknown sequence, which fills the gap, and replaces 500 bases of adjacent sequence. It contains sequence that is homologous to the first exon of human and mouse *HMGA2*. This sequence has been submitted to National Center for Biotechnology Information (NCBI) GenBank under the accession number KC529659. We experimentally defined the 5' and 3' end of *HMGA2* by RACE-PCR (SMARTer RACE Kit, Clontech-Takara), using total RNA isolated from Whippet testis tissue. The mRNA sequence of *HMGA2*, including the sequence of the exon 1 of the canine *HMGA2*, has been submitted to NCBI GenBank under the accession number KC529658.

*Association studies and haplotype analysis*

The quantitative trait association analyses were carried out using the Wald test in PLINK (Purcell et al. 2007). LD blocks were defined by Haploview 4.2 software (solid spine of LD, D' > 0.8) (Barrett et al. 2005). Haplotypes were inferred using the PHASE software with the default settings (Stephens et al. 2001; Stephens and Scheet 2005).

## Supplemental Figure Legends

**Supplemental Figure S1.** Amino acid alignment of the sequence of the exon 5 of the canine *GHR* gene and the orthologous exon in other species *(A)*. The amino acids at positions 177 and 191 are highlighted. In *(B)*, the heights of the letters illustrate the sequence conservation at each position across species (Crooks et al. 2004).

**Supplemental Figure S2.** Fine-mapping of size locus on CFA4 near *STC2*.
The strongest association was observed in the left peak in the CanMap association analysis (A). In the present study, 75 dogs from 15 breeds with a standard breed weight (SBW) < 9.1 kg (20 lbs) and 77 dogs from 14 breeds with an SBW > 24.9 kg (55 lbs) were genotyped at 44 SNPs in the region by SNPlex. The left peak remained stronger (B). The left peak was further surveyed in 226 dogs representing the full range of canine body weight (112 dogs from 22 breeds with an SBW < 12.2 kg (27 lbs) and 114 dogs from 26 breeds with an SBW ≥ 18.1 kg (40 lbs)) at 39 SNPs by SNPlex (C) and in sets of 12 dogs by capillary sequencing (D). The most highly associated SNP was identified (indicated by the asterisk in C and E). Haplotype analysis was performed on the markers depicted in red in (C; SNPs with a MAF > 0.05). Haplotypes with a frequency greater than 3% are depicted in (E) (total chromosomes = 452; 3% corresponds to 14 or more chromosomes). The top haplotype is the only one containing the derived allele at the most highly size-associated SNP. This haplotype was observed on 131 chromosomes, 92% of them in breeds with an SBW < 13.6 kg (30 lbs).

**Supplemental Figure S3.** *In silico* analysis of the *SMAD2* deletion variant predicts loss of a highly conserved ECR and transcription factor binding cluster among small breed carriers.

(A) Plot of the Phastcons4way scores across the *SMAD2* deletion variants for multiple alignments of the following assemblies to the dog genome: human (hg17), mouse (mm6) and rat (rn3) (Siepel et al. 2005). All scores were obtained from the UCSC Genome Browser's Phastcons4way track using CanFam2 (May 2005) coordinates chr7:46,763,901-46,781,570. Scored positions were lifted over to CanFam3.1 (Sep. 2011) prior to plotting. Values closer to 1 indicate that the base is more highly conserved across species. The positions of the two deletions are represented by red rectangles above the plot. Conserved elements identified by phastCons are displayed under the plot (aqua rectangles) with their respective logarithm of the odds (LOD) scores (phastConsElements4way track).

(B) Conservation profiles of the *SMAD2* deletion region in dog, cat, horse, cow, human, marmoset, mouse and opossum genomes as generated by Mulan (http://mulan.dcode.org/). Evolutionary Conserved Regions (ECRs) are displayed as dark red blocks on top of each species conservation layer.

(C) The upper half of the panel depicts all transcription factor binding sites, represented by tickmarks, detected with the multiTF utility. The lower half of the panel indicates that a cluster of multispecies conserved transcription factor binding sites is located within the 280 bp conserved element (lod=162) that is removed in Deletion 2 (A).

**Supplemental Figure S4.** Comparisons of mean log body weight suggest modes of inheritance.

The Standard Breed Weight (SBW) of each dog, transformed by natural log, is plotted by genotype at each marker, and the quantile boundaries and median are indicated by the superimposed boxplots. The mean log SBW of each group is represented by a black square.

