

Analysis of large versus small dogs reveals three genes on the canine X chromosome associated with body weight, muscling and back fat thickness

Jocelyn Plassais, Maud Rimbault, Falina J. Williams, Brian W. Davis, Jeffrey J. Schoenebeck, Elaine A. Ostrander

▶ **To cite this version:**

Jocelyn Plassais, Maud Rimbault, Falina J. Williams, Brian W. Davis, Jeffrey J. Schoenebeck, et al.. Analysis of large versus small dogs reveals three genes on the canine X chromosome associated with body weight, muscling and back fat thickness. PLoS Genetics, 2017, 13 (3), pp.e1006661. 10.1371/journal.pgen.1006661 . hal-01508065

HAL Id: hal-01508065

https://univ-rennes.hal.science/hal-01508065

Submitted on 21 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PLOS | GENETICS

RESEARCH ARTICLE

Analysis of large versus small dogs reveals three genes on the canine X chromosome associated with body weight, muscling and back fat thickness

Jocelyn Plassais[☯], Maud Rimbault[☯¤a], Falina J. Williams, Brian W. Davis, Jeffrey J. Schoenebeck[¤b], Elaine A. Ostrander*

Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, United States of America

☯ These authors contributed equally to this work.
¤a Current Address: UMR 6290 CNRS, Université de Rennes 1, Rennes, France
¤b Current Address: The Roslin Institute and the Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush Campus, Midlothian, United Kingdom
* eostrand@mail.nih.gov

Check for updates

OPEN ACCESS

Citation: Plassais J, Rimbault M, Williams FJ, Davis BW, Schoenebeck JJ, Ostrander EA (2017) Analysis of large versus small dogs reveals three genes on the canine X chromosome associated with body weight, muscling and back fat thickness. PLoS Genet 13(3): e1006661. doi:10.1371/journal.pgen.1006661

Editor: Leigh Anne Clark, Clemson University, UNITED STATES

Received: December 16, 2016

Accepted: February 26, 2017

Published: March 3, 2017

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The work was done while investigators were at NHGRI and the project was funded by the Intramural Program of the National Human Genome Research Institute. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Domestic dog breeds display significant diversity in both body mass and skeletal size, resulting from intensive selective pressure during the formation and maintenance of modern breeds. While previous studies focused on the identification of alleles that contribute to small skeletal size, little is known about the underlying genetics controlling large size. We first performed a genome-wide association study (GWAS) using the Illumina Canine HD 170,000 single nucleotide polymorphism (SNP) array which compared 165 large-breed dogs from 19 breeds (defined as having a Standard Breed Weight (SBW) >41 kg [90 lb]) to 690 dogs from 69 small breeds (SBW $\leq$41 kg). We identified two loci on the canine X chromosome that were strongly associated with large body size at 82–84 megabases (Mb) and 101–104 Mb. Analyses of whole genome sequencing (WGS) data from 163 dogs revealed two indels in the *Insulin Receptor Substrate 4* (*IRS4*) gene at 82.2 Mb and two additional mutations, one SNP and one deletion of a single codon, in *Immunoglobulin Superfamily member 1* gene (*IGSF1*) at 102.3 Mb. *IRS4* and *IGSF1* are members of the GH/IGF1 and thyroid pathways whose roles include determination of body size. We also found one highly associated SNP in the 5'UTR of *Acyl-CoA Synthetase Long-chain family member 4* (*ACSL4)* at 82.9 Mb, a gene which controls the traits of muscling and back fat thickness. We show by analysis of sequencing data from 26 wolves and 959 dogs representing 102 domestic dog breeds that skeletal size and body mass in large dog breeds are strongly associated with variants within *IRS4*, *ACSL4* and *IGSF1*.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Modern dog breeds display significant variation in body size and mass resulting from selective breeding practices. A genome-wide association study (GWAS) of 170,000 SNPs genotyped on a panel of 855 dogs from 88 breeds revealed two loci on the canine X chromosome that were strongly associated with the large Standard Breed Weight (SBW >41kg [90lb]) in domestic dog breeds. Fine mapping of both loci using whole genome sequencing (WGS) data from 163 dogs highlighted three candidate genes, *IRS4* and *ACSL4* within the first locus (82–84 Mb), and *IGSF1* at the second locus (101–104 Mb). Associated variants were found in all three genes. Of interest, the *IRS4* gene is involved in the IGF-1/growth hormone pathway and *IGSF1* is associated with human obesity. *ACSL4* is associated with muscling and back fat thickness in pigs, a phenotype we observe in "bulky" dog breeds. This work identifies three new genes and associated variants that contribute to body mass in dogs, advancing our understanding of morphologic variability in domestic dog breeds.

Introduction

Body size variation observed across domestic dog (*Canis lupus familiaris*) breeds provides one of the most visual examples of human selection. Dogs are thought to have been domesticated about 15,000–30,000 years ago [1–10], with the grey wolf being the closest living ancestor [2,4,11–14]. The majority of modern dog breeds, however, were developed within the past 300 years with over 340 official breeds noted worldwide [15,16].

The creation of breeds requires codified standards that describe the physical characteristics of the dog. The breeding strategies used to create dogs with highly specific features have resulted in relatively isolated, pure breeding populations [17]. The same selective pressures that have reduced phenotypic and genotypic heterogeneity within breeds [6,10,18–21] result in long stretches of linkage disequilibrium (LD) in dogs [20,22–24]. Given these advantageous features, studies of dog breeds have led to the identification of disease genes of interest for human health and biology, including rare human disorders [25–28], *e.g.* cancer [29–33]. The same genomic characteristics have also produced a stellar system for identifying the genes underlying both simple and complex morphologic traits, including coat color and texture variation, tail curl, ear position, skull shape, chondrodysplasia and body size (reviewed in [34–39]).

Body size is the most striking of these traits, as the difference in skeletal size from the smallest to largest dog breeds is about 40-fold [15,16]. Initial studies of dog body size focused on the Portuguese Water Dog (PWD), a breed for which the American Kennel Club (AKC) permits about a 50% level of size variation amongst members of the breed [40]. A genome-wide association study (GWAS) of PWD representing a range of body sizes identified the *insulin-like growth factor-1 gene* (*IGF1*) [41], and additional studies of Miniature Poodles and Dachshunds implicated the *IGF1 receptor* (*IGF1R*) as well [42], both of which are important regulators of body size. The *IGF1* pathway has also been established as important in normal stature in humans, and mutations in *IGF1* have been shown to reduce body size in mice [43–46].

Four additional positional candidate genes contributing to variation in canine body size have been identified: the *Growth Hormone Receptor* (*GHR*) on canine chromosome 4 (CFA4); *High Mobility Group AT-hook 2* (*HMGA2*) gene on CFA10; *Stanniocalcin 2* (*STC2)* on CFA4; and *SMAD family member 2 (SMAD2)* on CFA7 [22,47,48]. The most closely associated variants have been reported for each [47]. This includes two non-synonymous SNPs in exon 5 of

GHR, a SNP in the 5'UTR of *HMGA2*, a SNP located 20 kb downstream from *STC2* and a 9.9 kb deletion 24 kb downstream from *SMAD2*, all of which are highly associated with lower Standard Breed Weights (SBW) [47]. Additional studies noted ten additional putative loci [23,48]. These studies did not, however, identify causal variants and did not ascertain the contribution of each gene, or a combination of genes, to overall size variance in dogs.

Variant haplotypes of the six genes described above are strongly associated with large versus small body size, although some exceptions exist. While we showed previously that *IFG1*, *IFG1R*, *SMAD2*, *HMGA2*, *STC*, and *GHR* variants account for about 60% of body size variance in breeds with a SBW $\leq$41 kg (90 lb) which are referred hereto as "small/medium breeds," the same genes account for <5% of variance in breeds with a SBW >41 kg, hereto referred to as "large breeds". We initially identified two loci on the X chromosome spanning several megabases (Mb) as contributors to body size in large breeds through a GWAS of 915 dogs representing 80 domestic dog breeds [22]. The result has since been replicated by several groups [22,23,47,48]. No study, however has explored the result in detail, in part because the lack of heterozygosity on the canine X chromosome can reflect popular sire effects, which may complicate fine mapping efforts.

In this study we investigate body size loci on the X chromosome using SNP chip data from $\geq$800 dogs [21], together with whole genome sequencing (WGS) data. We show that both of the previously identified loci are strongly associated with large breeds, and we perform fine mapping at each locus using WGS data from 163 breeds. Together, these data reveal associations with three excellent positional candidate genes: *Insulin Receptor Substrate 4* (*IRS4*) which interacts with multiple growth factor receptors such as *IGF1R* [49], *Immunoglobulin superfamily member 1* (*IGSF1)* which is involved in the biosynthesis of thyroid hormones [50–52] and *Acyl-CoA Synthetase Long-chain family member 4* (*ACSL4*) which plays a role in lipid biosynthesis and fatty acid degradation [53].

Results

Genome-wide association study

We initially genotyped a large dataset of 855 dogs representing 88 breeds on the Illumina 170k Canine HD Array [21]. For purposes of this analysis, large breeds included 165 dogs from the 19 following giant breeds: Akita, Anatolian Shepherd Dog, Bernese Mountain Dog, Black Russian Terrier, Bullmastiff, Dogue de Bordeaux, English Mastiff, Great Dane, Greater Swiss Mountain Dog, Great Pyrenees, Irish Wolfhound, Kuvasz, Leonberger, Neapolitan Mastiff, Newfoundland, Rottweiler, Saint Bernard, Scottish Deerhound, and Tibetan Mastiff. Using the array data, we compared the genotypes from the above large breeds to 690 dogs from 69 small/medium breeds (S1 Table). To correct for cryptic relatedness and sex, we used GEMMA [54,55], a linear mixed-model method which accounts for population stratification and relatedness.

A total of 81 SNPs were significant at a genome-wide level for the trait of body mass, and which passed the Bonferroni significance threshold ($-\log_{10}(P) > 6.48$) (Fig 1A). Among these, we identified two primary loci on the X chromosome (Table 1). The first locus (locus 1) included 23 SNPs, and spanned 82,296,039 to 84,376,308 bp (Fig 1B). A stronger signal ($P = 7.74 \times 10^{-14}$) was identified at a second locus on the X chromosome, which spans 101,646,292 to 103,984,352 bp, corresponding to 56 additional SNPs that passed the significance threshold (Fig 1C). Neither of these loci were within the pseudoautosomal region of the X. Two additional SNPs located on CFA6 passed the significance threshold, chr6: 38,284,916 ($P = 9.36 \times 10^{-8}$) and chr6: 67,350,922 ($P = 1.80 \times 10^{-7}$), but no additional associated SNPs were found in these regions and the result was not explored further at this time.