The red line connecting the SBWs of D/D and A/A dogs permits comparison of the heterozygote SBW. The mean SBW of A/D dogs at the *GHR*(1) marker is on the red line, and thus equidistant from the D/D and A/A means, suggesting an additive mode of inheritance. At the *IGF1*, *SMAD2* and *STC2* markers, the mean SBWs of dogs heterozygous for one marker are slightly higher than midway between the SBWs of the homozygotes for the same marker; for the *IGF1R* marker, on the other hand, the mean heterozygote SBW is slightly lower than midway between the SBWs of the homozygotes. The SBWs of heterozygotes at the *HMGA2* and *GHR*(2) markers dramatically depart from the midline, and those two loci are the only that have a statistically significant dominance component.

## Supplemental Tables

**Supplemental Table S1.** Dog breeds and weights used in this study.

SBWs were obtained from several sources: "The complete dog book" (AKC), "The Encyclopedia of the Dog" (Encyclopedia), "Atlas of Dog Breeds of the World" (Atlas) or owner-reported weights from our NHGRI database (Database). See Methods and References for more details.

| Breed abbreviation | Breed name | Standard breed weight (kg) | Standard breed weight (lbs) | ln(Standard breed weight (lbs)) | Source |
|---|---|---|---|---|---|
| AFFN | Affenpinscher | 3.4 | 7.5 | 2.015 | Encyclopedia |
| AFGH | Afghan Hound | 24.9 | 55 | 4.007 | AKC |
| AKBA | Akbash Dog | 47.9 | 105.5 | 4.659 | Encyclopedia |
| AKIT | Akita | 42.0 | 92.5 | 4.527 | Encyclopedia |
| AMAL | Alaskan Malamute | 36.3 | 80 | 4.382 | AKC |
| AMST | American Staffordshire Terrier | 29.2 | 64.3 | 4.164 | Database |
| ANAT | Anatolian Shepherd Dog | 52.2 | 115 | 4.745 | AKC |
| AUST | Australian Terrier | 8.4 | 18.5 | 2.918 | Database |
| BASS | Basset Hound | 22.7 | 50 | 3.912 | Encyclopedia |
| BEAG | Beagle | 10.9 | 24 | 3.178 | Encyclopedia |
| BEAU | Beauceron | 34.2 | 75.5 | 4.324 | Encyclopedia |
| BEDT | Bedlington Terrier | 9.1 | 20 | 2.996 | AKC |
| BICH | Bichon Frise | 4.3 | 9.5 | 2.251 | Encyclopedia |
| BLDH | Bloodhound | 43.1 | 95 | 4.554 | AKC |
| BMD- | Bernese Mountain Dog | 40.1 | 88.5 | 4.483 | Encyclopedia |
| BORD | Border Collie | 17.9 | 39.5 | 3.676 | Encyclopedia |
| BORT | Border Terrier | 6.1 | 13.5 | 2.603 | AKC |
| BORZ | Borzoi | 37.4 | 82.5 | 4.413 | AKC |
| BOST | Boston Terrier | 7.9 | 17.5 | 2.862 | Encyclopedia |
| BOUV | Bouvier des Flandres | 33.6 | 74 | 4.304 | Encyclopedia |
| BOX- | Boxer | 28.3 | 62.5 | 4.135 | Encyclopedia |
| BRIA | Briard | 34.0 | 75 | 4.318 | Encyclopedia |
| BRIT | Brittany | 15.9 | 35 | 3.555 | AKC |
| BRTR | Black Russian Terrier | 52.4 | 115.5 | 4.749 | Encyclopedia |
| BRUS | Brussels Griffon | 4.1 | 9 | 2.197 | AKC |
| BSJI | Basenji | 10.2 | 22.5 | 3.114 | Encyclopedia |
| BULD | Bulldog | 25.8 | 56.9 | 4.041 | Database |
| BULM | Bullmastiff | 52.2 | 115 | 4.745 | AKC |
| BULT | Bull Terrier | 27.2 | 60 | 4.094 | AKC |
| CAIR | Cairn Terrier | 6.1 | 13.5 | 2.603 | AKC |
| CARD | Cardigan Welsh Corgi | 14.4 | 31.8 | 3.460 | AKC |
| CHBR | Chesapeake Bay Retriever | 30.6 | 67.5 | 4.212 | AKC |