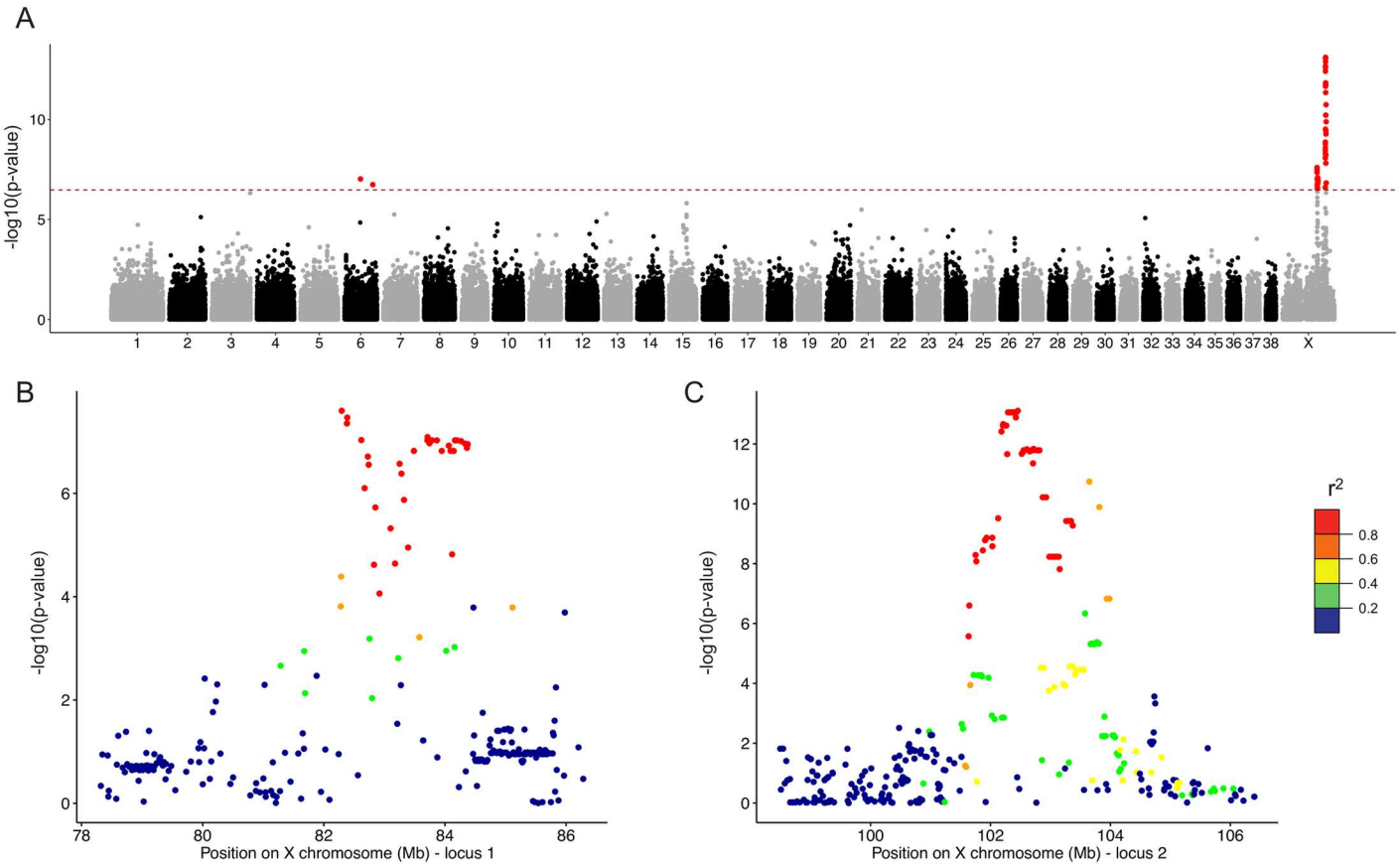


Fig 1. GWAS results for 165 large breed dogs *vs.* 690 medium/small breed dogs. (A) Manhattan plot for the GWAS is shown. The -$\log_{10}$ P-values for each SNP are plotted on the y-axis *versus* each canine autosome and the X chromosome on the x-axis. The red line represents the Bonferroni corrected significance threshold (-$\log_{10}$ (P) = 6.48) and SNPs passing this threshold are colored in red. (B and C) Regional plot for genome-wide significant association on the X chromosome for 78–86 Mb (B) and 98–106 Mb (C). Each plot spans the genomic regions from 4000 kb upstream to 4000 kb downstream of the most significant SNP at each locus. SNPs are colored based on the strength of LD values (r^2 values) considering the most strongly associated SNP and the other SNPs in the region.

doi:10.1371/journal.pgen.1006661.g001

We examined the loci of interest more closely by calculating pairwise linkage disequilibrium (LD) between SNPs within the 4,000 kilobase (kb) regions surrounding the most strongly associated SNPs at each of the two loci on the X chromosome. Thirty-five SNPs were highly correlated (pairwise $r^2 > 0.8$) at locus 1 (Fig 1B) while 53 were highly correlated at locus 2 (Fig 1C). We next investigated each locus by focusing on regions in which SNPs had pairwise r^2 values >0.5 and extending these regions by +/- 200 kb. The two refined intervals ranged from 82,079,576 to 84,576,308 for locus 1 and from 101,378,080 to 104,418,823 for locus 2. Locus 1 contains 17 annotated protein-coding genes and 11 annotated RNA genes (small RNAs and long non-coding RNAs) or pseudogenes (S1 Fig). Among these genes, the strongest candidate gene to emerge at locus 1 is *Insulin Receptor Substrate 4* (*IRS4)*, a gene involved in the thyroid hormone pathway, which is associated with IGF1R signaling and body mass index [49,56]. At locus 2, 20 protein-coding genes are annotated, including a cluster of olfactory receptor genes, and seven noncoding RNAs including microRNA, noncoding RNA and pseudogenes (S2 Fig). From these 20 genes, we identified one striking candidate, *Immunoglobulin Superfamily member 1 (IGSF1*), that encodes an immunoglobulin in the thyroid hormone pathway, and which was previously associated with obesity in IGSF1-deficient humans [50–52].

Table 1. Best SNP markers associated with SBW above 41 kg in dogs.

Rank	Chromosome	Position	Major Allele	Minor Allele	Allele Frequency	P-Wald	Closest Gene
1	X	102455197	G	A	0.425	7.74E-14	Olfactory Receptor Cluster
2	X	102292529	A	G	0.42	8.75E-14	Olfactory Receptor Cluster
3	X	102323911	A	G	0.42	8.75E-14	*IGSF1*
4	X	102333672	A	G	0.42	8.75E-14	*IGSF1*
5	X	102360497	C	A	0.42	8.75E-14	*IGSF1*
6	X	102381223	G	A	0.42	8.75E-14	*IGSF1*
7	X	102385409	G	A	0.42	8.75E-14	*IGSF1*
8	X	102415621	C	A	0.42	8.75E-14	Olfactory Receptor Cluster
9	X	102429959	A	C	0.42	8.75E-14	Olfactory Receptor Cluster
10	X	102423378	A	T	0.42	1.29E-13	Olfactory Receptor Cluster
…	…	…	…	…	…	…	…
54	X	82296039	A	G	0.391	2.51E-08	*IRS4*
55	X	82387385	A	C	0.395	3.41E-08	*IRS4*
56	X	82381884	G	A	0.408	4.41E-08	*IRS4*
57	X	83713308	A	G	0.386	8.10E-08	x
58	X	82619550	A	G	0.395	9.27E-08	*GUCY2F*
59	6	38284916	G	A	0.022	9.36E-08	*SRRM2*

doi:10.1371/journal.pgen.1006661.t001

Fine mapping strategy

To identify functional variants within these two critical intervals, notably in the two strongest candidate genes, *IRS4* and *IGSF1*, we used WGS data from 163 purebred dogs inclusive of 87 breeds representing the full range of body height and weight specified by the American Kennel Club (AKC). Each WGS had a mean read depth of at least 10x (S2 Table). Among these, 21 dogs from 14 breeds were considered large (SBW >41 kg [90 lb]), including the English Mastiff, Irish Wolfhound and Saint Bernard. We first filtered to retain biallelic variants including SNPs and small insertions or deletions <100 bp, with a minor allele frequency (MAF) >0.05. We next screened the remaining biallelic variants, keeping only variants for which the major allele frequency in large breeds was >0.5 and <0.5 in small/medium breeds (SBW ≤41 kg [90 lb]). A total of 6,809 variants remained for locus 1 and 1,997 variants for locus 2 (S3 and S4 Tables). Using these biallelic variant datasets, we performed a new association study for both loci using GEMMA, a linear mixed-model software [54,55], thus defining which alleles were the most strongly associated with large breeds, and in each case that allele was termed the "large allele".

Fine mapping results at locus 1

The 6,809 variants identified at locus 1 define a set of genotypes which correspond to a single large haplotype present in more than 90% of large breeds (S1 Fig). This spans the strong signal originally identified in the GWAS presented here. Among these variants, we identified one codon deletion (chrX.g.82288614-82288616delTCG) and one insertion (chrX.g.82288998-82288999insGCT) both in the exonic region of the *IRS4* gene that were in LD with one another (S1 Fig). Neither, however, are likely to be significant for this study as neither mutation changes the IRS4 protein size, distinguishes between various size breeds or is in a well-conserved region (Table 2). In addition, for each variant the "large alleles" were also identified in more than 20% of small/medium breeds.

While we discarded the above variants in *IRS4* from an association with body size, a re-analysis aimed at finding structural variants revealed a large 56 kb deletion (ChrX:82455513–

Table 2. Conservation of *IRS4* mutated codons (in bold) between mammals.

Species	IRS4 Protein Size	% Identity with dog	Codon Insertion (codon 1126)	Codon Deletion (codon 1253)
Large Dog (SBW >41 kg [90 lb])	1278	-	LERDLPASLASAA**A**AAAASAAAPALAL	PPEQQVSDNDDDDDD-DTYVRMDF
Medium & Small Dog (SBW ≤41 kg)	1278	-	LERDLPASLASAA-AAAASAAAPALAL	PPEQQVSDNDDDDDD**D**DTYVRMDF
Human	1257	82.6	LERDLSPS----SAPAVASAAEPTLAL	PPEREDSDNDDD-----THVRMDF
Mouse	1216	74.2	LDRDFPA----ASAVIAAPAEAPLLAV	RPQERADSEDDDDDD**D**DIYVRMDF
Cat	1124	86.0	LERDLSASL--------ASVAAPALAV	PRER-----EDSDDD**D**DTYVRMDF
Cow	1255	81.5	LERDLSASFAAAAAAV------PTLAL	PPERE------DSDD**D**DTYVRMDF
Horse	1051	82.4	LERDLSASLASAAALAA----APTLGV	PPEQE-----DSDDD**D**DTYVRMDF
Megabat	1260	76.3	LERDLYASFASAAVAVVV----PTFAL	LPERE-----DSDDD**E**DTYARMDF

doi:10.1371/journal.pgen.1006661.t002

82511744) located 150 kb upstream from the starting codon of *IRS4* (S1 Fig). The variant was only present in the Bernese Mountain Dog, Black Russian Terrier, English Mastiff, Greater Swiss Mountain Dog, Rottweiler, and Saint Bernard. Visualization of the deletion on an agarose gel indicated that it was also present in multiple other large breeds: the Alaskan Malamute, Bouvier des Flandres, Bullmastiff, Dogo/Presa Canario, Dogue de Bordeaux, Kuvasz and Leonberger.

ACSL4 gene and the large muscled phenotype at locus 1

Among the 6,809 biallelic variants identified at locus 1, we also found three variants, distinct from the above, which were themselves in LD (Fig 2), and which harbored the highest p-values (10^{-10}<P-value<10^{-15}, P-Wald test) (Table 3). One of the three is a SNP (chrX. g.82919525G>A) in the 5'UTR of *Acyl-CoA Synthetase Long-chain family member 4* (*ACSL4*), a gene which plays a role in lipid biosynthesis and fatty acid degradation [53]. This nucleotide is included in a highly conserved region also identified in the human and mouse genomes (S3 Fig). The other two SNPs were intergenic or intronic (in *AMMECR1*) (Table 3).

All three variants, the SNP in *ACSL4*, together with the two SNPs in the same LD block, were present in an interesting subset of large dogs. Specifically, variants were only identified in four of 19 large breeds: Bernese Mountain Dog, Greater Swiss Mountain Dog, Rottweiler, and Saint Bernard. The other 78 breeds, which included large, medium and small breeds, lacked all three variants. Of note, all three variants were missing in several large breeds that were skeletally quite large, but comparatively lean, including the Cane Corso, Great Dane, and Irish Wolfhound, among others (Fig 2). The breeds in which the variant is found are not simply skeletally large, but also considered "bulky," with considerable muscle and fat.

We next checked for the frequency of the 5'UTR *ACSL4* variant by testing a larger panel of 959 dogs from 102 breeds, which represented an additional 54 breeds (S5 Table). The "bulky allele" was present in several dogs with a bulky, heavily muscled body: Bullmastiff, Dogue de Bordeaux, English Mastiff, Greater Swiss Mountain Dog, Newfoundland, Rottweiler and Saint Bernard, where it appears fixed in nearly 100% of dogs from each breed (Fig 3). We found that the Alaskan Malamute, Bernese Mountain Dog, Black Russian Terrier, Bouvier des Flandres, Dogo/Presa Canario, Kuvasz and Leonberger breeds could be either heterozygous or homozygous for both alleles. In total, 48% of the large breeds shared the "bulky allele" (heterozygous or homozygous) (Fig 3). Sanger sequencing of a larger panel of dogs (≥10 dogs per breed) including the Anatolian Shepherd Dog, Great Dane, Great Pyrenees, Irish Wolfhound, Neapolitan Mastiff, and Scottish Deerhound confirmed the absence of the "bulky allele" in these breeds, many of which are long and lean rather than bulky. Of note, the *ACSL4* variant

21 dogs (SBW above 41 kg)	Sex	SBW (kg)	chrX:82919525	chrX:83149082	chrX:83231250	chrX:83430607	chrX:83509356	chrX:83718522	chrX:83907044	chrX:84062048	chrX:84069469	chrX:84291153
Scottish Deerhound 1	M	41.4	C	C	T	G	C	G	A	-	A	G
Great Pyrenees	M	42.0	T	G	T	C	T	G	T	A	ins/ins	G
Bernese Mountain Dog 1	F	43.1	T/T	G/G	T/T	C/C	T/T	G/G	T/T	A/A	ins/ins	G/G
Bernese Mountain Dog 2	F	43.1	T/T	G/G	T/T	C/C	T/T	G/G	T/T	A/A	ins/ins	G/G
Cane Corso	F	44.9	C/C	C/C	A/T	G/G	C/T	A/G	A/T	-/A	A/ins	A/G
Caucasian Ovcharka	F	47.4	C/C	C/C	A/A	G/G	C/C	A/G	A/A	-/A	A/A	A/A
Rottweiler 1	F	48.2	T/T	G/G	T/T	C/C	T/T	G/G	T/T	A/A	ins/ins	G/G
Rottweiler 2	M	48.2	T	G	T	C	T	G	T	A	ins	G
Rottweiler 3	F	48.2	T/T	G/G	T/T	C/C	T/T	G/G	T/T	A/A	ins/ins	G/G
Rottweiler 4	F	48.2	T/T	G/G	T/T	C/C	T/T	G/G	T/T	A/A	ins/ins	G/G
Black Russian Terrier	F	50.6	T/T	G/G	T/T	C/C	T/T	G/G	T/T	A/A	ins/ins	G/G
Irish Wolfhound	F	51.0	C/C	C/C	T/T	G/G	T/T	G/G	T/T	-/-	ins/ins	G/G
Landseer	M	56.7	C	C	T	G	T	G	T	A	ins	G
Greater Swiss Mountain Dog	M	60.1	T	G	T	C	T	G	T	A	ins	G
Great Dane	F	62.4	C/C	C/C	T/T	G/G	C/T	A/G	A/T	-/A	A/ins	G/G
English Mastiff 1	F	70.3	T/T	G/G	T/T	C/C	T/T	G/G	T/T	A/A	ins/ins	G/G
English Mastiff 2	M	70.3	T	G	T	C	T	G	T	A	ins	G
Saint Bernard 1	F	70.3	C/T	C/G	T/T	G/C	T/T	A/G	T/T	A/A	ins/ins	G/G
Saint Bernard 2	M	70.3	T	G	T	C	T	G	T	A	ins	G
Tibetan Mastiff 1	F	70.3	C/C	C/C	A/A	G/G	C/C	A/G	A/A	-/-	A/ins	A/G
Tibetan Mastiff 2	F	70.3	C/C	C/C	A/A	G/G	C/C	A/A	A/A	-/-	A/A	A/G
142 dogs (SBW under 41 kg)			**% of the control population with each genotype**									
Homozygous genotype for the rare allele	-	<41 kg	0.0	0.0	14.8	0.0	12.8	10.1	7.6	6.3	11.2	8.4
Heterozygous genotype	-	<41 kg	1.5	2.1	15.6	0.8	14.3	15.6	13.0	14.3	13.3	12.6
Homozygous genotype for the common allele	-	<41 kg	98.5	97.9	69.7	99.2	72.9	74.3	79.4	79.5	75.5	78.9
		P-value	**1.3E-13**	**5.7E-10**	**2.6E-08**	**2.4E-15**	**2.5E-08**	**3.1E-08**	**1.1E-08**	**2.0E-08**	**1.8E-08**	**2.7E-08**

Fig 2. Observed genotypes for the ten most strongly associated variants identified in large dogs using WGS data. The three first columns correspond to the dog, breed, sex and standard breed weight (SBW). The next 10 columns correspond to the 10 most strongly associated variants at locus 1, identified from WGS data. The first part of the table corresponds to large dogs (SBW >41 kg). Homozygous and hemizygous genotypes for the "large allele" are colored in red, homozygous/hemizygous genotypes for the "small/medium" allele are colored in blue and heterozygous genotypes are colored in yellow. The second part of the table shows the distribution of the "large allele" in the 140 dogs with a SBW ≤41 kg. Values correspond to the percentage of this control population showing each genotype by variant. The last row shows the respective p-value estimated (Wald test) for each variant.

doi:10.1371/journal.pgen.1006661.g002

mutation was never observed in medium or small breeds, even small muscled breeds such as American Staffordshire Terrier, Boston Terrier, or Bulldog. The results were the same with the two intergenic or intronic variants in LD with the 5' UTR *ACSL4* variant (S5 Table).

Sanger sequencing of the set of wild canids (24 grey wolves, two red wolves and two coyotes) confirmed that the three mutations, including the *ACSL4* variant, are absent from the wild canid population, leading us to consider these variants as derived alleles which were likely selected by humans to create large and muscled breeds. The *ACSL4* gene is associated with the traits of heavy muscling and "back fat thickness" in pigs, a phenotype that aptly describes the breeds carrying the mutation [57–61]. We conclude that *ACSL4*, potentially in concert with the upstream deletion in *IRS4*, is needed to create the large bulky/muscled phenotype observed in the breeds reported here (Figs 3 and 4).

Fine-mapping using Whole-Genome Sequencing (WGS) data at locus 2

The 1,997 variants at locus 2 also define a large homozygous haplotype found in all large breeds, except the Scottish Deerhound (S2 Fig). The haplotype found in the large breeds is also observed in 24 of 66 small/medium breeds including the Boston Terrier, Boxer, French

Table 3. Best variants associated with SBW above 41 kg in dogs, identified on the locus 1 (CFAX) using 163 WGS data.

Position	Major Allele	Minor Allele (Large Allele)	Allele Frequency	Non-missing Allele Count	P-Wald	Closest Gene	Details
83430607	G	C	0.08368	239	2E-15	*AMMECR1*	Intronic SNP
82919525	C	T	0.08475	236	1E-13	*ACSL4*	5'UTR SNP
83149082	C	G	0.09574	188	6E-10	*SH3D21*	7.6 kb upstream of the start codon
83907044	A	T	0.2426	235	1E-08	*CHRDL1*	Intronic SNP
84069469	A	ATAATAATAATAATAATAG	0.2947	190	2E-08	*PAK3*	4.9 kb upstream of the start codon
84062048	-	A	0.2392	209	2E-08	*PAK3*	13 kb upstream of the start codon
83509356	C	T	0.2833	240	2E-08	*PAK3*	13 kb upstream of the start codon
83232150	A	T	0.3259	224	3E-08	*TMEM164*	Intronic SNP
84291153	A	G	0.2655	177	3E-08	*PAK3*	Intronic SNP
83718522	A	G	0.2913	206	3E-08	*SNU13*	Intergenic

doi:10.1371/journal.pgen.1006661.t003

Bulldog, Irish Water Spaniel and Labrador Retriever. Not unexpectedly, we observe the heterozygous state in 19 additional small breeds. While WGS demonstrates that the haplotype is found in breeds of varying size, the fact that it is present in 18 of 19 breeds in the homozygous state suggest that it is necessary, but not sufficient, for large body mass. Within this region, we detected missense changes in three genes: *ARHGAP36* (*Rho GTPase Activating protein 36*), *IGSF1* (*Immunoglobulin Superfamily Member 1*) and *FRMD7* (*FERM Domain Containing 7*) (S4 Table).