| | | | | | |
|---|---|---|---|---|---|
| CHIH | Chihuahua | 1.8 | 4 | 1.386 | Encyclopedia |
| CHIN | Japanese Chin | 3.4 | 7.5 | 2.015 | Encyclopedia |
| CHOW | Chow Chow | 26.1 | 57.5 | 4.052 | Encyclopedia |
| CKCS | Cavalier King Charles Spaniel | 7.0 | 15.5 | 2.741 | AKC |
| CNCS | Cane Corso | 49.9 | 110 | 4.701 | Encyclopedia |
| COLL | Collie | 28.3 | 62.5 | 4.135 | AKC |
| CRES | Chinese Crested | 3.9 | 8.5 | 2.140 | Encyclopedia |
| DACH | Dachshund | 8.6 | 18.9 | 2.939 | Database |
| DALM | Dalmatian | 23.8 | 52.5 | 3.961 | Encyclopedia |
| DANE | Great Dane | 61.7 | 136 | 4.913 | Database |
| DDBX | Dogue de Bordeaux | 40.8 | 90 | 4.500 | Encyclopedia |
| DOBP | Doberman Pinscher | 34.9 | 77 | 4.344 | Encyclopedia |
| ECKR | English Cocker Spaniel | 13.6 | 30 | 3.401 | AKC |
| ESSP | English Springer Spaniel | 20.4 | 45 | 3.807 | AKC |
| FBUL | French Bulldog | 11.3 | 25 | 3.219 | Encyclopedia |
| FCR- | Flat-Coated Retriever | 31.8 | 70 | 4.249 | Encyclopedia |
| GLEN | Glen of Imaal Terrier | 15.9 | 35 | 3.555 | AKC |
| GOLD | Golden Retriever | 29.5 | 65 | 4.174 | AKC |
| GPYR | Great Pyrenees | 42.0 | 92.5 | 4.527 | AKC |
| GREY | Greyhound | 29.5 | 65 | 4.174 | AKC |
| GSD- | German Shepherd Dog | 38.6 | 85 | 4.443 | Encyclopedia |
| GSHP | German Shorthaired Pointer | 26.1 | 57.5 | 4.052 | AKC |
| GSMD | Greater Swiss Mountain Dog | 60.1 | 132.5 | 4.887 | Encyclopedia |
| GSNZ | Giant Schnauzer | 33.3 | 73.5 | 4.297 | Encyclopedia |
| HAVA | Havanese | 5.5 | 12.1 | 2.493 | Database |
| HUSK | Siberian Husky | 21.5 | 47.5 | 3.861 | AKC |
| IBIZ | Ibizan Hound | 21.5 | 47.5 | 3.861 | AKC |
| ITGY | Italian Greyhound | 5.1 | 11.3 | 2.425 | Database |
| IWOF | Irish Wolfhound | 66.6 | 146.9 | 4.990 | Database |
| IWSP | Irish Water Spaniel | 25.6 | 56.5 | 4.034 | AKC |
| KEES | Keeshond | 27.4 | 60.5 | 4.103 | Encyclopedia |
| KOMO | Komondor | 40.8 | 90 | 4.500 | AKC |
| KUVZ | Kuvasz | 42.5 | 93.8 | 4.541 | AKC |
| LAB- | Labrador Retriever | 30.6 | 67.5 | 4.212 | AKC |
| LEON | Leonberger | 42.0 | 92.5 | 4.527 | Encyclopedia |
| MALT | Maltese | 2.3 | 5 | 1.609 | AKC |
| MANT | Manchester Terrier (Standard) | 7.7 | 17 | 2.833 | AKC |
| MAST | Mastiff | 82.8 | 182.5 | 5.207 | Encyclopedia |
| MBLT | Miniature Bull Terrier | 12.9 | 28.5 | 3.350 | Encyclopedia |
| MNTY | Manchester Terrier (Toy) | 4.3 | 9.5 | 2.251 | Atlas |
| MPIN | Miniature Pinscher | 4.1 | 9 | 2.197 | Encyclopedia |
| MPOO | Poodle (Miniature) | 7.7 | 17 | 2.833 | Database |
| MSNZ | Miniature Schnauzer | 6.4 | 14 | 2.639 | Encyclopedia |
| NEAP | Neapolitan Mastiff | 59.0 | 130 | 4.868 | AKC |
| NEWF | Newfoundland | 56.7 | 125 | 4.828 | AKC |
| NOWT | Norwich Terrier | 5.4 | 12 | 2.485 | AKC |
| OES- | Old English Sheepdog | 29.9 | 66 | 4.190 | Encyclopedia |
| PAPI | Papillon | 2.8 | 6.1 | 1.808 | Database |
| PBGV | Petit Basset Griffon Vendeen | 16.1 | 35.5 | 3.570 | Encyclopedia |
| PEKE | Pekingese | 4.3 | 9.5 | 2.251 | Encyclopedia |
| PEMB | Pembroke Welsh Corgi | 11.8 | 26 | 3.258 | AKC |