Since the *IGSF1* gene is a strong candidate for body size [50–52], we examined it further, noting three variants in the canine sequence (S2 Fig). The first is a single nucleotide change in the 3'UTR (chrX.g.102360204G>A; rs24856221), but the distribution of the genotype in the dog population suggests that it was not associated with SBW (S4 Table). The second is a missense mutation in exon 12 (chrX.g.102364864T>G; rs852386368) that changes an aspartic acid to a glutamic acid (ENSCAFP00000027740.3:p.Asp768Glu). The codon is highly conserved in mammals (Table 4) and is an excellent functional candidate with a likely high impact on protein function (Polyphen score = 0.992). The third variant is an in-frame deletion (chrX.g.102369488-102369489insAAC; rs850984482) in exon 6 of the gene, which is in LD with the missense mutation. The deletion removes one polar amino acid, asparagine, in the conserved immunoglobulin-like domain (ENSCAFP00000027740.3:p.Asp376_Glu377insAsn) and is also a strong functional candidate.

The two potentially functional *IGSF1* mutations at locus 2 were considered further. To determine the ancestral allele for each, we used Sanger sequencing to ascertain genotypes from a set of wild canids, including 24 grey wolves from geographically diverse areas, two red wolves, and two coyotes. The two mutations (missense SNP at exon 12 and deletion at exon 6), while associated with large size in dogs, were never observed in the coyote, red wolves, or grey wolves, leading us to term these two large breed variants as "derived" alleles.

To determine the frequency of each candidate variant in domestic breeds, we used Sanger sequencing to analyze a large panel of 561 dogs encompassing 96 breeds (S6 Table). This panel included 10 additional large breeds and 36 more small/medium breeds. We observe both the exon 6 and 12 variations of *IGSF1* in the homozygous state in several large breeds of varying mass and skeletal size including the Bullmastiff, Great Dane, Great Pyrenees, Irish Wolfhound, Newfoundland, and Saint Bernard. Heterozygous genotypes were also identified in six

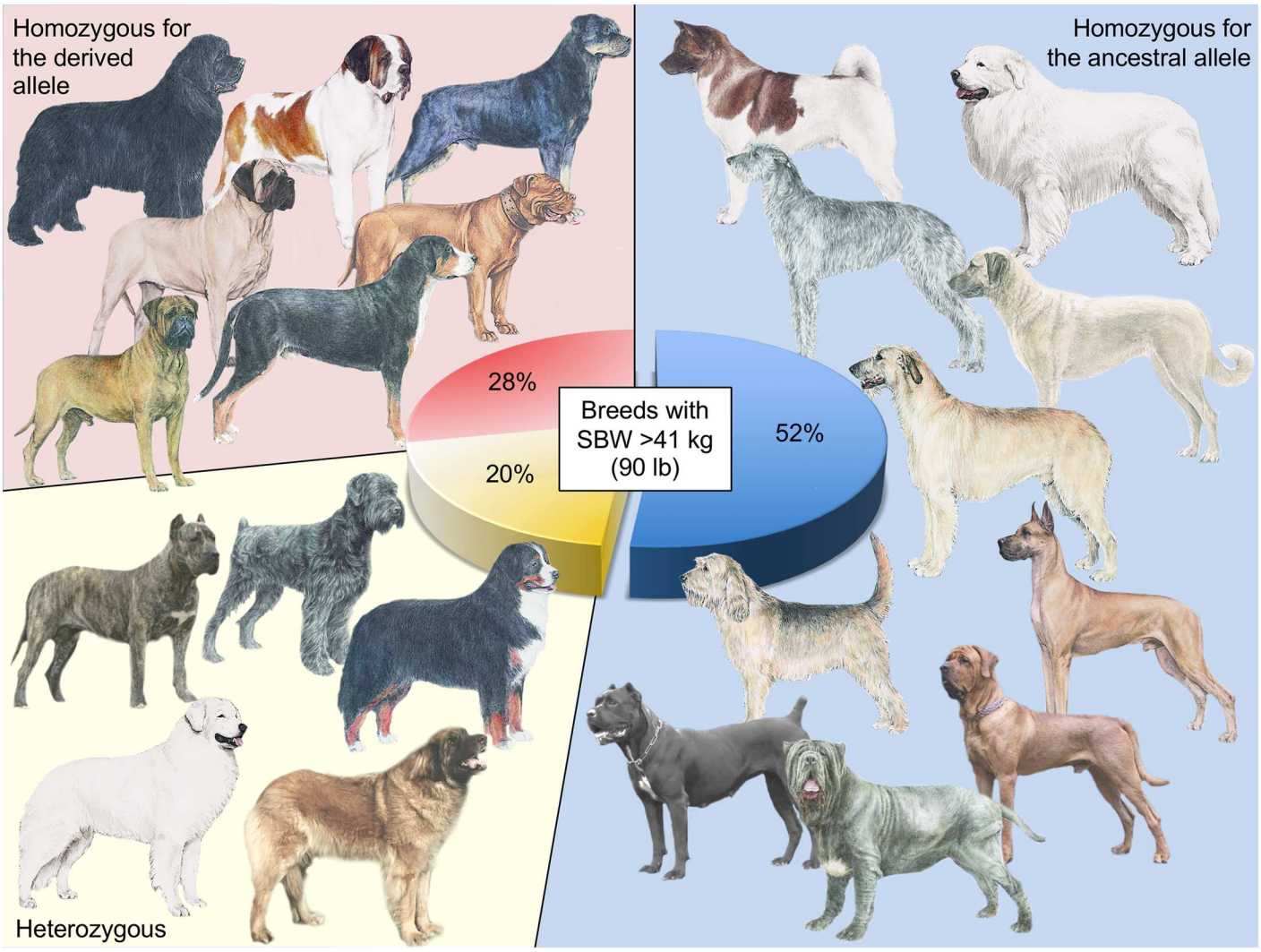


Fig 3. Distribution of the derived *ACSL4* allele in the large dog breeds. The Bullmastiff, Dogue de Bordeaux, English Mastiff, Greater Swiss Mountain Dog, Newfoundland, Rottweiler and the Saint Bernard are homozygous for the derived allele (red) while Bernese Mountain Dog, Black Russian Terrier, Dogo/Presa Canario, Kuvasz and Leonberger were either homozygous or heterozygous (yellow) at *ACSL4*. Other large dog breeds (*e.g.* Akita, Anatolian Shepherd Dog, Cane Corso, Great Dane, Great Pyrenees, Irish Wolfhound, Neapolitan Mastiff, Otterhound, Scottish Deerhound, Tibetan Mastiff, Tosa Inu), wild canids and all medium/small breeds carry the ancestral allele in the homozygous state (blue). Pictures provide by the American Kennel Club (AKC) and Larousse.

doi:10.1371/journal.pgen.1006661.g003

additional large breeds: Black Russian Terrier, Dogo Canario, Greater Swiss Mountain Dog, Mastino Abruzzese and Tibetan Mastiff. As expected, based on this pattern, we never found the "large *IGSF1* allele" for either variant in the five unrelated Sanger sequenced Scottish Deerhounds, which are included in our definition of large breeds. We identified 17 medium and small breeds for which all sequenced dogs were homozygous for the derived allele at both *IGSF1* variants, and 21 medium and small breeds that were heterozygous, recapitulating the pattern observed above. Interestingly, among the medium and small breeds, we found muscled breeds such as Boston Terrier, Boxer, Bulldog, French Bulldog, Miniature Bull Terrier and Shar-pei. The remaining 36 small/medium breeds with SBW ranging from 2.7 to 39.5 kg (6.1 to 87 lb), and corresponding to 48.7% of the medium/small breeds, were homozygous for the ancestral alleles for both variants (S6 Table). The "large alleles" were found in 95% of large

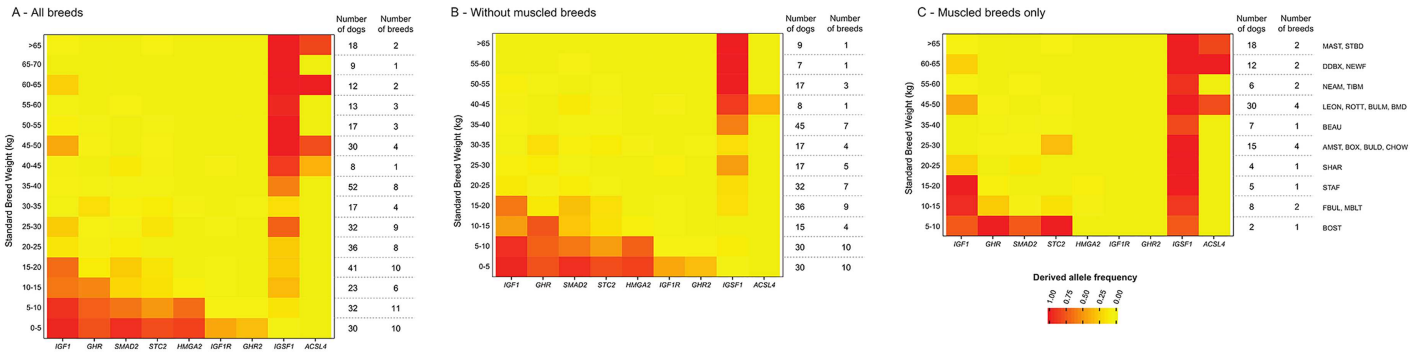


Fig 4. Derived allele frequencies at multiple loci involved in body weight and size variations in (A) modern dog breeds, (B) modern dog breeds without inclusion of muscled breeds, (C) only muscled breeds. The frequency of the derived allele in 5-kg weigh classes is represented on a color scale. Dogs with a SBW above 65 kg are collapsed in a single category (>65 kg) due to the lack of genotype variation in the group at these markers. Muscled breeds include the Boston Terrier (BOST), French Bulldog (FBUL), Miniature Bull Terrier (MBLT), Staffordshire Bull Terrier (STAF), Chinese Shar-pei (SHAR), Bulldog (BULD), Chow-Chow (CHOW), American Staffordshire Terrier (AMST), Boxer (BOX), Beauceron (BEAU), Bullmastiff (BULM), Bernese Mountain Dog (BMD), Rottweiler (ROTT), Leonberger (LEON), Tibetan Mastiff (TIBM), Neapolitan Mastiff (NEAM), Newfoundland (NEWF), Dogue de Bordeaux DDBX), English Mastiff (MAST) and Saint Bernard (STBD).

doi:10.1371/journal.pgen.1006661.g004

breeds, and the genotypes appeared fixed (homozygous for the "large alleles") in 76.2% of large breeds. By comparison, 51.4% of medium/small dogs carry what we considered to be "large alleles" (homozygous in 44.7%). This argues that while *IGSF1* likely plays a role in modulating weight variation in modern breeds, it is also, and more precisely, a contributor to the muscled phenotype in breeds spanning a range of body sizes (Fig 4).

Discussion

In this study, we identified two loci on the X chromosome associated with SBW in domestic dog breeds, using a panel of 855 dogs selected to represent the full range of canine body size, which we genotyped on the Illumina Canine HD SNP array. We showed that two large haplotypes at two loci were shared by the majority (>90%) of large breeds with SBW >41 kg (90 lb), for which derived alleles (not present in wolves) have been identified. Fine mapping using whole genome sequencing data from 163 dogs revealed candidate variants in *IRS4* and *IGSF1* that are strongly associated with large breeds. Interestingly, we also identified a phenotype of bulky or stocky build, which is also referred to as "heavily muscled," for which a third candidate gene, *ACSL4*, and variant were associated. The bulky haplotype was found post hoc and not detected by either our GWAS or any previously published GWAS, because no SNP on the canine HD SNP array is in LD with the variants. These particular allelic distributions in the

Table 4. Conservation of *IGSF1* mutated codons (in bold) between mammals.

Species	IGSF1 Protein Size	% Identity with dog	Codon Deletion (codon 377)	Codon changing (codon 768)
Large Dog (SBW >41 kg [90 lb])	1338	-	TSIDD-ESFFLNNVTY	PFKWSEPS**D**PMELVIKE
Medium & Small Dog (SBW ≤41 kg)	1339	-	TSIDD**N**ESFFLNNVTY	PFKWSEPS**E**PMELVIKE
Human	1327	90.1	TSIDD**N**TSFFLNNVTY	PFKWSEPS**E**PLELVIKE
Mouse	1317	85.1	SSNTG**N**NSFFLKNVTY	PFKWSEPS**E**PLELVIKE
Cat	1341	95.2	ASTDD**D**ESFFLNNVTY	PFKWSEPS**E**PLELVIKE
Pig	1340	92.9	TSIDD**D**KSFFLNNVTY	PFKWSEPS**E**PLELVIKE
Cow	1327	90.5	TSVYD**D**TSFFLNNVTS	PFKWSEPS**E**PLELVIKE
Horse	1316	93	TSIDD**N**ESFFLNNVTY	PFKWSEPS**E**PLELVIKE
Megabat	986	74.6	TNIND**D**ESFFLNNVTY	PFKWSEPS**E**PLEVVIKX

doi:10.1371/journal.pgen.1006661.t004

canine population highlight the strong impact of X chromosome genes in determining the weight and muscling of modern dog breeds.

Our previous studies identified alleles in the *GHR*, *HMGA2*, *SMAD2*, and *STC2* genes as major contributors to SBW [47]. When we included genotyping data from *IGF1* and *IGF1R*, which we had identified previously as body size genes in dogs [41,42], we showed that these six genes explain about 60% of body size variance in small/medium breeds, but <5% of variance in large breeds. This highlights a now recurring theme in dog genetics that a small number of genes of large effect control many complex phenotypes, as opposed to many genes of small effect as is observed often in humans.

We used two different approaches to identify variants associated with large body size. SNP chip data were used to identify large regions of LD. However, this strategy does not detect rare variants that are not in LD. WGS provides a complementary tool for these types of analyses. Indeed, this allows detection of rare mutations that would otherwise go unnoticed. In this study, the combination of dense SNP chip data (Illumina 170k) and WGS highlighted rare variants, such as the *ACSL4* mutation, which are specific to a subset of large breeds, a result not found with SNP chip data alone. This approach allowed us to define a new and very specific phenotype, the heavy muscling trait, which had not been previously described in dogs at a genetic level.

We found first that *IRS4* is strongly associated with large body size in dogs. The gene encodes a cytoplasmic protein that contains several potential tyrosine and serine/threonine phosphorylation sites. IRS4 interacts with multiple growth factor receptors such as IGF1R, enhancing IGF1-stimulated cell growth [49]. This gene is highly expressed in the hypothalamus which itself plays a primary role in regulation of body weight [56]. It is also estrogen-regulated [62], which may explain, in part, the established link between estrogen and body fat distribution [63]. Moreover, a double "knock-out" mouse model (*bIrs2-/-.Irs4-/y*) developed severe obesity suggesting that IRS4 synergizes and complements IRS2 [64]. In humans, six SNPs in *IRS4* have been identified that are associated with obesity, albeit in a cohort of patients with schizophrenia [56]. In our study, we identified three genomic variations in *IRS4*. Neither the codon deletion nor one codon insertion in the exonic region of *IRS4* appeared to be associated with disruptions in protein function. However, in large bulky/muscled breeds, we also detected an associated 56 kb deletion located 150 kb upstream of the start codon of *IRS4*. This deletion contained several repeated elements, and may contain regulatory elements that affect the expression profile of *IRS4* [65,66].

While no correlation was found between height and *IRS4* in the human study [56], in our canine study we observe a strong correlation between *IRS4* and SBW that extends to include standard breed height (SBH) (S4 Fig). SBH is the height range assigned by the AKC for a given breed. However, the addition of the SBH as co-variate in primary GWAS results in the loss of the locus 1 signal. Interestingly, the reverse analysis confirms the strong association between SBW and both loci. Indeed, the addition of SBW as a covariate for the SBH GWAS results in the loss of both signals on the X chromosome (S4 Fig). Overall this suggests that while both *IRS4* and *IGSF1*, the latter of which is the second candidate gene on X chromosome, are associated with variation in breed size, *IRS4* is necessary, but not sufficient, for increasing size.

We also showed that the *IGSF1* gene, positioned at a second locus on the X chromosome, is strongly associated with large dog breeds. This gene encodes a plasma membrane glycoprotein and is involved in the thyroid hormone pathway [50]. In large dog breeds, we identified two mutations, one single codon deletion and one missense mutation, both of which are located in a highly conserved immunoglobulin-like domain of IGSF1 protein. In humans, mutations in the same IGSF1 protein domain are associated with the X-linked IGSF1 deficiency syndrome [50–52,67–69]. Some patients show growth hormone (GH) deficiency during childhood, and

67% of male children are reportedly overweight while 21% are obese (Review in [70]). The general observation is supported by the fact that Igsf1-deficient male mice show diminished pituitary and serum thyroid-stimulating hormone (TSH) concentrations, reduced pituitary thyrotropin-releasing hormone (TRH) receptor expression, and increased body mass [50]. Measuring these hormone levels in dogs, while difficult, may confirm the parallels between dogs and mice.

We also detected a strong association between *IGSF1* and SBH (S4 Fig). Human studies used body mass index (BMI) as a measure of obesity given a particular height. To date, 97 loci are associated with human BMI [71]. It could be interesting to develop the same body mass index measure for dogs to better understand the results regarding *IGSF1*, *IRS4*, *ACSL4*, *IGF1*, *IGF1R*, *HMAGA2*, *GHR*, *SMAD2*, and *STC2*. This approach could explain why our study revealed that 50% of small/medium breed dogs have the "large alleles," mainly found in muscled breeds such as Boston Terrier or French Bulldog (Fig 4). Interestingly, the *IGSF1* locus also appears to be under selection in GWAS studies for other morphologic traits, such as brachycephalic (*e.g.* bulldog, pug) versus dolichocephalic (*e.g.* afghan hound, collie) skull shape [22,72] (Fig 4). In humans, patients with microduplication of the *IGSF1* locus present syndromic facial appearance [73]. The varying phenotypes associated with *IGSF1* illustrate the intermingling of genes and phenotypes regarding skeletal formation.

In addition to breed standard weight and heights, this study revealed a genetic association with a well-defined phenotype of bulkiness, due to heavy muscling and fat, which we found to be strongly associated with a highly conserved single nucleotide in the 5' UTR in canine *ACSL4* at locus 1 (S3 Fig). *ACSL4* belongs to the long-chain acyl-CoA synthetase (ACSL) family and five genes have been identified in mammals (*ACSL1*, *3*, *4*, *5*, and *6*) [74,75]. *ACSL4* binds specifically to longer chain polyunsaturated fatty acids. While *ACSL4* plays a role in many cellular processes [76–80], increased *ACSL4* expression in the liver likely promotes fatty acid uptake [53]. The relationship between the gene and body shape in dogs fits well with this observation. We did not observe the same relationship between *ACSL4* and stocky dogs from small breeds, suggesting that the genetic variant found in large dogs is not relevant in the absence of genes that increase body size.

In the pig, mutations in *ACSL4* are associated with a phenotype termed "back fat thickness (BFT)". There are 75 common breeds of pigs (http://www.thepigsite.com/) and large variation in adiposity between breeds has been described [58]. Pig breeds with considerable back fat are used to study human obesity as well as obesity-related diseases, such as metabolic syndrome [81]. Four QTLs on the porcine X chromosome were associated with the BFT, muscle mass, and intramuscular fat content [57,59]. Post-mortem studies reveal that polymorphisms surrounding the *ACSL4* gene are associated with BFT and muscle-associated traits in a pig breed-specific manner [57] as was observed in dog breeds within our study. Specifically, the canine variant (chrX.g.82919525C>T) was observed only in bulky dogs including, for instance, the Bullmastiff, Greater Swiss Mountain Dog, Newfoundland and Saint Bernard. All of these breeds are well-muscled breeds compared to the leaner Great Dane, Borzoi which, for example, lack the variant. The absence of the derived *ACSL4* allele in more than 97% of breeds which meet the definition of medium/small, as well as in giant thin breeds led us to define the "bulky phenotype" in dogs characterized by the traits of heavy muscling and back fat thickness which, together, are observed in 54% of the large breeds. We also notice an interesting correlation between the presence of the derived allele in some "large breeds" and their historic geographic distribution (S5 Fig). The "bulky allele" seems to have appeared in England-France (Dogue de Bordeaux, English and Bullmastiff), become fixed in these breeds, and then spread through Europe (Bernese Mountain Dog, Leonberg, Kuvasz). Mediterranean and Eurasian breeds (Cane Corso, Neapolitan Mastiff, Anatolian Shepherd) do not have this allele, likely

reflecting the recent geographic spread of the allele in Europe. Finally, additional studies in pigs describe two mutations in the *IRS4* gene, perhaps suggesting a second role for *IRS4* as a contributor to BFT as well as general body size [57,82].