| | | | | | |
|---|---|---|---|---|---|
| POM- | Pomeranian | 2.3 | 5 | 1.609 | AKC |
| PUG- | Pug | 7.3 | 16 | 2.773 | AKC |
| ROTT | Rottweiler | 45.4 | 100 | 4.605 | Encyclopedia |
| SALU | Saluki | 19.5 | 43 | 3.761 | Encyclopedia |
| SAMO | Samoyed | 26.3 | 58 | 4.060 | Encyclopedia |
| SCOT | Scottish Terrier | 9.1 | 20 | 2.996 | AKC |
| SHAR | Chinese Shar-Pei | 23.8 | 52.5 | 3.961 | AKC |
| SHIH | Shih Tzu | 5.7 | 12.5 | 2.526 | AKC |
| SILK | Silky Terrier | 4.5 | 10 | 2.303 | AKC |
| SPOO | Poodle (Standard) | 26.1 | 57.5 | 4.052 | Encyclopedia |
| SSHP | Shetland Sheepdog | 10.0 | 22.1 | 3.096 | Database |
| SSNZ | Standard Schnauzer | 15.0 | 33 | 3.497 | Encyclopedia |
| STAF | Staffordshire Bull Terrier | 14.1 | 31 | 3.434 | AKC |
| STBD | Saint Bernard | 70.3 | 155 | 5.043 | AKC |
| SUSX | Sussex Spaniel | 18.1 | 40 | 3.689 | AKC |
| TIBM | Tibetan Mastiff | 72.6 | 160 | 5.075 | Encyclopedia |
| TIBS | Tibetan Spaniel | 5.4 | 12 | 2.485 | AKC |
| TPOO | Poodle (Toy) | 2.8 | 6.1 | 1.808 | Database |
| TURV | Belgian Tervuren | 28.1 | 62 | 4.127 | Encyclopedia |
| TYFX | Toy Fox Terrier | 2.4 | 5.3 | 1.668 | Atlas |
| WHIP | Whippet | 12.9 | 28.5 | 3.350 | Encyclopedia |
| WHWT | West Highland White Terrier | 8.4 | 18.5 | 2.918 | Encyclopedia |
| YORK | Yorkshire Terrier | 2.5 | 5.5 | 1.705 | AKC |

**Supplemental Table S7.** Ancestral allele frequency in wild canids.

| | n | *IGF1* | *GHR*(1) | *SMAD2* | *STC2* | *HMGA2* | *IGF1R* | *GHR*(2) |
|---|---|---|---|---|---|---|---|---|
| Coyote | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Red wolf | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Gray wolf | 26 | 0.98 | 1 | 0.92 | 0.98 | 1 | 1 | 1 |

**Supplemental Table Legends** (Tables available at http://genome.cshlp.org/)

**Table S2.** SNPs, insertions and deletions identified by or selected for fine-mapping.

**Table S3.** Sanger fine-mapping genotypes.

**Table S4.** SNPlex genotypes and *p*-values.

All genotypes are derived from SNPlex assays except for genotypes for the 9.9 kb deletion downstream *SMAD2* gene on CFA7 (chr7:43,794,129); the latter was determined by gel electrophoresis of PCR products.

**Table S5.** Primer sequences and genomic positions for amplicons used for marker discovery at all loci.
Start and end position of the amplicons refer to the CanFam 3.1 dog assembly.

**Table S6.** Genotyping results of the 500 dogs from 93 breeds at the best size-associated markers.
The dogs randomly selected for some analyses are indicated (two males and two females per breed).

**Table S8.** PCR conditions and primers used to genotype the most highly associated markers.

**Table S9.** Average principal components (PCs) from genome-wide SNP profiles for each of the 65 breeds present in both our data set and the CanMap data set.

## References


American Kennel Club 1998. *The complete dog book*. Howell Book House, New York.

Barrett JC, Fry B, Maller J, Daly MJ. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**: 263–265.

Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res* **14**: 1188–1190.

Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8**: 186–194.

Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8**: 175–185.

Fogle B. 1995. *The Encyclopedia of the Dog*. DK Publishing, Inc., New York.

Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res* **8**: 195–202.

Nickerson DA, Tobe VO, Taylor SL. 1997. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res* **25**: 2745–2751.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**: 559–575.

Rozen S, Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**: 365–386.

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**: 1034–1050.

Stephens M, Scheet P. 2005. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am J Hum Genet* **76**: 449–462.

Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **68**: 978–989.

Venables WN, Ripley BD. 2002. *Modern applied statistics with S*. Springer, New York.

Wilcox B, Walkowicz C. 1995. *Atlas of dog breeds of the world*. T.F.H. Publications, Neptune City, NJ.