In this study, we utilized WGS and GWAS to identify genes highly associated with large body size in dogs. Modern dogs display a range of traits that have been easily mapped by taking advantage of the long LD observed in many breeds. That same LD makes it problematic to go from associated marker to gene. The availability of WGS represents a major advance for tackling this issue and, in this case, allowed us to disentangle the genetics of a complex trait on a relatively homogenous chromosome. While a large number of genes of small effect seem to control body size in humans, in dogs a surprisingly small number of genes of large effect explain the range in size observed across breeds. As dogs at the extremes of the body size continuum are studied, it will be interesting to note if genes previously identified from human studies are identified, or if an entirely new repertoire of genes are found which contribute to gigantism or miniaturization of breeds. Studies in domestic dogs, therefore, provide a mechanism for understanding the genetics that underlies traits of interest in both human and domesticated animals.

Methods

Sample collection and DNA extraction

Whole blood samples were collected into EDTA or ACD anticoagulant from AKC-registered dogs. Genomic DNA was extracted using a standard phenol-chloroform extraction protocol [83]. All procedures were reviewed and approved by the NHGRI Animal Care and Use Committee at the National Institutes of Health.

Phenotype assignment

Standard breed weights and height were obtained from several sources: weights previously listed in Rimbault et al. [47] were used, although they were updated if weights specified by the AKC [84] were different. If the AKC did not specify SBW and SBH, we used data from *Atlas of Dog Breeds of the World* [16]. SBW and SBH (male + female average) were applied to all samples from the same breed and the values used in this study are listed in S1 and S2 Tables. Analyses by sex did not change the results, thus we retained the genotypes as a single dataset.

SNP genotyping

Genotyping was performed using the Illumina 170K Canine HD SNP array containing approximately 170,000 SNPs distributed across the 38 canine autosomes and the X chromosome. Genotypes were called using Illumina Genome Studio software. In total, 855 dogs, 418 males and 437 females, were genotyped [21]. Dogs belong to 88 different breeds. Eighty-two breeds with nine to 11 dogs were genotyped and six large dog breeds with four to six dogs genotyped. All samples had a call rate greater than 93% (range: 93.57–99.98, average: 99.84). SNPs with a minor allele frequency $<1\%$ or the presence of $>5\%$ missing genotypes were pruned, resulting in a final dataset of 150,895 SNPs that were used for the subsequent GWAS. The GWAS was conducted using the software GEMMA v0.94.1 (Genome-wide Efficient Mixed-Model Association) [54,55] as a linear mixed-model software using a centered kinship matrix. Pedigrees of dogs used in the study were verified to avoid inclusion of close relatives, *i.e.* none shared a common grandparent. In the two regions of interest, pairwise r^2 values were calculated using Plink v1.07 [85].

Whole Genome Sequencing

Fine mapping at both loci used data from 157 individuals who had undergone WGS and for which the data were published or available online from the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra). Six new WGS recently produced by the NIH Intramural Sequencing Center (NISC) were also included. The latter were produced using the Illumina TruSeq DNAPCR-Free Protocol (Cat.FC-121-3001). Reads were aligned to the CanFam 3.1 reference genome (http://genome.ucsc.edu/cgi-bin/hgGateway?db=canFam3) using BWA 0.7.13 MEM [86] and sorted using SAMtools 1.3.1 [87]. PCR duplicates were marked as secondary reads using PicardTools 2.2.4 (http://github.com/broadinstitute/picard) for those libraries that were not PCR-free. GATK 3.5 [88,89] was used to perform local realignment around putative indels events using 714,278 variants published in [90] as the training set. A total of 172,254 Illumina Canine HD Chip positions and 2,738,537 dbSNP v131 variants were utilized for base recalibration with GATK 3.5. SNV were called per-individual in gVCF mode of HaplotypeCaller [91], with subsequent joint-calling across all individuals. Variant quality score recalibration was conducted with GATK best practices and default parameters for SNV and indels separately as follows. Indel recalibration: 714,278 variants as truth and training sets with a prior of six [90]. SNV recalibration: 172,254 Illumina Canine HD Chip variants (known, training, true, prior = 12); 2,738,537 dbSNP v131 variants (known, true, prior = 8); 3,627,539 published variants from [92] (known, training, prior = 6). We only used genomes with a sequencing depth >10X and retained only variants with a minimum of two alleles and a minor allele frequency >5%. For locus 1, 6,809 variants met the QC criteria, while 1,997 met the criteria for locus 2. These variants were analyzed using GEMMA v0.94.1 as a linear mixed-model software [54,55]. A centered kinship matrix was estimated extracting SNPs from the 163 WGS data using the positions of 147,740 SNPs of the Illumina Canine HD SNP array. DELLY and CNVnator were used to analyze structural variants, including indels, inversions and duplications that were >100 bp in length [92,93].

Sanger sequencing

To confirm the distribution of "large alleles" in the *IRS4* and *IGSF1* genes, we genotyped a panel of 512 dogs of 93 breeds and 24 wolves (S6 Table). Primer pairs were designed to target regions that included the variants of interest, and two pairs were specifically designed to reveal the absence/presence of the deletion (S7 Table). Targeted regions were assayed using polymerase chain reaction (PCR) with AmpliTaq Gold. PCR products were purified by ExoSap-It reaction (Affymetrix), and then Sanger sequenced using BigDye Terminator v3.1 (Applied Biosystems). Products from sequencing reactions were run on ABI 3730 DNA analyzer. Sequence traces were analyzed using Phred/Phrap/Consed package [94–96]. The absence/presence of the deletion was detected after migration of the PCR products on a 1% agarose gel followed by staining with ethidium bromide. To analyze the *ACSL4* variant, we sequenced a larger set of 985 unrelated dogs and wild canids, including 24 geographically diverse gray wolves from North America, Europe, and Asia, two coyotes and two red wolves (S5 Table). Three hundred and fifteen of these dogs were included in the dataset used for the initial GWAS.

Conservation between species

To estimate the conservation of mutated codons/nucleotides between mammals, we used both protein and gene sequences from *IRS4*, *IGSF1* and *ACSL4* which were available on Ensembl [97]. We selected proteins for dog, human, mouse, cat, pig, horse, cow, and megabat and we used SIM [98] and LALNVIEW [99] to align sequences.

Supporting information

S1 Fig. Summary of genetic investigations of X chromosome locus 1 provided by dog model, SNP chip and WGS data.
(TIF)

S2 Fig. Summary of genetic investigations of X chromosome locus 2 provided by dog model, SNP chip and WGS data.
(TIF)

S3 Fig. UCSC screenshot of the conservation between mammals of the 5'UTR variant in *ACSL4* (red square).
(TIF)

S4 Fig. Standard breed weight and standard breed height GWAS results. Manhattan plots using Standard Breed Weight without (A) and with Standard Breed Height as a covariate (B). Manhattan plots using Standard Breed Height without (C) and with Standard Breed Weight as a covariate (D).
(PDF)

S5 Fig. Historical distributions of the large breeds and the "bulky allele".
(TIF)

S1 Table. List of the 855 dogs genotypes on the Illumina Canine HD SNP array.
(XLSX)

S2 Table. List of the 163 dogs that underwent WGS for this analysis of the X chromosome.
(XLSX)

S3 Table. Variants identified at locus 1 by fine-mapping using WGS data from 163 dogs.
(XLSX)

S4 Table. Variants identified at locus 2 by fine-mapping using WGS data from 163 dogs.
(XLSX)

S5 Table. Lists of dogs used for Sanger sequencing of the 5'UTR mutation in *ACSL4* and the intronic *AMMECR1* variants.
(XLSX)

S6 Table. Lists of the dogs used for Sanger sequencing for the candidate mutations in *IGSF1*.
(XLSX)

S7 Table. Primers used for PCR and Sanger sequencing.
(XLSX)

Acknowledgments

We gratefully acknowledge the support from the Intramural Program of the National Human Genome Research Institute of the National Institute of Health, and the NIH Intramural Sequencing Center. We also thank dog owners and breeders who generously provided DNA samples for this study.

Author Contributions

Conceptualization: JP MR EAO.

Data curation: JP MR BWD.

Formal analysis: JP MR.

Funding acquisition: EAO.

Investigation: JP MR JJS EAO.

Methodology: JP MR EAO.

Project administration: EAO.

Resources: EAO.

Software: JP MR BWD.

Supervision: EAO.

Validation: JP MR FJW.

Visualization: JP MR.

Writing – original draft: JP MR FJW BWD JJS EAO.

Writing – review & editing: JP MR EAO.

References

1. Sablin MV, Khlopachev GA. The Earliest Ice Age Dogs: Evidence from Eliseevichi 11. Current Anthropology. The University of Chicago Press; 2002; 43: 795–799.
2. Savolainen P, Zhang Y-P, Luo J, Lundeberg J, Leitner T. Genetic evidence for an East Asian origin of domestic dogs. Science. American Association for the Advancement of Science; 2002; 298: 1610–1613.
3. Germonpré M, Sablin MV, Stevens RE, Hedges REM, Hofreiter M, Stiller M, et al. Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: osteometry, ancient DNA and stable isotopes. Journal of Archaeological Science. 2009; 36: 473–490.
4. Pang J-F, Kluetsch C, Zou X-J, Zhang A-B, Luo L-Y, Angleby H, et al. mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves. Mol Biol Evol. Oxford University Press; 2009; 26: 2849–2864.
5. Ovodov ND, Crockford SJ, Kuzmin YV, Higham TFG, Hodgins GWL, van der Plicht J. A 33,000-year-old incipient dog from the Altai Mountains of Siberia: evidence of the earliest domestication disrupted by the Last Glacial Maximum. Stepanova A, editor. PLoS ONE. Public Library of Science; 2011; 6: e22821.
6. Larson G, Karlsson EK, Perri A, Webster MT, Ho SYW, Peters J, et al. Rethinking dog domestication by integrating genetics, archeology, and biogeography. Proc Natl Acad Sci USA. National Acad Sciences; 2012; 109: 8878–8883.
7. Freedman AH, Gronau I, Schweizer RM, Ortega-Del Vecchyo D, Han E, Silva PM, et al. Genome sequencing highlights the dynamic early history of dogs. Andersson L, editor. PLoS Genet. Public Library of Science; 2014; 10: e1004016.
8. Skoglund P, Ersmark E, Palkopoulou E, Dalén L. Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. Curr Biol. 2015; 25: 1515–1519. doi: 10.1016/j.cub.2015.04.019 PMID: 26004765
9. Frantz LAF, Mullin VE, Pionnier-Capitan M, Lebrasseur O, Ollivier M, Perri A, et al. Genomic and archaeological evidence suggest a dual origin of domestic dogs. Science. American Association for the Advancement of Science; 2016; 352: 1228–1231.
10. Freedman AH, Schweizer RM, Ortega-Del Vecchyo D, Han E, Davis BW, Gronau I, et al. Demographically-Based Evaluation of Genomic Regions under Selection in Domestic Dogs. Webster MT, editor. PLoS Genet. Public Library of Science; 2016; 12: e1005851.
11. Wayne RK. Molecular evolution of the dog family. Trends Genet. 1993; 9: 218–224. PMID: 8337763
12. Vilà C, Savolainen P, Maldonado JE, Amorim IR, Rice JE, Honeycutt RL, et al. Multiple and ancient origins of the domestic dog. Science. 1997; 276: 1687–1689. PMID: 9180076

13. Thalmann O, Shapiro B, Cui P, Schuenemann VJ, Sawyer SK, Greenfield DL, et al. Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. Science. American Association for the Advancement of Science; 2013; 342: 871–874.
14. Leonard JA, Wayne RK, Wheeler J, Valadez R, Guillén S, Vilà C. Ancient DNA evidence for Old World origin of New World dogs. Science. American Association for the Advancement of Science; 2002; 298: 1613–1616.
15. Fogle B. *The Encyclopedia of the Dog*. New York: DK Publishing; 2000.
16. Wilcox B, Walkowicz C. Atlas of Dog Breeds of the World. 5 ed. Neptune City, NJ: T.F.H. Publications; 1995.
17. Dreger DL, Davis BW, Cocco R, Sechi S, Di Cerbo A, Parker HG, et al. Studies of the Fonni's Dogs from Sardinia Show Commonalities between Development of Pure Breeds and Population Isolates. Genetics. Genetics; 2016; 204: genetics.116.192427–755.
18. Ostrander EA, Giniger E. Semper fidelis: what man's best friend can teach us about human biology and disease. Am J Hum Genet. 1997; 61: 475–480. doi: 10.1086/515522 PMID: 9326310
19. Sutter NB, Ostrander EA. Dog star rising: the canine genetic system. Nat Rev Genet. Nature Publishing Group; 2004; 5: 900–910.
20. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. Nature Publishing Group; 2005; 438: 803–819.
21. Dreger DL, Rimbault M, Davis BW, Bhatnagar A, Parker HG, Ostrander EA. Whole-genome sequence, SNP chips and pedigree structure: building demographic profiles in domestic dog breeds to optimize genetic-trait mapping. Dis Model Mech. The Company of Biologists Ltd; 2016; 9: 1445–1460.
22. Boyko AR, Quignon P, Li L, Schoenebeck JJ, Degenhardt JD, Lohmueller KE, et al. A simple genetic architecture underlies morphological variation in dogs. Hoekstra HE, editor. PLoS Biol. Public Library of Science; 2010; 8: e1000451.
23. Vaysse A, Ratnakumar A, Derrien T, Axelsson E, Rosengren Pielberg G, Sigurdsson S, et al. Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. Akey JM, editor. PLoS Genet. Public Library of Science; 2011; 7: e1002316.
24. Sutter NB, Eberle MA, Parker HG, Pullar BJ, Kirkness EF, Kruglyak L, et al. Extensive and breed-specific linkage disequilibrium in Canis familiaris. Genome Res. Cold Spring Harbor Lab; 2004; 14: 2388–2396.
25. Merveille A-C, Davis EE, Becker-Heck A, Legendre M, Amirav I, Bataille G, et al. CCDC39 is required for assembly of inner dynein arms and the dynein regulatory complex and for normal ciliary motility in humans and dogs. Nat Genet. Nature Publishing Group; 2011; 43: 72–78.
26. Grall A, Guaguère E, Planchais S, Grond S, Bourrat E, Hausser I, et al. PNPLA1 mutations cause autosomal recessive congenital ichthyosis in golden retriever dogs and humans. Nat Genet. 2012; 44: 140–147. doi: 10.1038/ng.1056 PMID: 22246504
27. Drögemüller M, Jagannathan V, Becker D, Drögemüller C, Schelling C, Plassais J, et al. A mutation in the FAM83G gene in dogs with hereditary footpad hyperkeratosis (HFH). Barsh GS, editor. PLoS Genet. Public Library of Science; 2014; 10: e1004370.
28. Plassais J, Guaguère E, Lagoutte L, Guillory A-S, de Citres CD, Degorce-Rubiales F, et al. A spontaneous KRT16 mutation in a dog breed: a model for human focal non-epidermolytic palmoplantar keratoderma (FNEPPK). J Invest Dermatol. Elsevier; 2015; 135: 1187–1190.
29. Shearin AL, Hédan B, Cadieu E, Erich SA, Schmidt EV, Faden DL, et al. The MTAP-CDKN2A locus confers susceptibility to a naturally occurring canine cancer. Cancer Epidemiol Biomarkers Prev. American Association for Cancer Research; 2012; 21: 1019–1027.
30. Karyadi DM, Karlins E, Decker B, vonHoldt BM, Carpintero-Ramirez G, Parker HG, et al. A copy number variant at the KITLG locus likely confers risk for canine squamous cell carcinoma of the digit. Horwitz MS, editor. PLoS Genet. Public Library of Science; 2013; 9: e1003409.
31. Decker B, Parker HG, Dhawan D, Kwon EM, Karlins E, Davis BW, et al. Homologous Mutation to Human BRAF V600E Is Common in Naturally Occurring Canine Bladder Cancer—Evidence for a Relevant Model System and Urine-Based Diagnostic Test. Mol Cancer Res. Molecular Cancer Research; 2015; 13: 993–1002. doi: 10.1158/1541-7786.MCR-14-0689 PMID: 25767210
32. Karlsson EK, Sigurdsson S, Ivansson E, Thomas R, Elvers I, Wright J, et al. Genome-wide analyses implicate 33 loci in heritable dog osteosarcoma, including regulatory variants near CDKN2A/B. Genome Biol. BioMed Central; 2013; 14: R132.
33. Jónasdóttir TJ, Mellersh CS, Moe L, Heggebø R, Gamlem H, Ostrander EA, et al. Genetic mapping of a naturally occurring hereditary renal cancer syndrome in dogs. Proc Natl Acad Sci USA. National Acad Sciences; 2000; 97: 4132–4137.

34. Parker HG, Chase K, Cadieu E, Lark KG, Ostrander EA. An insertion in the RSPO2 gene correlates with improper coat in the Portuguese water dog. J Hered. Oxford University Press; 2010; 101: 612–617.

35. Parker HG, Shearin AL, Ostrander EA. Man"s best friend becomes biology"s best in show: genome analyses in the domestic dog. Annu Rev Genet. Annual Reviews; 2010; 44: 309–336.

36. Boyko AR. The domestic dog: man's best friend in the genomic era. Genome Biol. BioMed Central; 2011; 12: 216.

37. Schoenebeck JJ, Ostrander EA. Insights into morphology and disease from the dog genome project. Annu Rev Cell Dev Biol. Annual Reviews; 2014; 30: 535–560.

38. Salmon Hillbertz NHC, Isaksson M, Karlsson EK, Hellmén E, Pielberg GR, Savolainen P, et al. Duplication of FGF3, FGF4, FGF19 and ORAOV1 causes hair ridge and predisposition to dermoid sinus in Ridgeback dogs. Nat Genet. Nature Publishing Group; 2007; 39: 1318–1320.

39. Webster MT, Kamgari N, Perloski M, Hoeppner MP, Axelsson E, Hedhammar A, et al. Linked genetic variants on chromosome 10 control ear morphology and body mass among dog breeds. BMC Genomics. BioMed Central; 2015; 16: 474.

40. Chase K, Carrier DR, Adler FR, Jarvik T, Ostrander EA, Lorentzen TD, et al. Genetic basis for systems of skeletal quantitative traits: principal component analysis of the canid skeleton. Proc Natl Acad Sci USA. National Acad Sciences; 2002; 99: 9930–9935.

41. Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K, Zhu L, et al. A single IGF1 allele is a major determinant of small size in dogs. Science. American Association for the Advancement of Science; 2007; 316: 112–115.

42. Hoopes BC, Rimbault M, Liebers D, Ostrander EA, Sutter NB. The insulin-like growth factor 1 receptor (IGF1R) contributes to reduced size in dogs. Mamm Genome. 2012; 23: 780–790. doi: 10.1007/s00335-012-9417-z PMID: 22903739

43. Walenkamp MJE, Wit JM. Genetic disorders in the growth hormone—insulin-like growth factor-I axis. Horm Res. 2006; 66: 221–230. doi: 10.1159/000095161 PMID: 16917171

44. Rosenfeld RG, Belgorosky A, Camacho-Hubner C, Savage MO, Wit JM, Hwa V. Defects in growth hormone receptor signaling. Trends Endocrinol Metab. 2007; 18: 134–141. doi: 10.1016/j.tem.2007.03.004 PMID: 17391978

45. Fang P, Girgis R, Little BM, Pratt KL, Guevara-Aguirre J, Hwa V, et al. Growth hormone (GH) insensitivity and insulin-like growth factor-I deficiency in Inuit subjects and an Ecuadorian cohort: functional studies of two codon 180 GH receptor gene mutations. J Clin Endocrinol Metab. Endocrine Society; 2008; 93: 1030–1037.

46. Savage MO, Hwa V, David A, Rosenfeld RG, Metherell LA. Genetic Defects in the Growth Hormone-IGF-I Axis Causing Growth Hormone Insensitivity and Impaired Linear Growth. Front Endocrinol (Lausanne). Frontiers; 2011; 2: 95.

47. Rimbault M, Beale HC, Schoenebeck JJ, Hoopes BC, Allen JJ, Kilroy-Glynn P, et al. Derived variants at six genes explain nearly half of size reduction in dog breeds. Genome Res. Cold Spring Harbor Lab; 2013; 23: 1985–1995.

48. Hayward JJ, Castelhano MG, Oliveira KC, Corey E, Balkman C, Baxter TL, et al. Complex disease and phenotype mapping in the domestic dog. Nat Commun. Nature Research; 2016; 7: 10460.

49. Qu BH, Karas M, Koval A, LeRoith D. Insulin receptor substrate-4 enhances insulin-like growth factor-I-induced cell proliferation. J Biol Chem. 1999; 274: 31179–31184. PMID: 10531310

50. Sun Y, Bak B, Schoenmakers N, van Trotsenburg ASP, Oostdijk W, Voshol P, et al. Loss-of-function mutations in IGSF1 cause an X-linked syndrome of central hypothyroidism and testicular enlargement. Nat Genet. Nature Research; 2012; 44: 1375–1381.

51. Joustra SD, Wehkalampi K, Oostdijk W, Biermasz NR, Howard S, Silander TL, et al. IGSF1 variants in boys with familial delayed puberty. Eur J Pediatr. Springer Berlin Heidelberg; 2015; 174: 687–692.

52. Asakura Y, Abe K, Muroya K, Hanakawa J, Oto Y, Narumi S, et al. Combined Growth Hormone and Thyroid-Stimulating Hormone Deficiency in a Japanese Patient with a Novel Frameshift Mutation in IGSF1. Horm Res Paediatr. 2015; 84: 349–354. doi: 10.1159/000438672 PMID: 26302767

53. Westerbacka J, Kolak M, Kiviluoto T, Arkkila P, Sirén J, Hamsten A, et al. Genes involved in fatty acid partitioning and binding, lipolysis, monocyte/macrophage recruitment, and inflammation are overexpressed in the human fatty liver of insulin-resistant subjects. Diabetes. American Diabetes Association; 2007; 56: 2759–2765.

54. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. Nature Research; 2012; 44: 821–824.

55. Zhou X, Stephens M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat Methods. Nature Research; 2014; 11: 407–409.

56. Melkersson K, Persson B. Association between body mass index and insulin receptor substrate-4 (IRS-4) gene polymorphisms in patients with schizophrenia. Neuro Endocrinol Lett. 2011; 32: 634–640. PMID: 22167131

57. Ma J, Gilbert H, Iannuccelli N, Duan Y, Guo B, Huang W, et al. Fine mapping of fatness QTL on porcine chromosome X and analyses of three positional candidate genes. BMC Genet. BioMed Central; 2013; 14: 46.

58. Stachowiak M, Szczerbal I, Switonski M. Genetics of Adiposity in Large Animal Models for Human Obesity-Studies on Pigs and Dogs. Prog Mol Biol Transl Sci. Elsevier; 2016; 140: 233–270.

59. Cepica S, Bartenschlager H, Geldermann H. Mapping of QTL on chromosome X for fat deposition, muscling and growth traits in a wild boar x Meishan F2 family using a high-density gene map. Anim Genet. Blackwell Publishing Ltd; 2007; 38: 634–638.

60. Corominas J, Ramayo-Caldas Y, Castelló A, Muñoz M, Ibáñez-Escriche N, Folch JM, et al. Evaluation of the porcine ACSL4 gene as a candidate gene for meat quality traits in pigs. Anim Genet. 2012; 43: 714–720. doi: 10.1111/j.1365-2052.2012.02335.x PMID: 22497636

61. Mercadé A, Estellé J, Pérez-Enciso M, Varona L, Silió L, Noguera JL, et al. Characterization of the porcine acyl-CoA synthetase long-chain 4 gene and its association with growth and meat quality traits. Anim Genet. Blackwell Publishing Ltd; 2006; 37: 219–224.

62. Xu X, Coats JK, Yang CF, Wang A, Ahmed OM, Alvarado M, et al. Modular genetic control of sexually dimorphic behaviors. Cell. 2012; 148: 596–607. doi: 10.1016/j.cell.2011.12.018 PMID: 22304924

63. Frank A, Brown LM, Clegg DJ. The role of hypothalamic estrogen receptors in metabolic regulation. Front Neuroendocrinol. 2014; 35: 550–557. doi: 10.1016/j.yfrne.2014.05.002 PMID: 24882636

64. Sadagurski M, Dong XC, Myers MG, White MF. Irs2 and Irs4 synergize in non-LepRb neurons to control energy balance and glucose homeostasis. Mol Metab. 2014; 3: 55–63. doi: 10.1016/j.molmet.2013.10.004 PMID: 24567904

65. Nakanishi A, Kobayashi N, Suzuki-Hirano A, Nishihara H, Sasaki T, Hirakawa M, et al. A SINE-derived element constitutes a unique modular enhancer for mammalian diencephalic Fgf8. Batzer MA, editor. PLoS ONE. Public Library of Science; 2012; 7: e43785.

66. Tashiro K, Teissier A, Kobayashi N, Nakanishi A, Sasaki T, Yan K, et al. A mammalian conserved element derived from SINE displays enhancer properties recapitulating Satb2 expression in early-born callosal projection neurons. Brosius J, editor. PLoS ONE. Public Library of Science; 2011; 6: e28497.

67. Joustra SD, van Trotsenburg ASP, Sun Y, Losekoot M, Bernard DJ, Biermasz NR, et al. IGSF1 deficiency syndrome: A newly uncovered endocrinopathy. Rare Dis. Taylor & Francis; 2013; 1: e24883.

68. Nakamura A, Bak B, Silander TLR, Lam J, Hotsubo T, Yorifuji T, et al. Three novel IGSF1 mutations in four Japanese patients with X-linked congenital central hypothyroidism. J Clin Endocrinol Metab. Endocrine Society Chevy Chase, MD; 2013; 98: E1682–91.

69. Tenenbaum-Rakover Y, Turgeon M-O, London S, Hermanns P, Pohlenz J, Bernard DJ, et al. Familial Central Hypothyroidism Caused by a Novel IGSF1 Gene Mutation. Thyroid. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA; 2016;: thy.2015.0672.

70. Joustra SD, Heinen CA, Schoenmakers N, Bonomi M, Ballieux BEPB, Turgeon M-O, et al. IGSF1 Deficiency: Lessons From an Extensive Case Series and Recommendations for Clinical Management. J Clin Endocrinol Metab. Endocrine Society Washington, DC; 2016; 101: 1627–1636.

71. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015; 518: 197–206. doi: 10.1038/nature14177 PMID: 25673413

72. Schoenebeck JJ, Hutchinson SA, Byers A, Beale HC, Carrington B, Faden DL, et al. Variation of BMP3 contributes to dog breed skull diversity. Leeb T, editor. PLoS Genet. Public Library of Science; 2012; 8: e1002849.

73. Møller RS, Jensen LR, Maas SM, Filmus J, Capurro M, Hansen C, et al. X-linked congenital ptosis and associated intellectual disability, short stature, microcephaly, cleft palate, digital and genital abnormalities define novel Xq25q26 duplication syndrome. Hum Genet. Springer Berlin Heidelberg; 2014; 133: 625–638.

74. Soupene E, Kuypers FA. Mammalian long-chain acyl-CoA synthetases. Exp Biol Med (Maywood). SAGE Publications; 2008; 233: 507–521.

75. Coleman RA, Lee DP. Enzymes of triacylglycerol synthesis and their regulation. Prog Lipid Res. 2004; 43: 134–176. PMID: 14654091

76. Monaco ME, Creighton CJ, Lee P, Zou X, Topham MK, Stafforini DM. Expression of Long-chain Fatty Acyl-CoA Synthetase 4 in Breast and Prostate Cancers Is Associated with Sex Steroid Hormone Receptor Negativity. Transl Oncol. Neoplasia Press; 2010; 3: 91–98.

77. Wu X, Deng F, Li Y, Daniels G, Du X, Ren Q, et al. ACSL4 promotes prostate cancer growth, invasion and hormonal resistance. Oncotarget. Impact Journals; 2015; 6: 44849–44863.

78. Wu X, Li Y, Wang J, Wen X, Marcus MT, Daniels G, et al. Long chain fatty Acyl-CoA synthetase 4 is a biomarker for and mediator of hormone resistance in human breast cancer. Li J, editor. PLoS ONE. Public Library of Science; 2013; 8: e77060.

79. Chen W-C, Wang C-Y, Hung Y-H, Weng T-Y, Yen M-C, Lai M-D. Systematic Analysis of Gene Expression Alterations and Clinical Outcomes for Long-Chain Acyl-Coenzyme A Synthetase Family in Cancer. Shridhar V, editor. PLoS ONE. Public Library of Science; 2016; 11: e0155660.

80. Gazou A, Riess A, Grasshoff U, Schäferhoff K, Bonin M, Jauch A, et al. Xq22.3-q23 deletion including ACSL4 in a patient with intellectual disability. Am J Med Genet A. Wiley Subscription Services, Inc., A Wiley Company; 2013; 161A: 860–864.

81. Toedebusch RG, Roberts MD, Wells KD, Company JM, Kanosky KM, Padilla J, et al. Unique transcriptomic signature of omental adipose tissue in Ossabaw swine: a model of childhood obesity. Physiol Genomics. American Physiological Society; 2014; 46: 362–375.

82. Masopust M, Vykoukalová Z, Knoll A, Bartenschlager H, Mileham A, Deeb N, et al. Porcine insulin receptor substrate 4 (IRS4) gene: cloning, polymorphism and association study. Mol Biol Rep. Springer Netherlands; 2011; 38: 2611–2617.

83. Maniatis T, Fritsch E, Sambrook J. Molecular cloning: A laboratory manual. In: Cold Spring Harbor Laboratory Press, editor. New York; 1982.

84. American Kennel Club. The complete dog book, 20th edition. Ballantine Books. New York; 2006.

85. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81: 559–575. doi: 10.1086/519795 PMID: 17701901

86. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. Oxford University Press; 2009; 25: 1754–1760.

87. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. Oxford University Press; 2009; 25: 2078–2079.

88. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. Cold Spring Harbor Lab; 2010; 20: 1297–1303.

89. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43: 491–498. doi: 10.1038/ng.806 PMID: 21478889

90. Axelsson E, Ratnakumar A, Arendt M-L, Maqbool K, Webster MT, Perloski M, et al. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature. Nature Research; 2013; 495: 360–364.

91. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. Hoboken, NJ, USA: John Wiley & Sons, Inc; 2013; 43: 11.10.1–33.

92. Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. Oxford University Press; 2012; 28: i333–i339.

93. Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. Cold Spring Harbor Lab; 2011; 21: 974–984.

94. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 1998; 8: 175–185. PMID: 9521921

95. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 1998; 8: 186–194. PMID: 9521922

96. Gordon D. Viewing and editing assembled sequences using Consed. Curr Protoc Bioinformatics. Hoboken, NJ, USA: John Wiley & Sons, Inc; 2003; Chapter 11: Unit11.2–11.2.43.

97. Aken BL, Ayling S, Barrell D, Clarke L, Curwen V, Fairley S, et al. The Ensembl gene annotation system. Database (Oxford). Oxford University Press; 2016; 2016: baw093.

98. Huang X, Miller W. A time-efficient, linear-space local similarity algorithm. Advances in Applied Mathematics. Academic Press; 1991; 12: 337–357.

99. Duret L, Gasteiger E, Perrière G. LALNVIEW: a graphical viewer for pairwise sequence alignments. Comput Appl Biosci. 1996; 12: 507–510. PMID: 9021269